'92 KACC 1992. 10. 19~21

The Verbal Aspect of Concept Understanding

Y.Sakai and K.Nomura

Dept. of Mechanical Engineering
Yamaguchi University
Ube, Yamaguchi 755, Japan

Abstract

Communication using a language sometimes leads to the partner's misconception, as the content that a language can describe is not sufficient enough to accurately transmit one's idea in his mind. In order to supplement this difficulty, introduction of idea of imagery to link with verbal information applying the notion of prototype which is learned through experiencing and is a part of the idea of experience sequence to deal with experiencing.

1. Introduction

Recently, computerized automatization has made systems highly advanced. In order for a human to operate machines in a system, it is necessary to learn the special language for it and/or special knowledge and characteristics of it, which requires a human operator more skill than before. To decrease such human load, an intelligent man-machine interface is necessary. This sort of interface should support a human intelligent activities, such as communication using a natural language, accumulation of experience together with concept understanding and evaluation, and reasoning based on common sense[1]. A modelling and algorithm for this sort of interface will be described.

2. Image and Language

In understanding concepts, there are two aspects; image and language[2]. The former can be performed by matching, with empirical knowledge, some image obtained by extracting features from the visual information, and it is made in the level of perception rather than in the cognitive level. Such imagerial understanding is more concrete than verbal understanding. Here, the authors stand on the viewpoints that (1) visual perception occupies much part of human perceptive information, and that (2) image formation based on visual perception is easier in modelling compared with those of other sources of perception. And from these assumptions, only visual image or figures are concerned here among all kind of images. The latter aspect of concept understanding, language, is the one in which an organized idea is obtained by denoting concepts as nodes in an abstract manner and by drawing relationships held between them. The verbal aspect may include something broader than the imagerial aspect in that verbal understanding may sometimes be built indirectly by only verbal information without referring to image, and that a variety of reasoning are available. It can be said that a human achieves many-sided concept

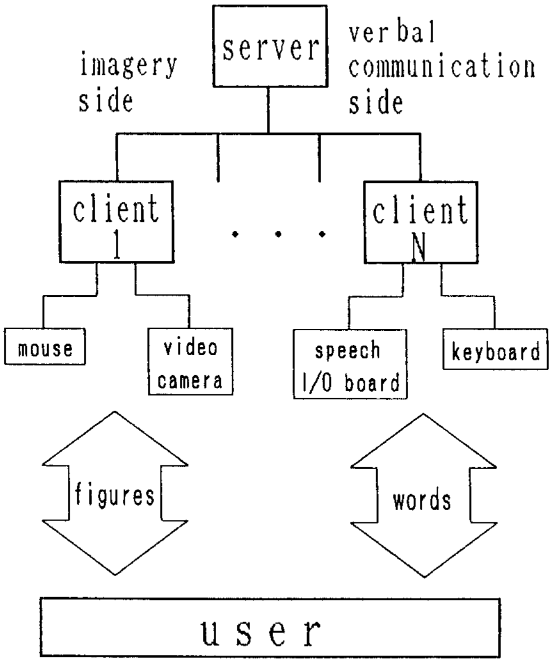

Fig. 1 Schematic of the interface used

understanding by utilizing bilateral translation between these two aspects. Shown in Fig.1 is the schematic of a LAN formed by personal computers. This system consits of clients (workstations) for the verbal aspect which deal with communication in Japanese through keyboards and speech I/O boards, those for the imagerial aspect which deal with visual information through a mouse and a video camera, and a server. Those inputted informations are analyzed and symbolized at clients, and then accumulated in the database. Referring to this database, utterances and images are interpretted interchangeably. This paper describes the methodology for obtaining images from inputted verbal information.

