Conceptual Lexicon Using an Object-Oriented Language

Shoichi YOKOYAMA
Electrotechnical Laboratory
Tsukuba, Ibaraki, Japan

Kenji HANAKATA
Universität Stuttgart
Stuttgart, F. R. Germany

Abstract

This paper describes the construction of a lexicon representing abstract concepts. This lexicon is written by an object-oriented language, CTALK, and forms a dynamic network system controlled by object-oriented mechanisms. The content of the lexicon is constructed using a Japanese dictionary. First, entry words and their definition parts are derived from the dictionary. Second, syntactic and semantic information is analyzed from these parts. Finally, superconcepts are assigned in the superconcept part in an object, static parts to the slot values, and dynamic operations to the message parts, respectively. One word has one object in a world, but through the superconcept part and slot part this connects to the subconcept of other words and worlds. When relative concepts are accumulated, the result will be a model of human thoughts which have conscious and unconscious parts.

1. Introduction

Semantic relations among words or concepts have been represented as a slot-filler definition in the lexicon [Bobrow 77]. These relations are usually represented as frame lists in LISP or Prolog. When the representation of a large-scale lexicon is needed, it is difficult to make and maintain the lexicon because the relation of super/subconcepts dramatically increases.

An object-oriented language has a strong inference mechanism. It has already been used for parsing sentences [Hayes 84, Nishida 84]. It can also describe the representation of semantic relations more simply than LISP. CTALK is such an object-oriented language and can dynamically handle the network using the concept of a world -- a group of objects -- and an object.

Now the systematization of semantic relations is made for the words which represent concrete objects. However, the organization of abstract words is difficult because it is not so clear how to obtain and deal with abstract concepts. On the other hand, research is being conducted to computerize or categorize the common dictionary [Yokoyama 77, Tsurumaru 84].

This paper describes a trial of systematization and categorization of abstract words. The procedure is as follows: first, the definition parts of abstract words in the dictionary are semantically analyzed, and the semantic relations are derived. Second, an object is established for an abstract word. Third, the super/subconcepts of these words are assigned to a father-son relation among the objects corresponding to the semantic network. Then the slot definitions are assigned to the object. From these slot values, new objects are made under the relative objects. Finally, the network is constructed from these abstract words.

The network formed by this method is considered to be a model of human consciousness. That is, when a human is reminded of a concept, other concepts relative to it are simultaneously extracted from the memory and stay beneath the main memory. This approach is considered to be an implementation of interactive activation model [Rumelhart 82] in abstract concept level.

2. Object-oriented language CTALK

The object-oriented language, CTALK, was developed at Universität Stuttgart [Hanakata 84]. In this section, a simple summary of CTALK is described.

The main elements of CTALK are worlds and objects. A world is a set of objects, and some active worlds stay in the main memory. When the number of worlds exceeds the constant defined by the system, the least recent worlds are gradually swapped into the secondary memory. The hierarchy or the network among the objects, that is, the father-son relation of the objects is simply set to every object, independent of the world structure.

An object consists of four parts, SUPERC, SONS, SLOTS, and METHODS. SUPERC means the superconcepts, that is, the fathers of the object. SONS, literally, means the sons of the object. Actually, in the input of the object, the names of sons are not specified explicitly. The control of these father-son relations are performed on the bit table. The use of this bit table enables the fast operation of creating and removing these relations, and also of treating inheritance mechanisms. The inheritance usually works inside a world, and can also be spread to other worlds. The direction of inheritance can be specified by VIEWED-AS operator [Selbmann 85], with which an object inherits the features of specified father(s).

SLOTS is a static property of an object. The form of SLOTS is as follows:

    [slotname (restriction) : default-value]
    [slotname slot-value]

The restriction and default-value are optional, and slot-value must fill the restriction in the same slotname in the antecedent object. If a conflict occurs at value assignment, no value can be assigned when the restriction is violated. When there is no restriction or the restriction is filled, another default-value can be assigned in a descendant object.

METHODS is the message part of the object. Sending the message invokes the operation written in the METHODS part. Of course, logical functions such as "if", "then", and "else", and numerical functions such as "equal", "greq" (greater than or equal) can be written in the METHODS part. User functions can be defined using system functions and other user functions.

