

TERMINOLOGY IN SCIENTIFIC AND TECHNICAL JAPANESE: PROBLEMS AND PROSPECTS

Ian T. M. Gow

Mitaka, Leamington Spa, United Kingdom

H.G. Wells once said of Japan:

"her secluded civilisation has not contributed very largely to the general shaping of human destinies; she has received much but she has given little."

For many this is still the case today. However, certain more perceptive observers have come to realise that Japan does have something to offer, not only to the industrialising countries but to post-industrial society. There are, of course, many who would deny that Japan may be part of the solution and would rather indulge in certain conspiratorial theories that would have Japan regarded as part of the problem. But attitudes are changing and many authorities on Japan would subscribe to Professor Ezra Vogel's belief that the traditional slogan *Wakon Yosai* (Japanese spirit and western means) now be reversed and that we should be talking of *Yokon Wasai* (Western Spirit and Japanese Means). There can be little doubt that in certain fields, especially applied science and technology, Japan already has much to offer the West. Moreover, should Japan's plans for "next generation technologies" guided by her Ministry of International Trade and Industry (MITI), reach fruition, especially with regard to the Fifth Generation Computer Project, then we can and must monitor and learn from the Japanese.

The major problem is the language but, before dealing with that aspect, one must mention the often under-rated problem of attitudes. There are two very important pieces of so-called conventional wisdom which those of us involved with processing Japanese language data are constantly confronted. The first is that the Japanese do nothing original and therefore there is no need to allocate vast resources to processing Japanese language data in science and technology. The second is that there is no need to process mountains of oriental script since all the best Japanese work appears in English. The first of these is palpably untrue today whilst the second, were it ever true, only applied to a very small select group of researchers whose peer reference group was an international one.

According to the latest research only seven per cent of scientific and technical periodical literature in Japanese is indexed or abstracted in the West. 1981 figures provided from the National Diet Library

indicate that 82.3% of all Japanese Scientific and Technical periodicals are published only in Japanese. Since Japanese is regarded as a 'difficult' language and since the number of competent translators in the Science and Technology fields is pitifully small, there are major problems to be overcome.

Japanese Studies in the West has a long, and in many ways, distinguished tradition. However, it is dominated by arts and social science graduates and, indeed, in many areas of Europe, it is still dominated by the so-called "Japanologists" specialising in premodern or, at best pre-war Japan. The result is that there simply are not sufficient numbers of competent translators from Japanese into Western languages with the kind of background to cope with the information required from Japan today, namely science and technology. If one takes, for example, the case of the United Kingdom, what provisions are made for the training of technical translators in this area? The Centre of Japanese Studies at Sheffield University has pioneered the training of scientists and technologists in Japanese through the annual summer course designed and taught by Dr Jiri Jellinek. Aston University offers science undergraduates the opportunity of beginning the study of Japanese as part of a general studies option. The Royal Society provides courses of the Japan Society for the Promotion of Science, a short spoken course in Japan for its visiting fellows. Unfortunately, in the latter case, few ever venture beyond this to actually master the written language.

What is clearly needed is the creation of the necessary tools for the job and that means, the training of technically or scientifically inclined people in Japanese, to a very high standard. So far, there is no evidence that western countries actually perceive this as a desperate need. Then there is the production of good scientific and technical dictionaries particularly from Japanese into western languages, written and compiled by practising translators with the needs of western translators foremost in their minds. Finally, there is the need to allocate far greater resources to the development of machines, from word processors to machine translation devised to aid those involved in processing scientific and technical Japanese.

The major barrier is, and will remain, the Japanese script. Although many have advocated the romanisation of the Japanese script, Japanese and Westerners, there are good reasons for the retention of the indigenous script other than the obvious political and cultural ones. Perhaps the most important and certainly the one most often mentioned is the large number of homonyms which would result. The Japanese script comprises a mixture of Chinese characters (Kanji) and two syllabic alphabets of 48 characters each (covering exactly the same sounds). These are called respectively Hiragana and Katakana.