3. Concept Understanding Based on Verbal Information

In everyday life, one communicates using some (natural) language; conversing and writing, etc. One can represent and transfer information properly by using a language. To understand verbal information, it is necessary to know what context is used for the information. Unless so, the information may have multiple meaning and be ambiguous. For example, consider the following sentences:

An automobile is running. ...(1)
An automobile is
running in the racecourse. ...(2)

In reading Sentence(1) one has to imagine a suitable context and the imagination of the scene depends on the individual or more directly his experience, while in the case of Sentence(2) it is easier to imagine the scene.

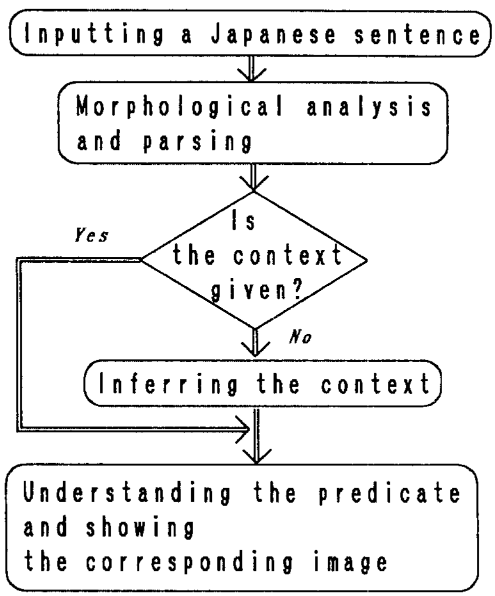

Fig. 2 Japanese language analysis

3.1 Natural Language Analysis[3]

Here in this paper, in order to simplify available expressions (Japanese) are restricted to those as follows:

(1) Parts of speech available in this system are nouns, pronouns, verbs, adjectives, "joshis" (= postpositional words functioning as an auxiliary to a main word), and participial adjectives.

(2) In the case of a conjugating part of speech, used is its end form or the form of a participial adjective.

(3) Commas are not used. Sentences of

"kanjis"(Chinese characters) with "hiraganas" are used.

3.1.1 Morpheme Analysis

A morpheme is a minimal unit of a sentence, which does not include any other meaningful character string in it. An instance of a morpheme is "hire" which means a fin of a fish. A morpheme does not always have a single form; for instance, "hire" and "bire." "Bire" appears as a part of, say, "o-bire" which is for a caudal fin in Japanese. They are different as character strings, but they indicate the same morpheme. A concrete form of a morpheme is called a morph. That is, a morpheme may have allomorphs like the above example, "hire" and "bire." A morph depends on tense, case, number, etc. Anyway some examples of morphemes in English are:

"in-" as in the word "indefinte,"
"ness-" as in "kindness,"
"-s" for plural which has allomorphs "-s" (including "-z") and "-es."

A morph can be taken into account just like the problem of conjugation. That is, if a morpheme has some allomorphs, then the morpheme can be regarded as the set of all its allomorphs. Hence the morpheme analysis need not directly be dealt with. Instead, it is sufficient to deal with identification of words.

For morpheme analysis, the following method is employed in the present system. An inputted sentence consists of "kanjis" and "hiraganas" as a string. A remarkable difference of Japanese from English is that Japanese need not require any blanks in a sentence. Let a sentence to be identified be denoted X_1 :

$$X_1 = c_1 c_2 \ldots c_n$$

And the following partial strings are also assumed:

$$c_1 c_2 \ldots\ldots c_n$$
$$c_1 c_2 \ldots c_{n-1}$$
$$\ldots\ldots$$
$$c_1 c_2$$
$$c_1$$

Let the set Z_1 be the set of all these partial strings. There is also considered the set A of words obtained through experiencing. Also, let the set C_1 be,

$$W_1 = \{x: x \in Z_1 \text{ and } x \in A\}.$$

W_1 is the set of candidates of the word which may be located at the top of the sentence X_1 . The longest string in the set W_1 will be employed as the best candidate. Let the word be denoted as w_1 . Then at this moment, the sentence is identified as

$$X_1 = w_1 X_2 .$$

In the same manner, X_2 , X_3 , X_m will be identified, one after another, as

$$X_2 = w_2 X_3$$
$$\ldots\ldots$$
$$X_{m-1} = w_{m-1} X_m$$
$$X_m = w_m .$$

Thus, the inputted sentence is denoted as

$$X_1 = w_1 w_2 \ldots w_{m-1} w_m .$$

If $W_j = \emptyset$, that is, proper word which should appear at the top of the partial string X_j is not found, then it means that it has not experienced the word to be identified (including its synonyms). So, the information is added to the system.

