

# AN UNDERSTANDING SYSTEM OF NATURAL LANGUAGE AND PICTORIAL PATTERN IN THE WORLD OF WEATHER REPORT

Eiji KAWAGUCHI, Masao YOKOTA, Tsutomu ENDO and Tuneo TAMATI

Faculty of Engineering, Kyushu University  
Hakozaki 6-10-1, Higashi-ku, Fukuoka, 812 Japan

An Information understanding System Of Basic weather Report (ISOBAR) is introduced first. It can adapt to both linguistic and pictorial input, and produce each output. The central part of this system is semantic processing, but ISOBAR's world is limited to weather report of Japan and Far East Asian area in terms of Japanese and corresponding weather charts. According to experimental results, the performance of the system proved to be satisfactory, except the computing time of picture processing procedures. Finally, the problems for the future research are pointed out.

## 1. Introduction

The present paper is based on the experimental research for understanding system of natural language and pictorial pattern. The prototype system is called ISOBAR (Information understanding System Of Basic weather Report). The model was designed and developed after the following philosophy:

1. In order to be able to understand natural language for human and machine, they need to have some correspondence to the knowledge or experience of external world (in our case visual world),

2. And also in order to make them understand pictorial patterns, it is necessary to extract and recognize some characteristic features of pattern, corresponding to human concepts which reflect to the structure and meaning of natural language man spontaneously developed in his long history,

3. Through these interactive mechanism, man can explain the pictorial pattern by using natural language and vice versa.

The ISOBAR's world is limited to only Japanese sentences of weather report and weather chart of Japan and Far East Asian areas. But from the view point of semantics, the fundamentals of processing algorithm or mechanism of this limited world is applicable to more general world.

ISOBAR has two operating modes in principle.

-----  
This work was financially supported in part by Grant-in-Aid for Special Project Research of the Ministry of Education, Science and Culture, 1978. Project Number 310233.

One is accumulation of meteorological database both from linguistic and pictorial inputs. The other is its retrieval. In the accumulation mode, man presents weather report sentences and charts to the machine. The machine interprets and understands the input information semantically by the aid of general knowledge about meteorology. It transforms the input into suitable data format and adds to the database if and only if the input is meaningful and conveying new information. If not, the machine either inquires about ambiguous points, or rejects the total input.

The semantics in ISOBAR in the final version is designed to have such hierarchy as in Fig.1. GKS is the symbol for General Knowledge about Semantics, and GKM for General Knowledge about Meteorology. SWM and KMP are Semantics in the World of Meteorology and Knowledge about Meteorological Phenomena respectively. Generally speaking, it is difficult to distinguish GKM from GKS strictly.

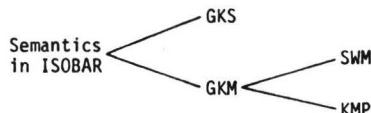


Fig.1 Hierarchy of semantics in ISOBAR

The present ISOBAR is not completed as a total system with full semantic hierarchy as in Fig.1. SWM was already implemented in the present system because it is the core of the semantics in the system, and fortunately it has rather simple structure. In the meantime, GKS must cover all field of our linguistic activities in

Japanese. So in our case, GKS is limited only to general linguistic knowledge which is essential to the field of weather report. Present ISOBAR is partly containing GKS in the form of "Event of Constituent" which will be introduced in the next chapter. KMP is regarded as the total contents of an "Encyclopedia of Meteorology", but this part is limited and is not explicitly separated from SWM in the present system.

## 2. Outline of ISOBAR

First, the weather report sentences and weather charts for ISOBAR are outlined. Then some important semantic notions are presented here.

### 2.1 Sentences of weather report

We limited the words and sentence patterns for ISOBAR only to those of weather reports in Japan which are broadcasted every day from NHK on its second radio program. We termed them basic weather reports. NHK weather reports consist of four parts:

- Part 1 General weather conditions
- Part 2 Local weather conditions
- Part 3 Reports from vessels
- Part 4 Information for fishermen

Among these, we omitted the sentences in Part 1 from our consideration. All of these materials were collected from real NHK program between June, 1976 and May, 1977, and these were stored in the database of ISOBAR. Examples of sentences from Part 2 to Part 4 are shown in Fig.2.

