Meta
Journal des traducteurs
Translators' Journal

Computerised Terminological Databases for Translators Who Use Word Processors

Peter M. Davidson

Volume 36, Number 2-3, juin 1991

URI: https://id.erudit.org/iderudit/002489ar
DOI: https://doi.org/10.7202/002489ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Davidson, P. M. (1991). Computerised Terminological Databases for Translators Who Use Word Processors. *Meta*, *36*(2-3), 424–430. https://doi.org/10.7202/002489ar

Article abstract

This paper describes the construction of a Japanese-English computer-based dictionary of financial terms for native-English speaking translators of Japanese, writing their English translation directly on to a word processor. The Japanese terms selected for entry were culled from Japanese international finance publications (books and journals) published since 1985, and are thus assured of being in current use. Two major terminological data or glossary producing software programs were tested, for "user-friendliness", minimum memory and disk space requirements. Both programs were designed for European languages, and not for ideographic Asian ones, and consequently adaptations also had to be made to mode of entry to accomodate the requirements of the Japanese language. These are discussed in the paper, as are the system of romanisation it was decided to use, and the modifications it was seen fit to make to simplify and rationalise look-up procedures.

COMPUTERISED TERMINOLOGICAL DATABASES FOR TRANSLATORS WHO USE WORD PROCESSORS

PETER M. DAVIDSON
Department of Japanese & Chinese Studies,
University of Queensland, St Lucia, Australia

RÉSUMÉ

On décrit l'élaboration d'un dictionnaire informatisé de termes financiers japonais-anglais pour les traducteurs de langue anglaise écrivant directement leurs traductions anglaises du japonais avec un traitement de textes. Les termes japonais choisis comme entrées ont été tirés de publications financières japonaises internationales (livres et journaux) publiées depuis 1985 et sont donc contemporains. Deux principaux programmes constitués de données terminologiques et de glossaires de documentation informatisée ont été analysés et conservés dans un minimum de mémoire et d'espace sur disquette. Ces deux programmes ont étés conçus pour les langues européennes, et non pour les langues à idéogrammes de l'Asie. On a donc dû aussi adapter le mode d'entrée pour répondre aux exigences de la langue japonaise. On étudie cette adaptation, de même que le système de romanisation choisi et les modifications qu'on a cru nécessaires afin de simplifier et de rationaliser les étapes de recherche.

ABSTRACT

This paper describes the construction of a Japanese-English computer-based dictionary of financial terms for native-English speaking translators of Japanese, writing their English translation directly on to a word processor. The Japanese terms selected for entry were culled from Japanese international finance publications (books and journals) published since 1985, and are thus assured of being in current use. Two major terminological data or glossary producing software programs were tested, for "user-friendliness", minimum memory and disk space requirements. Both programs were designed for European languages, and not for ideographic Asian ones, and consequently adaptations also had to be made to mode of entry to accomodate the requirements of the Japanese language. These are discussed in the paper, as are the system of romanisation it was decided to use, and the modifications it was seen fit to make to simplify and rationalise look-up procedures.

INTRODUCTION

There is a growing requirement in the English-speaking world for the translation of selected materials from the ever-increasing and already vast body of scientific and technical literature produced by the Japanese each year for their own consumption. One might expect much of it, and particularly straightforward technical material, to be translated by machines, but the current state of the development of machine translation is such that where translation from Japanese to English is required (and it still cannot be done effectively in the other direction), much time must still be devoted to human pre-editing of the Japanese text (insertion of missing subjects, simplification of nested clauses, etc.), and to human post-editing of the machine's English output. We thus still remain at the stage where, in the writer's view, it is more efficient for the human translator to undertake the whole task right from the outset. This has been already recognised by large companies like Unisys who claim that "the disappointing overall quality of machine translation output, poor productivity gains, and growing disenchantment among our translators has led us to abandon MT until the quality of its output is much better. We don't rule out the possibility of using it again, but we've come to the conclusion that existing commercial systems simply don't meet our standards. In short, MT is not yet a viable solution[1]."

