Concept and Structure of Semantic Markers for Machine Translation in Mu-Project

Yoshiyuki Sakamoto
Electrotechnical Laboratory
Sakura-mura, Niihari-gun, Ibaraki, Japan

Tetsuya Ishikawa
Univ. of Library & Information Science
Yatabe-machi, Tsukuba-gun Ibaraki, Japan

Masayuki Satoh
Japan Information Center of Science and Technology
Nagata-cho, Chiyoda-ku Tokyo, Japan

0. Abstract

This paper discusses the semantic features of nouns classified into categories in Japanese-to-English translation, and proposes a system for semantic markers. In our system syntactic analysis is carried out by checking the semantic compatibility between verbs and nouns. The semantic structure of a sentence can be extracted at the same time as its syntactic analysis.

We also use semantic markers to select words in the transfer phase for translation into English.

The system of the Semantic Markers for Nouns consists of 13 conceptional facets including one facet for "Others"(discussed later), and is made up of 49 filial slots (semantic markers) as terminals. We have tested about 3,000 sample abstracts in science and technological fields. Our research has revealed that our method is extremely effective in determining the meanings of *Wago* verbs (basic Japanese verbs) which have broader concepts like English verbs, "make", "get", "take", "put", etc.

1. Introduction

Semantic features are introduced to ensure the maximum possible accuracy of syntactic analysis, transfer and generation. We aim at a well-balanced usage of syntax and semantics throughout the whole process of machine translation.

The present paper introduces semantic concepts for nouns classified according to facets and slots which we called semantic markers. Then we show how these semantic markers are written in the respective lexicons for analysis, transfer and generation, and how effective they are in improving the quality and accuracy of the machine translation system in each phase of analysis and transfer. Therefore, semantic features are analyzed by the structure embedded into the case frame in Japanese syntactic analysis: these features play an important role when selecting words in the transfer phase from Japanese into English.

Semantic features are more word specific. Pairs of deep cases and nouns should be written in the lexicon. However, due to the huge number of nouns, it is more effective to include pairs of the deep cases and semantic markers in the lexicon instead of nouns.

The Mu-project is a Japanese national project supported by the STA(Science and Technology Agency) "Research on a Machine Translation System(Japanese - English) for Scientific and Technological Documents."*

2. Transfer Approach to Machine Translation

We are currently restricting the domain of translation to abstract papers in scientific and technological fields. The system is based on a transfer approach and consists of three phases; analysis, transfer and generation.

In the first phase, morphological analysis divides sentence into lexical items, then syntactic analysis is carried out by syntactic and semantic analyses of Case Grammar in Japanese. In the second transfer phase, lexical features are transferred and at the same time, the syntactic structures are transferred by making them match tree patterns between Japanese and English. Here, we use semantic features to select words for translation into English. In the final generation phase, syntactic structures are generated by the Phrase Structure Grammar and the morphological features of English.

The following describes the processing functions employed in our system.

Morphological analysis and generation program are described in LISP, which is adequate for morphological process in Japanese and English, while syntactic analysis, transfer and generation programs are written in GRADE (Grammar DEscriber). Such process written in GRADE are independent of natural languages in machine translation. GRADE allows a grammar writer to write grammars using the same expression in all three phases.

Grammatical rules written in GRADE(GRAmmar DEscriber) are translated into internal forms, which are expressed by S-expression in LISP. This translation is performed by GRADE translator.

3. Concept of a Dependency Structure based on Case Grammar in Japanese

In Japan, we have come to the conclusion that case grammar is the most effective one for Japanese syntactic analysis in machine translation systems. This type of grammar has been proposed and studied by Japanese linguists before Fillmore's presentation.

As the word order is heavily restricted in English syntax, ATNG(Augmented Transition Network Grammar) based on CFG(Context Free Grammar) is adequate for syntactic analysis. However, Japanese word order is almost unrestricted and *Kakujo-shi* (postpositonal case particle) play an important role as deep cases in Japanese sentences. Therefore, case grammar is the most effective method for Japanese

* This project was carried out with the aid of a special grant for the promotion of science and technology from the Science and Technology Agency of the Japanese Government.

syntactic and semantic analyses.

In Japanese syntactic structure, the word order is unrestricted except for predicates(verbs or verb phrases) which will be located at the end of sentences. In Case Grammar, verbs play a very important role in syntactic analysis, and the other parts of speech act only in partnership with or subordinate to verbs.

