CONCLUSIONS AND RECOMMENDATIONS

J. C. SAGER

DEPARTMENT OF EUROPEAN STUDIES & MODERN LANGUAGES -
The University of Manchester
Institute of Science and Technology

The organisers of this Congress, the Directorate general "Scientific and technical information and information management" of the Commission of the European Communities, have with their choice of title set a specific orientation and purpose for our deliberations and it is now our duty to assess the achievements of four days hard work, to evaluate the proposals made, and to draw up conclusions. The timing of this Congress is problem oriented. The creation of the European Information Network (EURONET) makes it desirable that the implementation of EURONET should benefit from expert advice and that its future users should not only see what will be offered to them, but that they should have an opportunity of influencing the shape of EURONET.

"Overcoming the Language Barrier" is a sine qua non for "Information Systems and Networks" if we conceive their function as extending the purely parochial dimension of small user groups in very narrow subject fields. The language barrier exists, however, not only on national or geographical levels. It has to be overcome between subjects, within subjects between scientists, engineers, and technicians, between indexers, between compilers of thesauri, between data banks, between translation systems, between all these groups of well-intentioned collaborators in a common cause and not least of all between the terminologists and grammarians whose efforts are dedicated to breaking down these barriers within languages and between languages.

Our theme was the language barrier. There are others, financial, political, cultural ones, but they can all be subsumed under our main concern. A satisfactory surmounting of the language barrier will take account of these considerations, and will be successful only to the extent that the solutions proposed are economically viable, politically acceptable and culturally respectful of the identity of the different language communities reflecting the historical, cultural, social and intellectual divergencies and stratifications of our societies.

The use of the verbal noun "overcoming" entails the will and the determination to succeed in this task and this was reflected in most of the papers presented. The formulation also implies an on-going process. Nevertheless we can, at the conclusion of this Congress, re-formulate our objectives more positively as "The Creation of Multilingual Information Systems and Networks".

This means that we accept the multilingual nature of Europe as a positive fact, and that we reject the alternative of purely mono-lingual systems and networks for supranational communication which put the burden of knowing the particular language chosen on the individual. It also means, and this again seems to emerge from our discussions, that each linguistic community is prepared to share its knowledge and accumulated data with other communities on terms of equality.

We set ourselves two specific objectives:
1)   to give those responsible for the Action Plan for the improvement of information transfer between European languages a good view of existing and developing systems and thus to enable them to take the most appropriate decisions.
2)   to make future users of EURONET acquainted with methods and tools that will soon be available to them. We are thus instrumental in shaping the future development of these multi-lingual systems.

The plan of the Congress reflected these two objectives: It presents a statement of the existing situation and achievements to date, and a projection of likely developments, i.e. a survey of research in progress on theoretical, applied and applicable topics. The Congress addressed itself to users and developers of information systems and networks who may, in fact, be the same people. But whether we are users or developers we are first of all users of language, that is, the languages we have spoken here, with all their imperfections and ambiguities, deliberate or accidental, which cause so much headache to our information scientists, computational linguists and logicians, and it is important that our first session should have been devoted to the use and learning of natural languages. While this is largely an individual preoccupation it affects publishing, teaching and data collecting policies, and some of our conclusions and recommendations consequently reflect the fundamental importance of our attitude to the learning and the use of foreign languages.

The second session looked at the nature of the language problem mainly as it affects translation. Reflecting our present state of knowledge the papers were largely concerned with terminology, its collection, processing and use, and hardly at all with the syntactic peculiarities of special languages which still confound our efforts in automatic translation.

Session three was difficult to place in the programme because translation has developed so much in recent years and is concerned with such a diversity of texts. It is on the one hand, despite much research, still virgin territory for linguistic and psychological theories, and the other, a vigorous process aided by machines. This session was more than a bridge between human and automatic translation. It was a reminder that human translation is still the most important means of interlingual communication and will remain so. It can be improved and assisted as we have seen but with our present

state of knowledge it is unlikely to be replaced for quite
some time if ever, except in strictly limited areas.

The session on thesauri was equally wide ranging. It is
linked to session two in that it shows the incorporation of
terminology into information systems via thesauri,and it
anticipates sessions five and six since thesauri can be
essential parts of translation systems. Two methods are
competing: the translation of monolingual and the creation
of multilingual thesauri. The old controversy between free-
text search and controlled vocabulary search seems to tilt
in favour of thesauri when we consider search strategies for
multilingual systems. The work already in progress in the
Commission shows the difficulties encountered.

