(data)

0. Abstract

In this paper, we focus on the features of a lexicon for Japanese syntactic analysis in Japanese-to-English translation. Japanese word order is almost unrestricted and Kokujo-shi (postpositional case particle) is an important device which acts as the case label (case marker) in Japanese sentences. Therefore, case grammar is the most effective grammar for Japanese syntactic analysis.

The case frame governed by Yomen and having surface case (Kokujo-shi), deep case (case label) and semantic markers for nouns is analyzed here to illustrate how we apply case grammar to Japanese syntactic analysis in our system.

The parts of speech are classified into 56 sub-categories.

We analyze semantic features for nouns and pronouns classified into sub-categories and we present a system for semantic markers. Lexicon formats for syntactic and semantic features are composed of different features classified by part of speech.

As this system uses LISP as the programming language, the lexicons are written as S-expression in LISP, punched onto tapes, and stored as files in the computer.

1. Introduction

The Mu-project is a national project supported by the STA (Science and Technology Agency), the full name of which is 'Research on a Machine Translation System: Japanese-English for Scientific and Technological Documents'.

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 consist of three phases: analysis, transfer, and generation.

In the first phase of machine translation, morphological analysis divides the sentence into lexical items and then proceeds with semantic analysis on the basis of case grammar in Japanese. In the second phase, transfer: lexical features are transferred and at the same time, the syntactic structures are also transferred by matching tree patterns from Japanese to English. In the final generation phase, we generate the syntactic structures and the morphological features in English.

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

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

As word order is heavily restricted in English syntax, ATNG (Augmented Transition Network Grammar) based on CFG (Context Free Grammar) is adequate for syntactic analysis in English. On the other hand, Japanese word order is almost unrestricted and Kokujo-shi play an important role as case labels in Japanese sentences. Therefore, case grammar is the most effective grammar for Japanese syntactic analysis.

In Japanese syntactic structure, the word order is free except for a predicate (verb or verb phrase) located at the end of a sentence. In case grammar, the verb plays a very important role during syntactic analysis, and the other parts of speech only perform in partnership with, and equally subordinate to, the verb.

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

3. Case Frame governed by Yomen

The case frame governed by Yomen and having Kokujo-shi, case label and semantic markers for nouns is analyzed here to illustrate how we apply case grammar to Japanese syntactic analysis in our system.

Yomen consists of verb, Kakujo-shi, adjectival and Kyoumatsu-shi, adjectival noun. Kokujo-shi include inner case and outer case markers in Japanese syntax. But a single Kokujo-shi corresponds to several deep cases: for instance, *'V!* indicates more than ten case labels including SPACE, SPACE TO, TIME, ROLE, MANNER, GOAL, PARTNER, COMPONENT, CONDITION, RANGE..... We analyze relations between Kokujo-shi and case labels and write them out manually according to the examples found out in sample texts.

* This project is being 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.
As a result of categorizing deep cases, 33 Japanese case labels have been determined as shown in Table 1.

Table 1. Case Labels for Verbal Case Frames

<table>
<thead>
<tr>
<th>Japanese Label</th>
<th>English Label Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 主体</td>
<td>Subject</td>
</tr>
<tr>
<td>(2) 客体</td>
<td>Object</td>
</tr>
<tr>
<td>(3) 与手</td>
<td>Recipient</td>
</tr>
<tr>
<td>(4) 与手</td>
<td>Origin</td>
</tr>
<tr>
<td>(5) 相手</td>
<td>Partner</td>
</tr>
<tr>
<td>(6) 相手</td>
<td>Opponent</td>
</tr>
<tr>
<td>(7) 時</td>
<td>Time</td>
</tr>
<tr>
<td>(8) 時・地点</td>
<td>Time-From, Time-TO</td>
</tr>
<tr>
<td>(9) 時・地点</td>
<td>Location</td>
</tr>
<tr>
<td>(10) 時間</td>
<td>Duration</td>
</tr>
<tr>
<td>(11) 場所</td>
<td>Space, Space-From, Space-To</td>
</tr>
<tr>
<td>(12) 場所</td>
<td>Site</td>
</tr>
<tr>
<td>(13) 場所・時点</td>
<td>Space-Time</td>
</tr>
<tr>
<td>(14) 場所・時点</td>
<td>Space-Time</td>
</tr>
<tr>
<td>(15) 始状態</td>
<td>Source</td>
</tr>
<tr>
<td>(16) 終状態</td>
<td>Goal</td>
</tr>
<tr>
<td>(17) 助格</td>
<td>Complement</td>
</tr>
<tr>
<td>(18) 明形・理由</td>
<td>Cause</td>
</tr>
<tr>
<td>(19) 手段・道具</td>
<td>Tool</td>
</tr>
<tr>
<td>(20) 材料</td>
<td>Material</td>
</tr>
<tr>
<td>(21) 場合</td>
<td>Context</td>
</tr>
<tr>
<td>(22) 方式</td>
<td>Manner</td>
</tr>
<tr>
<td>(23) 条件</td>
<td>Condition</td>
</tr>
<tr>
<td>(24) 目的</td>
<td>Purpose</td>
</tr>
<tr>
<td>(25) 役割</td>
<td>Role</td>
</tr>
<tr>
<td>(26) 内容規定</td>
<td>Constraint</td>
</tr>
<tr>
<td>(27) 規格</td>
<td>Range</td>
</tr>
<tr>
<td>(28) 提案</td>
<td>Proposal</td>
</tr>
<tr>
<td>(29) 視点</td>
<td>Viewpoint</td>
</tr>
<tr>
<td>(30) 比較</td>
<td>Comparison</td>
</tr>
<tr>
<td>(31) 伴属</td>
<td>Accompaniment</td>
</tr>
<tr>
<td>(32) 度</td>
<td>Degree</td>
</tr>
<tr>
<td>(33) 質</td>
<td>Quality</td>
</tr>
</tbody>
</table>