That is, syntactic analysis is made by checking the semantic compatibility between verbs and nouns. Consequently, the semantic structure of a sentence can be extracted at the same time as syntactic analysis.

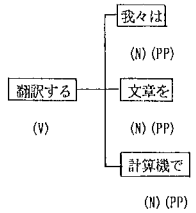
1) Morphological Analysis:Segmentation of a Japanese sentence by Lexicon Database

Ex. Input sentence'我々は計算機で文章を翻訳した.'is segmented as follows

'我々は/計算機で/文章を/翻訳した'.

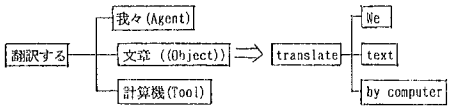
2) Syntactic Analysis: The item-to-item relationship of the sentence is analyzed to give syntactic features for the respective items.

Ex.



3) Lexical features are transferred and the syntactic structure are transferred by matching patterns between Japanese and English.

Ex.



4) Syntactic generation:The word order in English is converted according to Phrase Structure Grammar.

Ex. We translate text by computer

5) Morphological generation: Inflectional features such as tense, aspect, etc are attached.

Ex. We translated texts by computer.

Figure 2.1 Process of Machine Traslation in Mu-Project

4. Case Frame governed by *Yougen*

The case frame governed by *Yougen* using *Kakujo-shi*, Case Labels(deep cases) and semantic markers for nouns are analyzed to illustrate how we apply Case Grammar to Japanese syntactic analysis in our system.

Yougen consists of verb, *Keiyou-shi*(adjective) and *Keiyoudou-shi*(adjectival noun). *Kakujo-shi* includes obligatory case and optional case markers in Japanese syntax. But a single *Kakujo-shi* corresponds to several deep cases: for instance, <に> 'ni' corresponds to more than ten deep cases including SPAce, Space-TO, TIMe, ROLe, MANner, GOAl, PARtner, COMponent, CONdition, RANge, etc. We have analyzed relationships between *Kakujo-shi* and case labels, and written them out manually according to the sample texts of 3,000 abstracts.

As a result of categorizing deep cases, 34 Japanese case labels have been determined as shown in Table 3.1.

Table 3.1. Case Labels for Case Frames

Japanese Label	English	Examples
(1) 主体	SUBject	～が
(2) 対象	OBJect	～を
(3) 受け手	RECipient	～に与える
(4) 与え手	ORIgin	～から受ける，奪う
(5) 相手 1	PARtner	～と協議する，異なる
(6) 相手 2	OPPonent	～から保護する，独立する
(7) 時	TIMe	1980年に
(8) 時・始点	Time-FRom	5月から
(9) 時・終点	Time-TO	来年まで
(10) 時間	DURation	5分間加熱する
(11) 場所	SPAce	～に位置する，～で発生する
(12) 場所・始点	Space-FRom	～から帰る
(13) 場所・終点	Space-TO	～へ送る，～に到達する
(14) 場所・経過	Space-THrough	～を通る，上空を飛ぶ
(15) 始状態	SOUrce	5.5％から6％へ引き上げる
(16) 終状態	GOAl	英語から日本語に翻訳する
(17) 属性	ATTribute	適応性に富む，欠ける，乏しい
(18) 原因・理由	CAUse	事故で死ぬ，～から分かる
(19) 手段・道具	TOOl	イオン法で，ドリルで
(20) 材料	MATerial	ペーストで作る
(21) 構成要素	COMponent	～から成る，～で構成する
(22) 方式	MANner	並列に，10 m/secで
(23) 条件	CONdition	焦点深度で決まる
(24) 目的	PURpose	～に適する，備える，必要な
(25) 役割	ROLe	議長に選ぶ，～として用いる
(26) 内容規定	COnTent	～と呼ぶ，述べる，みなす
(27) 範囲規定	RANge	～について，～に関して
(28) 提題	TOPic	～は，～とは
(29) 観点	VIEwpoint	立場から，～の点で
(30) 比較	COmpaRison	～より大きい，～に劣る
(31) 随伴	ACCompaniment	～とともに，～に伴って
(32) 度合	DEGree	5％増加する，3キロやせる
(33) 陳述	PREdicative	～である
(34) その他	ETC	

Note: The capitalized letters are used as abbreviations

To write the semantic markers for nouns in the case frame of the verb lexicon, reference is made to the noun lexicon for these nouns.

