JAPANESE-ENGLISH TRANSLATION THROUGH INTERNAL EXPRESSIONS

Fujio Nishida   and   Shinobu Takamatsu

Faculty of Engineering, University of Osaka Prefecture
Sakai, Osaka, 591   JAPAN

This paper describes an approach to Japanese-English translation through internal expressions which are similar to those used in our recent approach to English-Japanese translation [2]. Attention is focused on construction of the internal expressions of Japanese sentences based on case structures of predicates and also conversion of the Japanese internal expressions to the English ones for generating good English sentences in conventional use. Finally, associated with translation, extraction of specified translated information from Japanese patent claim sentences is described briefly.

1. CASE STRUCTURES AND PARSING

In every Japanese sentence a predicate such as the main verb is located at the end of the sentence and takes a part of the governor of the preceding dependants as. follows:

$$t_1 \ j_1 \ t_2 \ j_2 \cdot \cdot \cdot \cdot t_n \ j_n \ p \cdot$$

In the above t_i (i=1,2,···,n) denotes a term depending on the predicate p and j_i denotes a postposition or a case suffix such as "Kakujoshi" and other auxiliary words.

Japanese is an allutinative language. A postposition j_i of a term t_i clarifies the syntactic role of the term t_i in a sentence governed by the main verb. Conversely, a set of these postpositions or syntactic case labels determines the syntactic function of the main verb in a sentence and is called a verb pattern. The verb pattern plays an important role to identify the syntactic structures of sentences in Japanese like Hornby's verb patterns in English.

From a semantic view point, Japanese predicates are classified to about 20 main categories and case structures are defined based on the verb categories and the verb patterns. The case labels and the categorical names used here are chosen the same as those of English case structures in the previous paper [2].

Table 1 shows case frames constructed on the verb pattern VP9 of a verb which takes syntactic cases of SUBJect, OBJect and DATIVE accompanied with representative postpositions "ga", "wo" and "ni" respectively.

Table 2 shows an instance of description of a verb in the word dictionary constructed on the case frame basis. The first and second columns show the names of verb patterns and semantic main categories taken by the verb respectively. They give the token of the case frame of the verb. The third and forth columns show the equivalents of the verb and the names of their Hornby's verb patterns. If there are some equivalents in a row of categories, some key tokens such as pairs of a case-label and a subcategory taken by their dependants are added in order that a better equivalent can be chosen.

The next example shows some Japanese sentences which have the main verb illustrated in Table 2, the internal expressions and the corresponding English sentences , where the postpositions in Japanese sentences are enclosed with parentheses, and the symbol " * " in the internal expressions denotes the term which is in front of the frame including the symbol.

Table 1 Case Frames of the Verb Pattern VP9

VP9	SUBJ (ga)	OBJ (wo)	DATIVE (ni)
PRED-PTRANS	AG-LIVING	OBJ-PHYS·OBJ	GO-PHYS·LOC ∪ PHYS·OBJ
PRED-POSS·TRANS	AG-HUM	OBJ-OBJ	RECIP-HUM
PRED-MTRANS	AG-HUM	OBJ-MENT·OBJ	RECIP-HUM
PRED-ASSOC·ACT	AG-HUM	OBJ-HUM	PARTIC-HUM

Table 2 A Part of the Item " 増す " in the Word Dictionary

Verb patterns	Categories	English equivalents (and the designated conditions)	Hornby's verb patterns
VP1 : SUBJ	ATTR·TRANS	increase (OBJ-NUMBer ∪ AMOUNT) gain (OBJ-FORCE ∪ WEIGHT) :	VP2A VP2A,B,C :
VP8 : SUBJ +OBJ	ATTR·TRANS	increase (OBJ-NUMBer ∪ AMOUNT) augment (OBJ-INCOME ∪ POWER) promote (OBJ-HEALTH) extend (OBJ-MENT·OBJ ∪ INFLUENCE ∪ TERRITORY) :	VP6A VP6A VP6A VP6A :

Example 1
(a) VP1 (1)　自動車　　の　　　数　　　が　　急速に　　増している

cars -(no) number -(ga) rapidly is increasing

(2) (PRED-ATTR·TRANS: increase, TENSE: present, ASPECT: progressive,
OBJ-NUMBer: number(NUM: * , OBJ-PHYS·OBJ: cars), MANN: rapidly)

(3) The number of cars is rapidly increasing.

(b) VP8 (1)　この　　辞書　　　の　　　語い　　　は　　　　5万語

this dictionary -(no) vocabulary -(wa)   fifty thousands words

に　　増された
-(ni)   was extended

(2) (PRED-ATTR·TRANS: extend, TENSE: past, VOICE: passive, OBJ-MENT·OBJ:
vocabulary(OBJ: *₁ , LOC: this dictionary), GO-QUANT: words(UNIT:
*₂ , NUM: fifty thousands))

(3) The vocabulary in this dictionary was extended to fifty thousands
words.

