Background to Mercury, a glossary building program for translators

Jean-Paul Krähenbuhl's article on glossary building in the February 1986 issue of Language Monthly brought in separately communications from reader Dwight Marsh, a user of a Mercury program; from Dr Leland D. Wright Junior, who has been associated with the same program (and with another product called WorldWriter which has already been reported in Language Monthly); and from Alan K. Melby, vice-president and technical director of the firm producing it, LinguaTech. The subject is evidently of great interest to translators everywhere, and extracts from the various communications are given below.

Dear Editor,
I am a freelance translator of French and Italian. For glossary building I had been using a modified BASIC program, with the drawbacks of not being able to look up directly from Wordstar, and limitation on file size. Having seen a demonstration of Mercury , at the introductory price of $85, I ordered it. I had to upgrade the RAM of my computer to accomodate *Mercury* and the word processor program (something over 300K RAM is needed). Also, I added a hard disk so I would not have to worry about watching the size of the glossaries.

One is always suspicious of the claims for a program. But *Mercury* does all that it advertises, and more. First, the most important: being able to look up terms right from the WP program without having to save or lose sight of the text, and with just an ALT and letter combination. The lookup menu appears in the lower half of the screen. Once you type in the term to be looked up, the process takes one to three seconds. Total time from beginning the process to returning to cursor position (including bringing the definition into the text if you wish) is less than ten seconds. Try looking up a word in the dictionary in that time! And that with an already substantial glossary of more than 2,000 terms.

Some other advantages. ACCENTS – Some 40 accents and symbols created by one or two key combinations, with a program to make your own keyboard changes if needed. LIST FORMAT – A line for the term (with plenty of room for a multi-word term) and an unlimited number of lines for definition fields. You can "chain", jump from one related term to another in the glossary marked with the same field symbol. You can combine your existing glossaries and later separate them or print out a portion of the glossary on the basis of a field symbol. LOOKUP – As I said, lookup is fast: one to three seconds. A big plus also is that you only need to type the first part, or root, of the term, and the response will tell you if there is also another term that has this root. The lookup ignores accents and initial capitalisation, so that you don't get a negative response for leaving out an accent or not knowing whether the term or abbreviation is capitalised. CHANGING AN ENTRY – This is quick an easy: either the term or one of the definitions, and as a precaution the program asks for a confirmation that you really want to save this corrected entry. Alphabetisation is automatic. There are several utility programs that help in copying files, combining, deleting etc. And disk capacity is the only limit on glossary size.

In summary I think this is an excellent job. It is so easy to store terms that I now store every term that I have to look up, thus saving time on the next occasion, and without any appreciable slowdown in lookup time because of the increased capacity. It is easy to change previous entries. And for abbreviations, I am beginning to appreciate the advantage of being able to bring a long expansion into the text.

Dwight Marsh
McLean, Virginia, United States.

Dear Editor,
This letter and the accompanying article are in response to the interesting article by Jean-Paul Krähenbuhl I thought that your readers might like to learn about the computer-based solutions for translators' terminology management problems with which I have been involved in developing here in the United States.

First, let me give you a little background information on who I am: I have been a full-time professional translator for over ten years. I am an active, accredited (Spanish-English) member of the American Translators Association, was on that organisation's Board of Directors for two three-year terms, and am currently editor of its official publication, the *ATA Chronicle*. Before entering the translation field, I was a university professor of Spanish for almost ten years. I worked as a salaried, i.e. in-house, translator for a major multinational engineering company for eight years, and since 1982 have operated my own translation business, Linguistic Information Specialists, located in the Cleveland, Ohio area.

Dr L. D. Wright

The Mercury glossary management software package

by Dr Leland D. Wright Junior

The article by Jean-Paul Krähenbuhl (*Using the microcomputer for glossary compilation*) described the author's ideas on ways to make effective use of standard micrcomputer software (DBaseII in this instance) for maintaining and managing terminological data.

I too have long been concerned with the problems discussed in Mr Krähenbuhl's article. Some years ago, I became seriously interested in the microcomputer's potential as a tool to help translators deal with the interminable, arduous and time-consuming job of compiling, storing and retrieving specialised terminological data. After several months of thought, research and planning, I eventually produced the conceptual design for a computer program that I felt would be ideal. I talked and wrote about my ideas every chance I got. However, the microcomputer industry was still very much in its infancy at the time, and it was apparent that most translators were not yet as enthusiastic as I about the possible applications of this technology to their daily tasks.

