PAPER

East-West Communication and Information Transfer -- Coordination of Specificity and Generality[†1]

Shizuo Fujiwara[†2]

Japanese cultural traditions underlie two recent projects in information science. First, Japan' s scholarly history has contributed to the development of improved information flow and resource sharing, in the form of a list of basic scientific and technical terms. Second, Japan' s unique three-part language, consisting of kanji, hiragana and katakana, has provided a model for future information management systems, which should seek to integrate the three elements of information: ideographic, symbolic and graphic.

Another area of tremendous societal importance, and therefore a major concern of information science, is safety control and risk management (SCRM). SCRM requires coordination between the various disciplines (the science of specificity) and information science (the science of generality). Coordination of specificity and generality is essential for further advancement in information science, and for finding the solutions to the challenges of SCRM.

1 Standardization of the basic scientific and technical terms

1.1 Historical background of the list of basic scientific and technical terms

The present effort to standardize the basic scientific and technical terms is of fundamental importance for presenting, transferring, and sharing information and information resources. This work began in the 17th century, when Japanese scholars translated the terms used in Western science, and then introduced these terms into the existing framework of traditional Japanese academic society. The first work of these scholars was to translate modern science into the Japanese language, i.e., to present Western science in kanji characters, a process which continues to this day.

Interestingly enough, these scholars often did not know the meaning of the scientific terms they were translating. However, the kanji characters they developed usually held the correct meaning. In fact, many of the characters are still alive and are used in text books and articles written in Japanese today. Thus, those early efforts have borne fruit for almost 400 years. Over the centuries, Japanese scholars have translated about 150,000 basic scientific terms in 32 disciplinary fields.

The scholars in each discipline have worked to identify the basic terms which represent the contents of their respective fields, and have encouraged others to use those terms when writing articles or teaching. Furthermore, English-Japanese and Japanese-English lists of these basic terms have been compiled. The administrative and publication costs of this work have often been borne by the Ministry of Education. Thus, due to intradisciplinary encouragement and financial support, each

†1 Presented as the plenary lecture at the 13th Congress of CODATA in Beijing on Oct. 19, 1992.

†2 Research Institute for Information and Knowledge, Kanagawa University, 2948, Tsuchiya, Hiratsuka, Japan

discipline has acquired standard, basic terms which represent the knowledge base of that discipline.

1.2 Production of a unified list of basic scientific and technical terms

However, these efforts have encountered serious difficulties in these days of computer-controlled databases and interdisciplinary research. The problem has arisen because the 32 learned societies have selected their basic terms independently of each other. Thus, different Japanese expressions often correspond to the same English term. This has produced serious difficulty in interdisciplinary communication, i.e., resource sharing. Accordingly, efforts have been made to combine these multiple Japanese expressions. Unfortunately, these efforts did not produce practical solutions.

Prof. Yuzuru Fujiwara and I have taken a practical approach to this problem. We have placed all of the terms chosen by the different learned societies into a single computer file: the 'Unified' list of the basic scientific and technical terms. This Unified list enables scholars to identify all of the different Japanese expressions which correspond to a single English term. The list is published by Kanagawa University and distributed by Kinokuniya Publishing Company.[1)] The list contains approximately 100,000 terms which have been chosen by 25 learned societies as the basic terms in each society's disciplinary field. Since this list was originally produced, more societies have compiled basic terms for their respective fields. About 32 societies currently have standardized lists of basic terms.

