Session 3:

Chairman's introduction:
Naming machine translation

Professor J.C. Sager

Department of European Studies and Modern Languages, University of Manchester Institute of Science and Technology, UK

Looking at the increasing diversity of the translation market, one feels a need for some clarity in product naming and description so that developers, producers, users and critics can continue to understand each other. The introduction of the concept machine translation or MT (as a noun describing a process) has forced us to widen the meaning of translation, and to coin the neologism human translation which only a few years ago was considered a tautology. One could also say that 'translation' has lost its distinctive semantic feature 'human activity', which is, of course, only true if one conceives the various activities a machine performs on a text as independent of human design and planning. When computers became regularly used for translation, this dichotomy 'human-machine' was broken down as it was realised that human intervention in the machine translation process was desirable or even necessary. A new concept was introduced with two expression forms, machine-aided and machine-assisted, which are synonymous. For some this was a reassertion of the essentially human nature of the art, craft or skill of translation, and therefore a more apt name for machine translation in general; for others it represents a sub-category of machine translation. We could possibly satisfy both viewpoints if we were to draw a distinction between machine-aided human translation and human-assisted machine translation, of which the former would be a conjunction of the concepts 'human translation' and 'machine translation' and the latter a particular type of 'machine translation'. It is, however, unlikely that such a distinction will in practice be made.

If these were all the names or concepts we have to sort out and clarify, our task would be easy. There are,

however, a number of other candidates for the status of technical terms, and there are a number of concepts of processes and products yet to be accommodated in our terminological field. Some of these tend to be specific to one or other of the specialist groups involved.

System developers speak of <u>generations</u> of machine translation systems. Such labelling is of academic and historical interest, as it refers to design criteria such as the presence or absence of certain modules, the use of certain techniques and the role they play in the system. A certain system is said to be of the first generation because it is designed in a certain way rather than because it produces translation of a lower quality than a system of the second or third generation. These names cannot, however, be applied rigidly as systems evolve and add features of other generations. Such labelling can, therefore, be controversial and it is advisable to leave it to the historians of MT.

Another categorisation based on linguistic strategies and processes is the distinction between <u>interlingua-based</u> and <u>transfer-based MT</u>. The difference is of considerable theoretical interest and is likely to remain a major design criterion. An interlingua-based system should, in principle, be the foundation for <u>multilingual MT</u>, as it could be completely neutral. Since, however, we have not yet achieved a satisfactory design of an interlingua for the representation of real texts, interlingua-based MT is either concerned with heavily restricted syntax, as in TITUS, or an ideal to be approached rather than achieved. In reality, therefore, a characterisation by linguistic strategy is of little interest to producers and users of translations.

On a practical level systems are designed to be used <u>online</u> or in <u>batch</u> mode. Since all early systems worked in batch mode, the name 'online' came into use when this technique was first introduced. On the surface, this is only a descriptive distinction of the mode of operation which need not affect the linguistic nature of the system, though online processing is essential for certain machine-assisted human processes and can then be subsumed. As the distinction is also dependent on the hardware used, the characterisation of being online is not likely to lead to the creation of a new set of sub-categories.

From a producer's point of view, we do, however, have to distinguish MT according to the type and point of human intervention in human-assisted machine translation: before the process of machine translation, after processing or in stages concurrent with processing. We thus speak of <u>pre-edited MT</u>, <u>interactive MT</u> or <u>post-edited MT</u>. When translators post-edit machine translation output we also speak of <u>machine pre-translation</u>, a term used by Veronica Lawson. Interactive MT has yet to be defined more closely

since it may involve providing the machine with missing dictionary items (which in some systems is done independently) or with syntactic disambiguation and may include both pre- and post-editing on a word processor. Equally pre-editing may involve as little as supplying close punctuation and identification of proper names and other items which require special treatment, or rewriting in a restricted syntax and vocabulary which the machine can analyse. As systems evolve and integrate word processing facilities the picture is changing. We may, therefore, have to redefine the concepts just enumerated in the light of new developments.

If so far we have seen a fair diversification of the terminology of MT to denote the process, we have yet to evolve a proper terminology for the products of MT. It is common to speak of <u>MT output</u> as a separate translation type which may or may not undergo further changes - editing and revision - before it reaches the user. The expression <u>raw output</u> is known and used to denote unedited translations, but then no clear distinction is drawn between such output which is directly usable and output which has to undergo editing before it can be used. While the end-product of any translation process is intended for the same purpose, namely that of communication, it is nevertheless desirable to differentiate among types of MT output in the same way as a sophisticated market differentiates among particular object-ives and types of human translation.

It is at this point where attempts at collecting the terminology of translation have to stop and speculation on the role of MT begins. If we use the computer for producing translations which are fully equivalent to and indistinguishable from human translation, the end-product must fit such human specifications as 'for information' or 'for publication'. If, however, we also produce machine output which does not pretend to be fully comparable with human translation, we need names to characterise such products as the unedited output used by the US Air Force, the rapidly post-edited output now being produced at the Commission of the European Communities or the multilingual abstracts produced by the Institute Textile de France. Appropriate names have yet to emerge as they should in the interest of honesty and accuracy of product description and naming.

AUTHOR

Professor J.C. Sager, Department of European Studies &
Modern Languages, University of Manchester Institute of
Science and Technology, PO Box 88, Manchester
M60 1QD, UK.