

## INTERACTIVE MACHINE TRANSLATION

### THE ALPS COMPUTER ASSISTED TRANSLATION SYSTEM

by Alan Danik Alps Systems

For the past three and one-half years, Automated Language Processing Systems Limited of Provo, Utah (U.S.A.), has been engaged in the commercial development and marketing of a range of multilingual document production tools including, but not limited to a computer assisted translation software package, designed with multicultural communication professionals in mind and intended to facilitate and accelerate the processing and management of documents in many languages at once.

#### AN HISTORICAL PERSPECTIVE

Automated Language Processing Systems Limited, otherwise known as ALPS, was formed as a private company on January 1st, 1980. Despite its apparent youth, the company nevertheless traces its roots back to the early 1970s, to a time when many within the organisation began their investigation of computer-aided translation techniques at Brigham Young University (Provo) on the basis of a rigorous theory of language being developed there which seemed to provide a means of representing linguistic structures in an unambiguous fashion and lent itself readily to computerisation.

The BYU-later-to-become-ALPS team opted for a set of compromises, reflecting the realities of translation in business, industry, and government. The ability to process unedited random text was a logical necessity, pre-editing is often very time-consuming, and writing protocols are generally difficult to enforce. Good quality output was a necessity as well, in order to minimise the amount of post-editing. It was therefore decided that human intervention during the translation process could significantly improve the quality of the unedited output by resolving ambiguities as they occurred.

## THE ALPS SYSTEM

### THE CHARACTER SET

The standard workstation provided with the ALPS system is composed of a video screen, and a detachable keyboard consisting of a typewriter-like keypad for typing, a cursor positioning keypad, and a function selection keypad.

The ALPS multilingual character set, comprised of more than 400 characters, includes the necessary alphabetical and numerical characters, diacritics, and punctuation marks for producing text in virtually all languages written with the Roman alphabet. Also included are many special characters for business, maths, and scientific applications.

The internal representations used for the ALPS multilingual character set maintain full compatibility with ASCII, EBCDIC, and Teletex multilingual character set. Therefore, no information loss occurs between those other representations and the ALPS representation.

### THE MULTILINGUAL WORD PROCESSOR

Of all the ALPS software packages, the word processor is capable of handling material in any of the languages that use the Roman alphabet, since the computer does not attempt any language related processing at this stage and no dictionaries are required.

Since the word processor offers the standard word processing capabilities of any dedicated word processor, it is possible to use the ALPS word processor in an office environment for tasks of a secretarial nature, such as preparing correspondence, transcribing dictation, or retyping text material that has been edited in any of the languages supported.

Translators and technical writers will appreciate the powerful electronic "cut and paste" facilities, available at the touch of a function key. Parts of other documents can be called up and incorporated into the file being worked on or, conversely, part or all of the text being edited can be tagged for automatic insertion in other documents, thus eliminating much retyping.

Since only the editing functions have been automated, no linguistic processing is carried out by the computer. The ALPS word processor used as a translation aid requires of the translator that he compose and type the entire translation, giving him complete and absolute control over his material and the translation process as well as greater freedom to explore different phrasings and different approaches.

#### SELECTIVE DICTIONARY LOOKUP

Selective Dictionary Lookup, or SDL, is an enhancement that may be added to the ALPS multilingual word processor, making user-built dictionaries available from within the text editor at the touch of the lookup function key on the keyboard. Creating or updating an SDL dictionary entry is barely more time consuming, and no more difficult, than typing the required characters. The user can create as many dictionaries as are necessary, all of which remain accessible at all times. Furthermore, the data selected from a dictionary is inserted by the computer precisely where it is needed in the text being worked on; therefore, no retyping is required.

Selective Dictionary Lookup involves no linguistic processing; therefore, there are no restrictions on language for dictionary entries, other than the use of the Roman alphabet. Consequently, these dictionaries may serve as translation dictionaries from one language into several others at once. They can be used as well to provide correspondence between regional variants of the same language, e.g. - British, American, and Canadian English.

SDL dictionaries need not be "bilingual". Monolingual dictionaries of synonyms, antonyms, correct spellings, etc. are also feasible. In fact, SDL dictionaries may contain any type of data.

All operations in the SDL mode are carried out interactively and strictly on command. Dictionary assistance is provided only when requested, so the user receives no unnecessary and potentially distracting information.