More importantly, even the least expensive microcomputer available at the time was still far more than I could afford to invest, and I certainly did not feel competent to undertake the programming work involved in developing a software package of this complexity.

Then, in late 1981, I made contact with another colleague (now a good friend), Dr Alan K. Melby, who just by chance had been thinking along similar lines and who also happened to be an experienced computer programmer and computational linguist. Unfortunately, Dr Melby and I were separated by considerable geographical space; he lived in Utah and I in Ohio. Early in 1982, however, we finally had the opportunity to spend several hours in face-to-face and fruitful discussions about our ideas, followed by numerous exchanges by letter and telephone.

To make a long story short, by the fall of 1982 our long-distance collaboration had resulted in the creation of a crude but working prototype of the software, which we demonstrated for the first time at the annual conference of the American Translators Association in October of that year.

The program was designed (and still is) to run on the IBM Personal Computer and its compatibles running under MS-DOS, and it was conceived to be used simultaneously with a word processing program. This latter requirement also meant that we had to spend a great deal of time and effort identifying a suitable word processor with which we could integrate our program. That is how Alan Melby and I ultimately made the acquaintance of Dr Mark J. Frederiksen, the third member of our informal "team" who supplied his expertise in the development of a powerful multilingual word processor (now capable of preparing texts in over 40 different langauges). This word processor, marketed in the United States under the name *World Writer* by Dr Frederiksen's company, Economic Insights, proved to be the best solution for our needs.

Initial response to our prototype among the translation community here in the United States seemed quite positive and encouraging, so for the next year we proceeded to develop that prototype into a marketable software package. Shortly afterward, I was able to acquire my first computer system (now replaced twice over) and could devote more time to testing out the operation of our nascent software in a typical translator's working environment. My wife also made extensive use of the program. In fact, between the two of us we have created over two dozen electronic glossaries since those early days of experimentation.

At first we called the experimental version of our program TAIM, an acronym standing for Translator Aids Integrated on a Microcomputer, but we later changed its name to *TerM* when the product was launched on the United States market. *TerM* first became available in late 1983, and we have sold a good number of copies here in the United States. However *TerM* did not do everything that we thought it ought to do, so we kept on trying to improve its capabilities (and lower its cost from the rather high initial price tag of $200).

Our original goal had nevertheless been realised. We had indeed created in *TerM* a simple, fast and flexible program that allowed any translator who owned a suitable microcomputer to store and retrieve specialised terminology and to transfer the appropriate target-language equivalent for a given term from the term file directly into the current project being prepared with the word processing software.

All term entries were automatically alphabetised by source-language term. Retrieval of a given entry took less than a second after typing in the appropriate source-language entry, and just one keystroke would transfer the target-language equivalent from the term file into the current text. Any one term file could contain over 30,000 entries (or "records", to use the appropriate computer jargon), depending on the average record length. Although there were (and still are) no restrictions on the size, format or content of any individual entry, we

did offer some practical guidelines to users of the program on how they could best organise their terminological data, especially if they wished to exchange that data with others.

Nonetheless, we felt that *TerM* had three major drawbacks: the first was that while using the program, the computer screen remained permanently divided into two halves (top and bottom), the top half being for text preparation and the bottom half being the "window" into the active term file; the second was that *TerM* only worked with one specific word processing program (World Writer); the third was that term files could only be exchanged on disk with other users of the program (in other words, there was no "universal" exchange format) or in hard-copy form.

Therefore, between 1984 and 1985 we set out to remedy those inadequacies. Last October our efforts culminated in the introduction of the "third generation" of the program, now renamed *Mercury* and priced at a very affordable $95. This program is now published by Dr Melby's company, LinguaTech International, and my own company is an authorised dealer.

Mercury retained or improved upon all the best features of *TerM*, while adding a number of new capabilities and eliminating all of the major disadvantages described above.

Once loaded into memory (occupying about 140 kilobytes of RAM), *Mercury* remains totally invisible to the user until it is needed. When the translator wants to consult any of his term files, a single keystroke displays the *Mercury* window in the bottom half of the screen, temporarily covering up any text in that portion of the screen.

If the desired term entry is present in the file and if an appropriate equivalent exists, a pair of keystrokes will take the selected target-language term (up to 200 characters long) from the glossary record and transfer it directly into the current text (a simple "cut-and-paste" procedure). The *Mercury* window then disappears from the screen until the next time the translator wants to consult the glossary.

If the desired term is not present in the glossary, the translator can immediately add that term, its target-language equivalent(s), and any other pertinent information, thereby making the entry available the next time it is needed. Similarly, if a given source-language term already exists in the file but no suitable target-language equivalent is included for the given context, the translator can update his existing entry by inserting the new equivalent and any pertinent information about it. Under normal circumstances, the program will not allow duplicates to be entered into any given glossary.