Much of the translation from Japanese into English is already effectively undertaken by professional translators, many of them native English-speaking translators working, nowadays, directly on a wordprocessor. They are quite likely to be using an IBM-compatible PC with a well-known commercial software program like WordPerfect, WordStar or Microsoft Word. If they are thus already using a computer for their wordprocessing requirements, how else can that computer assist or simplify their task of translation?

Martin Kay[2] suggests that although there have been considerable advances in the relevant areas of computer science, what advances there have been in linguistics have not been able to touch on the core problems of machine translation. He therefore proposes the development of cooperative man-machine systems, and, in particular, a "translator's amanuensis" which would incorporate "into a word processor some simple facilities peculiar to translation". Certainly the thorniest language problem facing multilingual wordworkers and peculiar to translation — especially those working in a variety of different disciplines — is that of specialist terminology. The constant stream of new coinages and the pressure of international standardisation place terminology, along with teenage slang, among language's most perceptibly fickle aspects.

Dictionaries, glossaries and word lists are the indispensable tools of a translator. Dictionary look-up time is, however, a significant time factor, and additionally "book-form" specialist dictionaries quickly become out of date after publication. Computers, on the other hand, lend themselves to the handling of databases, and thus to the building of technical dictionaries — databases of technical terms. Most translators will, of necessity, need to refer to "book-form" technical dictionaries for technical terms they are not familiar with. The act of consulting such dictionaries inevitably interrupts the translator's "keyboard rhythm" and train of thought, and, of course, requires more time. Computerised dictionaries obviate this problem, and also lend themselves to frequent updating of lexical items that developments in the sciences and technologies require, and permit translators to insert their own specific lexical requirements into them, something which "book-form" dictionaries are unable to do.

Unfortunately, there are no readily accessible international databases that Japanese into English translators can utilise: the well-known online multilingual termbanks like Eurodicautom, Normaterm and Termium, do not handle the Japanese language in either romanised or character form. Translators working from Japanese into English are thus likely to have to compile their own databases. This is not quite so bad as it may at first seem since there is some well-supported and frequently updated memory-resident software addressing translators' terminological problems which is now available on the Australian market at prices within the reach of professional translators who work with word-processors. This software which has, however, been specifically developed for European languages, enables translators to look up the term they require from databases stored on disks (hard or floppy) while still in the word-processor mode, and thence to "paste" the appropriate equivalent term in the language they are seeking directly into their on-screen translation. Additionally, such software permits translators to add and edit terms, and to create their own databases. And all of this can be achieved much more quickly and efficiently than the translator could do by other means.

Both software programs that we are currently using and investigating, Mercury/ Termex (LinguaTech Inc.) and INK TextTools (INK International BV, Holland), are not as yet able to handle the ideographic character sets required by languages such as Japanese and Chinese. Nor is there yet, to the best of our knowledge, any Japanese-language software available on the Japanese market designed to create databases of term-to-term equivalents for translators that will permit the direct on-screen entry of Japanese ideographs[3]. It thus becomes a matter of adapting the vagaries of the Japanese language to the Western software programs that are readily available at reasonable prices or else going without.

This is not too much of a problem since, unlike most native Japanese speakers, native English-speaking translators will have no particular "hang-up" about working with romanised dictionaries (even though they would probably prefer to use character dictionaries). This means that the European language-oriented software programs referred to above can be adapted to handle Japanese (and Chinese) in romanised form. Because of the large number of homophones in Japanese, however, it does become essential to devise a system that will enable accurate reference to be made to the ideographs with which each particular Japanese term is written.

It is against this general background that we decided to create a Japanese-English dictionary of technical terms in a format suited specifically to the needs of the many professional translators who are already using word-processors on IBM-compatible machines.

We particularly wanted to see how many Japanese-English terms could be contained on high-density floppy disks (1.2K or 1.4K), and thus whether or not translators would need to invest in a machine with hard disk for optimum usage. We also wanted to evaluate the "term search-and-paste" efficiency of the software, and the degree of ease with which such romanised dictionaries could be used by professional translators.

SUBJECT AND SOFTWARE

As the disciplinary area for the database, we selected one where there already exists a large demand for translations from Japanese into English — the area of international banking and finance. We decided to create our own *up-to-date* database by culling the relevant Japanese technical terms from Japanese-language journals, papers and books published in Japan in this area in the five years since 1985. The database could thus be assured of containing terms in current usage which would satisfy the lexical requirements of professional translators.