1.3 Preparation of a list of multi-field common basic scientific and technical terms

We then identified the approximately 15,000 terms in the Unified list which were used in more than two fields. These were called the "multi-field common" terms. The trivial terms were removed from these 15,000, resulting in a list of approximately 5,000 multi-field common basic scientific and technical terms. The terms in this list can still be considered "basic," because they were originally chosen by the learned societies as being basic for presenting the contents of each discipline. However, the list lacks terms which are unique to individual fields. This list was published in 1986 by FID, Federation Internationale de Documentation et Information, as FID publication No. 633.[2)]

1.4 Preparation of a multilingual list of multi-field common basic scientific and technical terms

Fujiwara then worked to translate the list of multi-field common basic terms into French, German and Spanish. Mr. Marcel Thomas of the Biblioteque Nationale, Prof. J. -E Dubois of the Universite de Paris VII, Prof. P. Canisius of Bundesanstalt fuer Strassenwesen, and Prof. E. Curras of Universidad Automona de Madrid, assisted in the translation. This work was published in March 1992 by RIIK of Kanagawa University.[3)]

Needless to say, standardization of scientific and technical terms is the key for database construction, and transfer, communication, and retrieval of information resources. The process of standardization entails the identification of the basic terms within each disciplinary field, the production of a unified list, and, finally, the preparation of multi-field common and multilingual list. The work begun by 17th-century Japanese scholars has continued to the present, when the task of handling these basic terms can be shouldered by machines, thanks to the development of the Japanese computer, with which my colleagues and I produced the first computerized file of these basic terms in early 1975.

2 Analysis of generality and specificity in terminology

2.1 Presentation of the problem

While performing the work mentioned above, particularly while preparing the multilingual list, we realized that we often encountered a problem: i.e., a single English term may be translated different fields. For example, a single English term can be translated differently in French, German, or Spanish depending upon the field. The situation is similar to what we Japanese have encountered while translating English into Japanese. We initially believed that, in Japan, this problem arose because the translators in the different fields were all working independently. However, this does not appear to be the case. Even in Europe, terms have been translated differently depending upon the field in which they are to be applied.

Thus, while preparing the multilingual, multi-field common list, I asked each of the collaborators to provide different translations of each term, if applicable, for the fields of mathematics, physics, chemistry, biology, technology and general science. This list of scientific and technical terms is the first which reflects a consideration for field-specific translations.

The nature of the problem cited above is that even some of the "basic" scientific and technical terms, which supposedly are general in nature still have specific meanings which depend upon the field in which they are used. This suggests that all scientific and technical terms have dual aspects of specificity and generality.

Good translation if essential if we are to realize better communication, transfer, and exchange of knowledge between people in different locations or different fields, and to ensure good translation, we must have some means of overcoming the obstacle of specificity. However, as was discovered in the preparation of the Unified list, this task is not an easy one.

Instead of combining the field-specific differences, as was done with the Unified list, it would be desirable to construct a computerized information management system. The system should contain all of the field-specific meanings and could provide either a general translation or one which responds to the specific needs of the user. A system with these abilities is indispensable for the future development of science and technology. The following analysis of existing material is an example of the type of work which should be performed to aid in the design of such a system.

2.2 Analysis of the keywords in CAS, Chemical Abstracts

Ten million keywords were collected from six hundred thousand CAS articles which appeared during 1977-1978, CA vol. 86 to 88.[4] These terms belonged to 25 sections in which cross-correlation analysis was performed. Sections 5 through 10, mostly organic chemistry, showed extremely high self-correlation with respect to the frequency of keyword appearance. The keywords from these sections are mostly technical, i.e., specific, terms. On the other hand, much less self-correlation was observed in the physical and analytical chemistry sections, suggesting that these two sections contain primarily terms of a general nature.

The results of this keyword analysis can be applied to fields other tan chemistry. Apparently, each field contains both general and specific terms. The specific terms are not likely to appear in other fields, while the general terms usually do appear in other fields, although not necessarily with the same meaning. The standardization of terms of general usage is the key for improved information flow. However, at this point, perhaps the best solution is to enter all of the different meanings for each term into a single computer file, from which they can be retrieved simultaneously.

2.3 Proposal

a) The full text databases of the various fields should be analyzed to identify those terms which appear most frequently. They are most likely terms of general usage.

b) Next, the terms collected in a) should be added to the Unified list, to produce a list of general terms of each field.

c) Finally, a translation team, representing at least Japanese, English, French, German and Spanish, should be formed to translate the list in b).

