

COMMISSION OF THE EUROPEAN COMMUNITIES

COM(80) 314 final

Brussels, 10th June 1980

SECOND PLAN OF ACTION FOR THE IMPROVEMENT OF THE TRANSFER OF INFORMATION BETWEEN EUROPEAN LANGUAGES

(Communication from the Commission)

THE CREATION OF A EUROPEAN ADVANCED MACHINE TRANSLATION SYSTEM (EUROTRA)

(Communication from the Commission to the Council)

PROPOSAL FOR A COUNCIL DECISION ON THE ADOPTION OF A EUROPEAN ECONOMIC COMMUNITY RESEARCH AND DEVELOPMENT PROGRAMME FOR A MACHINE TRANSLATION SYSTEM OF ADVANCED DESIGN

(presented by the Commission to the Council)

COM(80) 314 final

Second Plan of Action
for the improvement of the transfer of information
between European languages

1. Purpose

The second Plan of Action is primarily a continuation of the actions undertaken in implementation of the first Plan. However, the environment has changed as a result of the emergence of new needs and new techniques, and a certain number of new actions will be launched to take account of these developments.

1.1. Results of the first Plan of Action

The first three-year Plan of Action for the improvement of the transfer of information between European languages, approved by the Commission on 23 December 1976, produced the following concrete results :

- (a) The implementation of a system of automatic pre-translation of texts in natural language and its progressive improvement; its development is being increasingly oriented towards industrial and scientific applications.
- (b) The Commission's terminology bank has been expanded, particularly towards the less well represented languages, and extended to other areas of interest to the European institutions.
- (c) Three quadrilingual sectoral thesauri have been prepared and brought into use in the Member States, and a macrothesaurus for the agricultural field is currently being compiled.
- (d) Feasibility studies have resulted in the production of specifications for text processing equipment which will soon be brought into operation in support of computer-assisted translation.
- (e) The organisation in May 1977 of the Third European Congress on Information Systems and Networks, together with a large number of symposia and seminars has alerted public opinion to the various tools available for overcoming language barrier.
- (f) As a result of the Plan of Action, nine European universities have cooperated in planning a major R + D project designed to produce an advanced European machine translation system on which they will work together during the next five years *).

*) See the EUROTRA project, which is the subject of a separate proposal.

1.2. Preparation of the second Plan of Action

The following factors were taken into account in the preparation of the work programme :

- (a) While the Commission was setting up its initial projects in the field of multilingualism, countries throughout the world, particularly plurilingual countries (e.g. Canada, Switzerland) and those whose economies depend on exports to countries using other languages, were becoming aware of the problems caused by the multiplicity of languages.
- (b) Reflecting the downtrend in the cost of computer services, several computer-assisted translation systems have now become operational in specific fields, and there has also been an increase in the number of terminology banks, computerized dictionaries and multilingual thesauri.
- (c) The Community policy for networks has resulted in the start-up of the Euronet-Diane network, which has made a large number of monolingual documentary data bases available to the innumerable potential clients who now impatiently await the translation of these resources into their languages.
- (d) The Community policy for the telematics sector has brought the departments of the Commission into contact with a wide range of new resources and techniques capable of making a positive contribution to a large number of its projects.
- (e) In view of the need to develop multilingual tools for all the Community languages, the impending accession of new Member States to the European Communities will also increase the workload of the language services, and the list of problems requiring solution will include the transliteration of texts written in Greek.

2. Programme of work

As in the case of the first Plan, the various actions will be preceded by studies and surveys and concluded with evaluations, pilot schemes and trial applications.

The programme of work will be backed by continuous detailed monitoring of the research in progress and active support for the teams developing new systems in the member countries, with a view to ensuring that the work is appropriately oriented and tailored to the needs of the Community.

Much of the work will be entrusted to specialized firms and institutions in the Member States. As in the case of the first Plan of Action, the Community departments will seek the advice of the "Committee of Experts for the Transfer of Information between Community Languages (CETIL)", composed of specialists from the Member States and the Commission, selected for their technical know-how.

.../...

As far as the application of multilingual tools to scientific and technical information is concerned, however, the Commission will consult the Committee for Scientific and Technical Information and Documentation (CIDST) and its sub-group "Multilingual Aspects".

Action is planned on the following fronts :

- (a) general surveys on multilingualism and studies relating to new problems, including follow-up work ;
- (b) creation of multilingual terminological tools, including thesauri for information management, terminology banks and other computerized dictionaries and dictionaries for machine translation;
- (c) design and development of machine translation software to meet the Community's requirements;
- (d) creation of an infrastructure of data processing equipment for multilingual activities;
- (e) establishment of methods and standards for multilingual communication; studies on their implications, particularly from the organizational standpoint;
- (f) fostering of the practical applications of existing multilingual tools.

2.1. Studies and surveys

Specific surveys relating to the teaching and use of languages in the Member States will be supplemented by a general survey of multilingual communication in Europe and between Europe and the rest of the world.

Other surveys will focus on language comprehension and the mechanisms underlying the transfer of information between people and between man and machines.

The study of the translation market launched in 1979 will be supplemented by specific studies of trends in demand, in the light of the appearance of machine translation on the market and various other factors.