The last two sessions presented the problems and achieve-
ments of automatic translation to date. The scope of these
systems varies considerably as does their success in being
implemented on a larger scale. The Commission is committed to
sponsor application oriented research in this area and a major
concern will be the compatibility of the various systems
connected in EURONET. There is a happy coincidence in the
fact that the establishment of EURONET more or less coincides
with the wider use of automatic translation systems. The
benefits of automatic translation have been demonstrated in
some areas. Large bibliographic data banks or processors of
information will probably take pragmatic decisions on the basis
of the functions of the data base and existing methods of
information handling. The creation of the essential diction-
aries stresses the need for a common data exchange format,
which would benefit EURONET. These two sessions also showed
clearly the limitations of automatic translation and that they
will not obviate the need for human translation nor indeed the
need to learn and use foreign languages in everyday contacts.

The Congress progressed from general to specific topics,
and from broader to technically more complex questions. The

earlier sessions dealt with areas we are all concerned with as
speakers, writers, listeners and readers of natural languages.
By receiving translations of all the papers and by the provision
of an interpreting service we may have gained the impression
that there is no barrier or that it has been effectively over-
come. But this is, of course, only the surface; behind it lie
months of agonising translation, editing and all manner of
elaborate preparation which demonstrate human ingenuity in
dealing with this man-made barrier so cherished by each and
everyone of us. It must, of course also be said that we have
imposed upon ourselves restrictions in that only three out of
seven community languages were used and that the principality
where this Congress took place has modestly waived all claims
to its most widely used language becoming another official
community language.

Conclusions

1. The first and most important is that the congress has
succeeded in assembling under one roof a great number of the
most distinguished users, developers and researchers of multi-
lingual communication. This fact shows a desire to co-operate
and this co-operation has been greatly aided by this meeting.
The Infoterm Symposium (1975) in Vienna called for increased
co-operation among terminologists, linguists and information
scientists. The papers read at this congress show the extent
to which this has happened. I note, however, the absence of
researchers in artificial intelligence who, I believe, will
have significant contributions to make in future.

2. The second general conclusion I should like to draw, and
I can do this in all modesty, is that the collection of papers
presented in the two volumes of the proceedings is the most
comprehensive survey of the state of the art ever to have been
assembled. It is a report of a considerable amount of research
application over the last few years by private industry, uni-
versities, governments, research organisations, supranational
and international organisations. It covers work at all levels

of intralingual communication, translation, data-banks,thesauri
and information systems and has benefited from interdisciplinary
collaboration. In its scope I should like to compare this
volume to a book well known to most of us here: Professor
Wüster's <u>International Standardization of Language in Tech-
nology,</u> which since 1936 has not been equalled. I am sure
that he would have approved of this comparison.

3.   The third conclusion is a more pessimistic one and almost
acquires the character of a dilemma. Because of increased
interlingual contacts there is a greater volume of translation
of texts and summaries, abstracts, indices of documents from
one language into another; some of this work is even duplicated,
without, however, being equivalent. Human inventiveness is not
only manifested in the development of new theories, i.e. con-
ceptual systems, but also in the designation of conceptual
units. Not every interlingual information processor has the
self-restraint, or even the time to look for existing forms
or the ability to judge whether he has found the right one.
The large translation departments are particularly aware of
this problem and their terminological data banks can be con-
sidered as a safeguard against a wild proliferation of terms
and definitions. Despite the most careful control, in every
dictionary there lurk a fair number of designations of doubtful
authenticity which hinder communication. It is to be hoped
that the greater contacts between data banks of all kinds will
show up these differences and that ways can be found of elimi-
nating such linguistic mishaps. The danger of creating pseudo-
languages is considerable and must be checked.

