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

Research and Development of an Electronic Dictionary: Current Status and Future Plan

Naohiro KAKIZAKI

Executive Director, Japan Electronic Dictionary Research Institute

1. Introduction

The social application of computers has been continued to widen and diversify. A technology enabling the computer to understand and process the language people use, natural language processing technology, is considered fundamental among the technologies necessary to further expand computer applications, because it will optimize the man-machine interface.

The need for application of this technology is found in every language-using aspect of society, for example:

* Advanced document processing systems such as translation, information retrieval and abstracting
* Sophisticated interface system such as language-understanding input and sentence-generating output
* Advanced language education systems such as intelligent CAI, etc.

To make good use of current language processing technology, the Electronic Dictionary Research Institute (EDR) is developing a dictionary enabling a computer to accurately understand natural language data and in high volume, that is, a large-scale electronic dictionary.

2. Electronic Dictionary

An electronic dictionary is not a machine readable version of conventional dictionaries. It is a completely new dictionary designed from scratch. It is a dictionary developed for a computer to understand a natural language.

In order for human and the computer to understand a text written in a natural language, it is necessary to know the meaning of a word and the meaning within the context it is used. In an electronic dictionary, the concept expressed by a word (concept), grammatical characteristics of the word when it expresses the concept, expressing the particular use of the word, and the knowledge necessary for understanding the concept will be written in an form understandable by the computer. On the other hand, the computer will have the information necessary for understanding the use of words, in the form of grammar rules.

3. Composition of Electronic Dictionary

The electronic dictionary is composed of a word dictionary and concept dictionary. The word dictionary contains information concerning words themselves and the concept dictionary contains information about concepts.

The word dictionary contains the following elements: a headword which is the surface expression of the words, concepts to be expressed by the headword, and various grammatical characteristics of the word when the headword expresses certain concept.

The concept expressed by a word is defined by a sentence which so that humans can distinguish it from other concepts. This will be used as an entry in the concept dictionary. In *a* compound word, how the concepts expressed by the compound word is expressed by the concepts of component words.

The concept dictionary defined the possible relation between two

concepts and concept is defined by other concepts. These relation is expressed as a binary relationship or a set of binary relationships. As to the relation between concepts, there are case relations, causal relations, synonym, superior and inferior relations. The relationship between the word dictionary and concept dictionary is shown in Figure 1.

Word dictionary is divided into two dictionaries. Basic dictionary and terminology dictionary. And Japanese and English words are the subjects the word dictionary, there are 8 word dictionaries.

The concept dictionary is divided into two dictionaries: and concept classification dictionaries which defines superior and inferior relation, and description dictionary which defines other possible relations between concepts.

4. Development method

The biggest problem in the development of a large-scale electronic dictionary is unity and uniformity of the information and the accuracy of such information.

When making the word dictionary, a work sheet is used, as shown in Figure 3, description is made per concept contained in a word, and the description is input into the computer. Using the dictionary editor in the computer, several editing and correction will be done as well as establishing relation among the compound words, and the words comprising the compound words, and translated words. Unity and uniformity will be kept by using a person who is knowledgeable in dictionary editing to develop it. And accuracy will be kept by verifying the dictionary with the analysis results of the words by the computer, using a large-scale textbase.

The concept dictionary is developed by mainly using the data obtained by verifying the dictionary with the analysis results of the word by the computer, using a large-scale textbase. In order to evaluate the accuracy and effectiveness of an dictionary application systems, such as a machine translation system, voice recognition system, and information retrieval systems, are developed. Based on the evaluation data of these application systems, the dictionaries will be corrected and improved upon.

The goal, current status, and development schedule of each dictionaries are shown in Figure 3.

5. The Use of Dictionaries

The electronic dictionaries being developed at EDR contains information necessary for the computer to understand natural languages and they can be used in various applications such as: machine translation, question answering system, and information retrieval systems. However, of all the applications systems, we consider the machine translation systems using an interlingua is the main goal.

In the text analysis, morphological analysis and syntactic analysis is performed by using the grammatical characteristics of the word dictionary. The ambiguities arising in the analysis process will be dissolved or reduced by mapping the analysis result to the model given as concept description.

6. Conclusion

EDR is developing electronic dictionaries for Japanese and English. In order to extend the object language for the multi-language machine translation system using the interlingua method, we need to make electronic dictionaries for other languages.

In the machine translation system using our electronic dictionaries, the link between the languages is established via synonymous, superior, and inferior relationships, etc in the concept dictionary. From now on, we hope that when someone develops dictionaries, the interface will be kept, making the language link wider.