Note that we write only the semantic markers for these nouns appearing in the context of our samples.

Kakujo-shi as surface cases and case labels as deep cases are described for *Yougen*. Then semantic markers for nouns preceeding to *Kakujo-shi* are described.

5. Semantic Markers for Nouns

This section describes what the system of semantic information for nouns is and what the concept of semantic markers is and how semantic

markers are attached to nouns.

5.1 System of semantical information for nouns

1) Study

In the primary stage of our study, we thought that all nouns were symbols to display the following concepts recognized by humans. We set up four concepts in heighest level; "Concrete objects", "Abstract concepts", "Phenomena", and "Human actions". Concrete objects are the selfsame objects in the world. Abstract concepts are the standards which fix intellectual activities of humankind. Phenomena include both social phenomena and natural phenomena. Human actions are the selfsame acted by humans. We assigned facets to these four concepts. Then we further extracted the feature of a part from these facets and assigned a new facet "Parts".

Similarly, another concept of "Attribute" was extracted from "Phenomena" and "Human actions". This feature is crucial especially for action nous. Thus we added two facets; "Parts" and "Attribute". Nouns also include concepts of measurement, space & topography and time. So we added three facets; "Measurements" , "Spaces & Topography", and "Time".

We classified into more concepts as follows.

The concept of concrete objects are classified into "Nations & Organizations", "Animate objects" and "Inanimate objects" which constitute three independent concepts. The concept of Human actions was classified two facets, "Sense & Feeling" and "Actions".

We called the scope formed by the concept "conceptual category". It is difficult to define the conceptual scope explicitly. The concept which can be defined explicitly in the conceptual category is called a facet. The facet is subclassified into a number of semantic slots. This relation is illustrated in figure 5.1.

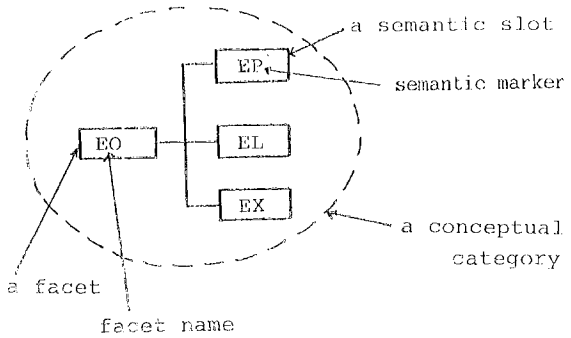


Figure 5.1 Relationship between Facet and Semantic Markers

2) Subclassification of Facets

Facets, for example, were subclassified into slots as follows, the facet of Animate objects was subclassified as semantic slots "humans", "animals", and "plants" . The facet of Phenomena was subclassified as slots "natural phenomena", "physical phenomena", "power and energies", "physiological phenomena", "social phenomena" and "social systems and customs". We then set up an "others" slot in each facet, for these words which cannot be assigned to any slot. The use of these slots is explained in section 5.3. We will study "others" slots through semantic analysis for nouns; new slots or facets may have to be assigned.

These semantic slots and facets are named semantic markers. The System of Semantic Markers for Nouns is shown in figure 5.2. The system of semantic markers for nouns is made up of 13 conceptual facets including "Others" markers, and 49 filial slots as terminals.

We also use Special Semantic Markers for "functional words" which represent some patterns, syntactic or semantic information. For example, the word "comparison" presupposes more than two nouns(arguments); comparison between "A" and "B". Then, "WK(Relation)" as a special semantic marker is attached to the word "comparison". The word "time" assumes time case. These features suggest an effective device for semantic analysis.

5.2 Concept of Semantic Markers

The following describes concepts of 12 facets in the System of Semantic Markers for Nouns ("Other"(ZZ) not included).

1) Nations and Organizations (OF)

This conceptual facet includes words related to such functional human groups as nations, parties, corporations and organizations. Words in this facet can occur with volitional verbs, when used as subjects.

2) Animate Objects(OV)

This conceptual facet includes such names as that of man, animal and plant. However, names of organs of the animate objects are included in the slot of "Organs or Components"(EL) under the facet of "Part". Names of diseases are included in the slot of "Physiological phenomena" (PB) under the facet of "Phenomena".

3) Inanimate Objects(OS)

This conceptual facet only includes words related to concrete objects in the inanimate objects, such as natural substances, parts and materials of products, artificial substances and institutions. The objects which do not exist as concrete objects are included in the facet of "Intellectual Objects".

