'95 KACC (1995. 10. 23 ~ 25)

IMAGING A SCENE FROM EXPERIENCE GIVEN VERBAL EXPRESSIONS

Y. Sakai,* M. Kitazawa,* and S. Takahashi**

* Dept. of Mechanical Engineering, Yamaguchi University, Ube, Yamaguchi 755, JAPAN

FAX: +81-836-35-9411

** Graduate School, Yamaguchi University, Ube, Yamaguchi 755, JAPAN

1. INTRODUCTION

In the conventional systems, a human must have knowledge of machines and of their special language in communicating with machines. In one side, it is desirable for a human but in another side, it is true that achieving it is very elaborate and is also a significant cause of human error. To reduce this sort of human load, an intelligent man-machine interface is desirable to exist between a human operator and machines to be operated. In the ordinary human communication, not only linguistic information but also visual information is effective, compensating for each other's defect. From this viewpoint, problem of translating verbal expressions to some visual image is discussed here in this paper. The location relation between any two objects in a visual scene is a key in translating verbal information to visual information, as is the case in Fig. 1. The present translation system advances in knowledge with experience.

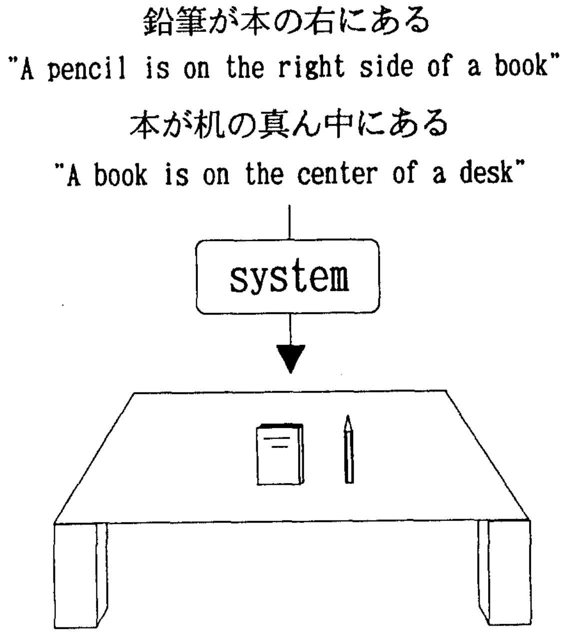

Fig. 1 Schematic of translating words into a visual image

It consists of Japanese language processing, image processing, and Japanese-scene translation functions.

2. JAPANESE PROCESSING

In analyzing Japanese language here, modification relations included in a sentence are found and the original sentence is described in a kind of predicate form which will be called "patternization." Morpheme analysis and parsing are utilized in doing the above processing. Fig. 2 shows the flow of the above procedures schematically by employing a sentence which goes "A pencil lies on the right side of a book." See Reference [1] for details. Here in this paper, the procedures are briefly described together with revised points.

