

The Transfer Phase of the Mu Machine Translation System

Makoto NAGAO

Jun-ichi TSUJII

Department of Electrical Engineering

Kyoto University

Kyoto, Japan 606

1 Introduction

The interlingual approach to MT has been repeatedly advocated by researchers originally interested in natural language understanding who take machine translation to be one possible application. However, not only the ambiguity but also the vagueness which every natural language inevitably has leads this approach into essential difficulties. In contrast, our project, the Mu-project, adopts the transfer approach as the basic framework of MT. This paper describes the detailed construction of the transfer phase of our system from Japanese to English, and gives some examples of problems which seem difficult to treat in the interlingual approach.

The basic design principles of the transfer phase of our system have already been mentioned in (1) (2). Some of the principles which are relevant to the topic of this paper are:

- (a) Multiple Layer of Grammars
- (b) Multiple Layer Presentation
- (c) Lexicon Driven Processing
- (d) Form-Oriented Dictionary Description

This paper also shows how these principles are realized in the current system.

2 The Construction of the Transfer Phase

The transfer phase of our system consists of the following three sub-phases (Fig. 1).

- Pre-Transfer Phase (Pre-TP)
- Main Transfer Phase (MTP)
- Post-Transfer Phase (Post-TP)

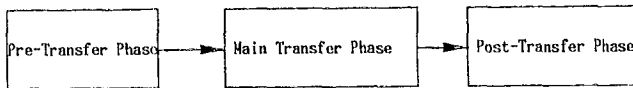


Fig. 1. Construction of the Transfer Phase

Analysis results of input sentences of the source language (SL) are represented in the form of annotated tree structures, from which one can retrieve various levels of information contained in the input sentences. The MTP is performed recursively from the top of the tree to the bottom (Fig. 2). At each step of the recursion, the sub-structure governed by a node (current node, node X in Fig. 2) is transferred to the corresponding structure of the target language (TL). This phase mainly performs lexical transfer, that is, it chooses an appropriate target lexical item for node X by examining the annotation parts of that node and its dependent nodes (nodes Y₁, .., Y₂, .., Y_n). The selection of a target lexical item may place certain constraints on the transfer of the dependent nodes. These constraints are expressed as property-value pairs and added to the

annotation parts of the dependent nodes, which are to be utilized in the subsequent recursion steps. For language pairs such as Japanese and English which belong to quite different language families, however, the lexical transfer is not so straightforward. It often happens that single lexical items of SL correspond to complex expressions of TL and vice versa. Furthermore, certain structural changes are also required.

Because the MTP recursively transfers sub-structures governed by single nodes, certain global structural differences cannot be naturally treated. Such global changes are dealt with by the Pre-TP and Post-TP sub-phases (See section 7).

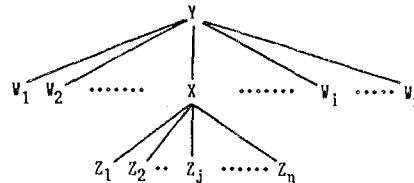


Fig. 2. Data Structure in the Transfer Phase

Japanese Lexical-Item	Translation Equivalent	Field Code	Part-of-Speech	
Column-A				
Japanese Surface Marker	Deep Case Relation	English Surface Marker	Translation Equivalent	Part-of-Speech
Column-B				
Japanese Surface Marker	Deep Case Relation	English Surface Marker	Translation Equivalent	Part-of-Speech
Column-C				
Japanese Modifier	Translation Equivalent	Part-of-Speech		
Column-D				
Japanese Predicate	1	2	3	4
JSC				
JDC				
MOD				
English Predicate				
ESC				
EDC				
JDC				

Fig. 3. Form-oriented Dictionary for Nouns (Bi-lingual Dictionary)

3 The Form-Oriented Dictionary and Lexical Transfer of Nouns

The following four columns are provided in the form-oriented bilingual dictionary of nouns (Fig. 3). The description of each column is expanded into corresponding types of lexical rule (i.e. GRADE rules which are stored in the dictionaries and invoked at specified processing phases: GRADE is the programming language for writing grammar rules used in our project).

(1) Column A: Transfer based on the Relationship to the Governor (Fig. 4)

Even major parts-of-speech such as nouns need not correspond between two languages. Certain Japanese nouns correspond to words of different parts-of-speech in English, depending on the surrounding context.

