Summary of FBIS Seminar on Machine-Aided Translation

S. R. Petrick
IBM T. J. Watson Research Center
Yorktown Heights, N. Y. 10598

In the first paper delivered at this conference Wallace Chafe presented the following model of translation: a source language sentence is first parsed to produce a surface structure. This is converted by some process of comprehension to a deeper, conceptual structure that reflects the meaning of the sentence in a more direct way. This conceptual structure may or may not be a language-independent universal structure. In those models where it is not universal but instead is tailored to the source language, it must be converted to a corresponding conceptual structure that is similarly specific to the target language. In any case, conceptual structures must be mapped by a verbalization process into corresponding target language surface structures whose debracketizations yield the required target sentence output.

Other speakers suggested extensions to this model, for example, to provide for context beyond isolated sentences. Basically, however, Chafe's model provides a good basis for discussing the translation efforts which were described by the other speakers at this conference. For example, one way in which different systems roughly based on Chafe's model can vary is in the relative depth of their conceptual structures. Actual systems that were discussed varied in this respect all the way from rather abstract structures that directly represented meaning to shallow structures whose relationship to corresponding sentence meanings was, at best, tenuous. All of the commercially intended MT systems which were described appeared to rely upon such shallow structures, in some cases on surface structure itself. In most cases this was explicitly stated by speakers at this conference, and in other cases it could be inferred from outright errors in exhibited sample output where intended meaning was not correctly determined. All of these MT systems, however, exhibited what might be

called extensive coverage of the source language, i.e., output was produced for every source language sentence (undoubtedly also for ungrammatical source language utterances).

In contrast to the commercially intended MT systems stand the Artificial Intelligence systems for natural language understanding, which in most cases have yet to be applied to MT. Their advocates point out the necessity for deeper conceptual structures as well as supplementary information and inference in order to adequately translate certain sentences. They pay a price, however, for their insistence on more adequate conceptual structures, because those structures are not easily obtained for unrestricted text input. It is equally true of the AI and Computational Linguistics systems, whether based on formal grammars or procedurally defined, that the coverage of the source language provided is currently very sparse. Due to the fact that most source language sentences are not processed by these systems, they are unsuitable for unrestricted text and have been applied only to question answering systems and to restricted toy-world domains. The amount of effort required to extend the coverage of, say English, to a state useful for MT while maintaining the adequacy of assigned conceptual structures might be variously estimated by different authorities, but it is my opinion that it is very large indeed, large enough to make such applications as question answering systems more attractive candidates for consideration in the next few years.

Another point to note in conjunction with all of the systems discussed at this conference is that their treatment of the process Chafe referred to as verbalization is rather primitive. Thus in spite of the fact that this aspect of a computational linguistic system is often referred to as uninteresting or trivial compared to the task of understanding an input utterance, and in spite of the fact that many normally difficult facets of verbalization do not present

a problem in MT, the current output of language processing systems is very unnatural and rough. This is true of AI systems as well as operational MT systems.

If, in fact, we examine the specific realizations of the components in Chafe's model which were reported to be included in the MT systems described at this conference, we find very few changes over the situation that prevailed ten years ago. The comprehension component is realized by such means as a context free grammar, a Q-System, or an analysis-based ad hoc procedural specification. Difficulties and shortcomings related to conceptual structures have already been noted. These have changed very little over the past few years. Similarly, we have already commented on that portion of the target language output inadequacy which is attributable to shortcomings in the treatment of verbalization. In summary then, currently operational or projected MT systems are only marginally different in their underlying organization and design than their predecessors.

If, then, there is little that is novel about the underlying models of current and projected MT systems, it is natural to ask how many hardware and software improvements have been made. Several claims were made about improvements in procedural programming languages. Although I am fully aware of the benefits which follow from the use of a well suited programming language, I don't think the improvements which are claimed are very significant. For one thing, many language processing tasks are still very difficult to program using the best programming languages. And for another, convenient programming is no substitute for the absence of satisfactory models and algorithms. Recent advances in editors and time sharing systems might, however, be significant factors in making the development of machine-aided human translation more attractive.

Hardware developments of the past decade include time sharing hardware, automatic photocomposition devices, larger primary and secondary storage, faster processing speeds, and lower costs. Optical character recognition was reported not to have advanced significantly in the past few years. There is still a limitation to a fixed set of fonts, and the only large scale applications at this point involve fonts carefully designed for OCR.

We have seen increases in computational power per unit cost and can expect to see more such increases. The question which arises, however, is what their effect is likely to be on MT. The key issue is how much of the total effort can be handled by a computer and how much must still be done by human labor. Text input, pre-editing, and postediting can take as much human time and effort as complete human translation.

Of critical importance is the evaluation of current MT systems to determine the quality of their unedited output, the uses for which such output is acceptable, and the amount of postediting that is required to meet well defined higher standards. No clear results of this type were provided at the conference and careful study is necessary to resolve certain seemingly contradictory claims. Thus, there were reports of translation output which was not postedited, other output which was only lightly postedited, and still other output that was extensively postedited. The implication was given that no more editing was required than was given, and, although there is a sense in which that claim is undoubtedly true, it fails to take into consideration the quality of the output, the purpose for which the translation was requested, and the degree of requestor satisfaction. Although I did not systematically examine large quantities of source language input and corresponding unedited target language output, the examples which I did examine

suggested a rather low level of performance with respect to both fidelity of meaning output and to smoothness and naturalness of the output. The overall quality of output produced strikes me as comparable to that of ten years ago, and a colleague of mine with more experience in MT than my own assessed the output I showed him as more ambitious in its attempt to achieve natural output than past systems but probably not any more successful. Attempts to produce natural target language word order and correct insertion of articles helped in some cases but just as often made the translation worse. Clearly, it is no simple task to evaluate the quality of output achievable through the use of a particular MT system, to determine the amount of post-editing necessary to bring it up to required standards of quality, and to estimate the likely cost of achieving that quality. Each prospective user of an MT system must carefully do this, but from what was presented at this conference I would not expect any current MT systems to compete economically with human translation except in those few cases where requirements for quality and accuracy are so low that unedited or very lightly edited output suffices.

In addition to postedited MT, this conference also discussed the use of hardware and software aids to human translation. There seemed to be a consensus that well-engineered systems can be produced now, that their use looks promising, and that they probably are limited to increasing the productivity of human translators by a factor of 2-1 or 3-1. Opinion was divided as to whether they might evolve into human-aided MT systems. It did appear clear that existing systems have not yet been carefully field tested, and that they do not contain all the aids to translation that have been suggested.

STANLEY R. PETRICK is President, 1976, of ACL. For biography, see AJCL Microfiche 37:3.