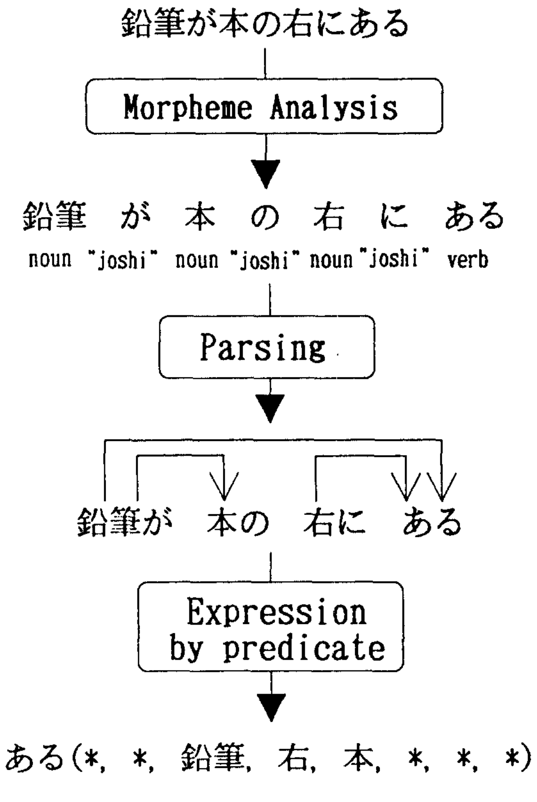

Fig. 2 Flow of processing words

2-1 Morpheme Analysis

Morpheme analysis deals with recognizing parts of speech used in a given sentence by extracting morphemes. In the preceding paper, "saichou-itchi method" was used. The disadvantage of this method is that it only extracts morphemes already experienced in the past. But sometimes a sentence includes unexperienced words and it is a fact that any word is unexperienced if it appears for the first time. So some inference is necessary for recognition of such a word (morphemes). And that word will be a morpheme already experienced in later analysis, if correctly included as part of knowledge in the analyzing system. In the present paper, a new word can be analyzed into a morpheme if the new word is located in between two known words as in Fig. 3. Revision of "saichou-itchi method" was made by incorporating the following algorithm:

(1) If the original algorithm fails in extracting a word from a string of characters, then the first character of that string is deleted.

(2) For the new string obtained in the above procedure (1), "saichou-itchi method" is again applied. If still unsuccessful, then go back to (1). If successful, the deleted characters in due order are to form a word not experienced in the past.

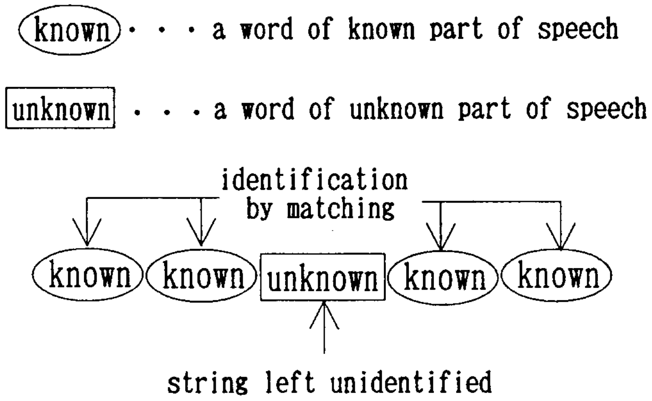

Fig. 3 Extraction of an unknown word

2-2 Inferencing the Part of Speech of an Unknown Word

If an unknown word is extracted as a morpheme, then its name of part of speech must be identified. Inferencing is made referring to rules of Japanese grammar. The parts of speech of two words put right before and right after that unknown word are all known from the above assumption for the location of an unknown word as in Fig. 4. This is the major information in identifying the part of speech of the unknown word. If the candidate of part of speech of the unknown word is perfectly identified, this is included in the knowledge of experienced part of speech as the one of that word. In the case of more than one candidates, they are also included in the knowledge together with that word. In both cases, the result of identification is available in later analysis.

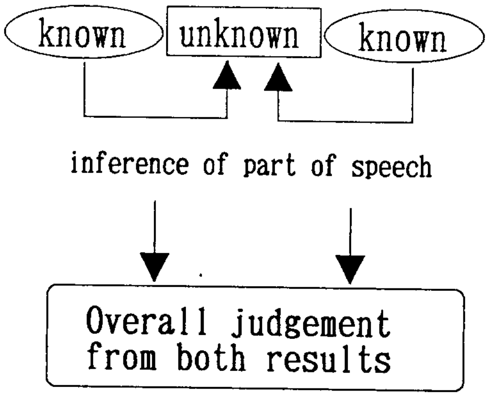

Fig. 4 Identification of the part of speech of an unknown word(s)

2-3 Parsing

In parsing, modification relations between any two words in the given sentence from the knowledge of morphemes identified in the above processing based on the knowledge of Japanese grammar (school grammar). Applicable parts of speech in this analysis was five (noun, "joshi," verb, adjective, and auxiliary verb) in the previous paper, and here in this paper the number of applicable parts of speech are increased to seven, revising the parsing rule. Newly included parts of speech are adverb and participial adjective.

3. IMAGE PROCESSING

An image here means a mental image. One can imagine a visual image of a situation from verbal explanation of that particular situation from experience. Since experienced image are memorized related to some verbal expressions, one can also find proper verbal expression of an image or a scene he/she is seeing. The image processing here is to provide predicates which reflect position relations between objects

One experiences similar scenes many times with position relations. Sometimes such position relations are also given some verbal expressions to identify them in the linguistic level. And, as already mentioned, this process is a kind of patternization because of the finiteness of the number of available words.

