



Deseret Language and Linguistic Society Symposium

Volume 8 | Issue 1

Article 16

3-26-1982

Integrating Word Processing, Term Management, and Machine Translation

Alan K. Melby

Follow this and additional works at: <https://scholarsarchive.byu.edu/dlls>

BYU ScholarsArchive Citation

Melby, Alan K. (1982) "Integrating Word Processing, Term Management, and Machine Translation," *Deseret Language and Linguistic Society Symposium*: Vol. 8 : Iss. 1 , Article 16.
Available at: <https://scholarsarchive.byu.edu/dlls/vol8/iss1/16>

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

INTEGRATING WORD PROCESSING, TERM MANAGEMENT, AND MACHINE TRANSLATION

Alan K. Melby

Linguistics Department
Brigham Young University

At last year's DLLS symposium (March, 1981), the author proposed on a "suggestion box" translator aid. In October 1981, the system became operational and was tested by the students in a translation seminar. Further consideration of the problem of computer aids for translation, together with the many good ideas put forth by the seminar students, has resulted in a proposal for a significantly expanded system which includes the "suggestion box" aid as one component. This new translator aid system integrates word processing, term management, and machine translation.

Traditionally, machine translation systems were designed with the long-range goal of replacing the human translator. The system proposed in this paper, on the other hand, is designed to be a tool for a human translator, **never** a replacement. The new system will have three levels. Level two corresponds to the "suggestion box" aid of last year. Level one is a lower level which does not even require the source text to be available in machine-readable form. Level three is the highest level and requires a remote machine translation system which can operate without the presence of a translator. Levels one and two are now being programmed on the IBM 370/138 computer at the BYU Humanities Research Center. Work on level three will begin next year.

THE "ALL OR NOTHING" SYNDROME

Originally, fully automatic high-quality translation was the only goal of research in machine translation. Until recently, there seemed to be a widely shared assumption that the only excuse for the inclusion of a human translator in a

machine translation system was as a temporary, unwanted appendage to be eliminated as soon as research progressed a little further. This "all or nothing" syndrome drove early machine translation researchers to aim for a fully automatic system or nothing at all. It is now quite respectable in computational linguistics to develop a computer system which is a **tool** used by a human expert to access information helpful in arriving at a diagnosis or other conclusion. Perhaps, then, it is time to entertain the possibility that it is also respectable to develop a machine translation system which includes sophisticated linguistic processing yet is designed to be used as a tool for the human translator.

If each sentence of the final translation is expected to be a straight machine translation or at worst a slight revision of a machine translated sentence, then disappointment is probable. After experimentation, Brinkmann concluded that "the post-editing effort required to provide texts having a correctness rate of 75 or even 80 percent with the corrections necessary to reach an acceptable standard of quality is unjustifiable as far as expenditure of money and manpower is concerned" (Brinkmann,1980). Thus, a strict post-edit approach must be nearly perfect or it is almost useless. Many projects start out with high goals, assuming that post-editing can surely rescue them if their original goals are not achieved. But even post-editing may not make the system viable.

A PROPOSED ALTERNATIVE

This paper proposes that an interesting alternative to the "all or nothing" approach is to anticipate from the beginning that not every sentence of every text will be translated by computer and find its way to the target text with little or no revision. Then an effort can be made from the beginning to provide for a smooth integration of human and machine translations. The proposed translator-aid system (TAS) will have three integrated levels of aid under the control of the translator. We will now describe the three levels.

Level one translator aids can be used immediately even without the source text being in machine-readable form. In other words, the translator can sit down with a source text on paper and begin translating much as if at a typewriter. Level

one includes a word processor with integrated terminology aids. For familiar terms that recur there is a monolingual expansion code table which allows the user to insert user-defined abbreviations in the text and let the machine expand them. This feature is akin to the "macro" capability on some word processors. The key can be several characters long instead of a single control character, so the number of expansion codes available is limited principally by the desire of the translator. Level one also provides access to a bilingual terminology data bank. There is a term file in the microcomputer itself under the control of the individual translator. The translator may also have access to a larger, shared term bank (through telecommunications or a local network). Level one is similar to a translator aid proposed by Leland Wright, a well-known professional translator. Ideally, the translator would also have access to a data base of texts (both original and translated) which may be useful as research tools.