Note: The capitalized letters form the English acronym for that case label.

When semantic markers are recorded for nouns in the verbal case frames, each noun appearing in relation to とく and とくに in the sample text is referred to the noun lexicon.

The process of describing these case frames for lexicon entry are given in Figure 1.

For each verb, とく and とくに and case labels able to accompany the verb are described, and the semantic marker for the noun which exist antecedent to that とくに are described.

4. Sub-categories of Parts of Speech according to their Syntactic Features

The parts of speech are classified into 13 main categories:

- Nouns, pronouns, numerals, affixes, adverbs, verbs, とく and とくに, とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers.

Action nouns are classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers. Action nouns can be classified into とく with conjunction feature, unknown nouns, mathematical expressions, special symbols and complementizers.

43
Adverbs are divided into 4 sub-categories for modality, aspect and tense. In Japanese, the adverb agrees with the auxiliary verb. *Chinatsu-fuku-shi* agrees with aspect and tense.

*Teito-fuku-shi* agrees with gradability.

Auxiliary verbs are divided into 5 sub-categories based on modality, aspect, voice, cleft sentence and others.

Verbs may be classified according to their case frames and therefore it is not necessary to sub-classify their sub-categories.

5. Semantic Marking of Nouns

We analyze semantic features and assign semantic markers to Japanese words classified as nouns and pronouns. Each word can give five possible semantic markers.

The system of semantic markers for nouns is made up of 10 conceptual facets based on 44 semantic slots and 38 plural filial slots at the end (see Figure 2).

5.1 Concept of semantic markers

The 10 conceptual facets are listed below.

1) Thing or Object

This conceptual facet contains things and objects: that is, actual concrete matter. This facet consists of such semantic slots as Nation/Organization, Animate object, Inanimate object, etc.

2) Commodity or Ware

This conceptual facet contains commodity and wares: that is, artificial matter useful to humans. This facet consists of such semantic slots as Material, Means, Equipment, Product, etc.

3) Idea or Abstraction

This conceptual facet contains ideas and abstractions: that is, non-matter as the result of intellectual activity in the human brain. This facet consists of such semantic slots as Theory, Conceptual object, Sign-Symbol, etc.

4) Part

This conceptual facet contains parts: that is, structural parts, elements and contents of things and matter.

Figure 2. System of Semantic Markers for Nouns
5 Attribute
This conceptual facet contains attributes; that is, properties, qualities or features representative of things. This facet consists of semantic slots such as Property Characteristic, Status Figure, Relation, Structure, etc.

6 Phenomenon
This conceptual facet contains phenomena; that is, physical, chemical and social actions without human activity. This facet consists of semantic slots such as Natural phenomenon, Artificial phenomenon, Power Energy, etc.

7 Doing or Action
This conceptual facet contains human doing and actions. This facet consists of such semantic slots as Action Deed, Movement Reaction, Effect Operation, etc.

8 Mental activity
This conceptual facet contains operations of the mind and mental process. This facet consists of semantic slots such as Perception, Emotion, Recognition Thought, etc.

9 Measure
This conceptual facet contains attributes; that is, the extent, quantity, amount or degree of a thing. This facet consists of semantic slots such as Number, Unit, Standard, etc.

