

[From: IFTT'89, April 26-28, Oiso Prince Hotel, Japan]

Electronic Dictionary: a Language Knowledge Base

Toshio Yokoi

Japan Electronic Dictionary Research Institute, Ltd.

Mita Kokusai Bldg. Mita 1-chome, Minato-ku, Tokyo 108 Japan

1. The role of an electronic dictionary

The role of dictionary in a machine translation system is decisive. Therefore, a large-scale and high-quality electronic dictionary is indispensable for the realization of more advanced machine translation. Improvement of the dictionary is the most effective means to advance the established system step-by-step. And, the development of a high-level dictionary will be the main theme in the R&D for a new and more advanced system.

Machine translation systems are the main theme in the field of information processing, and continuous efforts should be exerted in the R&D of this field in the future. There are many unknown problems in the R&D of more advanced machine translation technology. Therefore, we cannot avoid a wide variety of trials and errors. To undergo these trials and errors effectively and painlessly, a common R&D base should be prepared. This common base is the data base on linguistic phenomena - the electronic dictionary.

Among many subjects for R&D, the one that requires special focus is the establishment of a full-scale technology for semantic processing. For this purpose, the electronic dictionary is indispensable as a linguistic knowledge base describing the semantics of language.

2. The EDR electronic dictionary project

The Electronic Dictionary Project began in April 1986 with the target of the research and development of a large-scale electronic dictionary. The Japan Electronic Dictionary Research Institute, Ltd. (EDR) was established for the project's promotion. The institute is jointly funded by the Japan Key Technology Center (Japanese government) and eight major computer companies. The period for R&D is nine years and the budget is about ¥14 billion.

For the Word Dictionary on general vocabulary, the raw data has been prepared by subcontractors. The development of the dictionary has just commenced - mainly by the staff of the institute utilizing an integrated support system for electronic dictionary development. For the Word Dictionary of terminology for information processing and the Concept Classification, raw data preparation is nearly complete. The raw data preparation for Concept Description began in the 1988 fiscal year.

R&D at the institute is mainly done by the staff. It also includes, however, joint research with ETL, cooperative research with the Fifth-Generation Computer Project of ICOT, CICC's Machine Translation System R&D Project, and research entrusted to universities. To obtain the various data - the base for development - the preparation of the dictionary raw data is subcontracted to representative publishing companies; the next data is purchased from newspaper

companies, and so forth.

The decision on when and how the results of the project should be made public is scheduled to be made during fiscal 1989. The specifications of the dictionary, however, and the dictionary interface and academic papers are to be released widely not only in Japan but also overseas. It is noteworthy that the dictionary interface comprises a complete set of word and concept heads, and that a dictionary made in accordance with this interface can be integrated with the EDR electronic dictionary.

In any case, we at EDR are exerting every possible effort to make a better electronic dictionary. Our goal is to encourage and disseminate dictionary development activities and to promote diversified utilization of the electronic dictionary.