4) Intellectual Objects(IO)

"IO" includes words related to theories, abstract tools and materials, intellectual products that are created by human intellectual activities.

5) Phenomena(PO)

"PO" includes words related to natural phenomena, physical phenomena, power & energies, physiological phenomena, social phenomena and systems/customs. Words having causal properties are attached to words under this facet using plus-minus signs ("+" and "-"). Sign "+" denotes desirable conditions(e.g. success), while sign "-" indicates undesirable conditions(e.g. suicide)

6) Sense and Feeling(SO)

"OS" includes words related to human mental phenomena such as feeling, reaction, recognition and thinking.

7) Actions(DO)

"DO" includes words related to human activities such as human actions and movements.

8) Parts(EO)

"EO" includes such words related to parts and components of concrete objects as parts, components and organs.

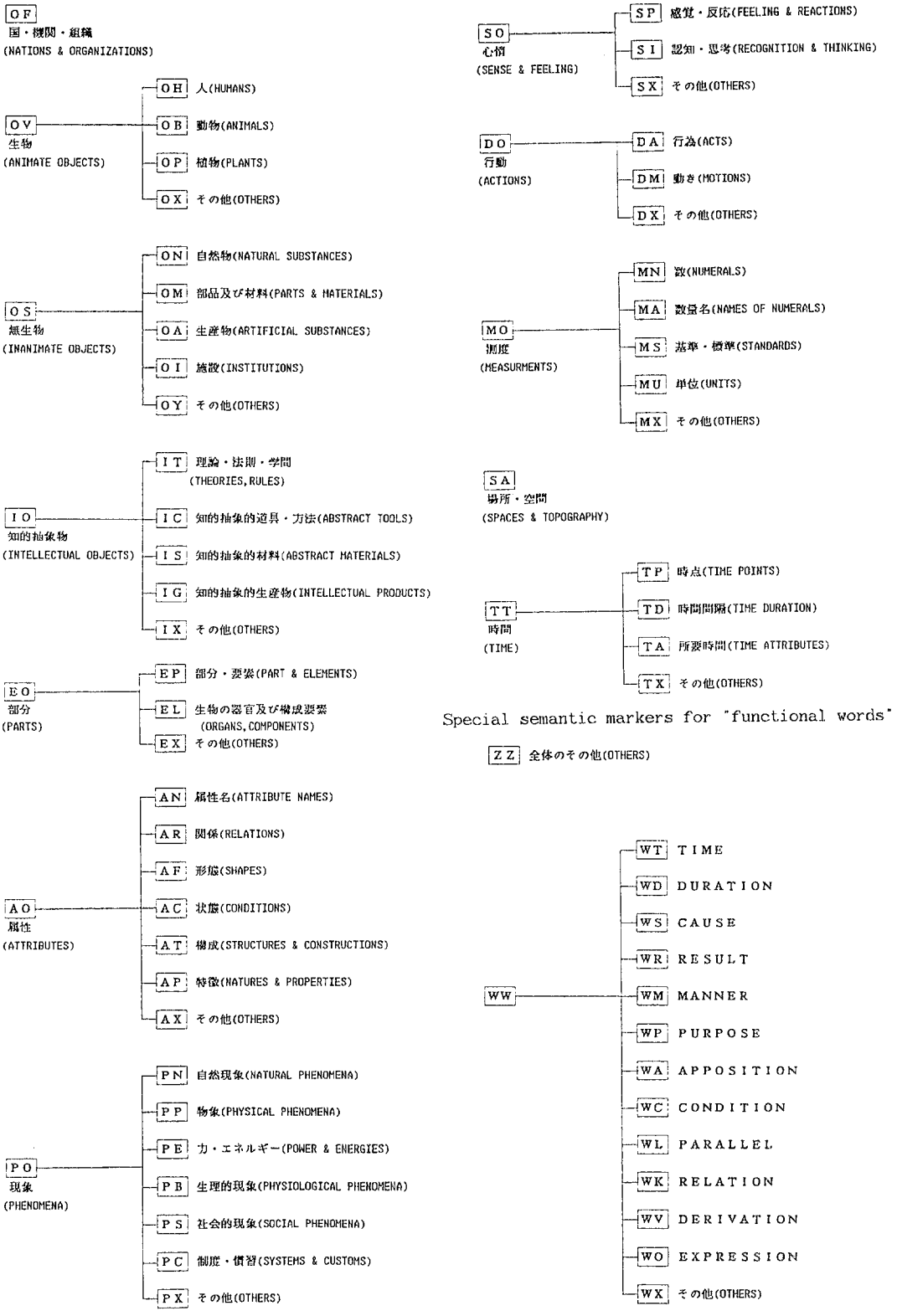


Figure 5.2 System of Semantic Markers for Nouns

9) Attributes(AO)

"AO" includes words related to attributes of concrete objects and abstract concepts. Their slots consist of attribute's names, and attribute's values with causal relations, shapes, structures, constructions and nature.