### 2.2 Weather charts

Weather charts for ISOBAR are illustrated in Fig.3. The meaning of each pictorial component will be found in Table 1.

### 2.3 Notions in the semantic structure of weather report[1],[2],[3]

We introduce three basic semantic notions to

#### Part 2)

O-SAKA DEWA NNE NO KAZE , FU-RYOKU 3 TEN-KI WA HARE DE , KIATSU 998MB , KION- 23°C . (OOSAKA area, NNE wind, wind force 3, fair, pressure 998MB and temperature 23°C.)

#### Part 3)

KAN-TO-TO-HO- NO N35°E146° DEWA W NO KAZE , FU-RYOKU 4 , TEN-KI WA HARE DE , KIATSU 05MB . (At N35°E146° east of the KANTO district, west wind, wind force 4, fair and pressure 1005MB.)

#### Part 4)

- (a) KAN-TO-HIGASHIKAIJO- NO N36°E149° NIWA , 1004MB NO TEIKIATSU GA ATTE , E NI 75KM DE SUSUN-DEIMASU . (A low pressure of 1004MB at N36°E149° east of the KANTO district, proceeds east at 75KM/hr.)
- (b) NIHON-HUKIN- O TO-RU 1020MB NO TO-ATSUSEN- WA , N60°E140° , N56°E136° , ... , N19°E113° NO KAKUTEN-O TO-TTEIMASU . (An isobar of 1020MB in and around Japan passes through N60°E140° , N56°E136° , ... and N19°E113°.)

Fig.2 Examples of weather report sentences

describe the semantic contents of the world of weather report. The first is "Constituent" of meteorological world in ISOBAR. The second, "Attribute" of Constituent, and the last, "Event" in Attribute-space. "Attribute-space" in this context means an abstract space whose bases are all Attributes essential to that Event.

After careful analysis of weather report sentences and weather charts, we postulated 17 Constituents(denoted by  $C_1, C_2, \dots, C_{17}$ ) in all. These are tabulated in Table 2. Attributes of Constituents are extracted to be 14 different kinds ( $A_1, A_2, \dots, A_{14}$ ) as in Table 3. Each Attribute is, in other words, a semantic phase or point of view of Constituents. Relation between Constituents and Attributes makes "Cross-Relation Table" of Constituents and Attributes. "Semantic Table" is defined to be such Cross-Relation Table that is filled with each "Attribute Value". A set of weather reports at a special data and time, or a sheet of weather chart, makes complete Semantic Tables, if all input sentences or pictorial components are meaningful, and all are recognized correctly. Table 4 shows an example of Semantic Tables.

"Event" in an Attribute-space is introduced as a semantic contents of "verb", "adjective", "adverb", etc.; those are all representing the conditions of Constituents or relation among Constituents. We visualize each event as a

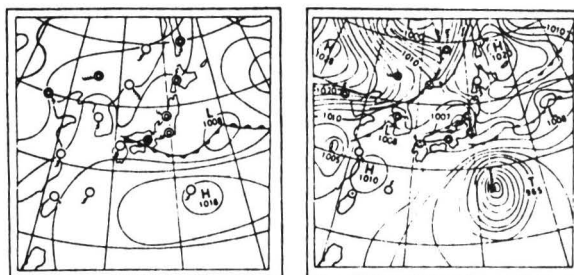


Fig.3 Examples of weather charts

Table 1 Picture components in weather chart

| Component           | Pictorial pattern | Information from the component       |
|---------------------|-------------------|--------------------------------------|
| Weather symbol      | ① ⊙ ●             | Location, variety of weather         |
| Wind symbol         | ↖ ↗ ↘ ↙           | Location, direction, force           |
| High pressure       | H                 | Location, central pressure           |
| Low pressure        | L                 |                                      |
| Tropical depression | TD                | Number, location, central pressure   |
| Typhoon             | T                 | Location, variety of front           |
| Front line          | —▲—               | Locus, pressure value                |
| Isobar              | —                 | No information (constant components) |
| Coast line          | —                 |                                      |
| Latitude, longitude | —                 | Typhoon number, pressure value       |
| Numerical           | 01.....9          |                                      |

locus of some Constituent in an Attribute-space. There are many patterns of Events in the meteorological world in ISOBAR. Some are temporal patterns, others are spatial patterns. In the

