[*Machine Translation Summit*, September 17-19, 1987, Hakone Prince Hotel, Japan]

ATLAS:

Fujitsu Machine Translation System

Hiroshi UCHIDA

Fujitsu Laboratories Ltd.

Due to the rapid advancement of both computer technology and linguistic theory, machine translation systems are now coming into practical use. Fujitsu has two machine translation systems. ATLAS-I is a syntax-based machine translation system. ATLAS II is a semantic-based system which aims at high quality multilingual translation. In this paper, the ATLAS II translation mechanism is explained.

1. Introduction

In 1984 Fujitsu marketed the automatic machine translation systems, ATLAS-I and ATLAS II. ATLAS-I was the world's first commercial English-Japanese translation system. Fujitsu is also conducting a joint project on research and development of a Japanese-Korean machine translation system in cooperation with Korean Advanced Institute of Science and Technology. ATLAS II aims at achieving multilingual translation. At present, however, the commercial version of the ATLAS II system translates Japanese to English. However, some effort has been directed toward achieving multilingual translation. In addition from 1983 to 1985, Fujitsu contributed technological support to the SEMSYN project, a German-Japanese translation system being developed at Stuttgart University, West Germany. At the Tsukuba Expo '85, we also conducted machine translation experiments, translating Japanese children's compositions into English, French, and German, English news texts into Japanese, French, and German, and also mutual translation of simple sentences between Japanese, Swahili, and Innuit (Eskimo).

2. ATLAS II

ATLAS II aims to simulate human translation, understanding a sentence written one language, then expressing it in another. Any language is based on the assumption that every person is able to understand a sentence from the meaning of the component words and context. Syntactic rules are also based on this assumption. To be able to translate naturally a computer should also be able to do this.

Human have their own world models, formed from linguistic knowledge, common sense, cause-effect relationships, and human characteristics. This is why humans can perform both semantic and contextual analysis with ease. The world model can be extended by inference, or narrowed according to the context. Humans also have a language model which guides our actual use of words.

ATLAS II is equipped with both a world model and a language model (see Fig.1). The world model is expressed as a semantic relation between concepts: The language model expresses the co-occurrence relation between words. Grammatical rules for analysis and generation, and transfer rules are provided for simulating the human translation process.

The conceptual structure is a semantic network representation of an inputted sentence. Fig 2 shows the conceptual structure which is equivalent to "John drunk beer yesterday." The network consists of nodes and arcs: a node represents a word conceptually, "John#1" of "John", "drink#l" of "drink", "beer#l" of "beer", "yesterday#1" of "yesterday", a binary arc denotes a deep case relation between nodes such as "agent", "object", "time". In addition to the above binary arcs, there are unary arcs which indicate additional information about a specific node, such as tense, sentence style, for example, in Fig.2 "past" indicates tense and "focus" indicates focus.

The system understands an input sentence in the form of a conceptual structure. Humans understand a sentence by using their knowledge. ATLAS II refers to its world model in the same way as humans. The world model defines every possible relation between concepts. For example, the knowledge, "animals drink" is expressed in the world model as follows:

(animal#1, drink#l, agent) = true

The lefthand side indicates a conceptual structure where an arc "conjoins the nodes" animal#l and node drink#l. This is found to be true by referring to the world model. The system checks whether the conceptual structure is included in the world model. If it is, the system accepts it; if it is not, the system rejects it and asks for an alternative sentence analysis.

Relation between concepts should be as universal as possible. However, it is not possible to apply this to all concepts, because each language is to some degree, unique. As a result, a conceptual structure produced by analyzing a Japanese sentence may remain Japanese to some extent; consequently, this structure may not be appropriate for English generation. For example, the sentence "Ningen niwa zunou ga aru." would ideally be translated as "Man has a brain." To do this, conceptual transfer is required; if not, the literal translation "There is a brain in man" will be produced.

Conceptual transfer is performed between conceptual structures: from a source language dependent structure to a target language dependent structure. The conceptual structure interface guarantees complete separation between analysis and generation. The interlingua approach serves for almost all translations and the transfer approach is used only for specialized ones, allowing a minimum number of transfer rules. As a result, this system is appropriate for multi-language translation.