Since the Japanese did not have an indigenous script their first idea was to borrow from China. At first the Chinese characters were borrowed only for their phonetic value and this has continued to a limited extent. However, since the Japanese language is polysyllabic, the result was that a whole series of cumbersome kanji were needed to represent a very simple Japanese word. The Japanese decided to simplify certain Chinese characters and create two identical syllabic alphabets: Hiragana and Katakana. Kanji today are used for nouns, adjectives and verb roots of Chinese and Japanese origin and basically are used to

provide the meaning in the sentence with the syllabic alphabet providing the syntax. The number of Kanji required by the average Japanese has greatly decreased due to the post war language reforms and the basic list, the Toyo Kanji, is 1,850 but a slight increase is expected in the near future. This is because there are actually about 3-4,000 in common use (compared to 7-8,000 before the war). The Toyo Kanji are used in textbooks, newspapers, etc.

Hiragana is used for the conjugation of verbs and adjectives as well as for particles, conjunctions, etc. It is sometimes used for very rare Japanese or even Chinese words, where the original Kanji is not known or not in the print set or possibly not in the official set of Kanji. Katakana is angular whilst Hiragana is more cursive. Katakana is used for foreign words (except Chinese), foreign names, some onomatopoeia and for emphasis (like italics). It is also used in telegrams. Together the two forms are collectively known as Kana.

When it comes to the dictionary look-up process, the first problem is the variety of ways in which a dictionary can be compiled. There is a bewildering variety of dictionaries available, some 6,500 are listed in the 1982 dictionary catalogue of the Shuppan Nyusu Company. However, they are nearly all compiled with the needs of a native speaker-reader of Japanese in mind. Most Westerners begin with Andrew Nelson's Japanese/English Character Dictionary, which converts Kanji into English by using an element of the Kanji, the radical, to locate the appropriate Kanji or Janki compound. Other dictionaries use a combined stroke count for the location of Kanji. The radical and the stroke count method are very common in Japanese/Japanese dictionaries, although the latter method is not particularly favoured by many Westerners, since it is relatively easy to miscount the number of strokes and also the look-up method is slower than with radicals. Then there are the Kana dictionaries where one simply looks up the Kana word represented, or the reading of the Kanji. However, for the latter one must of course know the correct reading of the Kanji and Westerners with a high ability to read characters in this way, while by no means rare, are not in the majority. In some dictionaries of this kind, the Kana are all together in syllabic order, but sometimes one can have Hiragana entries under a sound and then Katakana follows on. Finally, there are the dictionaries using romanised Japanese and whilst these are favoured by many Westerners even they are not without problems. The major one is that there are different systems of romanisation. For example the "Jiten" (dictionary) is often printed "Ziten"! The problems are not insurmountable, but they do irritate and lead to "translator fatigue" particularly since one may need to use five or six dictionaries, sometimes very large ones, in one's search for the elusive word. Sometimes even this ends in total failure when it is discovered that the author, or typesetter, has inserted the wrong Kanji with a similar shape and identical sound!

Inconsistencies such as those above however, are not really the main problem. It is the quality and range of available dictionaries which is the main impediment to effective scientific and technical translating from Japanese into Western languages. Many of the dictionaries on the market are not dictionaries at all, but merely word lists in which nouns are by far the most numerous. Many are prepared by academics who are not necessarily practising translators and many are simply translations of western dictionaries. In general, they are

prepared for the purpose of translating in Japanese and only contain a small, and often incomplete, section for translating out of the language. Some areas of Science and Technology are well covered such as plastics, chemistry, medicine and business/commerce but the gaps are most evident in metallurgy, textiles and botany. There are few westerners with the ability and inclination to compile dictionaries in these areas (or the time!). One of the most useful to both Western and, surprisingly even to native Japanese speakers is Professor P.G. O'Neill's "Japanese Names" a dictionary of use to all those involved with Japanese translating ancient or modern, technical or non-technical, since the reading of Japanese names is horrendously difficult. The University of Arizona has produced some small dictionaries in recent years and one useful dictionary of Patent Terms has been compiled by Thomas Wilde. Dr Jiri Jelinek has produced an Integrated Grammar Dictionary derived from his outstanding research in machine translation and is not engaged in producing technical dictionaries at the University of Sheffield

It may be that the problems faced by Japanese-Western language translators are no different in kind from that of their colleagues in other fields and that it is merely the degree of difficulty. The advent of the computer, the Japanese word processor and the term bank promise some help to us in our task, but even there, there are many problems. Clearly, the use of term banks and data banks, can be of tremendous benefit provided that the problems of input and operating can be overcome. There are, one is told, already machines on the market to translate, but, on investigation, one finds simply sophisticated calculators containing word lists, such as those produced by Sharps and Casio. For professional translators, these are nothing more than toys.