Table 2 List of Constituents

| Symbol          | Constituents        | Linguistic Expression | Pictorial expression     |
|-----------------|---------------------|-----------------------|--------------------------|
| C <sub>1</sub>  | Low pressure        | TEIKIATSU             |                          |
| C <sub>2</sub>  | High pressure       | KO-KIATSU             |                          |
| C <sub>3</sub>  | Typhoon             | TAIFU-                |                          |
| C <sub>4</sub>  | Tropical depression | NETSUTASEITEIKIATSU   |                          |
| C <sub>5</sub>  | Warm front          | ON-DAN-ZEN-SEN-       |                          |
| C <sub>6</sub>  | Cold front          | KAN-REIZEN-SEN-       |                          |
| C <sub>7</sub>  | Stationary front    | TEITAIZEN-SEN-        |                          |
| C <sub>8</sub>  | Occluded front      | HEISOKUZEN-SEN-       |                          |
| C <sub>9</sub>  | Isobar              | TO-ATSUSEN-           |                          |
| C <sub>10</sub> | Ridge               | KIATSUNOMINE          |                          |
| C <sub>11</sub> | Wind                | KAZE                  | No direct correspondence |
| C <sub>12</sub> | Fog                 | KIRI                  |                          |
| C <sub>13</sub> | Surge               | NAMI                  |                          |
| C <sub>14</sub> | Storm               | BO-FU-U               |                          |
| C <sub>15</sub> | Local weather       | KAKUCHINOTEN-KI       |                          |
| C <sub>16</sub> | Weather at sea      | KAIJONOTEN-KI         |                          |
| C <sub>17</sub> | Time of observation | KAN-SOKUJIKAN-        |                          |

Table 3 List of Attributes

| Symbol         | Attribute         | Symbol          | Attribute           |
|----------------|-------------------|-----------------|---------------------|
| A <sub>1</sub> | Location          | A <sub>8</sub>  | Temperature         |
| A <sub>2</sub> | Pressure          | A <sub>9</sub>  | Wind Direction      |
| A <sub>3</sub> | Number of Typhoon | A <sub>10</sub> | Time                |
| A <sub>4</sub> | Direction         | A <sub>11</sub> | Degree of Influence |
| A <sub>5</sub> | Area              | A <sub>12</sub> | Velocity of Motion  |
| A <sub>6</sub> | Wind Force        | A <sub>13</sub> | Wind Speed          |
| A <sub>7</sub> | Weather           | A <sub>14</sub> | Visibility          |

Table 4 Examples of Semantic Table

| Constituent        | Time           | A <sub>1</sub> | A <sub>2</sub> | A <sub>3</sub> | A <sub>4</sub> | A <sub>5</sub> | A <sub>6</sub> | A <sub>7</sub> | A <sub>8</sub> | A <sub>9</sub> | A <sub>10</sub> | A <sub>11</sub> | A <sub>12</sub> | A <sub>13</sub> | A <sub>14</sub> |
|--------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| C <sub>1</sub> (x) | t <sub>0</sub> | β              | γ              |                | α              |                |                |                |                |                |                 |                 |                 |                 |                 |
| C <sub>1</sub> (y) | t <sub>0</sub> | p              |                |                | α <sub>1</sub> |                |                |                |                |                |                 |                 |                 | β <sub>1</sub>  |                 |
|                    | t <sub>1</sub> | q              |                |                | α <sub>1</sub> |                |                |                |                |                |                 |                 |                 | β <sub>1</sub>  |                 |
| C <sub>6</sub> (z) | t <sub>0</sub> | α <sub>1</sub> |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |
|                    | t <sub>0</sub> | α <sub>2</sub> |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |
|                    | t <sub>0</sub> | α <sub>3</sub> |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |

N.B. All above are the semantic structures of the following sentences (i.e. S1-S3) which appear in the weather report at t<sub>0</sub>.  
 S1: KO-KAI(α) NO N34°E122°(β) NIWA 1000MB(γ) NO TEIKIATSU(C<sub>1</sub>(x)) GA ARIMASU.  
 S2: TEIKIATSU(C<sub>1</sub>(y)) GA MAIJI 20KM(α<sub>1</sub>) DE HIGASHI(β<sub>1</sub>) SUSUN-DEIMASU.  
 S3: KAN-REIZEN-SEN-(C<sub>6</sub>(z)) GA N34°E107°(α<sub>1</sub>), N35°E115°(α<sub>2</sub>), N37°E122°(α<sub>3</sub>) O TO-TTEIMASU.  
 The tense of S2 is "Progressive present tense", which we denote by the pair of times, i.e. t<sub>0</sub>(the present) and t<sub>1</sub>(some immediate future).