Figure 2 shows the translation flow of ATLAS II.

3.1. Analysis Process

The sentence analysis section analyzes an inputted sentence and expresses its meaning as a conceptual structure in the form of a semantic network. This section consists of three modules; SEGMENT for morphological analysis, and ESPER for syntactic analysis and semantic analysis. This section uses the word dictionary, word adjacency relations analysis rules, and semantic relations to analyze the sentence. Figure 2 shows how each module uses the dictionaries, and the rules, and the forms of processing results.

An input sentence is first analyzed morphologically and then divided into morphemes. SEGMENT performs this morphological analysis using the word dictionary and adjacency relations. Generally, morphological analysis and synthesis are highly language-dependent. This system, however, adopts a language-independent method for multilingual translation. This method uses an adjacency matrix which defines the adjacency possibility between morphemes.

Morphemes extracted by morphological analysis are output into an analysis node list. ESPER receives this node list and each morpheme is treated as a terminal node. The sequence of these nodes is the same as that of the input morphemes. Each node obtains grammatical and semantic information from the word dictionary. Grammatical information is a set of grammatical attributes. This allows each grammatical rule to cover a wide range of linguistic phenomena, thus reducing the number of rules. Each terminal node contains the most probable word selected from several possibilities.

ESPER consists of a status stack, analysis window, and control section. The status stack monitors the status during analysis; the analysis window view two adjacent nodes. ESPER performs simultaneous syntactic and semantic analysis using analysis rules which are based mainly on context-free grammar. The suitability of syntactic processing is verified semantically.

Semantic processing is performed with a series of semantic symbols which correspond to the conceptual structure. The applied rule attaches a semantic symbol to the new node and determines the semantic relation between two nodes in the analysis window. The semantic processing checks to find if the processing is consistent with general world knowledge.

Finally, ESPER arrives at a conceptual structure of the input sentence. This conceptual structure is verified by referring to the world model. If it is incorrect, ESPER reanalyzes the sentence and outputs another possibility.

3.2. Transfer Process

The transfer section is provided to fill the gap between the source language and the target language. Differences in languages stem from among other things, the cultural background of the people speaking each language. Superficially, it appears as a difference in words and grammar; internally, it appears as a difference in concepts and in the speaker's way of thinking.

ATLAS II compares these difference, not superficially, but internally; examining not the differences between words or grammar, but the differences between concepts and thinking. The differences, therefore, are treated at the level of the intermediate representation, and the conceptual structure is transferred. However, the pivot approach which does not require this transfer, is suitable in most cases.

Let's look at a few cases which would require such a transfer. For example, the sentence "Heya niwa mado ga futatsu aru" would be literally translated as "There are two windows in this room" but the natural translation would be "This room has two windows." Another case involves the causative. Japanese expresses it using the auxiliary verb 'saseru'; while English depends on an intransitive verb and word order.

3.3. Generation Process

Target language text is generated from the conceptual structure which is in the form of a semantic network. This conceptual structure is converted into a linear word string. This direct conversion eliminates the need for transformation, allowing not only the generation mechanism but also the rules to be language-independent. Using this approach, generation rules can deal with both syntactic structuring and morphological synthesizing at the same time, thus simplifying the generation mechanism.

The generation system consists of a generation window, output list and a rule interpreter. The rule interpreter traverses each node of the conceptual structure by moving the generation window and returns an output list containing the translation results.

The generation window is set at the first node of the conceptual structure and is then moved from node to node. This window is used to check the nodes and arcs. The contents of the output list indicate the surface-structure word order.

The rule interpreter interprets each generation rule, traverses each node by moving the generation window, and selects words from nodes and arcs by checking the co-occurrence relation and adjacency relation. Each selected word is added to the output list.

The co-occurrence relation between two words gives the true/false value indicating the likelihood of the two words co-occuring in the same sentence. Generally a concept covers several words. For example, a concept indicating 'sonzaisuru' in Japanese includes selection of a word from several candidates by checking the co-occurrence relation between the candidates.