Ex-1 木 の 机 ---> wooden desk
 (tree) (desk)
 対話形式 での 入力 ---> interactive input
 (interaction) (input)

However, the same nouns should be translated as nouns in English in contexts such as:

Ex-2. 木 の 成長 ---> growth of trees
 (tree) (growth)
 対話形式 の 利点 ---> advantage of interaction
 (interaction) (advantage)

Japanese Lexical-Item	Translation Equivalent	Field Code	Part-of-Speech
木	tree	GEN	N

Column-A

Japanese Surface Marker	Deep Case Relation	English Surface Marker	Translation Equivalent	Part-of-Speech
の	MATerial	null	wooden	ADJ

This indicates that:

if '木' is used as a noun modifier
 and the case relation between '木' and the governor is 'MATerial'
 then the translation equivalent is 'wooden', else 'tree'.

Fig. 4. Example of Descriptions in Column-A

(2) Column B: Selections of Prepositions (Fig. 5)

Certain nouns determines the prepositions which relate their dependents of certain deep cases to them, as predicates (verbs) usually do.

Ex-3. 教師 としての 資格
 (teacher) (qualification)
 -----> qualification for a teacher
 教師 としての 地位
 (teacher) (position)
 -----> position as a teacher

Ex-4. 結果 に対する 影響
 (results) (effect)
 -----> effects on the results
 家族 に対する 義務
 (family) (duty)
 -----> duty to the family

The semantic relationships are not sufficient for selecting such prepositions appropriately. Some postpositional case particles should be translated into different prepositions, even though the deep semantic relationships in Japanese are the same (that is, Japanese native speakers cannot intuitively distinguish the difference.) In Ex-4, for example, English native speakers can clearly distinguish the semantic relationship of 'results' and 'effects' (GOAL or OBJECT case) from that of 'family' and 'duty' (BENEFICIARY case), whereas in Japanese the same postpositional case particles are used and so the distinction of the semantic relationships is not so clear for Japanese natives. We cannot, therefore, even expect the stability of semantic relationships across the two languages.

Japanese Lexical-Item	Translation Equivalent	Field Code	Part-of-Speech
影響	effect	GEN	N

Column-B

Japanese Surface Marker	Deep Case Relation	English Surface Marker	Translation Equivalent	Part-of-Speech
に対する	GOAL	'on'	effect	N

Japanese Lexical-Item	Translation Equivalent	Field Code	Part-of-Speech
義務	duty	GEN	N

Column-B

Japanese Surface Marker	Deep Case Relation	English Surface Marker	Translation Equivalent	Part-of-Speech
に対する	GOAL	'to'	effect	N

These descriptions indicate: The same case marker should be translated differently depending on the governors. Note that the marker 'に対する' usually correspond to English preposition 'against'.

Fig. 5. Example of descriptions in Column-B

(3) Column D: Transfer based on Collocations with Predicates (Fig. 6)

Certain combinations of nouns and predicates (verbs, adjectives, etc.) in Japanese are translated into single words in English: the combinations are lexicalized by single English words. Combinations of nouns and adjectives, for example, are often lexicalized by single English adjectives.

Ex-5. 光沢 がある ---> lustrous
 (luster) (to exist)
 効率 が 高い ---> efficient
 (efficiency) (to be high)

These combinations can appear in the position of an embedded clause and cause structural changes between two languages (Fig.7). Combinations of verbs and their case elements may also correspond to single English verbs.

Ex-6. 焦点 を 合わせる -----> focus
 (focus) (to adjust)

It also happens that Japanese verbs with wide ranges of meaning should be translated into certain English verbs when collocated with specific nouns. Rules which choose appropriate English verbs by referring

to semantic classifications of case elements (nouns) are defined as lexical rules in the verb dictionary (See Section 4). However, certain collocations of verbs and nouns are so specific that choices of appropriate target verbs cannot be expressed by the lexical rules of verbs of this type. Column D in the form-oriented dictionary of nouns is used to specify such collocations between nouns and verbs.

Ex-7. 間隔 を あける -----> leave a space
(space) (??)

The description of this column is not only expanded into lexical rules of the nouns but also into the property (J-DIC-DERIV) of those nouns. The values of this property is a set of Japanese predicates which require special treatments when they are collocated with the nouns. When node X in Fig. 2 is a predicate, the rules in the main grammar check whether there exists a dependent node whose property J-DIC-DERIV contains the predicate, and if such a node exists, they invoke the lexical rules of the corresponding nouns.