As our distinguished speaker from Japan pointed out, the Japanese have had major problems with inputting Japanese as well as with aspects of machine translation. The Japanese typewriter is a truly cumbersome and awesome device to the foreigner familiar with Japanese. The normal Japanese typewriter used to have approximately 2,000 characters with a single point and press system. This was refined and a multishift system introduced and this speeded things up. Such a system is used on certain input devices for Japanese script word processors. Needless to say, a word processor using this input method would not be of much use to an individual Western translator unless it was permanently manned by a qualified Japanese native operator. In recent years, a two stroke typewriter system, the Rainputto system, has been introduced and significant increases in typing speeds were recorded. However, while possibly a little easier for a Westerner to operate, an input device using this would hardly seem the ideal answer and again, a permanent operator might well be needed. The Wang three corner coding system whereby all Kanji are broken into three principal component parts, is a very promising development and is being adopted by American libraries for inputting Japanese, Chinese and Korean for cataloguing purposes. This system may well have considerable success in the future and, one is told, can be used with some facility by Westerners.

The most promising developments however, for Japanese and Westerners alike, would appear to lie in the following areas: the Kana to Kanji conversion method, whereby one inputs the sound and by means of frequency and semantic analysis a certain number of appropriate Kanji, with meanings, are flashed on the screen. This phonetic to Kanji

processor, uses a Kana keyboard and Kanji and Kana are flashed onto the screen, etc. Then there is the Voice Input Device System whereby, using similar frequency and semantic analysis Kanji and Kana are flashed onto the screen; finally, there is the possibility of a machine which will simply recognise a Kanji, the so-called optical recognition method. Research here, in terms of both printed and handwritten Japanese, is being heavily subsidised by the government under the Pattern Information Processing System Project. Toshiba, amongst others, have already developed a complete voice input system for the Japanese script. These latter developments all promise to be of much greater importance than the Rainputto and Wang three corner system as far as teams of translators working with word processors etc., are concerned.

These new devices do show that the situation regarding the translation of scientific and technical Japanese into Western languages can and probably will improve. Machines such as word processors can be used as teaching aids, as basic tools for compilation of terms not found in dictionaries, as well as the monitoring of the constant changes in technical terminology. The Japanese have invested vast amounts of cash, time, effort and not inconsiderable skill in this area, but one would have to say that their main concern is for the processing of languages and data into Japanese. We will have to take the initiative regarding research on dictionaries and machine translation in the other direction and we should not be waiting for the Japanese to do it for us.

At present, there are a number of groups and individuals scattered throughout the West who are striving, with precious little support, moral or financial, to achieve something in terms of machine translation from Japanese. Jiri Jellinek at Sheffield is by far the best known and has greatly enlarged the available pool of translators as well as pioneering machine translation in Britain. Christian Galinski's work at the Technical University of Vienna is very promising indeed. Here he is working closely with Dr Simoncsics and they have now achieved a computerised terminological retrieval system for specialised vocabularies. At Grenoble in France, Dr Tsujii and others have apparently made considerable progress using the Ariane system but there is little data published on this. In Germany, a proposal for governmental cooperation between academic, governmental and private institutions as well as individuals, was being investigated in the late 70's. The coordinators were to be Professor Zimmerman of Saarbrücken and Professor Nagao of Kyoto University. Unfortunately, it never came to fruition. Apparently Professor Zimmerman is at present working on machine translation of scientific titles whilst Dr Hanakata at the University of Stuttgart is working on hardware research. In the USA the Stanford University group are doing considerable research in this area and are actively involved with the Asian Libraries in the USA.

A considerable amount of research, and the resources for that research, is now vitally important if we are to be able to take advantage of the vast amount of data available in Japanese in the fields of science and technology. Unfortunately, whilst many translators would dearly love the opportunity to go off and actually start work which would help to solve the problem, there is seldom, if ever the time. Research in these areas is generally regarded as a problem for the medium or longer term and translators are working to deadlines such as "by today" or worse still, "by yesterday please". Our own organisation has realised that medium term or longer term planning and research is