3 The concurrent use of Kanji, Kana and Hiragana in Japan, and the future development of information science

Unlike in most other countries, people in Japan use a three-part language, consisting of kanji, hiragana and katakana. Kanji characters are ideograms. The three thousand characters which comprise the first rank are taught in the primary school and are used in popular media, such as in newspapers. There are approximately three thousand second rank characters as well. As ideograms, they enable people to efficiently produce, or convey, the intended meaning. The character invokes the imagination of the reader immediately upon being seen.

The number of kanji characters currently in use, several thousand, is often accused of being monstrous. However, it is undeniable that thousands of images can easily be produced and shared by people who understand and use these characters.

The Western way of thinking has often been described as logical and methodical, whereas that of the Japanese is often considered illogical and intuitional. This contrast between the two cultures has formed two different styles of communication. The Western style uses a sequential array of words which presents, and develops, the intended meaning logically and systematically. On the other hand, the Japanese, or Chinese, style uses graphic representation to present the entire meaning at once, as shown in Fig. 1.

The Japanese have invented Katakana letters. While each Katakana letter originally had a unique symbolic meaning, today they are commonly used as articles and conjunctions, similar to Hiragana.

They have been used conventionally to represent some action, fact or other image (Fig. 2). However, the meaning of Katakana varies with the context in which it is used. The relationship between Katakana and Kanji is similar to that between numbers and words. Words can provide a meaningful context within which numbers are defined. If the numbers are not defined clearly, then they cannot convey the intended meaning. Katakana was originally used in a similar man-

Hiragana letters worked well for making clearer the presentation of the Japanese tongue. For example, the cases of the subject in Japanese are presented accurately in three ways as like as those in English, whereas those in Chinese are not identified from each other, because they are presented in only one letter.

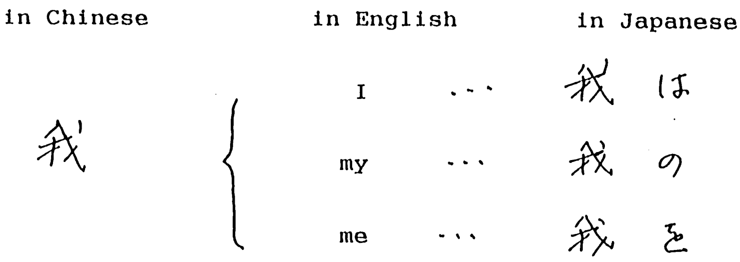

Fig.1

ner, and there may have been considerably more Katakana letters than the present forty eight. As time passed, however, Katakana's context-based meaning fell into disuse, and it is currently used as an alternative to Hiragana.

Hiragana fits between kanji and katakana. It is symbolic and also has an ideographic function.

The concurrent use of these three forms of language may have instilled in the Japanese people some unique characteristics which distinguish them from Westerners. Most personal characteristics of Japanese are fairly well westernized. However, there are some exceptions which often cause misunderstanding and frustration. Perhaps these differences may be attributed to the concurrent use of kanji, katakana, and hiragana. I would welcome the opportunity to discuss this topic at some other time.

For now, however, I would like to turn to the design of computer software for Kanji characters.

When computers were first introduced to this country, we found that the Japanese language was difficult to present with a computer. Programs must be presented alphabetically, that is, in a sequential array of alphabetic letters. This constraint does not lend itself to presentation of the Japanese language, which is formed primarily of ideograms. Of kanji, katakana, and hiragana, katakana symbols, or letters, were the most promising for this application. Therefore, Japanese computers have been made by using katakana letters to represent alphabetic letters.

Transformation of the katakana letters into kanji is a problem which remains to be solved.