If the translator has created and stored a number of different term files and if the desired term cannot be found in the active file, he can quickly instruct the program to change to a different file and proceed with his search. On a hard-disk system (NOT required to run the program but recommended for anyone planning to create large glossaries), dictionary look-up time can be measured in milliseconds, while adding or updating term entries is extremely easy and fast. Most importantly, at no time does the translator have to exit the word processor in order to work with his term file(s).

Mercury now works with almost ten popular word processing programs. One of the features of *Mercury* permits the user to programme the keyboard in any way desired for generating special and accented characters on the screen (as well as for handling other "macros"). Glossary creation can take place either in "interactive" mode (that is, while preparing a translation with the word processor) or separately with a special file-builder utility.

Mercury also features a special conversion utility that will allow the user to take any term file and turn it into a text file in what we call the "exchange format". Similarly, a glossary prepared in this exchange format with any standard word processor that uses the ASCII character set can be converted for use with *Mercury* in the interactive mode. The conversion utility also provides for merging two or more smaller term files into a larger one and for extracting a smaller term file from a large one (for example, all terms relating to a specific field or preferred by a specific client). Files in the exchange format can be edited and printed with any word processor, thereby making it possible for the translator to "publish" glossaries if desired.

But of greatest interest to translators is the fact that the *Mercury* exchange format offers a viable and universal standard for transferring terminological data files which contain entries in languages using a wide variety of special characters or diacritical marks.

The theoretical maximum size of any single term file produced with *Mercury* is 5 megabytes (five million entries at an average per-record size of 1000 bytes) – a considerable increase over the 32,000-byte maximum of a term file produced with its predecessor, *TerM*. The only limitation on the number of term files is the computer's available mass-storage capacity. For systems with floppy-disk drives only, the term files will by necessity be restricted to disk capacity (typically 360,000 bytes), but smaller glossaries can be kept on individual diskettes (by subject matter, for instance). In other words, with *Mercury* a translator can create any number of highly specialised small glossaries, yet the option also exists for creating one or more extremely large term files when a hard-disk system is used.

As was the case with *TerM*, all *Mercury* glossary entries are completely free-form and of variable length. Source-language terms can be up to 50 characters long, but over 1,200 characters of space is available for the target-language equivalent(s) and any other pertinent information, such as the subject field(s), source(s), contextual example(s), usage notes, and so forth. The sole requirement imposed on the user is that each piece of information about a given source-language entry (target-language equivalent(s) plus all other relevant data) must be preceded by some number/letter code enclosed in curly brackets (e.g., {1}, {f}, {s} etc.); in computerese these are called *field delimiters*.

Mercury also features a special cross-referencing capability, so that multiple-word term listings can appear both under their individual constituents as well as the complete phrase. When any of the constituents is looked up in the glossary, the program can be instructed to move directly and automatically to the full entry using that constituent (we call this feature *chaining*).

Besides its major intended application as a tool for translators who need immediate on-line access to terminological data, *Mercury* lends itself very well to the preparation of research notes, subject indexes for books, and other uses involving the organisation and management of information.

To summarise the advantages of *Mercury*:

1. It works interactively with the translator's word processor, giving direct and immediate on-screen access to the needed terminological information, including many terms not normally available in printed form.

2. It allows the translator to speed up his production considerably by using the "cut-and-paste" feature to transfer terms directly from the term file into the current text.

3. It completely eliminates the need to spend time on traditional manual "term-file management" chores, since *Mercury* automatically alphabetises all entries and makes them accessible with a single keystroke.

4. It provides multilingual capabilities in all common European languages plus Russian, both for the glossary entries and for text production (with the appropriate word processor).

5. It provides an extensive amount of free-format space for both the target-language equivalent(s) and any pertinent information.

6. It allows translators anywhere to create glossaries in a "universal format", to exchange them with colleagues, and to print complete or extracted versions of any glossary prepared with *Mercury*, whether for sale or for consultation in hard-copy form.

7. It is extremely affordable, even for a part-time freelance translator.

8. It can be purchased with low-cost "starter" glossaries in a number of subject/language combinations.

In short, about the only thing that *Mercury* will not do for the translator is the terminological research needed to identify the target-language equivalents and the other pertinent information about them. But once that essential work has been completed and the individual terms have been entered into the appropriate glossary, the translator is free to concentrate on the major task of preparing his translations.