Parsing is performed by using the case frame. First, all the words involved in
a sentence are retrieved from the dictionary. Then all the possible parsings
under respective frame conditions are carried out in parallel from the left of the
sentence. The case labels of dependants are determined by referring to the syn-
tactic and semantic categories of the dependants and the case frame of the
governor.

The internal expressions sometimes lack the agent case or the object case cor-
responding to the original Japanese sentence. When the internal form has a pred-
icate affecting on some objects and lacks the agent in the active voice, it is
expanded into an English sentence in the passive voice. When the main predicate
belongs to category of state such as existence and attribute and lacks the object
case, an appropriate pronoun to the context such as "it", "we" or "they" is sub-
stituted for the thematized object case as the general environment or the general
experiencer.

Japanese noun words are not obliged to have articles which give a measure of
definiteness to an object indicated by a noun word. Since English noun words take
determiners such as articles obligatorily in many cases, the articles must be re-
stored from the context of the Japanese sentences in construction of their inter-
nal expressions. The restoration is carried out in many cases by anaphoric
analysis and reference to information about the conventional use of articles de-
scribed for every noun phrase in the word dictionary.

2. CASE CONVERSION

The case structures of Japanese and English are partly different from each other at the level of the internal expressions currently used. Though constituents of case structures such as categorical names and case labels are chosen the same in both languages, there are different expressions due to logically possible combinations of a governor and the dependants for the same event or the same action, and the preferable combination depends on the individual language.

On the other hand, it is desired that a thematic term is kept unchanged through translation. Syntactically the thematic term takes the front part of a sentence in both English and Japanese, and furthermore, takes the subject case in English.

Hence if the internal expression obtained from a Japanese sentence does not satisfy these conditions, some conversions are tried to yield a more suitable English internal expression. Some of them are shown in the following.

2.1 Conversion of Existence and Attribute Expression

As well known, Japanese is a BE language while English is a HAVE language. In Japanese the possessive expression is almost confined to the case where a human has something in his hand, and the other possessive expressions in English are generally described in existence and attribute expressions.
Example 2
  Let us consider the following sentences:
 (a) He has a daughter.   (b) Copper has high electric conductivity.
  They are usually expressed in Japanese as follows:
 (a) (1) 彼   には   一人 の   娘   が   いる

     he -(ni wa) a -(no) daughter -(ga) is
    (2) (PRED-EXIST: is, OBJ-HUM: a daughter, LOC-HUM: <u>with him</u>)
    (3) <u>With him</u> is a daughter.
 (b) (1) 銅   は   導電率   が   高い

     copper   -(wa)   electric conductivity -(ga) is high
    (2) (PRED-ATTR: is high, OBJ-PHYS·QUANT: electric conductivity,
       LOC-PHYS·OBJ: <u>as for copper</u>)
    (3) <u>As for copper</u> electric conductivity is high.

In each illustration above, (1), (2) and (3) denote a Japanese sentence, the internal expression with terms replaced with English equivalents and a literal translation obtained by rewriting the internal expression respectively. The part with a double underline denotes a thematic part in the source sentence.

As seen from the above the literal translation does not preserve the thematic term of the original Japanese sentence within the standard English sentential form.

The translation with the thematic term unchanged requires replacement of the main predicate by HAVE verb and the accompanied case conversion as shown in Table 3. Rules 1 and 2 are conversion rules of Japanese EXISTence and ATTRibute expressions to English POSSessive expressions.

Table 3 Case Conversion between BE and HAVE Verbs

Rule 1	Japanese	(PRED-EXIST: be , OBJ-THINGS: t_1 , LOC-PHYS·OBJ ∪ PHYS·LOC: t_2 $)^1$
	English	(PRED-POSS: have , OBJ-THINGS: t_1 , POSSESSOR-PHYS·OBJ ∪ PHYS·LOC:^{1} t_2)
Rule 2	Japanese	(PRED-ATTR: t_0 , OBJ-PHYS·QUANT: t_1 , LOC-PHYS·OBJ: t_2)
	English	(PRED-POSS: have , POSSESSOR-PHYS·OBJ: t_2 , OBJ-PHYS·QUANT: t_1 (PRED-ATTR: t_0 , OBJ: *))

Example 3

The applications of the above rules to the internal expressions in Example 2 yield the following expressions.

(a) (PRED-POSS: has, POSSESSOR-HUM: he, OBJ-HUM: a daughter)
(b) (PRED-POSS: has, POSSESSOR-PHYS·OBJ: copper, OBJ-PHYS·QUANT: electric
conductivity(PRED-ATTR: be high, OBJ: *))

These internal expressions can be rewritten to the sentences (a) and (b) near the heading of Example 2.