10 Time and Space
This conceptual facet contains space, topography and time.

5.2 Process of semantic marking
The semantic marker for each word is determined by the following steps.
1) Determine the definition and features of a word.
2) Extract semantic elements from the word.
3) Judge the agreement between a semantic slot concept and extracted semantic element word by word. and attach the corresponding semantic markers.
4) As a result, one word may have many semantic markers. However, the number of semantic markers for one word is restricted to five. If there are plural filial slots at the end, the higher family slot is used for semantic featurization of the word.

It is easy to decide semantic markers for technical and specific words. But, it is not easy to mark common words, because one word has many meanings.

6 Lexicon Formats for Syntactic Analysis
Lexicon formats for syntactic and semantic features are composed of different features classified by part of speech.

1) Features of verb:
   Subject code: verb used in specific field.
   Part of speech in syntax: verb
   Verb pattern: classifying the verbal case frame. A categorized marker like Hunny's case pattern is planned to be used.

2) Features of Keiou-shi and Keioudu-shi:
   Features of Keiou-shi and Keioudu-shi: both syntactic features are described in the same format.
   Sub-category of noun: proper, common, action, adverbial, etc.
   Lexical unit for transfer lexicon, semantic markers, thesaurus code, and usage.

3) Features of adverb:
   Sub-category of adverb (Jodai-shou, Teido, Chinjitsu, Syoun) considering modality, aspect, tense and gradability

4) Features of Keiou-shi:
   Sub-category of Keiou-shi (demonstrative, interrogative, definitive, or adjectival) and conjunction (phrase or sentence)

5) Features of Jodai-shou (auxiliary verb):
   Jodai-shou are sub-classified by sub-category on semantic feature:
   Aspect: past, perfect, perfective, stative, progressive, continuing, experimental...
   Voice: passive, or causative
   Cleft sentence (purpose and reason)

6) Features of Jo-shi:
   Sub-category of Jo-shi: case, conjunctive, adverbial, collateral, final or Junitori
   Case: features of surface case (ex. ‘GA’ ‘WO’ ‘N’ ‘TU’ ...), modified relation (Roni or Ryou modification).
   Conjunctive: sub-category of semantic feature (cause reason, conditional/provisional, accompaniment, time/place, purpose, collateral, positive or negative conjunction, etc)

7) Data Base Structure of the Lexicon
As this system uses LISP as the programming language, the lexicons are punched up as...
S-expressions and input to computer files (see Figure 3).

For the lexicon data base used for syntax analysis, only the lexical items are held in main storage; syntactic and semantic features are stored in VSAM random access files on disk (see Figure 4).

\[ (S\text{-expression ("実験サイト")}) \]

\[ (S\text{-expression ("合計する")}) \]

\[ (S\text{-expression ("した")}) \]

\[ (S\text{-expression ("こと")}) \]

\[ (S\text{-expression ("あらかじめ")}) \]

\[ (S\text{-expression (" Поэтому")}) \]

\[ (S\text{-expression ("できる")}) \]

The head character of the lexical unit is used as the record key for the hashing algorithm to generate the addresses in the VSAM files.

### 8. Conclusion

We have reached the opinion that it is necessary to develop a way of allocating semantic markers automatically to overcome the ambiguities in word meaning confronting the human attempting this task.

In the same thing, there are problems how to find an English term corresponding to the Japanese technical terms not stored in dictionary, how to collect a large number of technical terms effectively and to decide the length of compound words, and how to edit this lexicon data base easily, accurately, safely and speedily.

In lexicon development for a huge volume of Yougen, it is quite important that we have a way of collecting automatically many usages of verbal case frames, and we suppose it exist different case frames in different domains.

### Acknowledgment

We would like to thank Mrs. Mutsuko Kimura (IBS, Toyo Information Systems Co. Ltd., Japan Convention Service Co. Ltd.) and other members of the Mu-projecl working group for the useful discussions which led to many of the ideas presented in this paper.

### References

### Parts of Speech in Japanese Syntax

<table>
<thead>
<tr>
<th>P.O.S.</th>
<th>Japanese</th>
<th>English</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>名</td>
<td>名詞</td>
<td>Noun</td>
<td>(組織、人名、地名等)</td>
</tr>
<tr>
<td></td>
<td>形容詞</td>
<td>Adjective</td>
<td>(長さ、数、感情、命令、評価、感動等)</td>
</tr>
<tr>
<td></td>
<td>動詞</td>
<td>Verb</td>
<td>(動作)</td>
</tr>
</tbody>
</table>

The table above lists various parts of speech in Japanese, along with their English equivalents and examples of usage. The columns include P.O.S. (Parts of Speech), Japanese, English, and Usage. The usage examples provide context for how these parts of speech are used in sentences.