For example, the word "color" is the attribute's name(then, marker in AO), words such as "red" and "white" are attribute's value(AC).

10) Measurements(MO)

"MO" includes words related to numerals, name for numerals, standards, and units for measurement. Examples are "argument", "fee", "standards", and "kilometer".

11) Space and Topography(SA)

"SA" includes words referring to spatial extention of concrete objects and abstract concepts. Examples are direction, area, orbit and Brazil

12) Time(TT)

"TT" includes such words related to time points, time duration and time attributes, as "autumn", "for a week", "every day" and "life time".

5.3 How to attach semantic markers to words

The semantic markers for nouns are determined in the following steps.

1) Attach semantic markers to the following nouns.

Proper noun
Common noun
Action noun 1(Sahen-mei-shi)
Action noun 2(except action noun 1)
Adverbial noun(only when the words include the concept of "Time" or "Location")
Interrogative pronoun
Personal pronoun
Demonstrative pronoun(only when the words include the concept of "Location")

2) Attach semantic markers to the words according to the definition, semantic scope and examples given in the "definition table" of semantic markers.

3) Do not attach semantic markers to the following words:

Molecular formulas
Arithmetic expressions
Names of product models

4) If a word belongs to multiple slots in the same facet, attach all relevant markers.

5) If a word belongs to a facet but this word not belongs to any appropriate slot in the facet, attach "others" marker in this facet to that word.

6) If a word is equal to a facet name itself, attach the semantic marker of that facet name to the word.

7) If the concept of a word is not included in any facet, attach "Others" facet(ZZ) to that word.

8) For compound words consisting of more than one word, attach the markers putting into consideration semantic information of the compound words themselves; do not always attach the marker only to the last element of the compound.

6. Semantic Information for Adverbs

In our system, adverbs are subclassified as follows.

1) Adverb of condition(Joukyou fuku-shi)
2) Adverb of degree(Teido fuku-shi)
3) Adverb of statement(Chinjutsu fuku-shi)
4) Adverb of quantity(Suuryou fuku-shi)

Besides, the aspects of verbs are classified; "Mood", "Aspect", "Tense", and "Degree". They contrast specific adverbs. Then semantic information for adverbs is used to ensure more accurate translations. Semantic information for adverbs are defined according to the concept aspects as follows.

1) Semantic information on mood determined by the adverb of statement

Subjunctive(e.g. if), Interrogation(e.g. when), Negation(e.g. not always), Desirability(e.g. possibly), Entireness(e.g. entirely), Concession(e.g. kindly)

2) Semantic information on aspects of verbs determined by the adverb of condition

Completion(e.g. finally), Progression(e.g. rapidly), Repetition(e.g. repeatably), Convention(e.g. accordingly)

3) Semantic information on tense determined by adverb of condition and statement

Past(e.g. yesterday), Present(e.g. now), Future(e.g. tomorrow)

4) Semantic information on degree when the adverb or adjective can be modified by the adverb of degree and quantity

Scale(e.g. seriously), Degree(e.g. fairly)

7. Examples of Semantic Markers Used in Analysis

7.1 Determination of the Usage of Verbs by Case Patterns

Case patterns are used to determine the usage of verbs having broader concept. This is especially an effective method in determining the meanings of *Wago* verbs(basic Japanese verbs) having broader concepts like English verbs, "make", "do", "take", "put", etc.

We take *Wago* verb "当たる" as an example and show the difference of the meanings of verb "当たる" by mean of case pattern (a), (b), (c). Furthermore we show the semantic markers which co-occur to each case.

e.g. Wago verb 当たる(hit, strike, understand, treat, be engaged in, be equal, correspond, be appropriate, etc.)

(a) 当たる in the concept of ぶつかる, 接触する(hit, strike, reflex, collide, etc.)