2.2 Conversion of State-Oriented Expressions

Japanese is a state-oriented language and often takes a type of description that thing A changes to thing B owing to thing C in an event.· If C is a non-living object, C is interpreted as a cause or an instrument rather than an agent from the standpoint of translation, and usually does not take the subject case even if C is emphasized.

On the other hand, English is an action-oriented language and often uses expressions such that thing C makes thing A change to B even if the category of C is a non-living object.

The following shows some conversion rules between the above expressions.

Rule 3   (PRED: t_0 , OBJ: t_2 , MEANS∪CAUSE∪INSTR: $\underline{t}_1$)
$\longrightarrow$ (PRED: t_0' , AG: $\underline{t}_1$, OBJ: t_2)

Rule 4   (PRED: t_0 , MODAL: capable, AG: t_2 , [OBJ: t_3],
MEANS∪CAUSE∪INSTR: $\underline{t}_1$)
$\longrightarrow$ (PRED-ENABLE: t_c , AG: $\underline{t}_1$, OBJ:
(PRED: t_0 , AG: t_2 , [OBJ: t_3]))

, where the left sides of the rules are Japanese case structures and the right sides are those of English, categorical names are omitted except the case particularly required, and the contents enclosed with brackets denote some optional items. t_0' is a transitive verb such as VP6A corresponding to an intransitive verb or a transitive verb in the passive voice t_0 in Rule 3.

Example 4

(1)   この   信号   により   機械   が   確実に   始動する

this signal   -(niyori)   machine   -(ga)   reliably   start

(2) (PRED-PTRANS: start, OBJ-PHYS·OBJ: machine, INSTR-THINGS: this signal,
MANN: reliably)

(3) By this signal the machine starts reliably.

(4) (PRED-PTRANS: start, VOICE: active, AG-THINGS: this signal, OBJ-PHYS·OBJ:
machine, MANN: reliably)

(5) This signal starts the machine reliably.

3. SOME TRANSLATION RESULTS

Along the line described in the preceding sections, some experiments on translation have been carried out in an interactive mode. The average translation time excluding word retrieval and interaction is about 0.35 seconds per a Japanese word. Some experimental results are shown in this section, where names of categories in the internal expressions are omitted for simplicity.

Example 5

(A) Input Japanese sentences

2つ[1] の   信号[2] を   比較する[3]   ことにより、   流体[4] が   加えている[5]   力[6]
の   目安[7] が   与えられ[8]   次いで[9]   電子回路[10] で   この[11]   測定値[12] が
流量[13] の   尺度[14] に   変換される[15]

(B) The internal expression

(PRED:give[8], VOICE:passive, OBJ:measure[7](DET:indef, NUM:SINgular, CHAR:*[1], OBJ:
force[6](DET:indef, NUM:UnCountable, OBJ:*[2])(PRED:exert[5], AG:fluid[4](DET:indef,
NUM:UC, OBJ:*[3]), OBJ:*[2])), MEANS:(PRED:compare[3], AG:__, OBJ:signal[2](DET:indef,
NUM:two[1], OBJ:*[4])))   , and then[9]

②(PRED:convert[15] VOICE:passive, INSTR:electronic circuit[10](DET:indef, NUM:SIN,
OBJ:*5), OBJ:measured value[12](DET:this[11], NUM:UC, OBJ:*6), GO:scale[14](DET:indef,
NUM:SIN, CHAR:*7, OBJ:flow rate[13](DET:indef, NUM:SIN, OBJ:*8)))

(C) The output English sentences

①′Comparing two signals gives a measure of force exerted by fluid, and then an
②′electronic circuit converts this measured value to a scale of a flow rate.

In the above translation, Rule 3 was applied to the underlined parts ① and
② in paragraph (B) to yield ①′ and ②′ in paragraph (C) respectively in order to
keep the thematic or emphatic terms unchanged.

Example 6

(A) Input Japanese patent claim sentences

仕事関数[1] が 高い[2] 金属電極[3] と、 その[4] 金属電極[5] 上の セレン層[6] と、
その[7] セレン層[8] 上に 在り[9] 格子定数[10] が セレン[11] と 整合した[12]
結晶性[13] 半導体層[14] と、 その[15] 半導体層[16] 上に 形成した[17] 金属電極[18]
より 成る[19] 半導体装置[20]   (a) [_____↑_____] (b)
─────────↑ (c)