3.1.2 Parsing

For the sentence

$$X_1 = w_1 w_2 \ldots w_{m-1} w_m ,$$

Let the part of speech of w_i , $i=1,\ldots,m$, be p_i , $i=1,\ldots,m$. From the knowledge of p_i , the sentence is divided in to "bunsetsus." "A bunsetsu" is a fragment of a sentence which consists of

a "jiritsugo" + (0 or more)"fuzokugo"

where a "fuzokugo" is a "joshi" or "jodoushi; "a "jiritsugo" is an independent word which can be defined as words other than "fuzokugo." A "fuzokugo" is always used after a "jiritsugo" and then it adds some meaning to the "jiritsugo" as its main word. Or sometimes it indicates the relationship between "jiritsugos." Paying attention to the restrictions imposed for allowable sentences, a "bunsetsu" can have the following forms.

(1)noun+"joshi(s)" (2)noun+"jodoushi"
(3)verb (4)adjective.

Applying the knowledge above, X_1 can denoted as

$$X_1 = d_1 d_2 \ldots d_l\ ,\ l \leqq m.$$

As is well known, a sentence can have a predicate expression. A predicate, in this case, can be a verb, or an adjective, or a "jodoushi (=some special ones)." A complex sentence may be a complex sentence which includes more than two predicates. This kind of sentence structure can be analysed by the present system. The restriction (2) makes it possible. Anyway in that case, a (complex) sentence is divided into a set of simple sentences. Thus a sentence can be denoted as

$$X_1 = e_1 e_2 \ldots e_h \ldots e_q\ ,\ q \leqq l,$$

and each e_h is equivalent to a sentence. The last "bunsetsu" of e_h is the predicate of e_h as a sentence. The rest of the "bunsetsus" in e_h have the form (1)noun+"joshi" listed above. Finally, q predicate expressions are obtained from X_1.

4. Verbal Information Understanding Applying the Idea of Context

4.1 Experiencing and Concepts

A context can be given as an independent sentence which describes the situation, such as "The situation is ..." Or it is expressed using prepositions like "...in some place." From this viewpoint, a set of predicate expressions P_h as the alternative form of real inputted passage are examined if one of them includes an expression for context. If they do not include any information about context, then inference is necessary to obtain a proper context for those inputted sentences. For this purpose, accumulation of experience is included in the present system. The methodology for accumulation of experience here has its basis in the idea of experience sequence.[4,5] An experience sequence is, as its name indicates, a sequence of objects (or scenes or patterns or words or something) experienced and hence it indicates some order in which those objects are evaluated. The last term of an experience sequence is said to be a representative of that particular experiencing[6] in case where a particular ordering is employed. A natural ordering is the time order in which those pieces of experience are obtained. Other ordering can be available, if any, such as the frequency of occurrence. The notion of typicality can be expressed by the present notion of representative employing the frequency of occurrence; that is, the representative element of an experience sequence in the order of the frequency of occurrence implies the typicality. (This issue can be dealt with from the viewpoint of measure for experience sequence.[6]) The frequency ordering may give a partial order instead of a total order. In that case, the time ordering can be combined with the frequency ordering in order to avoid the difficulty. A database is