Two methods of producing Kanji with computers have appeared. One method was developed in Taiwan (Formosa) and the United States. The principle behind this method was to produce all of the elements of Kanji and then to combine the elements to produce a single Kanji character (Fig. 3). This involved a logical, systematic analysis, and subsequent recombination, of the Kanji elements.

On the other hand, the Japanese method, which was developed after the Taiwanese method, accepted the Kanji character in its entirety as a pattern and produced it graphically (Fig. 3). The Taiwanese method is analytical and fol-

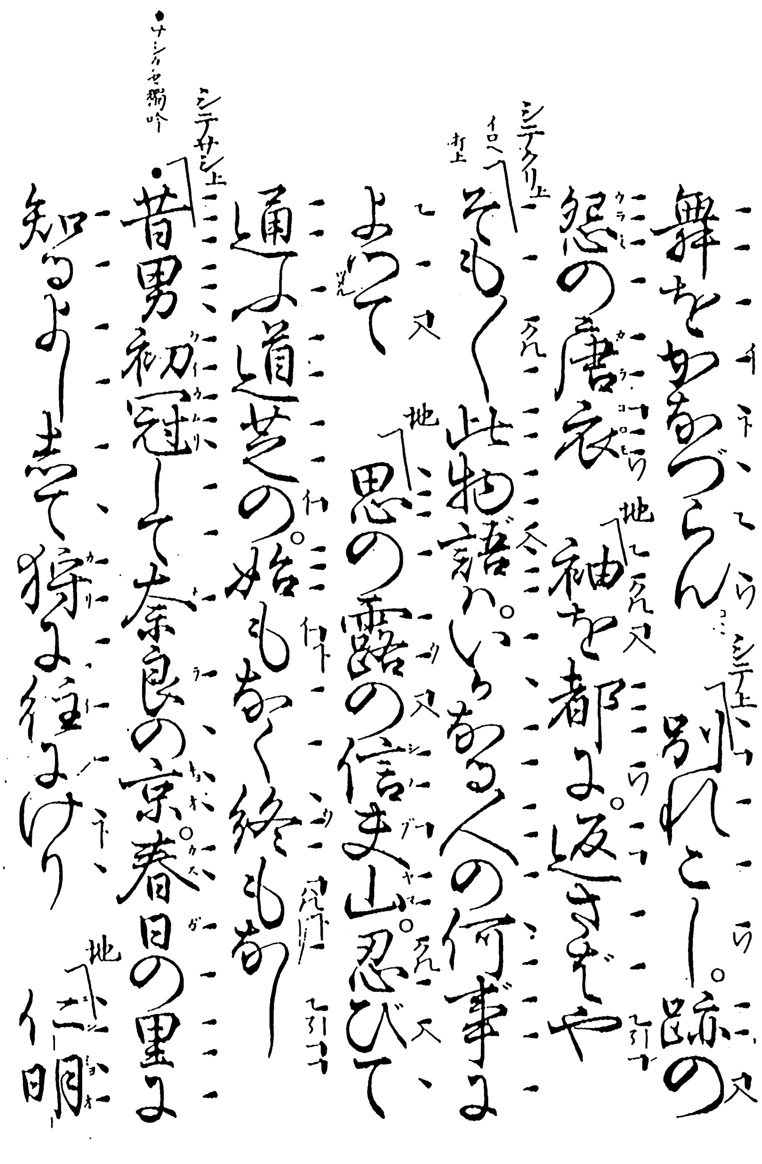

Fig.2

lows the Western way of thinking, while the Japanese version is presented instantaneously, and involves pattern recognition.

The former is logical, but the latter is more efficient. Note that pattern recognition, which is consistent with the Japanese tradition, worked well for computer program design, which is most often thought of as an analytical and logical process.

Proposal:

The effects of the concurrent use of Kanji, Hiragana and Katakana in Japan should be studied, and the results should be applied to future computer design.