Case pattern(a): A<object, physical phenomenon> が B<object, place> に ---(verb)

Ex.1 A<石(stone)> が B<硝子(glass)> に 当たる(A stone hits glass)
Ex.2 A<光(light)> が B<斜面(slope)> に 当たる(Light hits the slope)
Ex.3 A<電波信号(electric wave signal)> が B<山(mountain)> に 当たって反射する(An electric wave signal hits the mountain and reflexes)

(b) 当たる in the concept of 従事する(undertake, be engaged in, deal with, etc.)

Case pattern(b): A<object with "will"> が B<action> に---(verb)

Ex.1 A<巡視艇(patrol boat)> が B<救助(life-saving)> に 当たる(A patrol boat is engaged in life-saving)
Ex.2 A<A氏(Mr. A)> が B<指揮(command)> に 当たる(Mr. A is in charge of commanding)
Ex.3 A<A社(Company A)> が B<点検修理(inspection and repairs)> に 当たる(Company A deals with inspection and repairs)

(c) 当たる in the concept of 該当する(be equal, correspond, be appropriate, etc.)

Case pattern(c): A<human> が B<human> に---(verb)
A<place> が B<place> に---(verb)
A<time> が B<time> に---(verb)
A<measurement unit> が B<measurement unit> に---(verb)

A and B can take variable values, but should not take different values in the same sentence.

Ex. 1 サイゴン(Saigon)の B<銀座並木通り(Ginza Namiki Avenue)> に 当たる A<チュドー通り(Chudo Avenue)> (Chudo Avenue in Saigon corresponds to Ginza Namiki Avenue)
Ex. 2 A<今日((today)> が ちょうど(just) B<一年目(first year)> に 当たる (Today just corresponds to the first year)
Ex. 3 A<1インチ(1 inch)> が B<2.54cm> に 当たる(1 inch is equal to 2.54 cm)
Ex. 4 B<女のブラウス(women's blouses)> に 当たる A<開襟シャツ(open-necked shirts)> (Open-necked shirts corresponding to women's blouses as office wear)

In this way, we can determine the usage of verbs by means of case patterns and the semantic markers.

7.2 Interpretation of Optional Cases

One *Kakujo-shi* (surface case) often plays the role of different deep cases in Japanese. Often, various optional cases are included in this deep case. Each optional case is determined by the combination of *Kakujo-shi* (surface case) and the semantic marker of the noun which co-occurs with it in the dictionary. In the process of transfer, appropriate English prepositions must be specified according to each determined optional deep case.

For example, take *kakujo-shi* "で". The optional case is determined from the semantic marker of noun and the semantic marker which co-occurs with the surface case of the verb. Then the case frame in English is selected and the preposition "in" is determined. This process is shown as follows.

Ex. (1) Source sentence: 主として現在市場で活躍している数値制御超高速穴あけ機について説明。

Translated sentence: The numerically-controlled superhigh speed drilling machines are explained which are active mainly now in markets.

Explanation of Example (1): Let us observe the Japanese Analysis Dictionaries for Noun and Verb shown in Figure 7.1. The noun <市場 (market)> has semantic markers (SA and PS) according to Noun Dictionary, while verb <活躍する (be active)> has (SA or OF) for the noun according to case slot2 of case pattern in the verb dictionary. "Market" and "be active" match with each other with respect to semantic marker "SA". Thus, the surface case <で> in the Japanese input sentence is determined to have the surface case "SA", which corresponds to case label <場所> SPAce.

Ex. (2) Source sentence: 逆制動放射によるレーザ吸収項について二つの取り扱い方で問題を解き、数値解を求めた。

Translated sentence: Problems are solved by two handling methods about laser absorption terms by the inverse damping radiation, and numerical solutions are obtained.

Explanation of Example (2): Using these Dictionaries, the noun <取り扱い方 (handling methods)> has semantic markers (IC and AN) while verb <解く (solve)> has (DA, IT, IC or TS) for the noun preceding the *Kakujo-shi*. "handling methods" and "solve" match with each other with respect to "IC".

```
((＄見出し番号 "０００００００５６０５")
 (＄更新年月日 "８５０１２２")
→(＄見出し情報 (＄見出し語 "市場") (＄読み "しじょう"))
 (＄形態素情報 (＄形態品詞 名) (＄前接情報 1) (＄後接情報 1))
 (＄構文－意味情報
→(＄構文品詞 名詞)
  (＄品詞細分類 普通名詞)
  (＄名詞意味マーカ ＳＡ ＰＳ)))
```