In the dependency structures produced by the Japanese analysis grammar, simple modification of nouns by adjectives such as

Ex-8. 強い 雨 -----> heavy rain
(to be strong, (rain)
to be big,
to be heavy, etc.)

are also represented as embedded clauses (Fig. 8), and so column D is used to choose adjectives in English. Note that most of the combinations between nouns and appropriate adjectives are highly conventionalized in each language and so it is hard to imagine the conceptual primitives for adjectives that could be language independent.

Japanese Lexical-Item	Translation Equivalent	Field Code	Part-of-Speech
効率	efficiency	GEN	N

Column-D					
Japanese Predicate		1	2	3	4
高い	JSC	が			
	JDC	SUBJ			
	MOD	#del			

English Predicate				
	ESC		SUB	
efficient	EDC		OBJ	
	JDC		#	

JSC: Japanese surface case marker JDC: Japanese deep case
ESC: English surface case marker EDC: English deep case
MOD: Additional specification ('#del' shows the case element should be deleted in English, i.e. the meaning of the case element is lexically included by English predicate.)

This shows :If '効率'(efficiency) is collocated with '高い'(high), then the combination should be translated into 'efficient', else the translation equivalent is 'efficiency'.

Fig. 6. An Example of Descriptions in Column-D

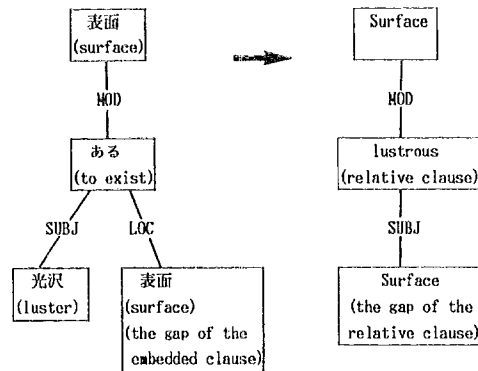


Fig. 7. Structural Changes Caused by Lexical Transfer

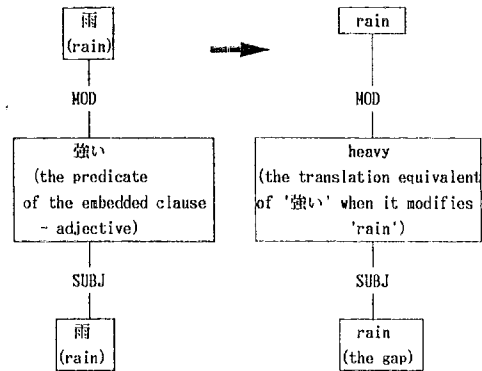


Fig. 8. Lexical Transfer of an Adjective

(4) Column C : Transfer based on Collocation with Modifiers

We have several words of minor parts-of-speech in Japanese which roughly correspond to determiners, quantifiers, etc. in English. Almost the same kind of information as column D is specified in this column for these words.

4 The Form-Oriented Dictionary and Lexical Transfer of Verbs

When the node X in Fig. 2 is a verb or an adjective, the transfer grammar for predicates is invoked. If it does not exist, the other rule in the grammar invokes the lexical rules derived from the form-oriented dictionary of the predicate (node X). By examining the annotation parts of that node and its dependents, the lexical rules of a predicate may determine

- (1) the translation equivalent in English
- (2) the deep case relationship and surface case maker of those dependents strictly governed by the predicate.

If a Japanese predicate is really ambiguous, a separate case frame is provided for each of its meanings. The Japanese analysis grammar disambiguates the meaning and a separate lexical entry is prepared in the bilingual dictionary, i.e. different meanings of a predicate are treated as different lexical items

in the transfer phase. However, the real problems in translation is not caused by ambiguity such as this but by vagueness of meaning. When we treat a language pair such as Japanese and English, it is not exceptional but rather usual that a verb of SL has to be translated into different target lexical items, even though the native speakers of SL cannot clearly realize the meaning difference.