4. Conclusion

The biggest problem with any machine translation system is the quality of the translation. Unfortunately, current technology has not produced perfect results. We have to provide support systems for pre-editing, post-editing, and dictionary compilation. And also we have to study how to use machine translation system effectively. The quality of translation depends on the accuracy of both rules and dictionaries, as well as the amount of information contained in the dictionary. But this presents another problem: the greater the amount of information, the longer the processing time. It is also difficult to guarantee the accuracy of a large amount of information. These problems cannot be solved by one company alone. We must ask for assistance from users, especially in the compilation of dictionaries. We believe, however, that machine translation will eventually prove superior to manual translation in terms of speed and consistency, and will play an important role in promoting international communication.

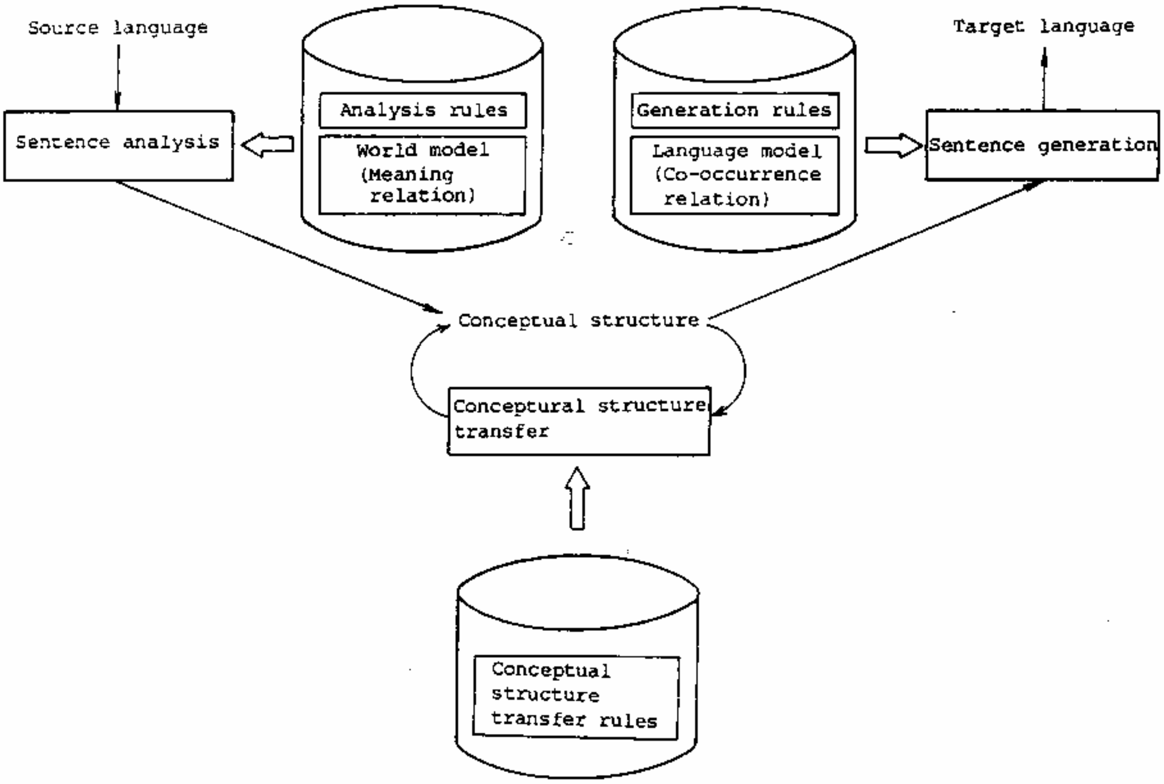
Source language
Sentence analysis
Analysis rules
World model (Meaning relation)
Generation rules
Language model (Co-occurrence relation)
Target language
Sentence generation
Conceptual structure
Conceptural structure transfer
Conceptual structure transfer rules