Part of speech ↑ *Semantic marker*

Lexical unit

(a) Contents of the Noun Dictionary for <市場 (market)>

```
((＄見出し番号 "Ｖ０００７６００－０１")
 (＄更新年月日 "８５０４０４")
 (＄見出し情報                                  } Lexical information
   (＄見出し語 "活躍する")
   (＄語尾字数 2)
   (＄漢字部 1 2)
   (＄読み "かつやくする"))
 (＄形態素情報                                  } Morphological information
   (＄形態品詞 動)
   (＄動詞活用型 サ変)
   (＄前接情報 2)
   (＄後接情報 13))
 (＄構文－意味情報
   (＄分野コード 電気)
   (＄構文品詞 動詞)
   (＄格パターン                                ← Japanese case pattern
     Ｖ１
     (＄アスペクト 継続)
     (＄意志 無)
     (＄格支配情報                              ← Case slot1
       ((＄表層格 が)
Surface case →  (＄深層格 主体)
Deep case ——→   (＄名詞意味マーカ ＯＶ)          } ←
Semantic marker (＄必須性 1))
       ((＄表層格 で)
        (＄深層格 場所)
        (＄名詞意味マーカ ＳＡ ＯＰ)             } ←
        (＄必須性 0))))))                       ← Case slot2
```

(b) Contents of the Verb Dictionary for <活躍する (be active)>

Figure 7.1 Contents of the Japanese Analysis Dictionary

Thus, the surface case <で> in the Japanese sentence is determined to have the surface case "IC", which corresponds to case label < 手段・道具 > TOOl.

8. Examples of Semantic Markers Used in Transfer Process

In the transfer phase the English verb is chosen by examining the semantic markers for nouns filling the deep case slot of the condition part of the verb transfer dictionary.

Two examples of selecting the traslation for the Japanese verb, "含む" are as follows:

Ex. (1) Source sentence: ナトリウムおよびよう化スカンジウムを含む鉛直水銀アークの放射および対流模型。

Translated sentence: The radiation and convection models of the vertical mercury arcs which contain the sodium and the scandium iodide.

Ex. (2) Source sentence: 与えられた正規双直交系を含む正規双直交基底の構成法に関して、具体的な構成法を示すことにより補題を証明。

Translated sentence: Lemmas are verified by showing specific constitution methods about constitution methods of the normal double orthogonal bases which include given normal double orthogonal systems.

Explanation of Example (1):

Based on the Japanese to English Transfer Dictionary, Figure 8.1, both <ナトリウム (sodium)> and <よう化スカンジウム (scandium iodine)> in Example (1) have OM. According to the conditions of the dictionary, OM matches one of the semantic markers (OS OM PN PB PP PE) in correspond to the appropriate case slot (in this case <対象 (object case)>) of Japanese case frame in the dictionary for the verb <含む> (contain, include). So verb "contain" is selected.