linguistic analysis of weather report, we need an explicit expression of each event pattern. ISOBAR has all lists of variations of event patterns, which appear in the world of weather report. These event patterns work as both GKS and SWM in the present system. Semantic Tables, in the meantime, exclusively work as SWM.

## 2.4 General performance of the system

The total configuration of ISOBAR is illustrated in Fig.4. It consists of seven processing units and three databases. Linguistic input, i.e., weather report sentences in the accumulation mode, and questions in the retrieval mode, are sent to STA(Syntactic Analyzer) unit first. If the input is syntactically correct, it is stored in the LDB(Linguistic DataBase) only in case of weather report sentences, and at the same time, it is transferred to SSG(Semantic Structure Generator) unit. SSG generates the semantic structure of the input sentence. Then SAS(Semantic Analyzer and Synthesizer) analyzes the structure. The output from SAS makes a Semantic Table which is to be stored in SDB(Semantic DataBase) in the accumulation mode, or to be hold in SAS itself in the retrieval mode.

On the other hand, pictorial input, i.e., the weather chart, both in the accumulation and retrieval mode, are given to PPP(Pictorial Pre-Processor) first. Pictorial processing consists of digitization of analog image, noise elimination, data compression, etc.. In the accumulation mode, all original data and their second data are stored in PDB(Pictorial DataBase) in a compressed form according to a special coding scheme. Output from PPP is sent to PR(Picture Recognizer) unit in every mode, where each component in the chart is recognized. Then its output is fed to SAS unit. SG(Sentence Generator) and CG(Chart Generator) work only in the retrieval mode. That is, linguistic answer to the question is synthesized by SAS using SDB and generated by SG, while pictorial answer is synthesized by SAS using SDB and PDB,

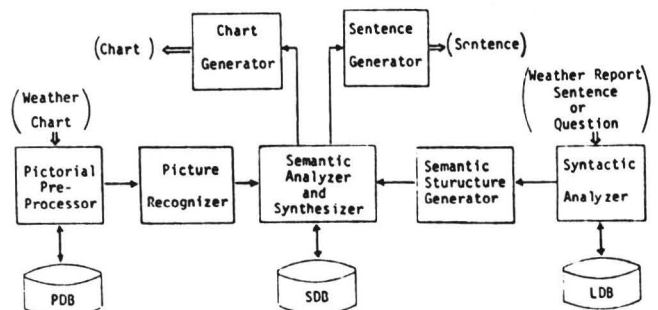


Fig.4 Configuration of ISOBAR

then generated by CG. More details about these units and databases will be presented in Chapter 3 and 4.

Fig.5 shows the hardware system of ISOBAR. The main system has special I/O devices, all of which are peripherals of either T-40(TOSBAC-40C) mini-computer or 0-4300(OKITAC-4300b) multi-CPU mini-computer system.

### 3. Sentence understanding

The outline of the processings in ISOBAR for linguistic input is presented in this Chapter.

#### 3.1 Restriction for input sentences

Original weather report sentences have mostly regular sentence patterns. In ISOBAR, we admit any sentence which satisfies the following restrictions.

[Restriction 1] The vocabulary for input sentence is about 200 words, most of which are adopted from original weather report sentences, and some specific words are supplemented for interrogative and imperative sentences, e.g., "ITSU(when)", "DOKO(where)", etc..

[Restriction 2] The forms of input sentences are either (1)imperative form like, "...GATA NO TEN-KIZU O SYUTSURYOKUSEYO(Output all weather chart of ... type.)", or (2)such forms that are arbitrarily transformed from original kernel

sentences in weather report by the optional rules R1-R5.

R1:Alternation of words in kernel sentences.

R2:Addition to such a word-group as "1976 NEN- 9 GATSU 11 NICHU 12 JI (WA)", which denotes a special data and time.

R3:Substitution of interrogatives to words or word groups of kernel or transformed sentences.