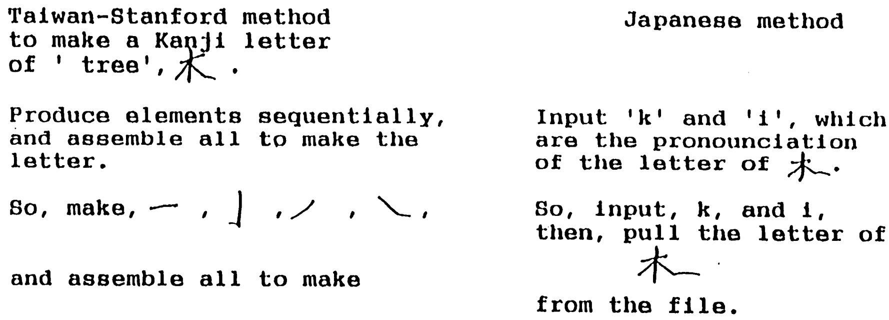

Fig.3

4 Application of information science to safety control and risk management (SCRM)

4.1 Introductory remark

SCRM is a focus of social interest. We, as science researchers, should devote our energies to it for the sake of society. Furthermore, we, as information science researchers, should consider it a major field for the application of information science. SCRM encompasses the following areas:

A) R & D of methods for evaluating the normality/abnormality of the systems being analyzed,

B) construction of a database of the results of A), and

C) application of A) and B).

R & D of the sensors and sensing systems should be a primary concern, and we should rely heavily on Fourier transform analysis of the data measured by the sensors, etc.

Furthermore, both text and numerical data should be the focus of database design. The first step is to design an intelligent database management system for handling all of the textual and numerical/experimental data.

Finally, with regard to application of information science to SCRM, measurements of the fluctuations of cylinders, buildings, and so on, have begun.

4.2 Current status of R & D on SCRM

Seven years ago, a research group was formed by about 25 researchers from universities, government institutions and industry. This group meets for discussion bimonthly, and has held three international meetings. The papers presented at the international meetings have been published in RIIK's 1988 and 1989 annual reports. Cooperative research projects in R & D and database construction are currently underway, and some of the results will be reported at a later meeting.

References

1) S.Fujiwara and Y. Fujiwara, Eds. "The integrated term list of science and technology," Kinokuniya Publishing, 1987, Tokyo, ISBN 906279-00-7.

2) S.Fujiwara, "The list of multimeaning and multidiscipline-common terms in science and technology,"

FID publication number 633, Hague, 1984.

3) S.Fujiwara, "The list of multifield-common and multilingual basic terms in science and technology," Kanagawa University, Hiratsuka, 1992.

4) S.Fujiwara, J. Chem. Information and Computer Sciences, 1981, 21, p.66.

(Received December 16, 1992)

著者紹介

藤原鎮男(正会員)

(現)国文学研究資料館客員教授, 東大名誉教授, FID 名誉会員, 理学博士. 昭和 19 年 9 月東京帝国大学理学部化学科卒, 昭和 24 年電気通信大学助教授, 同教授をへて, 昭和 35 年東京大学理学部教授. 同 56 年定年退職, 千葉大学理学部教授. この間, 東大図書館長, IUPAC 化学情報特別委員, FID 理事, 副会長, 昭和 61 年神奈川大学教授, 理学部創設学部長. 知識情報研究所創設所長. 平成 5 年 3 月同学退職現在にいたる. 研究歴: NMR から電子計算機利用にふれ, 以後, 計算機による情報処理の情報学に入る.

この間, 学術会議の学術文献情報研究連絡委員会, 日本化学会化学情報部会, 日本情報知識学会の創設に関わり, また自身, 大学情報処理システム (TOOL-IR), 図書館情報システムの構築, 学術用語の組織化のなどに参画した. 著書: 科学大辞典 (丸善 昭 60) ほか, 所属: 情報知識学会, 日本化学会, 日本分析化学会, 米国化学会など.