Ex-8. 性能 を 上げる ----> improve the performance
 (performance)
 速度 を 上げる ----> increase the speed
 (speed)
 水面 を 上げる ----> raise the water surface
 (water surface)

The above three usages of the verb '上げる' are continuous in the sense that all of them have the core meaning in common. If we considered them as different meanings, the Japanese verb '上げる' would have plenty of different meanings and be a highly ambiguous word. Note that the English verb 'to wear' should be translated into more than five different Japanese verbs depending on what is to be worn. 'To wear shoes', 'to wear a watch', 'to wear spectacles', etc. should all be translated differently. These facts show that it is impossible to establish the set of lexical items of the interlingua. In our system, then, these 'translationally ambiguous' lexical items are treated in the transfer phase.

The form-oriented dictionary of verbs is shown in Fig. 9. Each pair of condition and a transformation is expanded into a tree transformation rule of GRADE.

Japanese Predicate

	JSC	が	を		Translation
上げる	JDC	SUBJ	OBJ		Equivalent
Condition-1			#AP		improve-2
Condition-2			#NO		increase-1
Condition-3					raise-1

transformation-1

improve	ESC	SUB	OBJ1		
	EDC	CPO	OBJ		
	JDC	SUBJ	OBJ		

transformation-2

increase	ESC	SUB	OBJ1		
	EDC	CPO	OBJ		
	JDC	SUBJ	OBJ		

transformation-3

raise	ESC	SUB	OBJ1		
	EDC	CPO	OBJ		
	JDC	SUBJ	OBJ		

Fig. 9. Form-oriented Dictionary for Verbs and an Example of Descriptions

Because certain Japanese verbs are translated into complex expressions in English, one can specify not only the English main verbs but also arbitrary phrases governed by the verbs as constants.

Ex-9 Xを 試作する
 ----> to produce trans[X] on a trial base

Note that X in the above example indicates a variable an arbitrary Japanese noun phrase can fill in. TRANS(X) shows the result of transferring the noun phrase into English.

5 Recursive Processing and Multiple Layer of Grammars

Fig.10 shows that the schematic construction of the MTP. The substructure governed by the node X is transferred recursively by an appropriate subgrammar, depending on the properties of that node. For simplicity, only the subgrammars for main clauses and noun phrases are shown in detail. The blocks surrounded by bold lines in this figure are those which are performed by lexical rules described in section 3 and 4. The lexical rules are invoked at the beginning of the subgrammars, because they are the rules for treating idiosyncratic phenomena specific to lexical items and are applied before the other, more general rules are applied. Constraints on some of the dependent nodes such as syntactic forms (a prepositional phrase with a specific preposition, etc.) are determined by these lexical rules, and added to their annotation parts. These constraints are utilized at the next recursion step where the substructure under each dependent node is