Fig. 1 Translation Process of ATLAS Ⅱ

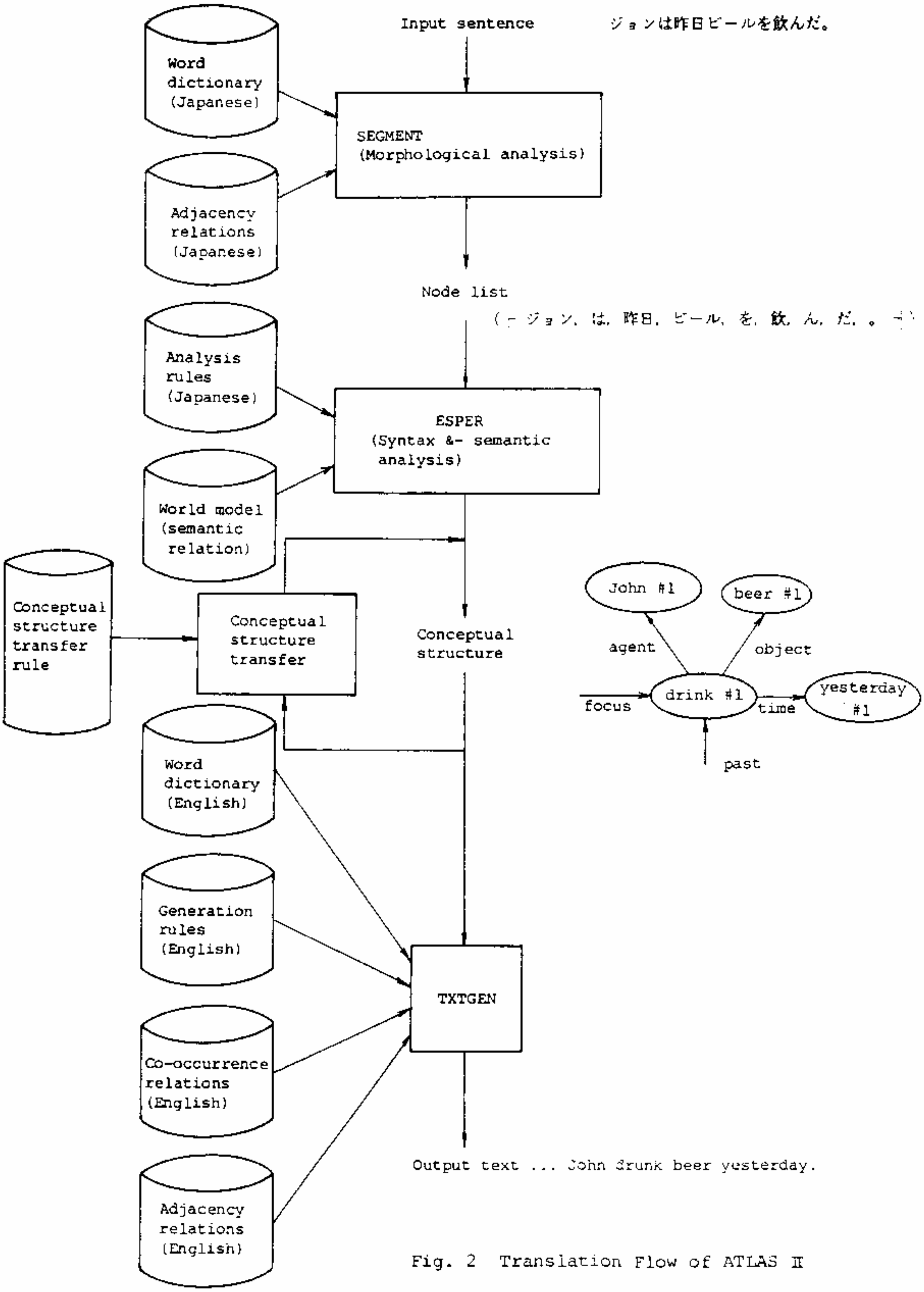
Input sentence
ジョンは昨日ビールを飲んだ。
Word dictionary (Japanese)
Adjacency relations (Japanese)
SEGMENT (Morphological analysis)
Node list
(⊢ ジョン, は, 昨日, ビール, を, 飲, ん, だ, 。 ⊣)
Analysis rules (Japanese)
World model (semantic relation)
ESPER (Syntax &- semantic analysis)
Conceptual structure transfer rule
Conceptual structure transfer
Conceptual structure
John #1
beer #1
agent
object
focus
drink #1
time
yesterday #1
past
Word dictionary (English)
Generation rules (English)
Co-occurrence relations (English)
Adjacency relations (English)
TXTGEN
Output text ... John drunk beer yesterday.

Fig. 2 Translation Flow of ATLAS II