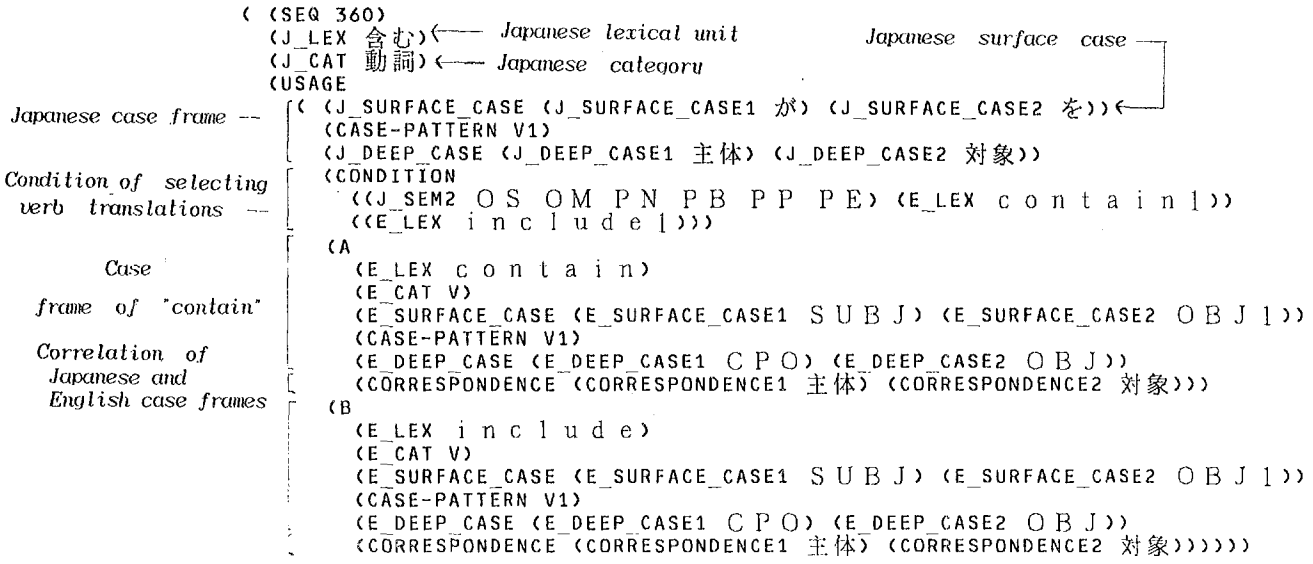


Figure 8.1 Contents of the Japanese to English Transfer Dictionary

As for Example(2), the semantic marker for < 正規双直交系 (normal double orthogonal system)> is IC, which does not match any of the semantic markers in the appropriate case slot (in this case < 対象 (object case)>) in the dictionary for the verb <含む>. Thus, verb "include", which is the default value of the English word, is selected.

9. Conclusion

1) When semantic markers are attached by human operation, several problems arise. The first problem is simple mistakes made by humans. The second problem is a fluctuation of semantic analysis due to a large amount of data. So it is necessary to develop an automatic marking system to save time and to improve efficiency.

2) When assigning semantic markers to nouns, we attached them without considering the relationship between nouns and verbs. That is, we attached semantic markers simply based on noun concepts. This is not adequate to handle nouns which are intrinsically related to verbs. One of the solutions to this problem will be to study the correlation matrix of the semantic markers for nouns in relation to the case frame of verbs.

3) Our system of semantic markers for nouns has been designed for Japanese nouns. We have to design a system of semantic markers for English nouns. Since recognition for its concept in an Enlish word is very difficult for the Japanese, we are also studying a method of evaluation test to handle these data.

4) Our system of semantic markers for nouns simply consists of the tree structure of facets and slots. Subclassification for these structures with deep tree structure is a significant problem in order to analyze the concept of nouns more in detail, but such a semantic marking operation will become more complex and difficult for 1), 2) problems.

5) We suppose that a concept system for words is not a static structure, but various semantic networks constructed dynamically according to a given story.

We must give thorough consideration to this prospective problems.

Acknowledgment

We would like to thank Prof. Makoto Nagao, Prof. Junichi Tsujii, Mr. Junichi Nakamura(Kyoto University), Mrs. Mutsuko Kimura (Institute for Behavior and Science), and Miss Masako Kume(Japan Convention Services, Inc.) and the other members of the Mu-project working group for the useful discussions which led to many of the ideas presented in this paper.

References

In English

(1) Sakamoto, Y., Satoh, M. and Ishikawa, T.: Lexicon Features for Japanese Syntactic Analysis in Mu Project-JE, COLING84, Stanford, 1984.

(2) Nagao, Makoto:Structural transformation in the generation stage of Mu Japanese to English machine translation system, Theoretical and methodological Issues in machine translation of natural language, Hamilton, New York, 1985.

(3) Tsujii, Jun-ichi:Science and Technology Agency's Machine Translation Project, Proceeding of the International Symposium on Machine Translation, JIPDEC, Tokyo, 1985.

In Japanese

(4) Nakai, H. and Satoh, M.; A Dictionary with Taigen as its Core, Working Group Report of Natural Language Processing in Information Processing Society of Japan, WGNL 38-7, July, 1983.

(5) Nagao, M.; Introduction to Mu-Project, WGNL 38-2, 1983.

(6) Sakamoto, Y.; Yougen and Fuzoku-go Lexicon in Verbal Case Frame, WGNL 38-8, 1983.

(7) Ishikawa, T., Satoh, M. and Takai, S.; Semantical Function on Natural Language Processing, Proc. of 28th CIPSJ, 1984.

(8) Nagao, M., Tsujii, J., Nakamura, J., Sakamoto, Y., Toriumi, T. and Satoh, M: Outline of machine traslation project of the Science and Technology Agency, Journal of Information Processing, Vol.26, No.10, 1985.