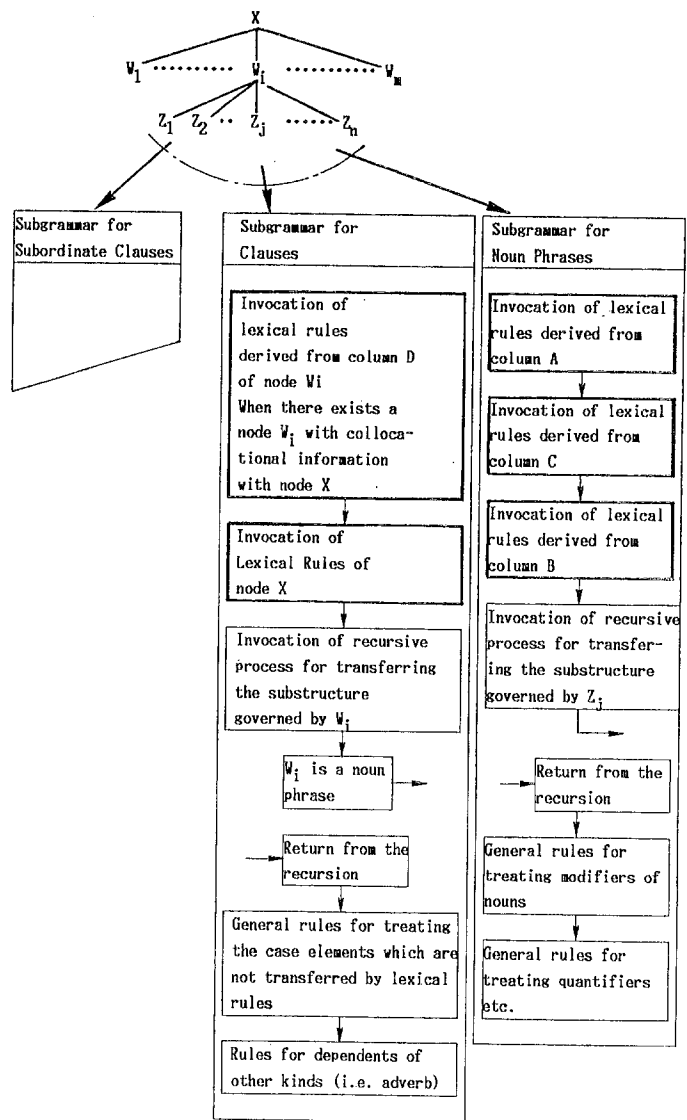


Fig. 10. Schematic Construction of subgrammars of MTP

transferred. Only the dependent nodes which are not treated by the lexical rules are dealt with by general rules. That is, general rules are applied to the dependent nodes of a predicate, only when the node W_i in Fig. 10 is not treated by

- (1) lexical rules from the column D of the node W_j
- (2) lexical rules of the predicate, the node X
- (3) lexical rules from the column A of the node W_i
- (4) lexical rules from the column C of the node W_i

The general rules are further divided into the following two categories :

(1) Rules in the case when the deep case relation between the node W_i and the governor (the node X) is uniquely determined by the analysis grammar : They determine the surface preposition of English by referring to the deep case interpretation, the surface case marker of Japanese, and the lexical item and the semantic markers attached to the node W_i . (55 semantic markers are used in our system to classify Japanese nouns).

(2) Rules in the case when the deep case relation is not uniquely determined in the analysis phase : They determine the preposition just by referring to the surface case marker of Japanese. The preposition roughly corresponding to the Japanese surface case marker is selected.

The above two categories of rules are just rough classifications and, in reality, more detailed rules are prepared for each Japanese case marker. In particular, there are several type 2 rules for each surface case marker which take into account the set of possible case interpretations given by the analysis phase.

Thus, the dependent nodes of a predicate are transferred by several kinds of rules with differing preferences in order to utilize as much of the information extracted during the analysis phase as possible. The same multiple layer construction of subgrammars is also adopted for the other grammatical units such as subordinate clauses, embedded clauses, etc.

6 Lexical Rules and the Form-Oriented Dictionaries

The form-oriented dictionaries are provided just for preparing dictionary descriptions for a large number of ordinary lexical items. The forms given in the preceding sections are used by lexicographers who do not have any detailed knowledge about the transfer grammar. At execution time, these descriptions are expanded into GRADE lexical rules (Fig. 11).

If one wants to specify complicated transfer rules which cannot be expressed in these forms, one can provide lexical rules directly in GRADE, and attach them to specific lexical items. Such lexical rules are also invoked at the appropriate time during the transfer phase. The rules in the transfer grammar checks whether the lexical items have lexical rules of certain types, and if they have, the lexical rules are invoked whether they are derived from the form-oriented dictionaries or prepared directly in the GRADE form. One can specify arbitrary tree-transformations in GRADE so that one can prepare very flexible transfer rules specific to lexical items. Lexical rules for certain items such as Japanese verbs with wide ranges of usages are given directly in GRADE.

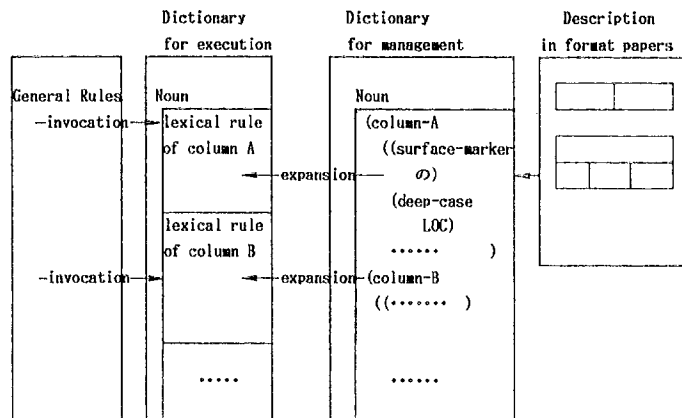


Fig.11 Relationships between Format Dictionaries and Dictionaries for Execution

Though most Japanese verbs and nouns are respectively translated into English verbs and nouns, certain parts-of-speech in Japanese do not have such direct equivalents in English, and the word of such parts-of-speech need to be translated into various kinds of expressions. The transfer rules are highly specific to each lexical item, and therefore, we do not prepare any dictionary formats for these words. The transfer rules are directly prepared in the GRADE form. The adverbial postpositional particles in Japanese, for example, are translated into adverbs, prepositions, styles of sentences, etc., depending on each lexical item and its surrounding context.