Level two translator aids require the source text to be in machine-readable form. Included in level two are utilities to process the source text according to the desires of the translator. For example, the translator may run across an unusual term and request a list of all occurrences of that term in that text. Level two also includes a "suggestion box" option (Melby,1981) which the translator can invoke. This feature causes each word of the current text segment to be automatically looked up in the term file and displays any matches in a field of the screen called the suggestion box. If the translator opts to use the suggested translation of a term, a keystroke or two will insert it into the text at the point specified by the translator. If the translator desires, a morphological routine can be activated to inflect the term according to evidence available in the source and target segments.

Level three translator aids integrate the translator work station with a full-blown machine translation (MT) system. The MT component can be any machine translation system that includes a self-evaluation procedure. The system uses that procedure to assign to each of the translated sentences a problem rating (e.g. "A" means no detected problems, "B" means some uncertainty about parsing or semantic choices made, "C" means probable flaw, and "D" means severely deficient).

The actual machine translation for level three is done remotely on a separate computer without the direct involvement of a human translator. Then the segmented source text and the machine translation for each segment, together with its self-assigned "grade", are placed on a diskette and sent to the translator. The translator works at a small station which, ideally, is a self-contained microcomputer which is programmed to support all three levels of aid. Level one, as mentioned previously, requires no diskette containing source text. This means that at level one, the translator can get straight to work on a new document. At level two, a diskette containing source text is needed before the translator can begin work. And at level three, a diskette containing source text and machine translation is needed before work can begin.

At level three, on any segment, the translator may request to see the machine translation of that segment. If it looks good, the translator can pull it down into the work area, revise it as needed, and thus incorporate it into the translation being produced by the translator. Or the translator may request to see all those machine translations that have a rating above a specified threshold (e.g. above "C"). Of course, the translator is **never** obliged to use the machine translation unless the translator feels it is more efficient to use it than to translate manually. No pressure is needed other than the pressure to produce rapid, high-quality translations. If using the machine translations make the translation process go faster and better, then the translator will naturally use them.

A positive aspect of this three level approach is that while level three is dramatically more complex linguistically and computationally than level two, level three appears to the translator to be very similar to level two. Level two presents key terms in the sentence; level three presents whole sentences. At level three, any segment which does not have a qualifying machine translation will cause a smooth, automatic shift to level two for that segment and back to level three for the next qualifying segment. So, when good level three segments are available, it can speed up the translation considerably, but their absence does not stop the translation process or even greatly hinder it. Thus, a multi-level system can be put into production much sooner than a conventional post-edit system. And the sooner a system is put into production, the sooner useful feedback is obtained from the users.

CONCLUSION

The multi-level approach described in this paper is designed to please (a) the sponsors (because the system is useful early in the project and becomes more useful with time), (b) the users (because they are in control and choose the level of aid), and (c) the linguists and programmers (because they are not pressured to make compromises just to get automatic translation on every sentence).

Future papers will report on progress and problems in the design and implementation of the translator aid system described in this paper.

REFERENCES

- (1) Andreyewski, Alexander, Translation: Aids, Robots, and Automation, META Vol. 26, No. 1 (March 1981) 57-66.
- (2) Baudot, Jean, Andre Clas, and Irene Gross, Un modele de mini-banque de terminologie bilingue, META, Vol. 26, No. 4 (1981) 315-331.
- (3) Boitet, Ch., P. Chatelin, P. Daun Fraga, Present and Future Paradigms in the Automatized Translation of Natural Languages, in: COLING80 (Tokyo, 1980).
- (4) Brinkmann, Karl-Heinz, Terminology Data Banks as a Basis for High-Quality Translation, in: COLING80 (Tokyo, 1980).
- (5) Kay, Martin, The Proper Place of Men and Machines in Language Translation, Xerox Palo Alto Research Center Report (October 1980).
- (6) Lippman, Erhardt, Computer Aids for the Human Translator, Report presented at the VIII World Congress of FIT, Montreal (1977).
- (7) Melby, Alan K., Melvin R. Smith, and Jill Peterson, ITS: Interactive Translation System, in: COLING80 (Tokyo, 1980).
- (8) Melby, Alan K., Linguistics and Machine Translation, in: James Copeland and Philip Davis (eds.), The Seventh LACUS Forum 1980 (Hornbeam Press, Columbia, SC, 1981).
- (9) Melby, Alan K., A Suggestion Box Translator Aid, in: Proceedings of the annual symposium of the Deseret Language and Linguistic Society, (Brigham Young University, Provo, Utah, 1981).