The effects of Community enlargement on multilingual communication will be carefully analysed particularly from the data processing standpoint.

There will also be an in-depth examination of the economic and technical feasibility of extending the Community's multilingual tools to non-Community languages.

.../...

2.2. Multilingual terminological tools

2.2.1. Scientific and technical information management

Multilingual thesauri are the basis of the most highly reputed international retrieval systems and particularly of those now available to users via the Euronet - Diane network.

Having helped to develop highly advanced methods for the construction of thesauri, the Commission is ready to assist all who wish to make their documentary resources available to all the potential users in the Community by means of multilingual thesauri.

The Commission will assist in the setting up and management of a system for the updating and utilization of multilingual thesauri and will contribute towards the efforts to bring about standardization in this field.

2.2.2. Translation aids

The Commission has established for its translators a terminology bank whose reliability is steadily being improved by a strict validity check. It will continue to extend the bank's contents by systematically liaising with similar projects in hand elsewhere in the Member States and in various international organizations, thus avoiding the unnecessary duplication of current efforts.

In particular, the Commission will, in the very near future, create multilingual terminology glossaries for the fields of telematics and telecommunications and any other areas of advanced technology of interest to the Community.

It will also set up a special data bank for abbreviations; at a time when a growing number of new organizations are known to the institutions mainly or solely by their initials, such a data bank is sorely needed.

2.2.3. Dictionaries for machine translation

The Commission will continue to create dictionaries for fields of interest to the Community. It will reduce their cost by using terminology banks.

Initially, these dictionaries will be tested and used in the SYSTRAN project; their linguistic coding will be gradually tailored to the requirements of the more advanced systems (see 3.3.).

2.2.4. Computerized dictionaries

Publishers of dictionaries, whether general or specialized, are now increasingly using computerized maintenance and up-dating methods. The Commission will study the scope for using these resources in the Community terminology bank and in machine translation dictionaries.

2.3. Machine translation programmes

2.3.1. Pre-translation of text (Systran)

In the present Plan of Action, the Commission intends to transfer the responsibility for running the Systran system to whose development and improvement its departments have made a substantial contribution, to external organizations. It will facilitate the application of the system, mainly in the industrial and scientific fields and in connection with Eurohet - Diane.

Efforts to improve and extend the system will henceforth be guided principally by these applications.

2.3.2. Translation based on controlled syntax

The emergence of networks of the kind typified by videotex has been reflected in the need for systems of translations without post-editing. The Commission will undertake a study of the existing systems and the elaboration of a set of rules for writing in controlled syntax with a view to facilitating such applications.

2.3.3. European translation system (EUROTRA)

The EUROTRA project, the technical details of which were finalized in 1978 and 1979, thanks to the enthusiastic cooperation of the universities in the Member States, will be launched as a separate programme immediately following Council approval of the project.

In the present Plan of Action, the Commission will be responsible for the preparatory work.

2.4. Technical infrastructure

2.4.1. Information management

The Commission will set up the computer equipment needed to establish the links between the multilingual thesauri and data bases which they are to control.

.../...

2.4.2. Aid to translation

Appropriate available terminology will be placed at the disposal of the linguists (translators, interpreters and terminologists) of the Community institutions by the creation of a complete infrastructure using the most recent developments in the field of telematics.

2.3.4. Text processing

The Commission will acquire a pilot interactive text processing installation for the input, the preparation of texts to be translated and the post-editing of translated texts; it will use these facilities to conduct technical and ergonomic studies.

2.4.4. Transliteration

The Commission will study the introduction of equipment permitting transliteration from and into languages which use characters differing from the European alphabet.

2.5. Methodology and standardization

2.5.1. Evaluation methods

The Commission will carry out studies and tests to devise evaluation methods and criteria which are objective and free from the personal bias of the assessors.

Evaluation criteria will be developed for application to information retrieval in terminology banks.

2.5.2. European alphabet

The processing and transmission of texts in all the European languages will require a larger variety of characters than that available with current equipment.

The Commission will participate in the work intended to lead to the adoption, by standards organizations, postal administrations and equipment manufacturers, of a standardized European character set and of a new transmission code permitting the input, transmission and printout of texts in all the European languages.

2.6. Promotional activities

A broad Plan of Action of this type requires the participation of everyone directly or indirectly interested in multilingualism, including teachers, publishers, documentalists, data processing experts, linguists, translators, interpreters and terminologists.

.../...

This means identifying, contacting, informing and interesting all such persons. This will be done not only by organizing courses, seminars and demonstrations, but also by publishing a periodical containing critical reviews, abstracts of recent books and reports and short articles on work in progress.

3. Budget

The expenditure involved in implementing the Plan of Action is of the order of 3 750 000 EUA, broken down as follows :

	1980	1981	1982
Studies and surveys	70 000	150 000	180 000
Multilingual thesauri	30 000	40 000	60 000
Terminology banks	150 000	150 000	150 000
Dictionaries for machine translation	250 000	280 000	300 000
Computerized dictionaries	60 000	120 000	100 000
Making Systran operational	100 000	80 000	60 000
Systran pilot operations	200 000	320 000	250 000
Preparation of Eurotra	80 000	p.m.	p.m.
Technical infrastructure	80 000	80 000	20 000
Methodology and standards	30 000	30 000	30 000
Promotional activities	100 000	100 000	100 000
	<hr/>	<hr/>	<hr/>
	1 150 000*)	1 350 000	1 250 000

*) All or part of this programme will be carried out, depending on the budgetary allocations for 1980.