4.   The establishment of EURONET is not a rash and unpremedi-
tated action. The organisers of EURONET and the Action Plan
seek and need the widest possible support. This has been
offered by the participants of this Congress many of whom
represent existing data banks or organisations like Infoterm,
UNISIST or IEC concerned with co-ordination on a world-wide
scale. EURONET can only be as good as the data banks it

connects or the access tools it can provide for these banks. The more heterogeneous the data the more complicated the access tools. We have heard of its likely limitations and the intellectual effort yet required. Smaller meetings of operators of various data banks are needed. We have also heard that EURODICAUTOM has access to the Terminology Data Bank of the University of Montreal, and other organisations. We are looking forward to reading or hearing about the practical experience gained in the incorporation of one system into another which should also allow evaluations of the quality of either corpus.

Of greater interest to EURONET will be the direct access to other corpora, be they terminological or documentary. Very little could be said about this here but it will be the hard test for the future.

5. It was frequently pointed out that thesauri and terminologies not only serve different functions, but are also compiled or constructed along different principles. At the same time common elements were referred to, i.e. the fact that thesauri can be used as dictionaries and that both exercise control over vocabulary. Traditionally these two listing devices were developed quite separately by people with different training who ignored each other's existence. This Congress has shown how much this position has altered. We look forward to yet closer collaboration since the finding of the maximum of common ground is one of the cornerstones of a successful collaboration between translation and documentation and between terminological data banks and bibliographic information systems.

6. Data banks of words or terms - this distinction is not always made clearly - are fundamental tools of multilingual communication. They should be designed with the maximum flexibility so that they can grow beyond their traditional areas of application. They should allow the extraction of

alphabetic, subject and systematic dictionaries in print, microfiches, or for conversational mode and they should also provide the basic material for standardization processes and thesaurus compilation. In this way duplication of effort is avoided, as well as divergency of designations. The concentration on a common data base should ensure close collaboration between linguists and information scientists, interaction among modes of communication, and most importantly consideration of user requirements. The user after all is the same person. He specifies different requirements. We have accepted that he is best served by information in his own language. We should not, then, impose on him the need to learn a diversity of documentary languages in which Humpty-Dumpty arbitrarily decides the meaning of a word.

7.   There are different translations for different purposes, but there are no perfect translations, as there are no absolute synonyms. All translations whether partly human, or machine-aided to varying extent, are expedients and should be judged by, their appropriateness to the situation. We have an inalienable right to demand respect for our language even though small user groups may adopt the temporary expedient of a reduced language in order to gain quick access to new information.

8.   Work on documentary languages is well advanced in mono-lingual systems, but like any intellectual construct capable of infinite variation and improvement. The reluctance of experts to declare a system pragmatically satisfactory and to compromise on minor matters of notation creates a conflict with users who need workable systems and are prepared to accept a certain degree of imperfection, noise, or inaccuracy if it means direct access via one documentary language only. With the increase of interdisciplinary studies users no longer restrict themselves to one thesaurus and in a multilingual environment they want to consult thesauri on one subject in several languages, at least until more good multilingual ones are available. The compilation of thesauri implies standard-

ization. Standards are powerful and even dangerous instruments.
Specialists must exercise self-restraint in two directions.
They should agree on matters of presentation, structure and
notation as far as possible, while at the same time acknow-
ledging their limitations in matters of translation and in
the terminological presentation of different conceptual
organisations in different languages. It may be too early to
roll on the heavy machinery of a tightened ISO standard, but
this should not prevent indexers in countries connected via
Euronet from agreeing on some of the alternatives offered in
ISO 2788 (Guidelines for monolingual thesauri).

9.   There is no specific conclusion to be drawn on Automatic
Translation - we should perhaps speak more modestly of machine-
aided translation. We are in the middle of an exciting phase
of developments. One thing we can surely agree upon; there
are many systems for diverse purposes with varying degrees of
sophistication which should be judged only in terms of their
cost-effectiveness and the user satisfaction. The Commission
takes an interest in those developments as we have heard from
one evaluation. Others are to follow. These evaluations may
lead to experimental adoption of several systems for several
purposes which will be determined by its existing data and by
the several functions these translation systems are to have
within its information system. We are looking forward to being
widely informed of the Commission's action in this respect
since such a step is likely to have far-reaching repercussions.

10. Construct languages were referred to in a number of papers.
There seems to be little support for their introduction and
use as a primary tool of communication, but their role as an
intermediate language in a translation system appears to offer
some interesting possibilities.