The software used in the initial production of our database has been the latest available updates (initially version 1.46 and currently version 2.0) of the American software program Mercury/Termex (Linguatech International, USA), which is commonly known by the acronym MTX.

SYSTEM OF ROMANISATION

Since Japanese terms have been entered into the database in a non-ideographic, romanised form, it has seemed necessary to represent the actual Japanese ideographs employed to write the terms by 4-digit codes drawn from a source familiar to most, if not all, English-speaking Japanese-English translators, Nelson's *The Modern Reader's Japanese-English Character Dictionary (Revised Edition)*. It had been suggested that we should consider using the JIS Kanji Codes used by computer programmers, but our view was, and is, that there is no point in using an identification code unless it is readily accessible to all translators. Nelson is, but the JIS codes are not.

A major problem in the development and preparation of this database has been to refine and standardise a system of romanisation for the Japanese lexical entries in terms of spelling, capitalisation or not of proper nouns and phrases, and hyphenation (of compound words, etc.) that will both prove satisfactory and acceptable to end-users as well as satisfying the alphabet-ordering requirements of the computer software. To take the latter first, letters such as â and ê and ô are placed in a different section of the ASCII Code (which determines the computer's sorting and ordering of letters and words) than the letters a, e and o. A similar separation exists of upper-case and lower-case letters. Similarly, words separated by a space and those separated by a hyphen are given different positions in the overall ordering scheme[4].

There are several different systems of Japanese romanisation in existence. Those not acquainted with the two most common systems would be hard put to recognise that *Fuji* (Hepburn system) and *Huzi* (*Kunrei-shiki* system) represent the same word. We have opted for a *modified* Hepburn system as used in the 4th. edition of Kenkyusha's *New Japanese-English Dictionary*, which, apart from using macrons to indicate long vowels, adopts a spelling system readily acceptable to English speakers. Our justification for this has been that this, more than any other single dictionary, is the one all Japanese-English translators will be familiar with. Initial letters of the romanised forms of Japanese proper nouns have not, however, been capitalised, so as to allow a more natural alphabetisation.

HYPHENS AND SPACES

The Japanese language is written with a mixture of Chinese ideographs and Japanese-invented phonetic syllables, not with the Roman alphabet. Within the Japanese writing system, Japanese words are not separated by spaces as they are in European languages, and hyphens are not used to separate words into their morphemic components. The problem of hyphenation and spacing is, however, a vexing and arbitrary one to argue over for non-native Japanese who wish to write and record the language using the Roman alphabet. While the Japan Style Sheet5 suggests that the use of hyphens should be restricted to not more than one per word, some dictionaries use neither hyphens nor spaces even in the production of compound nouns (and thus produce monstrosities like *kakeichôsafutaichôsa* ("supplementary inquiry about the family income and expenditure survey") and *meimokukokuminsôshishutsu* ("gross national expenditures at current prices"). Others over-use hyphenation. We have decided to indicate syntactic bracketing[6] or structure by hyphenating prefixes and suffixes so as to separate them from the terms to which they are attached, and also to hyphenate three character words where two of the three characters form an accepted term to which the third character happens to be added. We use spaces, and not hyphens to separate discrete words in a term. Thus, the above two terms would, in our database, appear as *kakei chôsa futai chôsa and meimoku kokumin sô-shishutsu* respectively.

SELECTION AND FORMATTING OF TERMS WITHIN MTX

The terms selected for this Japanese-English dictionary of international finance have been culled from 12 books and 33 journal articles on the subject published in Japanese since 1985. The generalist and specialist Japanese language journal articles we chose to utilise were selected from the wide holdings of the main library of the University of Queensland on the basis of relevance and germaneness to our subject area. Our selection of book titles focused on that very popular Japanese genre aimed at educating and informing intelligent Japanese businessmen and bankers about their field. Some were already held in the university library, while others were purchased straight off the shelves of some Tokyo bookshops. This intentionally random approach was taken since there was no possible way in which we could have examined every single book and journal article in the area published during the period. The terms in the dictionary thus justify their presence exclusively on the grounds that they occurred in the random selection of current printed materials we chose to examine. It is inevitable that many translators will find terms they require to translate missing from this dictionary, but that is the nature of random selection. It is, however, not a major problem for the translator, since there is no difficulty in adding such "missing" words to his/her personal version.