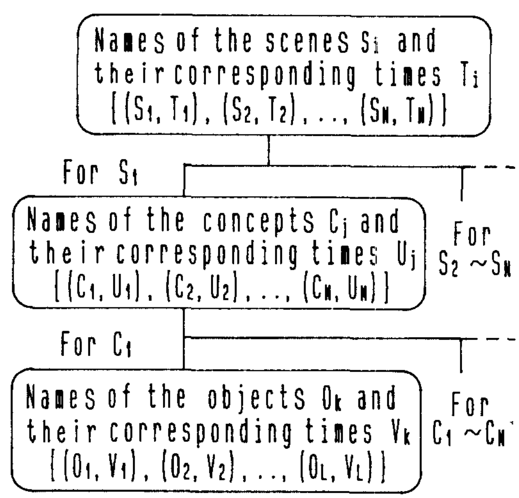

Fig. 3 Schematic of the database

built in the present system required in implementing the function of verbal information understanding by appying the notion of context. Fig.3 shows the schematic of the database. Concepts (objects) which constitute a scene together with the number of frequency experienced is filed for every context in which the objects are experienced. In its subordinate level of hierarchy, the instances of those particular concepts are stored also with the frequency of occurrence.

4.2 Context Inference and Predicate Interpretation for Correspondence between Image and Language

For incomplete information without any context, inference of context is necessary, as is already mentioned. Basically required thing for this processing is to draw some information about context from the concepts used in the sentence (or passage). For deeper analysis, the predicate(s) of the sentence(s) inputted must be taken into account. Here for a basic level, only the information from the concepts are dealt with. Let the set of the objects in the inputted sentence collected by the above analysis be

$$B=\{b_1, b_2, ..., b_K\},$$

and let the set of the names of contexts in the database as in Fig.3 be

$$C=\{c_1, c_2, ..., c_m\},$$

and further let the set of concepts experienced under the contexts in C be

$$D_{c_i}=\{d_{c_{i1}}, d_{c_{i2}}, ..., d_{c_{il}}\},$$

$i=1,...,m$. The number of times $e_{c_{ij}}$ of experiencing with respect to each of $b_1, b_2, ..., b_K$, can be found by referring to the database. Then the following value for evaluation is calculated:

$$E_{c_i}=\sum e_{c_{ij}},$$

where j ranges 1 to k. Among E_{c_i} thus obtained, let the largest one be E_{c_p}. The system makes decision that the proper context for the inputted sentence be c_p. An example is shown in Fig.4. If b_j's are instances, then the

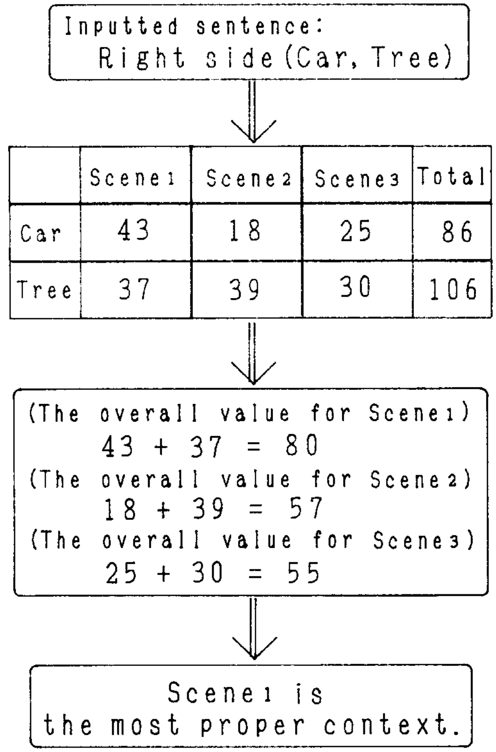

Fig. 4 Inference of context

inputted passage is sufficiently concrete. Unless so, b_j's indicate terms. (A term is defined to be the name of a concept.) In that case, one needs typical instances for those b_j's of which processing can be made by the methodology for typicality mentioned above. Fig.5 shows that a predicate P(x) can have some correspondence when the argument x is replaced by some concrete instance. In Fig.5, only the case of 1-place predicate is shown, but