R4:Abbreviation of words.

R5:Addition of interrogative particle "KA".

[Restriction 3] Input Japanese sentences are segmented word by word with spaces.

#### 3.2 Syntactic analysis, semantic structure generation and semantic analysis

The functions of STA and SSG in ISOBAR are as follows.

Input sentences are transformed in STA from surface structures into dependency structures by steps (1)-(4) below, and then into semantic structures in SSG by step (5).

Step(1) Parsing into individual words.

Step(2) Forming word groups.

Step(3) Construction of inter-word dependencies within each word-groups.

Step(4) Construction of dependencies among word-groups.

Step(5) Transforming dependencies into semantic structure(i.e., Semantic Table) by referring the Dictionary of Conceptual Connection(DCC).

At the end of step(4), each input sentence is identified with one of the template patterns, where verbs are essential keys. DCC as above is the most important dictionary that indicates which word-groups are representing "Constituents" or "Attribute Values", and in which slots in the Table they are to be stored.

In the course of above steps, syntactically anomalous sentences are rejected in STA. Several types of anomaly are detected in SSG, and in this case no semantic structures are generated. The semantic structures just generated are analyzed by SAS, and semantic contradictions are detected which can not be interpreted or understood in the world of weather report. This is realized as follows. SAS examines every slot of the Semantic Table whether stored Constituents and Attribute Values are correct. If not, the semantic structure is contradictory and rejected.

#### 3.3 Access to SDB and output generation

The response of the system to a legal input sentence is determined according to kinds of input sentences. It is as follows. Case(1) is the only case that opens the accumulation mode of ISOBAR, while Case(2)-(4) open the retrieval

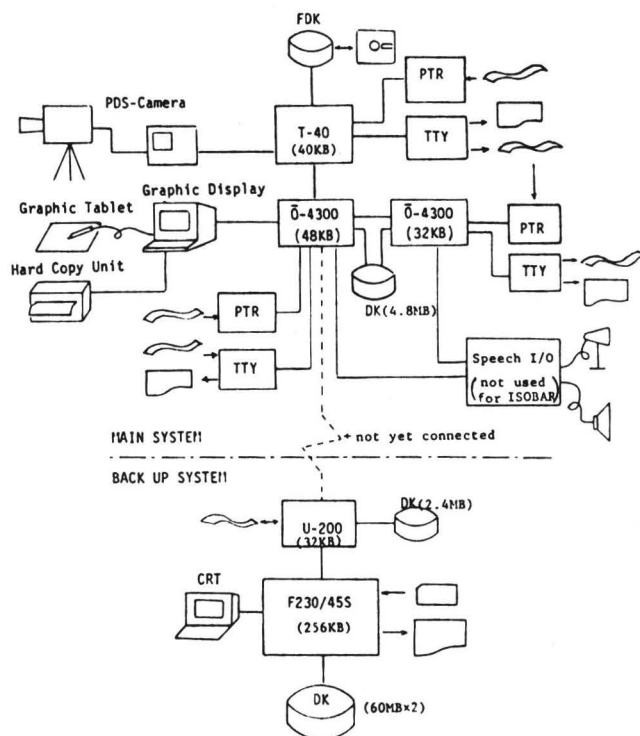


Fig.5 Illustration of the hardware system

mode.

Case(1) Declarative sentence

Every declarative sentence is assumed to be asserting some facts about weather. SAS compares the generated semantic structure of the input with every structure that was already stored in SDB, where comparison is executed on every Attribute. If the input is conveying new information to SDB, it is stored there, otherwise SG comments that it is already stored or contradictory.

Case(2) YES-NO question

Case(3) WH question

Case(4) Imperative sentences

If an imperative input sentence demands all charts of a specific type of pressure configuration, the system interprets the input as "When did the pressure configuration of...type occur?". Examples are shown in Appendix A.

#### 4. Picture processing

In this chapter we describe the operation of PPP and PR in ISOBAR.

