<u>REPORT OF SESSIONS 5 and 6</u>

<u>AUTOMATIC   TRANSLATION</u>

Jacques GOETSCHALCKX
Head of the Terminology Bureau
Commission of the European Communities - Luxembourg

Machine translation as the answer to the growing need for translations has proved highly controversial in recent years, particularly since the publication of the ALPAC report.

From the reports which we have heard, it is clear that many people have not allowed themselves to be discouraged. Mr. BRUDERER's extensive review provided ample proof of this.

While some have persevered along the same lines as the University of Georgetown, others have made great efforts in other directions in the hope of avoiding certain snags.

One approach is controlled syntax translation, which in theory offers interesting possibilities, although the preparatory work is rather arduous. It requires above all a high measure of integration, as a number of conditions must be met, even by the authors of the documentary texts. Clearly, this appreciably simplifies the syntactic analysis.

One system based on more or less the same considerations, imposes limitations at the lexical rather than the syntactic level. This approach eliminates problems of homography and ambiguity.

The dispute as to what constitutes a first or second generation system should perhaps not be taken too seriously at this stage, for both still have a long way to go, and others are already

claiming third generation systems.

Among users, the effort is focused on the precisest possible evaluation of the results obtained. It emerges quite clearly that the criterion remains unchanged: what is wanted is a good quality translation, but agreement ends at that point, because when asked to define a good translation some find it quite impossible to do so, while others have diametrically opposed views. Some go so far as to insist on results 'at least as good as the average translator could provide', while others are satisfied with a rough translation.

If we want to discuss the merits of machine translation, we must first agree on the criteria of evaluation. But here again these depend on the differing aims and the particular working conditions within the organizations which use a computer for translation work. A commercial enterprise will prefer cost effectiveness, while public administrations will - if I may run the risk of irony from the cynics among us - prefer speed. Those who produce operators' handbooks, maintenance guides, or safety standards will insist on quality and accuracy. The documentalist, and above all the user of documentation systems, will place the highest value on intelligibility,which is not always a function of the stylistic quality of machine trans-lations.

The question of costs was raised several times when methods of evaluation were being discussed, particularly by the private sector, so much concerned with cost effectiveness. There again we have to take into account the aims and the particular situation of each user. However, some applications are en-couraging from the economic point of view, even though there are doubts as to their quality.

The generation question may be considered in terms of hardware development, changes of approach, and translation requirements. Linguistic criteria are expressed in terms of morphology.

syntax, semantics, and pragmatics. Some people simply distin-
guish between systems which actually translate and those which
do not, or which do not claim to do so.

Very often one finds that the sophistication of the software
is inversely proportional to the volume of data required at
the input stage.

As to the question of whether systems are 'of our generation'
in the sense of bringing results in our time, this must clearly
depend on the extent to which it is felt advisable to allocate
the necessary funds.

On the question of metalanguages and machine languages there
seems to be no agreement. On the one hand there are arguments
of flexibility and transportability, and on the other hand the
capacity of the memory.

The ideas put forward regarding pre-editing, post-editing, and
co-editing should be considered in the context of their place
in the overall chain of translation, recourse to terminologists,
and the quality of machine translation.

The views expressed on pre-editing, and the various approaches
which may be used in conjunction with it no doubt explain the
lack of agreement on this point. We have even heard that a
technical text is tantamount to a pre-edited document because
the author will usually aim for clarity and simplicity. One
basic principle would seem to be that human intervention is to
be avoided whenever possible because it adversely affects cost
and time, which are the main advantages of machine translation.
Above all, do not refer back to the terminologist, who already
has a major part to play.

The amount of pre-editing should of course not be out of pro-
portion to the savings on syntactic analysis or subsequent
treatment of the text.

Post-editing is all the work involved in revising the translation and producing the final text. Co-editing is very often part of an overall word processing system.

As a conclusion from the foregoing, I would say that costs and cost effectiveness are in general closely linked with the degree of integration of the machine translation chain. We have to replan the administrative work which comes before translation, not so much by imposing lexical limitations as by using recently developed methods of processing texts.

It will be necessary to produce texts on media permitting direct digitalization, with no intermediate stage. Likewise, the editing of acceptable material could be part of the output procedure.

It is also clear that apart from the basic problems, such as homography and the removal of ambiguity, which are really difficult, machine translation often has to overcome difficulties of apparently secondary importance such as the problems of capital letters, abbreviations, and the full stop at the end of a sentence. Even more irritating are the problems connected with input operations, particularly at the organizational level, We must hope that these too can be solved by an integrated approach.

It emerges very clearly from the papers on the different translation systems that dictionary or vocabulary preparation is one of the crucial points, no doubt mainly because this is the only link in the chain which cannot be automated.

This means that an effective link-up with terminology data banks is of fundamental importance. It also follows that brilliant advances at the level of linguistics, grammar, or data processing may remain fruitless unless there is a parallel and equally well funded effort on the terminology side.

From the opinions expressed by qualified people it seems that a

machine translation system which reaches a level of accuracy
and quality such that a reviser can check the work without
cursing its inventor, and which is also reasonable from the
standpoint of cost effectiveness, is an acceptable system.
'Reasonable' here means that cost effectiveness must be viewed
in the context of future developments.

Writing and translation are not recent acquisitions, and all
the energy expended on their acquisition was written off long
ago. Evidently this is not yet the case with machine trans-
lation.

The future prospects of machine translation cannot be dissoc-
iated from the prospects of translation as such.

At the theoretical level, it is not unreasonable to consider
the usefulness of a universal intermediate language, whether
it be a numerical language, esperanto, or something else. When
the developing countries take their place, there will be needs
out of all proportion to the intellectual and linguistic capital
needed to cope with the work by human translation alone.

In the light of the reports which we have heard, it seems
desirable - within the limits set by available funds - in
machine translation as in other advanced techniques, to continue
equally with fundamental research and with experimental or
pilot systems, so that the two may enrich each other mutually.

It may well be that multidimensional trees are a sign of
springtime for further generations of automatic translation
systems.