(B) The internal expression

semiconductor device[20](DET:indef, NUM:SINgular, OBJ:*1)(PRED:comprise[19], OBJ:*1,
SO: { metal electrode[3](DET:indef, NUM:SIN, OBJ:*2)(PRED:be high[2], OBJ:work func-
        tion[1](DET:indef, NUM:SIN, OBJ:*3), LOC:*2),   ①
    selenium layer[6](DET:indef, NUM:SIN, OBJ:*4, LOC:metal electrode[5](DET:def,
        NUM:SIN, OBJ:*5)),
    semiconductor layer[14](DET:indef, NUM:SIN, OBJ:*6, CHAR:crystal[13])
        (PRED:exist[9], OBJ:*6, LOC:selenium layer[8](DET:def[7], NUM:SIN, OBJ:*7))
    ②(PRED:conform[12]to, OBJ:lattice constant[10](DET:indef, NUM:SIN, OBJ:*8),
        PARTIC:selenium[11](DET:indef, NUM:UnCountable, OBJ:*7), LOC:*6),
    metal electrode[18](DET:indef, NUM:SIN, OBJ:*10)(PRED:form[17], VOICE:active,
        AG:_, OBJ:*10, LOC:semiconductor layer[16](DET:def[15], NUM:SIN, OBJ:*11)) ③ })

(C) The output English sentences

A semiconductor device comprising
    ①′ a metal electrode having a high work function,
       a selenium layer on the metal electrode,
       a crystal semiconductor layer on the selenium layer
       ②′ having a lattice constant which conforms to selenium, and
    ③′ a metal electrode formed on the semiconductor layer.

In parsing the system asked whether the underlined part (a) in paragraph (A)
depended on the parts (b) or (c) and obtained the answer from the user. For con-
struction of a better English internal expression, Rule 2 was applied to the
attribute expressions of the underlined parts ① and ② in paragraph (B) to yield
the possessive expressions ①′ and ②′ in paragraph (C) respectively. Furthermore,
the active voice expression of the underlined part ③ in paragraph (B) which lacks
the AGent case was converted into the passive voice expression shown in the under-
lined part ③′ in paragraph (C).

4. EXTRACTION OF SPECIFIED TRANSLATED INFORMATION

By using a method similar to the above case conversion, specified structural
information written in English can be extracted from Japanese texts in parallel
with their translation. Each specification table for information extraction used
here consists of several case frames, and each case frame has a standard case
structure corresponding to a simple sentence or a phrase. Table 4 is an example

of specification tables used for patent claim sentences.

Table 4   A Specification Table for Semiconductor Devices

COMPOSITION:	(OBJ-Device: $\underline{t}$, COMPonent: $\underline{t_i}$)
PROCESS:	(PRED-ACT:__, OBJ-D: $\underline{t \cup t_i}$, AG:__, INSTR:__, SO:__, GO:__,
	LOC:__, MANN:__, MEANS:__)
QUALITY:	(PRED-ATTR:__, OBJ:__, LOC: $\underline{t \cup t_i}$, COMPAR:__, PARTIC:__,
	DEGR:__)
LOCATION:	(OBJ: $\underline{t_i}$, LOC-on: $\underline{t_j}$)

The internal expression obtained by parsing is standardized according to the normal form of the frame determined by the category of the predicate. The standardization consists of case structure conversions such as clausal to phrasal structure conversion by removal of a kind of copula predicate and also case-set conversion such as (OBJ,USED) versus (INSTR,OBJ).

Table 5 shows the information extracted from the internal expression (B) in Example 6 by the specification table shown in Table 4. The extracted information is moved to a relational data base and used for relational retrieval and others.

Table 5   The Information Extracted from (B) in Example 6

COMPOSITION

OBJ	COMP
semiconductor device	metal electrode1
semiconductor device	selenium layer
semiconductor device	semiconductor layer
semiconductor device	metal electrode2

QUALITY

PRED	OBJ	LOC	PARTIC
high	work function	metal electrode1	—
conform	lattice constant	semiconductor layer	selenium

LOCATION

OBJ	LOC-on
selenium layer	metal electrode1
semiconductor layer	selenium layer

PROCESS

PRED	OBJ	LOC
form	metal electrode2	semiconductor layer

5. CONCLUSION

Japanese and English are fairly different languages from each other. However, if the object field of processing is confined to some technical fields, it is expected from the above consideration that semi-automatic multilingual translation and extraction of specified structural information are realizable though there are left various refinement problems such as restoration of articles.

REFERENCES

[1] Nagao, M., Tsujii, J., Mitamura, K., Hirakawa, H. and Kume, M., A machine translation system from Japanese into English - another perspective of MT system - , COLING-80, (1980) 414-423.

[2] Nishida, F., Takamatsu, S. and Kuroki, H., English-Japanese translation through case-structure conversion, COLING-80, (1980) 447-454.

[3] Nishida, F. and Takamatsu, S., Structured-information extraction from patent -claim sentences, Information Processing & Management, (1982) (to appear).