##### 4.1 Pictorial pre-processor(PPP)

An analog image is converted into binary-valued picture with 1,024×1,024 pixels by PDS-camera. The noises caused by quantizing operation, are eliminated by simple masks. The pre-processed pictorial data are sent to PR. At the same time, they are stored in PDB. In this case, we took a strategy to reduce the amount of

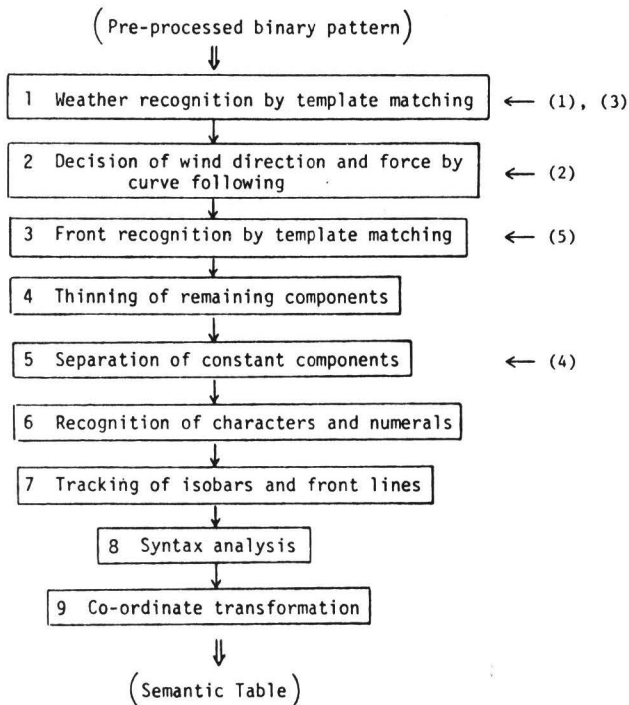


Fig.6 Flow diagram of PR

storage. In our working system, a picture is represented by a series of black or white square regions of various sizes. The square size, its location, and the gray level(either black or white) are encoded in the structural information of "0", "1" and "(" series. We call this expression DF-expression[4],[5]. DF-expressions of original charts, together with their second data, are stored in PDB. The second data include such items as complexity of picture, spectrum of primitives, headings of weather patterns, etc.. These are used for chart searching operation in the retrieval mode.

##### 4.2 Picture recognition(PR)

The major function of this unit is to extract and recognize each picture component as shown in Table 1. We have some a priori knowledges about weather chart in PR. They are as follows.

- (1) Every weather symbol is enclosed with a circle of a constant size. The variation of symbol is limited only to its interior pattern.
- (2) Other components are not connected to the circle except wind symbol.
- (3) There are 15 weather symbols in a chart. The location of respective symbol is fixed.
- (4) The location of coast lines, latitudes and longitudes are fixed, but they are partly erased or interrupted by other components.
- (5) Only front symbols and a few kinds of weather symbols have areas. The others are line drawings.
- (6) Numerals are located near the characters or isobars. They specify the number of Typhoon or pressure values.
- (7) The shape and size of numerals and characters are normalized.

The flow diagram of PR is shown in Fig.6, where each number within parentheses corresponds to above a priori knowledge which is used there. The final output from PR makes the Semantic Tables. Thus, all information from a weather chart is recognized as the semantic content of the world of weather report at the given data and time. We illustrate the results of step(1)-step(4) in Appendix B. These were experimented on a part of an entire picture covering Western Japan. The picture size is 256×256.

#### 5. Discussion

According to our experience, the performance of ISOBAR in the accumulation mode for linguistic input was very satisfactory. Also in the retrieval mode for linguistic input, the system worked successfully. In other words, we can say present restrictions for input sentences are rather severe. So do the vocabulary and sentence patterns. We should remove such restrictions as much as possible in the next version of

## ISOBAR.

On the other hand, its performance for pictorial input was not so much successful. There are no fundamental problems left for the performance of pictorial output, that is, the retrieval operation. But in the accumulation mode, or in recognizing operation of the system, some problems are left unsolved. For instance, computation time is the most intolerable problem. This is partly because of our small (in terms of the size of the main memory) and slow (cycle time) computer system, and partly because of large size of input pictures in spite of rather high compression. Another problems left are improvement of curve following algorithm. Present ISOBAR works in the man-machine interactive manner, but we must find some good idea of automatic gap tracer. Any solution, if ever, of those and other unsolved problems seems to require a high performance machine.

## 6. Conclusions

The conclusions obtained from this research are as follows.

(1) The semantic model in ISOBAR was proved to be appropriate.