11. There is growing interest in special language teaching
and in teaching languages for special purposes. The Council
of Europe is active in this direction and may produce prag-

matically oriented syllabuses. Registers of on-going research are being compiled and maintained and should soon be available on a European basis. This Congress stressed the importance of reading comprehension and the need for learning-vocabularies differentiated by subject and, more important, by level of complexity. It has also been suggested that greater "language guidance teaching" should precede the teaching of foreign languages. The conclusion to be drawn from these papers is that understanding of the different communicative functions of the mother tongue and its special subject languages as well as understanding the structure of the communication process greatly assist in the purposeful learning of foreign languages and that pragmatic text-analysis should be taught in the mother tongue as well as in the foreign language.

12. Multilingual communication requires multi-disciplinary professional education for the future mediators between documents and users.

    More specifically the formation of future translators and interpreters, terminologists and information scientists must be modified to enable them to cope with the new tasks before them and to use the new tools provided for this purpose.

    Translators in particular must understand the principles of lexicography, indexing and abstracting.

    Terminologists must be capable of using automatic data-processing devices and of working side by side with information scientists.

    Information scientists in their turn must be familiar with foreign languages, linguistic semantics and pragmatics.

13. Finally the Congress has affirmed its interest in and its commitment to maintaining the multilingual nature of technical and scientific information and even strengthening it. This

decision represents a realistic acceptance of the nature of languages as instruments and vehicles of thought and that we are deriving positive benefits from the different ways of conceptual organisations of our languages.

Specialists may find these conclusions platitudinous, but I believe this Congress can and should only take a broad overall view, and leave detailed discussions to smaller working groups.

Note: The Rapporteur received the following communication signed by 19 Congress participants:

"The subscribers heard at the "3rd.European congress about documentation systems and networks" several, partially contradictory opinions about the adequateness of planned languages, especially of the Internacia Linguo (Esperanto).

In order to receive a practical basis of judgement (but also at least as a symbolic politeness to participants whose language is not accepted as congress language) at the next congress the Internacia Linguo should be accepted as 4th congress language."

The Rapporteur presented this proposal to the meeting which did not express its support.

Recommendations

These conclusions lead to a number of specific recommen-dations which should be grouped into two categories: one set addressed to the Action Plan for the improvement of information transfer between European languages, the other specifically in relation to the creation of EURONET.

I. <u>Recommendations for the Action Plan</u>

1) There seems to be a need for a European research register to link and if possible to coordinate research in all types of Translation, Terminology, Standardisation, Information Science and Artificial Intelligence in so far as they are concerned with multi-lingual problems.

2) Interdisciplinary projects dealing with practice oriented problems traditionally have difficulties in attracting funds. These should be sponsored by the Action Plan.

3) The recommendations of the Seminar on "Scientific and Technical Publishing in a Multilingual Society" should be implemented: they are

- a) A greater number of good review and abstract journals in languages other than English.
- b) A greater number of translations of good technological documents into European languages.
- c) the setting up of a fund to assist with the translation and publication of technological texts.
- d) Collaboration among publishers and editors to improve the standards of scientific writing which may partly be done by means of guidelines elaborated on a European scale.

4) The efforts of the various centres concerned with the collection of information about existing translations should be strengthened and coordinated into an integrated computer data-base.

5) The learning of second and third languages should be encouraged with particular emphasis on reading comprehension. Research in this direction should be sponsored.

6) Pragmatic, problem oriented developments arising from user experience should be encouraged.

II . The remaining recommendations have implications for EURONET, but are not necessarily dependent on it.

7)   Hardware harmonisation should be aimed at in unified character sets, which are also respectful of graphemic patterns of individual languages.

8)   There is a need for wider distribution of diversified subject glossaries, both monolingual and multilingual ones in order to improve the use of terminology. Central and feed- back should improve the individual data banks. The aim should be harmonisation rather than standardisation which should be seen as a longer term goal. In any case not every term can nor should be standardised.

9)   The creation of preferred terms in thesauri is a process akin to standardisation. Where they do not enter into conflict with the conceptual organisation of a language, multilingual thesauri should have compatible formats and presentation.

10) Multilingual thesauri should be compiled simultaneously in cooperative ventures rather than translated.

11) The parallel structure of terminological data banks, dictionaries for automatic translation and thesauri should be more widely recognised despite their teleological differences. A greater coincidence between these three types of lists can be a major factor in the easier accessibility of a foreign language.