Ex-10 ジョンさえも ...する ---> Even John does ...
 (John) (to do)
 ジョンこそが ...する ---> It is John who does ..
 (John) (to do)
 A, B, Cなどの D ---> D such as A, B and C
 A, B, Cなど ---> A, B, C and so on

Note also that, because the roles of these particles are to give additional meaning to the core meaning (i.e. the propositional contents) of the sentences and some of them are related to the attitude or the intention of the speaker, the interlingual representation for these particles can be very complicated. It is also the case for the Japanese auxiliary verbs which follow the main predicates and express various kinds of modality, aspect, politeness, and so on. The correspondence of these expressions is highly dependent on language pairs and so should be treated in the transfer phase.

7 The Structural Transfer

The MTP transfers Japanese structures into corresponding English structures recursively by traversing the Japanese analysis trees (dependency trees) from the top to the bottom. The order of processing is pre-determined by the dependency trees given by the analysis phase. If the substructures are changed unexpectedly during the recursion, the upper level of recursive processing cannot respond to the change ; if it could, the rules which check the results of the transfer of the substructures and take appropriate actions would need to be provided, making the construction of subgrammars unnecessarily complicated.

There are pairs of stereotypical sentence styles in the two languages which cannot be translated by phrase-to-phrase translation.

Ex-11 Aだけが~されるのではなく、Bも...される
 ---> not only A is ~ but also B is ...

Because the clues of such stereotypical sentence patterns are embedded deeply in the analysis trees (see Fig.14) and resultant tree structures of the transfer phase become quite different from those of the analysis phase, it is difficult for the recursive processing in MTP to treat them.

The two phases, Pre-TP and Post-TP, are provided for such global structural changes. In these two phases, each rule traverses the whole tree independently in order to find where the rule is applicable.

8 Pre-Transfer Phase

In this phase, certain structures in Japanese which do not have direct English equivalents are transformed into more English-oriented structures. For example, the following structural transformation (Fig. 12) is taken place, because Japanese quantifiers such as 多い (many, much), 少ない (few, little) etc. can be used as predicates but English quantifiers cannot.

Ex-12 Xが 多い -----> 多くの Xが 存在する
 (many) (many) (to exist)

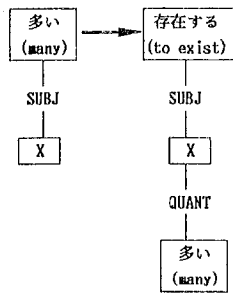


Fig.12. Structure Changes in Pre-TP

Furthermore, the following structures in Japanese are treated.

(1) Subordinate clauses which should be expressed by adverbs or prepositional phrases in English :

Ex-13 注意して 機械を 制御する
 (to pay attention to) (machine) (to control)
 -----> to control (the) machine carefully

A subordinate clause with a predicate of 'TE'-form (注意して - to pay attention to) is usually expressed in English as a subordinate clause with a conjunction such as 'while', 'when' etc. or a participle clause. However, when the predicate of the subordinate clause is a verb of a certain type, the clause expresses the 'manner' case of the main clause and it should be

translated into an adverb or a prepositional phrase in English. Rules are prepared in Pre-TP, which transform clausal constructions into phrasal constructions.

Such structural transformations are often required in the translation of Japanese into English. Some complex expressions which contain subordinate clauses or embedded clauses are used in Japanese only for expressing deep case relationships.

Ex-14 Aが 原因と なって、
 (A with a subject marker)(cause) (to become)
 ----->(direct translation)
 Because 'A' becomes the cause,
 (deep interpretation)
 'A' fills the deep case 'cause' of the main clause --> 'by'

Because direct translations of these expressions result in clumsy English, they are reduced into much simpler structures in the transfer phase (Fig. 13).

