

Time-sharing and computer-aided translation

E. O. LIPPMANN and W.J. PLATH

IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y.

The potential of time-sharing systems for aiding human translation

In the ALPAC report 'Language and Machines' [1], certain machine aids for human translators are described which were developed for use within a batch processing framework – that is, the traditional mode of computer facility operation in which individual jobs are batched prior to run time and processed sequentially using one input and output unit. The work referred to is that of the German Federal Armed Forces Translation Agency in Mannheim and the Terminological Bureau of the European Coal and Steel Community in Luxembourg. These organizations are cited as using computers to look up words and phrases unfamiliar to human translators working on individual technical documents, with a resultant increase in translator productivity on the order of 50 percent. (In both operations, expressions underlined by the translator in the input language document are keypunched and fed into a computer, to be looked up automatically in a technical dictionary. The computer prints out the results in the form of a text-related glossary which the translator employs as an aid in translating the document in question.)

The advent of time-sharing in electronic data processing has opened up promising new avenues for providing computational aids to the human translator. In a time-sharing environment, many people communicate with a computer simultaneously via terminals which look and operate much like regular typewriters. Since such terminals are linked to the computer by telephone lines, a user has considerable flexibility with respect to his choice of work location, provided that he has access both to a portable terminal and an ordinary telephone set. Compared to batch processing, a time-sharing system offers considerably greater potential for enhancing the productivity of human translators in two significant respects:

1. The computer operations can be smoothly integrated with the translation process. Instead of having to wait for a batch run to be prepared, submitted, and completed, the user is able to call on the resources of the computer as needed in the process of translation and get immediate response. Rapid iteration toward the desired goal (i.e., a finished translation) can be achieved by switching back and forth as many times as required among human translation, direct dictionary lookup, editing and printing via terminals.

2. An on-line system lends itself naturally to provision of a range of translation and manuscript creation aids much wider than dictionary lookup alone. Within such a system, revising, editing, and formatting, as well as dictionary lookup and updating, can be carried out conversationally by the user.

Experimental computer-aided translation

As part of a preliminary investigation of computer-aided translation in a time-sharing environment, various techniques of computerized dictionary lookup and text file management have been explored by E. O. Lippmann [2] of the Linguistics Group at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., under two time-sharing systems: TSS/360 and CP67/CMS. Under these systems, a user currently has an IBM 2741 typewriter as a link to the computer with an IBM 2260 display screen available as a special option. Without the screen, all computer responses (e.g., dictionary output) are transmitted to the typewriter.

Entering, and revising text at the terminal

Initially the user dials up the system on the telephone and signs on by typing in his credentials at the keyboard. The system prompts him to begin his work, which may be either creating new translation output or editing any existing text file. Even if the translator is not a skilled typist, he can enter translated text at his best speed, knowing that corrections can easily be made.

As the user types the first draft of his translation, it is simultaneously recorded by the computer. Typographical errors spotted before a line is completed may be corrected by simply backspacing over the incorrect characters and then retyping them as they should be. An entire line can be cancelled by striking the backspace key followed by a carriage return. If any question arises, immediate printout of the text entered so far can be requested. In order to issue such a request, the user must first enter 'edit mode', which he does by merely hitting the carriage return twice. From this moment on, any information typed, such as a print command, is treated as an editing request. To retrieve a single line, the user types 'retrieve' followed by a unique character or string of characters occurring in the line in question; alternatively he may refer to the line number in the text. The system will respond by printing the specified line on the terminal.

In edit mode, any string of characters (e.g., an untranslated expression or a misspelled word) can be replaced by another string (e.g., the corresponding translation or the correctly spelled word), either in single instances or wherever the original string of characters occurs in the text. If only a single occurrence is to be changed, the line containing the occurrence must first be retrieved as described above. It is then changed by typing the original character string followed by a delimiter and the new character string. The machine responds by making the requested substitution and typing out the corrected line at the user's terminal. If all occurrences of a given string in a text file are to be changed, it is not necessary to first retrieve any single occurrence; the system locates all occurrences by scanning the entire file starting at the top and types out each line that is altered. Verification of the changes may also be suppressed at the user's option.

Another double carriage return causes resumption of the text input mode, i.e., all typed information is again added to the original text. In fact, the user may flip back and forth at will between typing text and editing text simply by issuing a double carriage return in his current mode of operation.