#### \*AUTOMATIC DICTIONARY LOOKUP

One step beyond SDL, the Automatic Dictionary Lookup, or ADL, frees the translator from having to specifically request the translation equivalents for each technical or specialised term in his document by automatically retrieving the corresponding target language vocabulary from the dictionaries of his choice.

No syntactic processing is attempted by the computer in the ADL mode. However, a morphological analysis routine enables it to associate inflected forms (such as the English computer terminals) with the appropriate uninflected or base form (computer terminal) found in the dictionary.

The ADL programs compare the source language document pertinent to the material to be translated. The translator may then request a Keyword in Context Listing (KWIC) listing, which provides him with an alphabetized printout of all the words in the document along with context, and a Words-not-Found (WNF) listing in order to complete his dictionary before beginning to translate.

In the ADL mode, the video screen is split horizontally in three. The upper section contains a source language segment. This segment is repeated in the middle section, where each technical or specialised term is numbered and highlighted, along with its target language equivalent. The bottom section is a work area where the translator types his translation, pulling vocabulary down as he needs it, without retyping, by hitting the key whose number corresponds to that of the appropriate word or phrase. In this manner, typing is kept to a practical minimum, as is the risk of typing mistakes, and lexical consistency is ensured through automatic consultation of the computerised dictionaries. ADL is available or under development, for English, French, German, Spanish and Italian.

## COMPUTER TRANSLATION SYSTEM

To the automatic dictionary lookup function, the Computer Translation System (CTS) adds linguistic processing at the levels of grammar and syntax, computer consultation with the translator, as well as supervision and editing facilities.

The procedure is as follows: after the material meant for translation has been entered in the system, it is compared with the entries contained in the dictionaries. A working or document dictionary is automatically assembled by the computer in which are found only those source language words or phrases, along with their target language equivalents, actually occurring in the document to be translated.

Once the working dictionary has been set up by the computer, the translator can ask for a KWIC listing, with or without words not found and frequency of occurrence information. Meanwhile, the computer has divided the source text into manageable segments, according to pre-established rules, in preparation for the translation.

Again in CTS mode, the video screen is partitioned horizontally. The uppermost strip contains a segment of source text below some reference information. The middle strip will contain a draft translation of the segment appearing in the area above once processing of that segment has been completed. The bottom strip is a work area.

As the computer analyses the source language segment, it may encounter ambiguities relating to segment definition, spelling, lexical choice, or syntax. It would then ask the translator to answer specific questions regarding the ambiguities. For example, if there were more than one target language equivalent for a given source language word or phrase, the system would display all possible choices it had in the working dictionary, with each choice having a number. The translator would then select the appropriate translation by typing the corresponding number. Or, there may not be a target language equivalent in the working dictionary; the computer would then ask the translator to provide one.

After a segment has been processed in its entirety, a suggested rendering of that segment appears in the middle third of the video screen with each separate item (number, word, punctuation mark, etc.) numbered. Inflection, conjugation, agreement, and word ordering have been taken care of by the system.

The translator must now decide whether the proposed translation needs editing, in which case he would probably choose to do so immediately. The CTS software enables him to carry out editing quickly and effortlessly by using a powerful text editor. If the required editing is substantial, as sometimes happens, it is more efficient to rewrite the entire target language segment than to attempt to deal with component items individually.

Once a satisfactory rendering has been achieved, the translator calls for the subsequent segment, in the process sending the current one to final editing and printing stage and for any preparatory work such as formatting before typesetting. After completion of the final editing step, the text can be printed out on the systems letter-quality printer, recorded on magnetic tape for electronic typesetting and printing or output in any other form requested.

Language pairs currently available or under development are English-to-French, English-to-German, English-to-Spanish, English-to-Italian, and reversals (ie. French-to-English, German-to-English, etc.).

#### HIGHLIGHTS OF THE ALPS SYSTEM

##### COMPLETE USER-FRIENDLINESS

Despite the complexity and the processing power of the software, the ALPS automated translator aids are not difficult to learn, nor are they difficult to use. The software package includes self-paced on-screen tutorials, enabling the user to become familiar with the essential features and functions of the system and to try them out in a controlled environment. All process selection is accomplished through menus, as are the translation management tasks.

The number of dictionaries created, and their size, are limited only by the ultimate amount of memory space available and by practical considerations concerning the use of very large dictionaries. Similarly, there are few restrictions on the accessibility of dictionaries. The complete set of permanent and working dictionaries are available at all times from within any process to all translators working on the system. The same dictionary can be accessed by many translators at once, conflict arising only if more than one user wishes to consult a given entry at the same instant.