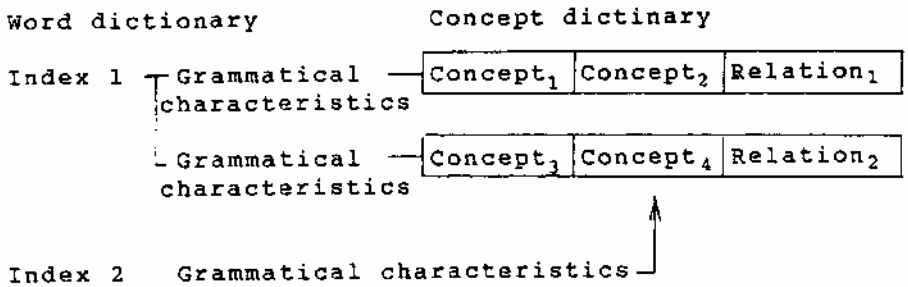

Fig. 1 Composition of elelctronic dictionary

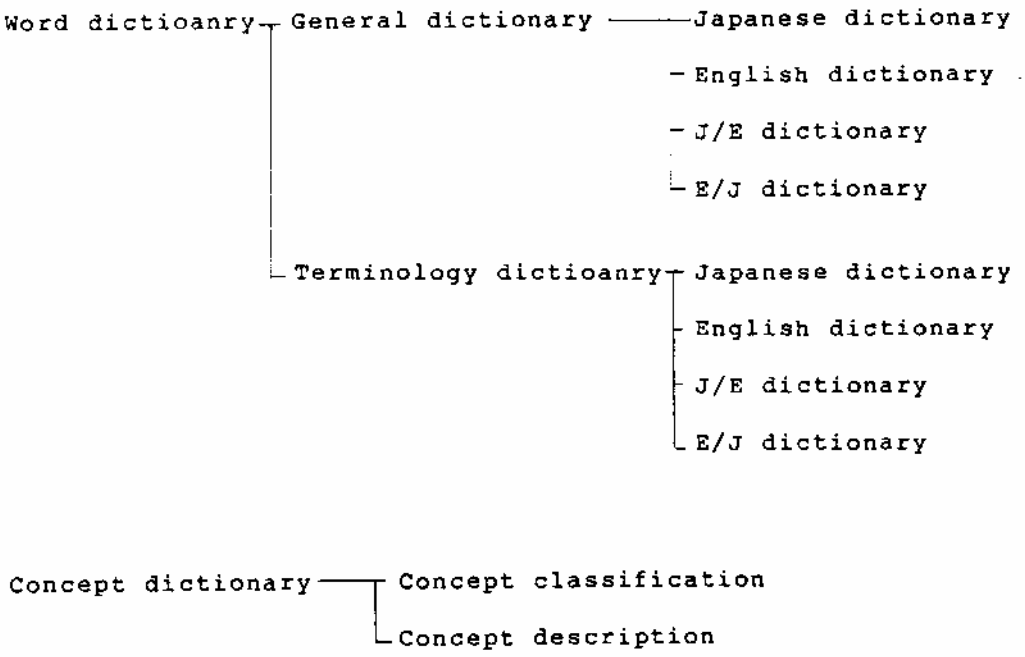

Fig. 2 Types of elelctonic dictionaries

Fig. 3 Research and Development Schedule

Sub-theme \ Year	1986 fiscal year	1987 fiscal year	1988 fiscal year	1989 fiscal year	1990 fiscal year	1991 fiscal year	1992 fiscal year	1993 fiscal year	1994 fiscal year
General dictionary (J,E,J/E,E/J)	Prototype 22,000 words	Prototype 100,000 words	Prototype	Evaluation	Improvement /Expansion	Improvement /Expansion	Improvement /Expansion· Evaluation	(Goal) 200,000 words for each dictionary	
Terminology dictionary (J,E,J/E,E/J)		Prototype	Prototype	Prototype	Evaluation	Improvement /Expansion	Improvement /Expansion· Evaluation	(Goal) 100,000 words for each dictionary	
Concept Classification	Experiment /Design	Primary prototype	Primary prototype	Evaluation	Secondary prototype	Improvement /Expansion	Improvement /Expansion· Evaluation		
Concept description	Experiment /Design	Primary prototype	Evaluation	Secondary prototype	Secondary prototype	Secondary prototype	Evaluation	Improvement /Expansion· Evaluation	Improvement /Expansion· Evaluation
Data management system	Design/ Prototype	Improvement /Expansion	Design/ Prototype	Prototype	Improvement /Expansion	Improvement /Expansion	Improvement /Expansion		
Testing and evaluation system		Design/ Prototype	Primary prototype	Evaluation/ Prototype	Secondary prototype	Secondary prototype	Actual use /Evaluation	Actual use /Evaluation	Actual use /Evaluation