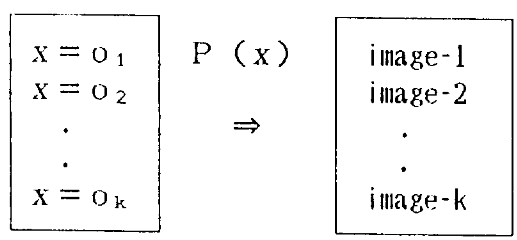

Fig. 5 Predicates relating to images

in general, predicates can be n-place ones for which a single argument is replaced by an ordered tuple. A database system can be offered by a network system.[7]

Conclusion

Communication using a language may be clearer than that by something schematic or imagerial. But at the same time the expression may be abstract and may include multiple interpretations of it. This fact sometimes leads to misunderstanding. Bilateral translation of information between image and language[7] is crucial in order to greatly reduce this sort of inconvenience. Incorporating the mechanism for reasoning by common sense for better reasoning is also effective. In the present paper, described is the basic methodology for the intelligent interface based on the authors' theory.

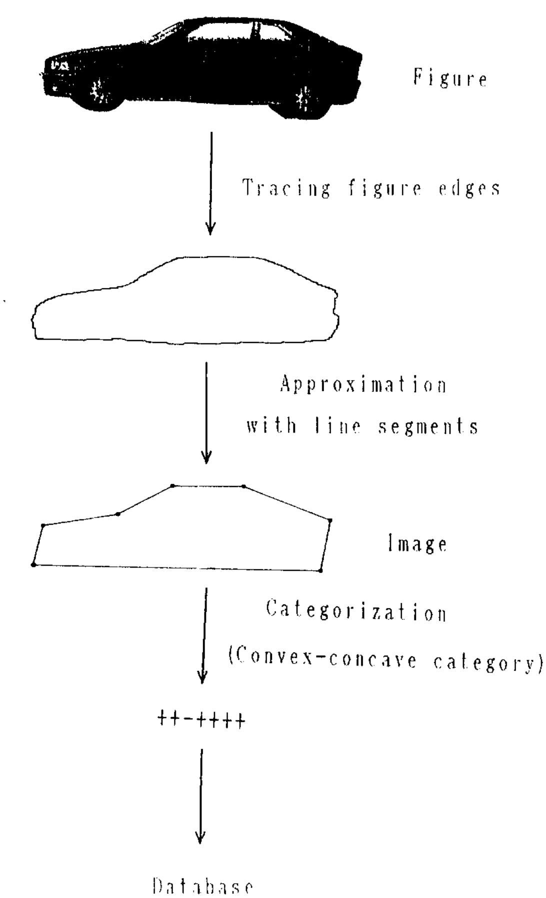

Fig. 6 An example of image processing

Acknowledgment

This work is supported in part by Japanese Government subsidy for aiding scientific researches. The authors greatly acknowledge this support.

References

[1] Y.Sakai, and K.Ohkusa, "On a Control System for Cutting process," Manufacturing systems, Vol.14, No.3, 1986, pp.287-293.

[2] For example, D.A.Norman, "Learning and Memory," W.H.Freeman and Company, Chapt.12, 1982.

[3] Y.Sakai, T.Murahashi, and K.Nomura, "Figure Understanding from Linguistic Information," Proc. of 35th ISCIE, May22-24, (5001), 1991.

[4] Y.Sakai, "The idea of Experience Sequence for Man-Machine Interfaces," Proc. of Int. Conf. on Manufacturing Systems and Environment, May28-June1, Tokyo, 1990, pp.319-323.

[5] Y.Sakai, "Inference and Estimation Using Experience-Based Knowledge," This Conference(IS 20), 1992.

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

[7] Y.Sakai, and M.Kitazawa, "Personal Computer Networking for Concept Understanding," This Conference, 1992.