4. TRANSLATING VERBAL EXPRESSIONS TO AN IMAGE FROM EXPERIENCE

Corresponding conceptual idea of geometrically extracted position relations to words enables the translation from verbal expressions to an image or scene. The key in getting a verbal idea for position relation is experiencing others' usage of that particular words which describes a position relation together with scenes in which that word is used. Automatic acquisition of position relations in words here utilizes this fact, applying the notion of experience sequence. Experience sequence is an idea for simulating human experience and acquiring a subset of extension of a concept. [2, 3, 4, 5] Suppose experiencing a position relation P. An experience sequence of vectors $V_i = (x_i, y_i)$ which means $x_i P y_i$ is a list in the time order of experience.

$$(V_1, V_2, \ldots, V_n). \qquad \ldots(3)$$

This sequence is reorganized by using some (partial) order relation such as frequency of use. The last element (or a set of such elements) of the sequence thus obtained is called the representative. As an example, suppose a number of similar experiences as in Fig.6 are experienced by the present system, followed by the sentence "A pencil lies on the right side of a book." Then, the word processing gives the predicate form of the given sentence. The image processing also patternizes position relations given in the images as:

$$(\text{pencil, book}) = (x_1, y_1), \ldots, (x_n, y_n). \qquad \ldots(4)$$

Thus through the above reorganizing, representativity for the position relation in predicate form is extracted. A simulated result is shown in Fig.8. This result shows that the relation "A pencil is on the right side of a book." ranges the area of the graph shown for the location of the area of attention relative to a book. The darkest area is the typical area for the maintenance of the relation.

REFERENCES

[1] Y.Sakai, and K.Nomura, "The Verbal Aspect of Concept Understanding," Proc. of '92KACC, Seoul, Korea, Oct.19-21, 1992, pp.630-635.

[2] Y.Sakai, "The idea of Experience Sequence for Man-Machine Interfaces," Proc. of International Conf. on Manufacturing Systems and Environment, Tokyo, May.28-June.1, 1990, pp.319-323.

[3] Y.Sakai, and T.Murahashi, " Experience-Based Realization of Conceptual Functions for Man-Machine Interfaces, " Proc. of IECON'91, Kobe, Japan, Oct.28-Nov.1, 1991, pp.1605-1610.

[4] Y.Sakai, " Inference and Estimation Using Experience-Based Knowledge, " Proc. of '92KACC, Seoul, Korea, Oct.19-21, 1992, pp.636-641.

[5] Y.Sakai, "A Human-Oriented Mechanism for Building Expertise, " Proc. of 2nd IEEE International Workshop on Robot and Human Communication, Tokyo, Nov.3-5, 1993, pp.84-89.

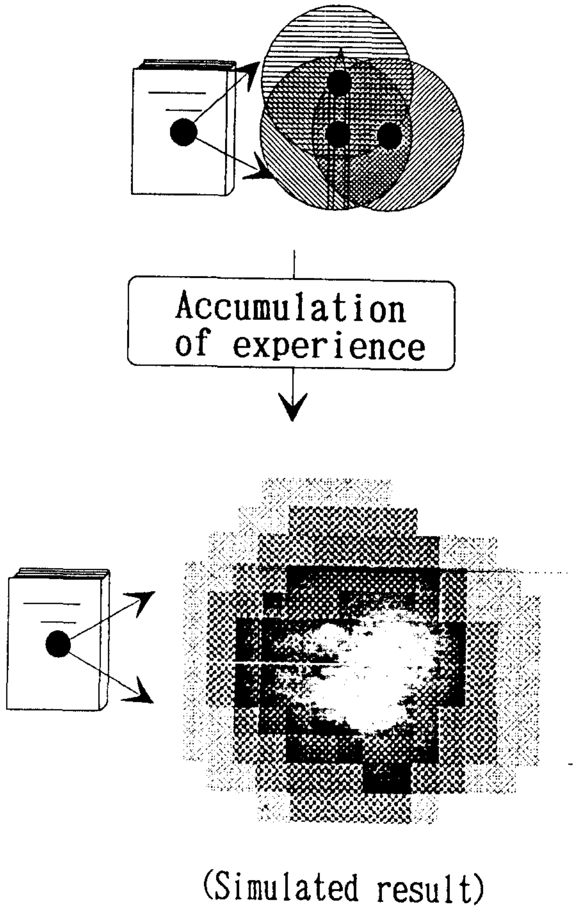

Fig.8 Allowable locations of a pencil for "on the right side of a book" from the system's experience

included in a given scene. Fig. 5 shows the flowchart of the image processing. The position relation between any two objects in a scene has much freedom while verbal expressions are finite and hence very restricted. But one can select proper descriptions of any specific position relation from experience. This is a process of patternizing geometrically given situations.