(2) The system performance for linguistic input was satisfactory.

(3) The present ability of ISOBAR for pictorial input is not necessarily sufficient, because of the machine size and some unsolved problems.

(4) In order to make ISOBAR more flexible and powerful, it is desired to remove restrictions for input sentences as much as possible.

(5) All picture processing procedure should be improved to be more time saving ones. But now, authors have already started such improvement.

## ACKNOWLEDGEMENT

The authors are deeply indebted to Mr. Y.Ogawa, R.Taniguchi, K.Tsuzuki and K.Noda for their considerable assistance with the experimental work.

## APPENDIX A Examples of Q-A

1978/12/30/12/ WA NIHON-KAIGAWA DEWA TEN-KI WA DAITAI HARE DESU KA. (Was the weather on the side of the Japan Sea almost fair at noon on Dec. 30, 1978?)

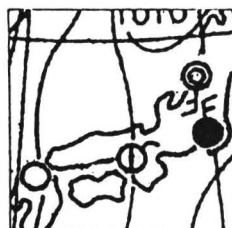
SPATTERN=7 QUEST=1 PLACE=12 WEATHER=2 DEGREE=2  
HAI, NIHON-KAIGAWA DEWA TEN-KI WA DAITAI HARE DESU. (Yes, the weather on the side of the Japan Sea was almost fair.)

1977/1/12/12/ WA NAHA DEWA TEN-KI WA S KA. (In NAHA area at noon on Dec. 12, 1977, was the weather south?)  
ZOKUSEI(TEN-KI) TO ZOKUSEI(S) TOGA HUTEKIGO-. (The weather has not the attribute of direction.)

1976/9/12/12/ WA L NO CHU-SHIN- KARA ON-DAN-ZEN-SEN- GA DOKO NI NOBITEIMASU KA. (Where did a warm front extend from the center of the low pressure at noon on Sep. 12, 1976?)  
SPATTERN=4 QUEST=1 TIME=1976/9/12/12/ ORG=L FNAME=5 LOC=#  
L NO CHU-SHIN- KARA ON-DAN-ZEN-SEN- GA N26DE140D NI NOBITEIMASU. (A warm front extended to N26°E140° from the center of the low pressure.)

ITSU SEIKO-TO-TEI NO KIATSUHAICHI TO NATTEIMASU KA. (When did the pressure configuration of typical winter type occur?)  
SPATTERN=6 QUEST=1 TIME=# TYPE=1  
1976/11/29/18/ WA NIHON- HUKIN- WA SEIKO-TO-TEI NO KIATSUHAICHI TO NATTEIMASU. (At 6:00p.m. on Nov. 29, 1976, the pressure configuration of typical winter type occurred around Japan.)

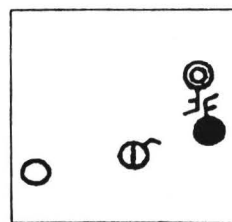
## APPENDIX B Experimental result of PR



Original picture.



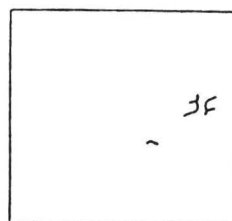
Cross shows exact position of weather circle.



Extracted weather symbol.



Remaining part.



Result of thinning of wind symbol.



Result of thinning of remaining part.

## REFERENCES

- [1] YOKOTA, M. and TAMATI, T., "A Study on Formal Description of Word Meaning(I)", Technology Rep. of Kyushu Univ., Vol.48, No.6, pp.809-814, Dec. 1975.
- [2] YOKOTA, M. and TAMATI, T., "Description of Word Meaning and Its Application", Tech. Rep. of IECE Japan, AL77-41, 1977.
- [3] YOKOTA, M. and TAMATI, T., "On Syntactic and Semantic Analysis of Weather Report Sentences", Tech. Rep. of IECE Japan, AL77-78, 1978.
- [4] ENDO, T. and KAWAGUCHI, E., "Some Properties of DF-Picture Expression and Its Application to Data Compression", Trans. IECE Japan, Vol. J62-D, No.2, Feb. 1979.
- [5] KAWAGUCHI, E. and ENDO, T., "On a Method of Binary Picture Representation and Its Application to Data Compression", IEEE Trans. on PAMI, to be published.