3. Construction of the conceptual network

3.1. Procedure

The procedure to make the conceptual network is as follows:

1. The definition part for an abstract word is derived from the corresponding entry word in a dictionary [Kindaichi 74], and semantically analyzed.
2. A new world is made for an entry word.
3. The top-level object named TOP is made. This is the control object for message passing and slot handling. The father object of TOP is now CTALK_IN_CTALK (default highest object in the whole world). All objects in the same world are the sons of TOP.
4. An object corresponding to the entry word is made as a son of TOP.
5. If the evident superconcept is derived, it is explicitly written in SUPERC part. In the case of an isa relation (a change to another word), the slot value "isa" is filled for the corresponding word.
6. Objects of the words written in the definition part of the entry words are made in the world of the entry word. These objects have an object with the same name as the father object in another world, that is, a superconcept.
7. The arc relations relative to the son object are assigned to the slot name, and the slot value is the corresponding word at the other side of the arc.

Finally, one world is formed for one entry word as a "lump" of the concepts. In the following sections, the details of this procedure is described.

3.2. Making the objects

Fig.1(a) shows the top-level object TOP in the world "mokuteki" (object or aim). Its superconcept is now CTALK_IN_CTALK, but if two or more conceptual words can be sublated ("aufheben"), the integrated concept becomes a new superconcept. The sons are the entry word itself and the objects (words) which are derived from the definition part of the entry word. These form the components in the semantic network. In the slot "mean", the literal meaning cited from the dictionary is written, and used as a normal search (i.e. consultation) of the dictionary. In order to consult the dictionary, the following message should be sent to CTALK_IN_CTALK:

[CTALK mean mokuteki].

The operation written in METHODS part in CTALK_IN_CTALK:

```
[mean ?x => (* do: (x chw:)
                   (TOP gv: mean))]
```

is invoked. The variable x is connected to "mokuteki", and the literal meaning is output. "* do:" is a PROGN-like function in LISP, and the return value is the evaluation of the last part of the following functions. "Chw:" is a system function changing a world, and "gv:" takes a slot-value.

```
TOP_IN_mokuteki
 SUPERC
    ( CTALK_IN_CTALK )
 SONS
    ( koto_IN_mokuteki )
    ( mono_IN_mokuteki )
    ( mokuteki_IN_mokuteki )
    ...
 SLOTS
    ( mean "kodosuru mokuhyotosite kangaerareta
    sositai nanigotoka, sonaritai nanimonoka" )
  /* something considered as a behavioral aim */
 METHODS
```

Fig. 1 (a) Object TOP_IN_mokuteki.

```
mokuteki_IN_mokuteki
 SUPERC
    ( TOP_IN_mokuteki )
    ( koto_IN_mokuteki )
    ( mono_IN_mokuteki )
 SONS
    ( mokuteki_IN_syusi )
    ( mokuteki_IN_meate )
 ...
```

Fig. 1 (b) Object mokuteki_IN_mokuteki.

```
koto_IN_mokuteki
 SUPERC
    ( TOP_IN_mokuteki )
    ( koto_IN_koto )
 SONS
    ( mokuteki_IN_mokuteki )
 SLOTS
    ( JBO isi )
    ( OR mono )
    ( YNA nani )
 METHODS
```

Fig. 1 (c) Object koto_IN_mokuteki.

Fig.1(b) shows the object of the entry word "mokuteki" (object or aim). Many of the entry words in the dictionary are explained as the subcategory of "mono" or "koto". "Mono" represents a concrete thing or object in many cases, but also represents the abstract things in some cases like in "mokuteki". On the other hand, "koto" represents abstract things or matters in almost all cases. According to the dictionary, both "mono" and "koto" are assigned as SUPERC's here, but it could be possible to make only "koto" a SUPERC. For the problems concerning the definition power of the dictionary, see **Conclusion**. The entry object can be connected to its relative words as SUPERC relations. In SONS part, the words referring to "mokuteki" are described. Here the link from "mokuteki"_IN_"syusi" (purpose) and "mokuteki"_IN_"meate" (guide or aim) is connected. The more complex the semantic network is, the greater the number of SONS objects is. In SLOTS part, there is no slot name nor value because no "isa" relation exists in the definition part in the dictionary.

Fig.1(c) shows a part of objects which should be made on the world "mokuteki". "Koto" is one of the superconcepts of "mokuteki", and by itself a subconcept of "TOP" and "koto_IN_koto". 3 slot names are assigned at SLOTS part, that is JBO, OR, and YNA. The strange names such as JBO and YNA indicate the reverse direction of arcs. This guarantees the two way relation of objects. Slot value is, of course, the corresponding word name. In the case of "kodo" (behavior) the structure is similar to "koto".

3.3. World "mokuteki"

Fig.2 shows a part of a conceptual diagram for the world "mokuteki" after all relative objects have been made. --→ indicates the arc from father objects in other worlds, --→ the arc to son objects in other worlds, and —→ the connection inside the world. When new terms or example sentences for the word are added to this world, the structure inside the world will change. On the other hand, when new entry words are added to the system, the number of worlds increases. If these words are relative to the world "mokuteki" or refer to it, connection arcs are newly formed, and the network becomes more complex.