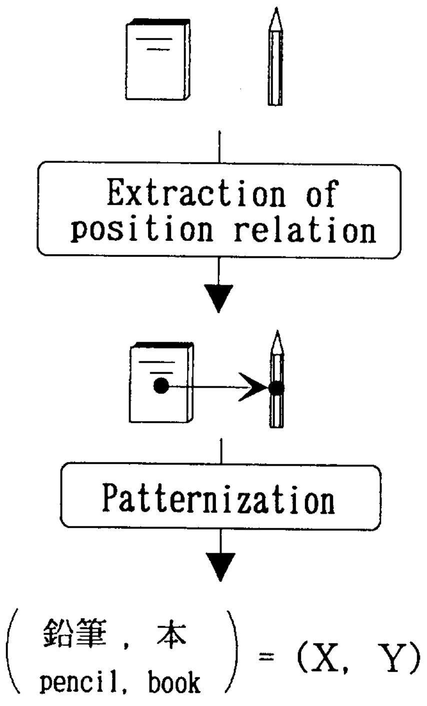

Fig. 5 Flow of the processing from imagerial side

3-1 Extracting Position Relations

In order to extract a position relation, some index points are necessary which represent positions of objects. For such an index point, any point in a particular object is applicable if it is chosen in order to satisfy some fixed conditions for index. In the present paper, the figure center is employed as index point. Suppose two objects A, B in a scene. An index point is given for each object in the form of coordinates (X, Y). Basically, the position relation P from the object A to B, APB, is described as:

$$APB = (x, y) \qquad \ldots (1)$$

where $x=X_B - X_A$; $y=Y_B - Y_A$; the index point of A is (X_A, Y_A), and that of B, (X_B, Y_B). This way of description is not good enough as a human-like use of position relation. A human may identify all position relations shown in Fig. 6 which are the same from the standpoint of expressing them in words. Moreover, a human does not employ such a point-to-point idea in evaluating the position relation between two objects. He/she utilizes more blurred information in order to grasp the position relation in some total fashion as the position relation between materials each of which has dimension. In other words, any point of some area of attention in his view is not differentiated from each other. Suppose such area of attention is shaped circular. Then the situation can be schematized as in Fig. 7. Thus letting an area of attention be R as in Fig. 7 and (X_B', Y_B') be the coordinate of any point in the area R, the following notation instead of Equation (1) is employed:

$$(A, B) = (X, Y) \qquad \ldots (2)$$

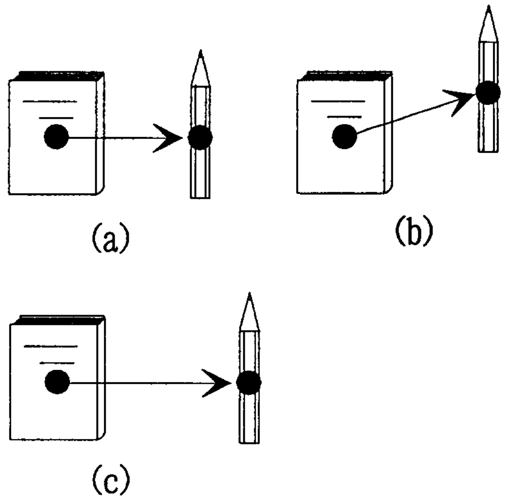

Fig. 6 Pairs of objects of different position relation from geometrical viewpoint

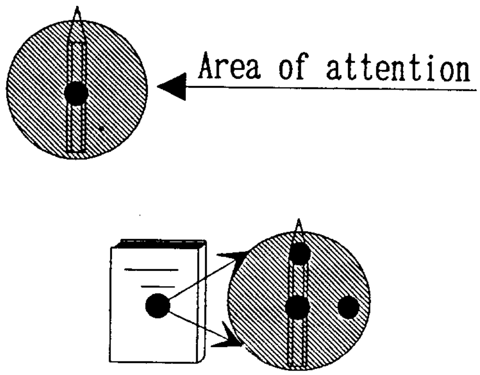

Fig. 7 Position relation as a set of vectors