ABSTRACT

Many individuals, companies, and government agencies need to have a large volume of foreign language printed matter translated for their use. Computer-aided translation offers the advantages of speed and volume over the normal non-assisted human translation process. With a shortage of trained language specialists, the volume of material to be translated, and the speed of translation which is desired, the computer-aided translation market becomes the most feasible and desirable option in the translation field.

Weidner is one company which recognized the needs in this particular area and has developed both hardware and software to fill this need. Constant improvement and attention to detail is needed to keep such a system operating at top accuracy. It is Weidner’s goal to stay on top of all advances in the field as well as to offer a complete line of language and translation services to the community.

KEYWORDS: machine-assisted translation, computer-aided translation, translation, Weidner, software, MicroCAT, MacroCAT, hardware, English, Arabic, French, German, Portuguese, Spanish, Japanese, Italian, custom translation

Since 1977, Weidner Communications Corporation (WCC) has had as its primary goal the development of an effective, efficient computer-aided language translation system. WCC’s approach has been one which recognizes both the advantages and limitations of computer-aided language translation, using them to create a cost-effective product which can be used on its own or as a link in a complete language-processing tool for individual translators, companies, and government agencies with steady volumes of translation.

State of the art computer-aided translation systems have not achieved 100% accuracy in reproducing human speech, or more correctly, human styles of writing. Therefore, WCC made the decision early on to incorporate into its product the most accurate rates the system can achieve. The company is aware that improvement of the translation routines will certainly be part of the future for this new technology, and is already making advances in that direction. However, the success of the WCC system is not contingent wholly upon linguistic accuracy.

The success WCC has enjoyed to date is due to a combination of factors. Foremost is the variety of hardware and software configurations in which the WCC Computer-Aided Translation (CAT) system is available. These range from multi-user minicomputer-based configurations which utilize the Digital Equipment Corporation PDP-11 and VAX computers (WCC’s MacroCAT) to single-user systems on IBM PC XT and ITT Xtra micros (the MicroCAT line). The MacroCAT systems offer a batch multiprocessor, which allows execution of a variety of jobs from one terminal. This batch processor and the deferred modes available on both the MacroCAT and MicroCAT optimize use of the computers during off hours. The WCC stand-alone multilingual word processor is an integral part of both WCC systems. For the multi-user system, MacroCAT, it has the specific advantage of bringing the editing and text entry functions off-line, thereby increasing the capacity of the central processing unit to handle on-line jobs.

A considerable number of language directions are available on the MacroCAT systems: English to Arabic, French, German, Portuguese, and Spanish; French, German, and Spanish to English; Japanese to English. The single-user MicroCAT system offers the above language directions with the exception of English to Arabic, German to English, and Japanese to English. English to Italian will be released on both the MicroCAT and the MacroCAT later this year.

Flexibility in system configuration includes the ability to interface with other types of input and output hardware. WCC systems presently interface with a variety of terminals, printers, typesetting machines, and optical character readers. It is WCC’s aim to constantly increase the number of interfaces it offers, thereby facilitating installation of CAT systems in a variety of environments.

The modular design of WCC’s software is another factor in the success of the CAT systems. Computer-aided translation technology is new; improvements and product enhancements are identified frequently. WCC has chosen to develop a product which can be updated often and to which new features can be added with relative ease. These improvements include enhanced translation
routines, new word processing features, improved dictionary building menus, and improved file maintenance features. After improvements are identified and tested by WCC, they can then be released to the customer for incorporation into the existing system. The software is menu-driven, so that new or enhanced features generally require no or very little additional learning by the user.

WCC provides a Core dictionary that can be expanded to suit the customer’s needs. Specific dictionaries can be created to accommodate the glossaries of different subject areas. Additionally, the MacroCAT system offers a lookup dictionary which is accessible from the word processor. Combined, these dictionaries give the translator an extremely high degree of control over lexical output.

**WORKING WITH A CAT SYSTEM**

The translation workflow using a CAT system differs somewhat from the traditional translation process. Five basic steps are involved.

**Input**

Before they can be processed on a CAT system, documents must be input. Depending on the system’s configuration, data can be entered in a number of ways: via word processor keyboard, by scanning with optical character reading devices, by transmission over telephone lines via modem, or on previously prepared nine-track tape, or diskettes. Once the data has been input, it will never need to be typed again. This in itself is a tremendous advantage when compared to traditional methods of translation in which texts are handwritten and typed in final format, or typed and retyped to achieve a perfect copy.

**Dictionary Management**

After texts have been entered into the system, they must be compared to the system’s dictionary to identify words the dictionary does not contain. This is done by performing a vocabulary search of the source text. Once identified, unknown words and phrases are then entered into the dictionary by a translator or terminologist who has received training from WCC in this process. They remain there until the user chooses to modify or remove them. Entries to the dictionary can be checked for accuracy with the line-by-line immediate translation feature.

**Translation**

The translation process, like a vocabulary search, can be carried out during off hours. This allows the most efficient use of both computer and translator time. Computer-generated translation is produced at a rate of 6000 to 8000 words per hour on the MacroCAT system; on the single-user MicroCAT the rate is approximately 1500 words per hour. What occurs in the central processing unit during translation is described in greater detail below.