An ALPS dictionary entry comprises a source language word or phrase, and one or more target language equivalents with word category and grammatical feature information, supplemented in some cases with dictionary management information or translation selection information meant for the translator. The coding of entries does not require the expertise of a programmer and a linguist.

Dictionary entries can be transferred from one dictionary to another without retyping and any dictionary can be printed out in whole or in part.

System dictionaries can be extracted from, or copied whole into, a working dictionary at the user's option. The translator may also choose to retain a list of words not found, since this type of updating can be attended to before starting the translation or in the course of translating. Finally, it is the translator's decision as well to have the working dictionary assembled at once or entered in a queue for batch processing along with other distractions overnight or at a convenient time.

Computerised dictionaries ensure terminological accuracy and consistency as well as spelling consistency from one document to the next, indeed from one translator to the next, while simplifying and accelerating the lexical data base maintenance operations.

#### FLEXIBILITY OF OPERATION

The selection of the most appropriate translator aid or level of assistance must take into account all the factors affecting a given job, such as translator preference, source and target languages, type of material to be translated, and the ultimate purpose of the translation. It is up to the translator to decide what involvement he wishes the computer to have in the translation process. Furthermore, depending on which source and target languages are needed, the range of automated aids at the translator's disposal is more or less complete.

Where information content or vocabulary is of prime interest, ADL or unedited CTS output may be quite appropriate, whereas a translation which must be highly polished could be produced in CTS mode with subsequent editing, or using either ADL or the word processor (with or without SDL) if the translator preferred complete freedom in composing his text.

Never is the translator excluded from the translation process and never does the translator relinquish his decision-making prerogative, whatever the extent of computer assistance chosen. On the contrary, his expertise is used to good advantage in organising the work and in guiding and supervising the computer as it goes about the task. The interactive approach adopted by ALPS enables the translator to retain full control over the target language material being produced and consequently full responsibility for it.

#### PROCESS CUSTOMIZATION

Processing needs differ from user to user and some of them may not be met by the ALPS system. The ALPS programs and the Data General operating system under which they run can be modified only by ALPS personnel. Nevertheless, it is possible for a user to make the system more responsive to his particular requirements without ALPS involvement by devising his own programs, called macro-instructions.



With no previous programming experience and no knowledge of a programming language, a user is able to create his own sophisticated programs using facilities made available to him by the ALPS software, more specifically by the word processing software. The user need only be familiar with the various text editing functions and have a clear perception of the problem facing him and of the solution he seeks. By "teaching" the computer how he wishes the problem to be solved, pressing, in an appropriate sequence, any number (up to 100) of function keys, cursor positioning keys, etc., the user writes a small program which he may execute at once or save in the computer's memory for execution at some later time, with as many repetitions as is necessary. Therefore, "learning" macro-instructions constitutes an elegant automated solution to repetitive and cumbersome editing problems.

The more adventurous user who has acquired some familiarity with operating system commands and how they execute has at his disposal a more powerful type of macro-instruction, useful for system management or data input and output tasks. Such a user, with a little practice, should be able to write his own programs allowing him to automate operational procedures or handle text editing of a type not possible using the standard software. It will also be possible for personnel not conversant with computers to assist with the daily operation of the ALPS system loading material off or on to magnetic tape, establishing communication with word processors or other computers to exchange files, etc.

#### • INTERFACING CAPABILITIES

With the appropriate software interfaces, the ALPS system is capable of communicating with virtually any other computer, word processor, terminal, printer, or piece of typesetting, photo-typesetting and photo-composition equipment. Source language text can be input by downloading from another computer, over communications lines, or from magnetic tape which in turn may have been created via diskette conversion or Optical Character Recognition (OCR). Target language material, including any typesetting codes present in the original, can be output in the desired form, irrespective of the ALPS process which produced it.

#### EXTENDING APPLICATIONS

ALPS software runs on a general purpose mini-computer. Through selection of hardware components, and with the addition of appropriate software packages offered by Data General and other vendors, the user can design a system made to order for his particular requirements, capable of meeting not only his document production needs in many languages, but his other information processing and management objectives as well, such as general accounting, payroll, electronic filing, electronic mail, planning, scheduling and so on.

For Spanish into English (SPANAM), and, especially in the past year, development of machine translation in the other direction, i.e. English into Spanish (ENGSPAN). We believe very strongly that the project's viability comes from the opportunity we have had to work on different aspects of the system at the same time.