Dictionary lookup

When in the course of translating (text input mode) the user encounters one or more source language expressions he wants to look up, he first returns to edit mode by

striking the carriage return key twice. Then he enters the name of the stored dictionary and the expression(s) to be looked up. If an expression is found, its dictionary entry is immediately typed out (or flashed onto the display screen if one is attached). If a sought-for term is not contained in the dictionary, a message is issued to that effect and the alphabetically nearest entries in the file are displayed.

In the case of a missing term, various options are open to the user. He can mark a source language term in the text with special characters (e.g., asterisks) for easy recognition in future editing operations. He can also send a message to any other user who may be on line with the computer at another terminal, perhaps miles away, requesting assistance. Entering 'talk to' followed by a user name and a message will cause the system to immediately transmit the message to the appropriate party. If he is not certain that the other party is also on line to the computer, the user may type 'who?', to which the computer will respond with the names of all users currently on line.

In returning to text input mode from dictionary lookup or edit operations, the user issues a double carriage return, which automatically positions him at the line where he previously left off. All new text will be inserted after the line in question. If he happens to be in the middle of the text file but wants to add lines to the bottom of the file, the user types 'bottom'. The system will respond with the last line in the file as a reminder and then append all subsequent input following that line.

Automatic document assembly

While typing or editing text, the user may intersperse special format control words in the text to control subsequent layout printing. Format control words take care of problems such as line spacing, justification, page numbering, and embedding of standard phraseology, tables, or other text segments. By typing in appropriate control words, the user can cause spaces up to the length of a full page to be left blank to provide for subsequent insertion of graphics. When a text file is printed on the computer printer or the typewriter in formatted layout, control words are not printed, but are instead interpreted by the system as specifying layout changes which are effective directly below the points at which the control words occur. Since printing may be stopped at any desired point, paper forms or type elements may be changed in midstream as required for the final product.

After the text of the draft translation has been recorded by the computer, it is saved by a) entering the edit mode via a double carriage return, and b) typing 'store' followed by whatever name the user chooses to assign to it. This name must be employed in accessing the stored file at a later time, when the user may wish, for example, to reprint part or all of the file on the typewriter for review and possible revision. To do this, the request 'printout' must be issued, followed by the name of the file and any additional information concerning the print format.

The edit facilities already used during the rough-draft creation can now be employed again for further text revisions based on the printout. The user can reprint replace, insert or delete text at any point in the original translation (or in any other text or dictionary file). If material is inserted into the file, the system will slide everything else back to make room for it. If material is deleted, the system will contract the text correspondingly. In either case, pagination will be adjusted automatically at printout time.

The user may wish to devote a terminal session exclusively to editing, to dictionary lookup, or to dictionary updating. He can always proceed at his own rate of speed. He need not be concerned about under-utilizing the computer during periods of slow progress or idle time on his terminal, since other users are being served (or 'background' programs are being run) while his connected typewriter or display screen is inactive.

Team of translators

Since the on-line system allows many users to work either independently or collectively on a variety of applications, a team of translators with differing responsibilities could use such a system to work cooperatively on large translation tasks. As the work progressed, translators could consult with each other and share text and dictionary files via their terminals. Since many files can be accessed concurrently, a user with an editorial function could edit an entire document (consisting of text portions processed by different translators) for consistency of format and terminology. The users would be entirely independent of secretarial or print-shop work schedules, since texts or dictionaries can be updated or inspected at will any time the system is on the air. The security of any file could be maintained by taking advantage of existing system provisions for rejecting access attempts by unauthorized users.

Status

The resources of the time-sharing computers at the IBM Research Center in Yorktown Heights, N.Y., are primarily employed by scientists, engineers and programmers to run scientific jobs from remote terminals. The edit, print and file handling capabilities of these systems are also increasingly being used for the creation of newsletters, manuals and library files. In addition to the experience gained in the use of interactive editing and manuscript formatting as publishing tools, which can largely be expected to carry over to the publication of translations, the feasibility of using specifically translation-oriented experimental software for performing such functions as on-line dictionary lookup, maintenance, and display has been demonstrated on a pilot basis. What is still critically lacking, however, is the sort of experience that can only come from having professional translators use such a system over a period of time while developing [the requisite data bases and dictionary files. It is precisely this sort of experience which will ultimately be decisive in determining the extent to which on-line aids to human translation live up to their considerable potential.

REFERENCES

- [1] Automatic Language Processing Advisory Committee (ALPAC), 'Language and Machines', Publication 1416 National Academy of Sciences, National Research Council, Washington, D.C., 1966.
- [2] E. O. Lippmann, 'An approach to computer-aided translation', Research Report RC 2588, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., August 20, 1969.