**Post-editing**

Once a text has been translated, it becomes the job of the translator to post-edit the translated text. This task includes correcting imperfect grammar, identifying and correcting semantic discrepancies, and adding style. The on-line dictionary accessible from the word processor allows MacroCAT users to access alternate translations while editing. After working with machine-generated text over a period of time, translators learn to anticipate the accent of the machine, and generally reach speeds of 600 to 1200 words per hour editing raw translation. Translators also provide feedback about dictionary entries to the dictionary building responsibility. This allows the dictionaries to constantly be improved and fine-tuned.

**Output**

The final step in the process, once the translation has been perfected, is to output the documents to some type of device. The WCC systems allow texts to be printed on a variety of letter-quality and dot-matrix printers, moved to diskette or nine-track tape for storage, telecommunicated to other locations, or sent to typesetting equipment for printing. The choice of output method depends upon the system’s hardware and software configuration, as well as the customer’s needs.

**THE CAT TRANSLATION PROCESS**

The WCC translation software, although easily accessible to the user, employs some of the most sophisticated routines available. These fall into four basic categories, based on function.

**Pre-analysis**

In pre-analysis, the text to be translated is broken into sentence units. Formatting characters are inserted, embedded typesetting commands are identified, and contractions and possessive forms are found and expanded into their full forms. This information is stored for use later in the translation process.

**Dictionary Lookup**

During this stage of the translation process, source words are morphologically analyzed to identify their base forms (singular nouns, infinitive forms of verbs, positive forms of adjectives and adverbs). This information, like the data collected during pre-analysis, is preserved for later use. The target translations are then located in the system dictionary.

**Analysis**

Two different types of constructions are morphologically and syntactically analyzed by this module of the program: verb strings, which are analyzed and compacted into single units, and multiple-word idiomatic phrases, which are analyzed and compacted under a key word chosen during the dictionary entry process.

Take for example the sentence “The life insurance company has been offering lower rates.” The phrase “life insurance
company” could be entered to the dictionary as an idiomatic expression. In this instance, the phrase would be compacted as a noun under the key word “company.” The verb string “has been offering” would be compacted to the single word “offer,” and information relative to tense, mood, auxiliary verbs, etc. would be retained for later expansion into the target language equivalents.

Homographs are also resolved in this phase of translation. The WCC system defines a homograph as a word which functions as more than one part of speech in its base or inflected forms. “Offer” in the sentence above is an example of such a homograph, as it can function both as a noun and a verb. Although the function of “offer” in this sentence is clear, in some cases the distinction is less evident, as is the case with –ing forms of English verbs. These can be verbs, nouns, gerund nouns, or adjectives, and generally their function is less easy to discern.

Lastly, the phrase and clause structure of the sentence is analyzed using WCC’s own parsing routines. Just as with the morphological analyses performed earlier, information relative to the function of the different phrases and clauses is stored for use during the transfer stage.

Transfer

By the time this point is reached, the system has syntactically and morphologically analyzed all elements of the sentence: individual words, verb strings, and idioms, phrases, and clauses. Translations for the single words, verb strings, and idioms have also been located in the system dictionary. During transfer, these sentence elements are reordered and expanded into the correct target language structures. Insertion and deletion routines add or remove elements which are present in the source language, but not in the target language, and vice-versa.

Synthesis

In the final stages of the translation process, the morphological information gathered from the source language during the dictionary lookup stage is applied to the target language. Agreement between elements in the target language is ensured, and minor adjustments are made to certain target constructions. For example, contractions commonly used in the target language are created at this point. Finally, the entire translation is output to a file in the same format as the source text, based on the formatting information gained during the preprocessing stage.

This translated file can then be brought into the WCC multilingual word processor for post-editing, be printed, stored, typeset, or otherwise processed according to the customer’s needs. And of course, all dictionary building procedures that are performed in order to translate a particular text are permanently stored in the system’s dictionary, where it can be accessed by future translations.

THE WCC TRANSLATION SERVICE BUREAU

Proof of the cost-effectiveness of the WCC Computer-Aided Translation System can be seen in its Translation Service Bureau (TSB). The bureau was established in 1982, and has already grown into one of the largest (by revenue) in the country. The primary business of the service bureau is translation of large volumes of texts for a variety of customers using the WCC translation system and software, although jobs for which no translation software exists are also accepted. Already the TSB has experienced expansion into the area of typesetting, and this facet of the bureau continues to grow. The ultimate goal is to become a total service language processing center with the capability of performing all aspects of document publication, with the WCC translation software as a major link in the chain.

At present, approximately fifty translators are employed by WCC’s TSB. They represent numerous fields and include native speakers of all languages translated by the WCC system. With this outstanding staff, a high-powered central processing unit, and some twenty-five terminals (many of which are off-line editing stations), WCC is producing translations in a variety of subject areas at rates of up to 800,000 to 1,000,000 words per month. Additional terminals in the homes of some staff members allow text entry and editing to take place outside the office, giving those who choose to work in their homes that opportunity.

The TSB also functions as the test site for new releases of WCC’s translation software. All software releases are given to the TSB prior to their distribution to customers. The software is given an examination more thorough than any laboratory setting could reproduce, and the TSB actually generates income while the testing is being done. The TSB and WCC’s research and development team work together closely to ensure that the software released to the customers is of high quality.

Weidner Communications Corporation intends to continue to perfect and expand its product line, as well as offer an ever wider array of translation and language processing services for the benefit of those needing such services.

Author’s Address

Margaret M. Perscheid
928 South Alfred St.
Alexandria, VA 22314