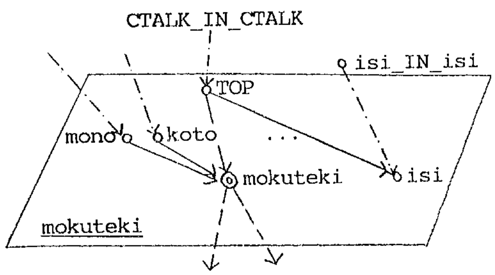

Fig. 2 Part of the structure of world "mokuteki".

4. Conclusion

The world shown in Fig.2 is considered to be a model of the human conscious process. When the human considers a certain concept, the central concept is fetched to the main memory. That concept corresponds to a world in this system. Relative concepts are connected to this concept and exist near (or beneath) the main concept. Some of these concepts often go to the surface of consciousness, but usually stay beneath the consciousness. These concepts correspond to other worlds connected to this world.

In the current system, relatively static properties are defined, and only a few dynamic connections are used. Dynamic functions in CTALK including restriction conditions in SLOTS part should be considered in the next version.

If a similar structure is found between two or more worlds, integration of these worlds can be considered. On the other hand, when one world is too big to be treated, the separation of the world would occur. Separation and integration processes of worlds are also considered as a model of human thinking. However, the automatization of the process is very difficult.

Knowledge for the world is now derived from a Japanese dictionary. One reason for using the dictionary is that it is difficult to deal with the words having abstract concepts, and the dictionary is one of the most convenient clues for treatment. The other reason is that this dictionary has been computerized [Yokoyama 77] and will be used for automatic semantic analysis. However, definitions of entry words in a dictionary have various problems such as definition, semantic elements, and co-reference [Nakano 85]. Description of other dictionaries should be discussed, and the efforts to find suitable representation and connection will be continued.

Acknowledgements

This research started at Universität Stuttgart when Yokoyama stayed there as a visiting researcher sent by the Japanese Government. The authors express their appreciation to researchers in Universität Stuttgart and Electrotechnical Laboratory, as well as staffs in both governments.

References

[Bobrow 77] Bobrow, Daniel G. and Winograd, Terry: "An Overview of KRL, a Knowledge Representation Language", Cognitive Science, Vol.1, No.1, pp.3-46 (1977).

[Hanakata 84] Hanakata, Kenji: "A Powerful Agent for the Control of Inheritance and Message Passing in the Object-Oriented Language CTALK", GWAI-84 (1984).

[Hayes 84] Hayes, Philip J.: "Entity-Oriented Parsing", COLING-84, pp.212-217 (1984).

[Ishiwata 84] Ishiwata, Toshio: "Features in Japanese Representation" (in Japanese) in Tanaka, Akio et al: "Performance II", Asakura Publishing Co. (1984).

[Kindaichi 74] Kindaichi, Kyosuke et al. (eds.): "Shin-Meikai Kokugo Jiten" (Second Edition) (Concise Japanese Dictionary) (in Japanese), Sansei-do (1974).

[Nakano 85] Nakano, Hiroshi: "Problems about the Description of Semantic Definitions" (in Japanese) in Kusanagi et al.: "Grammar and Semantics II", Asakura Publishing Co. (1985).

[Nishida 84] Nishida, Toyoaki and Doshita, Shuji: "Combining Functionality and Object-Orientedness for Natural Language Processing", COLING-84, pp.218-221 (1984).

[Rumelhart 82] Rumelhart, David E. and McClelland, James L.: "An Interactive Activation Model of Context Effects in Letter Perception: Part 2", Psychological Review, Vol.89, No.1, pp.60-94 (1982).

[Selbmann 85] Selbmann, Frank: "Entwurf und Implementierung eines kontrollierbaren Backtrack Mechanismusses so wie einer universell verwendbaren undo-Funktion für die objekt- und botschaftsorientierte Sprache CTALK" (in German), Diplomarbeit der Universität Stuttgart (1985).

[Tsurumaru 84] Tsurumaru, Hiroaki et al.: "A Consideration about the Development of a Semantic Dictionary" (in Japanese), 30th Conference of IPSJ, 5K-7 (1985.3).

[Yokoyama 77] Yokoyama, Shoichi: "Preparation for the Database Management of a Japanese Dictionary" (in Japanese), Bul. Electrotechnical Laboratory, Vol.41, No.11, pp.855-863 (1977).