All the relevant technical terms (both words and phrases) appearing in the materials surveyed were immediately entered into a database created with the MTX glossary management software (Version 2.0) on the hard disk of an IBM-AT compatible computer. The rationale for this method of entry was that the ultimate selection or rejection of any particular term would be made from the computer after initial compilation. The terms entered include nouns and noun phrases, verbs, adjectives and adverbs, but it was decided not to indicate the grammatical function of any terms, as this would be immediately apparent to the professional translator. Japanese verbs are given in their plain ("dictionary") form, and additionally, the English translations of Japanese adjectives are given as adjectives, verbs as verbs, etc.

As each Japanese term was added to the MTX dictionary being constructed, its Nelson character code was inserted as entry {0}, the first entry in the data field, with the number for each of the characters in the term being separated from the next by a semicolon, but no space. English meanings were thereafter appended as entries {1}, {2}, {3}, etc. (as many as required, but not exceeding 50 in number, or a total of 1 900 characters). Articles, definite and indefinite, have been omitted when they are the initial word of the English term. English nouns have been entered in the singular form, proper nouns have their initial letter capitalised, and verbs are given in the infinitive, but without the preposition "to". Spellings are given in British English, but can be easily changed to American English if required.

As has already been mentioned, final editing took place when all collected entries had been added. Since the dictionary covers international finance, the major purpose of the editing process was to remove any general, non-technical entries which might have slipped in, unless such entries possessed a specialist meaning as well.

USING THE SOFTWARE TO LOOK UP TERMS

One of the distinct advantages of the MTX software is its "user-friendliness". Once the program has been loaded into memory, it can be brought up on screen within the word-processing translation document file by typing <Alt>-M. Only on the first occasion when the software has been transferred to the hard disk of the translator's computer does it have to be given the path and name of the dictionary files to be loaded. Thereafter, after loading into memory (via a batch file) before the word-processing software is loaded, it is available for immediate use whenever required by tapping the appropriate "bring-up" or "hot" key.

To look up a Japanese term, <Alt>-L is typed, and the program comes up on the half of the screen not occupied by the word-processing cursor. The program asks for the term or key the translator is searching for to be typed in. It is not necessary to enter any macrons the term may have, and in the case of long words it is not even necessary to enter every single letter. The first four or five letters are normally sufficient to locate the term. Before selecting a specific term, the translator can elect to "zoom" through that particular area of the database to see what terms and compounds are to be found. This brings up a smaller window in which surrounding terms are also listed, and the appropriate selection can be made from there.

If the Japanese term in question exists in the database, it will appear within a window on screen once selected. Immediately underneath it will appear a fixed entry {0}, which gives the references to the Chinese characters with which the term is written. Immediately underneath this entry will be found a list of its English meanings (not exceeding fifty in total!). The translator selects the term most appropriate to the context of the translation by entering the number listed beside it, presses the return key, and is returned to his document. Here he sets the cursor to the position where he wished to "paste" (enter) the term, presses <Alt>-P (meaning "Paste") on the keyboard, and the selected English term is immediately "pasted" into his translation document file.

If a Japanese term the translator is searching for is not to be found, then the translator will have to consult a "book-form" dictionary. He can then simply elect to add it to his computer database with its character codes and English meanings. The database can thus grow on a daily basis as its deficiencies are revealed. The only limit to the number of entries so added is the size of the floppy disk or the hard disk. Similarly, should the translator wish to amend or edit any of the existing entries in the database, this can be done very easily as required.

MTX MEMORY REQUIREMENTS

The glossary management program, MTX Version 2.0, occupies up to 130K of memory, while the popular word-processing program, WordPerfect 5.0, occupies 330,096 bytes, making a total of almost 460K for these programs alone. Microsoft Word 4.0 and MTX requires a lesser total of 380K, while WordStar 5.0 requires a total of 400K. The other glossary-making software program, Text-Tools, and its "Lookup" program, is slightly larger than MTX at 180K, thus requiring the addition of approximately 50,000 bytes.