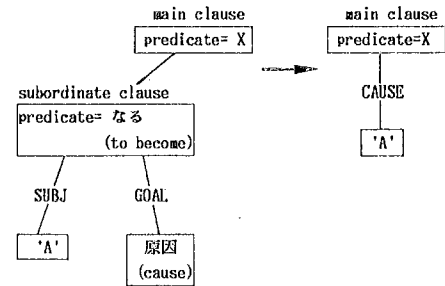


Fig. 13. Reduction of a Structure in Pre-TP

(2) Stereotypical expressions of Japanese which have corresponding English stereotypical expressions : The correspondence of stereotypical expressions in the two languages such as given in Ex-11 cannot be recovered appropriately by the phrase-to-phrase transfer in MTP. Rules are prepared in Pre-TP, which discover the existence of such stereotypical expressions and insert special nodes in the dependency trees to facilitate the recursive processing of MTP (Fig. 14).

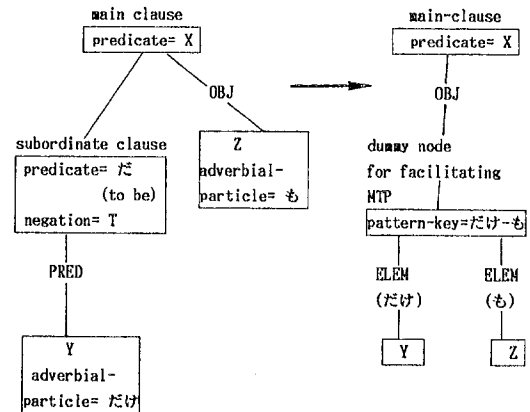


Fig. 14. Correspondence of stereotypical patterns

(3) Complex Expressions following predicates :
 Though auxiliary verbs in Japanese follow main predicates to express various kinds of information such as modality, aspect, politeness, etc., some complex expressions which contain embedded clauses are also used to express the same kind of information and should be expressed in English by auxiliary verbs, aspect verbs (ex : to begin, to continue, keep --ing, etc.), etc. (Fig. 15).

Ex-15 Xは _____ ~する 傾向に ある
 (embedded clause) (tendency) (to be)
 ---->(direct translation)
 X is in the state of tendency ~
 (natural translation) X tends to ~

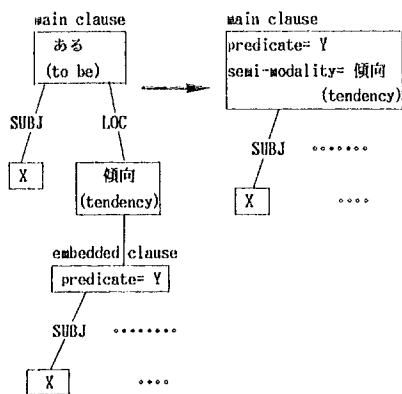


Fig. 15. Reduction of Structures in Pre-TP

Note that what kinds of information can be expressed by simple auxiliary verbs and so by simple linguistic constructions is dependent on each language, and the adjustment of the difference requires certain structural changes.

9 Post-Transfer Phase

The structures which reflect the characteristics of Japanese are transformed into more natural structures in English at this phase. Some of the typical transformation are given below.

(1) Deletion of causative verbs : Causative constructions are often used in Japanese to express events which can simply be expressed by using transitive verbs in English.

Ex-16 Aが Bを 回転 させる。
 (to rotate) (auxiliary suffix
 for causative construction)
 --->(direct translation) A makes B rotate.
 (natural translation) A rotates B.

(2) Transformation from constructions with intransitive verbs into constructions with transitive verbs : It happens quite often that an event is naturally expressed by an intransitive verb in Japanese but by a transitive verb in English.

Ex-17 データ処理 進歩 で 自動化船が 増加した
 (data (advance) (case (automated (to increase
 processing) particle ships) -past tense)
 -'by')

---->(direct translation)
 Due to advances of data processing,
 the number of automated ships increased.
 (natural translation)
 Advances of data processing increased
 the number of automated ships.

10 Conclusions

We explained in this abstract the construction of the transfer phase of our Japanese-English translation system. The implementations of the whole system has already been completed and the translation results (10,000 sentences) are now being evaluated by professional translators and native speakers of English. The evaluation results obtained by now are quite satisfactory.

The description in the form-oriented dictionaries discussed in this abstract are currently prepared for about 74,000 nouns and 5,000 verbs and adjectives.

(Reference)

- (1) Nagao, M. et.al. : <Dealing with the Incompleteness of Linguistic Knowledge in Language Translation', Proc. of COLING 84, Stanford University, California, 1984
- (2) Nagao, M. et.al. : <The Japanese Government Project for Machine Translation', Computational Linguistics, Vol. 11, no. 2-3, 1985