These figures indicate that the translator's IBM-compatible computer will require a full 640K of RAM to allow for configuration, for the memory-resident glossary management software, and for the translator's word-processing program and document files. The machine itself will need to have DOS 3.0 or higher. Translators will need to experiment with any other programs they may wish to have resident in memory, so as to arrive at a satisfactory compromise with the memory of their particular computer.

DICTIONARY SIZE AND FLOOPY DISKS

It had been initially hoped that it might be possible to use the dictionary directly from disk, but as indicated the figures presented in the section above the size of modern word-processing programs and the glossary management software is such that a hard disk becomes almost a prerequisite. Our finance dictionary has ended up containing 5 274 Japanese terms after editing, which currently takes up approximately 1K. For an IBM-compatible computer with two 1.2Mb floppy disk drives and 640K of RAM, it would be possible, though not too efficient, to have the bare minimum of the word-processing program and the document file on Drive A, and to have the database on Drive B. But for efficiency and speed, investment in a hard disk becomes essential. Most modern word-processing programs have useful spelling and hyphenation dictionaries and thesauri which can only be efficiently utilised from hard disk.

PUTTING THE DICTIONARY INTO THE INK TEXTTOOLS FORMAT

Once the MTX international finance database had been finally edited, we converted a copy to the format required for use by the Dutch glossary management program, INK TextTools. This was a fairly time-consuming process achieved by means of utility programs included in MTX and INK TextTools software. The colourful on-screen format of the TextTools software is neither quite so neat nor immediately appropriate to Japanese language requirements as is MTX, but it does possess the possible advantage of being able to convert the Japanese-English dictionary into an English-Japanese one by means of a utility program supplied with the software.

The INK Textools software will just fit on to a 1.2M high-density floppy disk, which can be removed after loading to allow insertion of the word-processing software, while the English-Japanese and Japanese-English databases each require about 900K. Thus, again, a hard disk really becomes necessary to make the most effective use of the database, since computers without a hard disk would require at least **two** high-density disk drives (1.2K or 1.4K).

TESTING OF THE DICTIONARY

This is obviously the most important aspect of the project, and the stage at which we have currently arrived. We shall be asking the final-year students of The University of Queensland's Master of Literary Studies course in Japanese Conference Interpreting and Translation to test the database while translating appropriately selected Japanese passages into English as part of their course work in 1990. These students are already required to use word-processors for all their translation assignments, and they will be trained to use the MTX software and the international finance database. We shall also be seeking the assistance and cooperation of professional translators in assessing the value and effectiveness of the database. Their feedback will be taken into account in the final editing of the dictionary, particularly in regard to the current format in which term translations are presented and the appropriateness of their form for immediate pasting into their on-screen translations. Needless to say, we hope that translators will find this first computerised Japanese-English terminology dictionary helpful and useful.

NOTES

[1] *Electric World*, 15, sept./oct. 1989, p. 11.

[2] Kay, Martin (1980): *The Proper Place of Men and Machines in Language Translation*, Palo Alto, Xerox Palo Alto Research Centre. 3. Linguatech have informed the writer that oriental language version is scheduled for completion around the end of March 1990.

[4] Version 2.0 of MTX does, however, allow for "collation" of non-standard IBM characters to determine alphabetical order.

[5] Published by the society of Writers, Editors and Translators (SWET), Tokyo, 1983.

[6] For an interesting general discussion of the problem of compound nouns, see: Jones, Karen Sparck, "Compound Noun Interpretation Problems". In Fallside, F. and Woods, W.A. (Ed.), *Advanced Course on Computer Speech Processing*, London, Prentice-Hall, 1985, pp. 363-381.

BIBLIOGRAPHY

Electric Word, #15, Sept./Oct. 1989.

KAY, Martin (1980): *The Proper Place of Men and Machines in Language Translation*, Palo Alto, Xerox Palo Alto Research Centre.

SOCIETY OF WRITERS, EDITORS AND TRANSLATORS (SWET)(1983): *Japan Style Sheet*, Tokyo.

SPARCK, Karen (1985): "Compound Noun Interpretation Problems", Fallside, F. and Woods, W.A. (Ed.), *Advanced Course on Computer Speech Processing*, London, Prentice-Hall.