8

COMMUNICATION FROM THE COMMISSION
TO THE COUNCIL

Subject : The creation of a European advanced machine translation system
(EUROTRA)

9 1.

1. Reasons and aims

The European Community is a multilingual group of people who are striving for ever closer unity. Many different languages are spoken in Europe, and as a reflection of Europe's cultural diversity, this variety represents something of great value.

Unfortunately, however, it is also a major hindrance to communication between persons and closer ties between nations. By hampering trade and the transfer of technology, it is detrimental to the Community's competitiveness in the world. Even within our own institutions, it is a constant source of misunderstanding and friction.

We must act quickly to break down these barriers, by providing the Community with a tool for efficient multilingual communication. This goal is now within our reach.

Much has already been achieved over the last few years. Multilingual thesauri for documentation centres have been set up in many places. Terminological data banks have been produced to aid translators. Simple translation systems have been introduced in a number of industries. The Commission has participated in this work, in the context of its first Multilingual Action Plan*.

Using the results so far achieved we must now set about developing a machine translation system which will provide high-quality translations simultaneously in all the Community languages rapidly and at a low cost.

This system, which is known as EUROTRA,

- is based on the latest findings, in particular in linguistics,
- is based on modern word processing techniques,
- is adaptable to other languages,
- is compatible with all information processing equipment, in particular that of European origin, because of the portability of its software, the use of high level programming languages and compliance with international standards in this field, and finally
- is capable of incorporating future developments in the field of linguistics, remote access information processing, and even artificial intelligence.

* Communication from the Commission to the Council of 31 December 1976.

2. History

In the United States consideration was given to machine translation as soon as the first computers started operating. But in that monolingual country there were too few clients for such a multilingual tool, and consequently it was never developed beyond an elementary stage and was used only for rough translation from Russian and, later on, Vietnamese. All in all, the American concepts have progressed but little over the last twenty-five years and it has been in Canada rather than America that research has been conducted and followed up by some ad-hoc applications.

In Japan, where the computerization of society is much discussed, people are far too concerned with the remote transmission of ideograms to devote much thought to the development of a complete machine translation system.

In France, Germany, Italy and England, and also at the JRC in Ispra some basic research has been carried out and, in particular, complex linguistic models have been set up, but owing to a lack of funds an operational informational system has never been implemented. Moreover this scattering of talent and money resulted in knowledge acquired going to waste.

The Commission's decision to invite leading personalities from the relevant institutes in the member countries to meet in Luxembourg in February 1978 led to a unanimous agreement on the need to create and operate at Community level a truly European machine translation system for which, thanks to the active and ready cooperation of these institutes, the technical specifications have now been drafted.

3. The background and conditions for success

After three years of work as part of the first Commission action plan, for the purposes of which the Commission was unfalteringly encouraged and supported by CETIL (1) on the one hand and CIDST and CREST (2) on the other hand:

- multilingual thesaurus management software developed by the Commission has been widely adopted in Europe,
- the Community terminological data bank will soon be accessible via EURONET to thousands of users in the Member States and

(1) Committee of Experts for the Transfer of Information between European Languages

(2) Scientific and Technical Research Committee

- the system of automatic pre-translation Systran, despite its intrinsic conceptual limitations, is already in use in industry and in the Euronet-Diane network, and it is anticipated that it will be developed accordingly in these fields.

Computerized information processing has meanwhile been generally adopted thanks to the spectacular drop in unit costs, and this method, which today is reliable and capable of operating over long distances, has now become the remote processing of information.

All the necessary technical and economic conditions are therefore right for a large-scale Community action programme to be launched. Against a background of indisputable logic and genuine - indeed immense and pressing need - particularly in Europe, the next step is to concentrate the efforts of the top European specialist and technicians and to deploy them in a five-year working programme. Thus, the action which the Commission is proposing to the Council can create the appropriate conditions at Community level for this programme of innovation which no single country could possibly implement on its own.

4. Cost/benefit analysis

In view of the fact that Community funding amounts to 7.7 millions EUA, to which the national contributions of the participating centres, amounting to 5 millions EUA, must be added, a description of the anticipated advantages and details of the various groups who will benefit are clearly mandatory:

- at the end of 5 years of sustained effort the Community and the Member States will have at their disposal the most advanced multilingual translation system which will be suitable for numerous industrial applications and will produce royalties on a world scale as a result;
- those who will benefit directly by the system will be the various branches of industry and in particular the exporting industries, and various governmental and quasi-governmental institutions in the Member States, as well as thousands of research workers, and ultimately ordinary people, businessmen, tourists, students, etc.

- European universities and specialist institutes will have an unsurpassed tool for research and teaching which will place them in a position of world leadership;

Furthermore, the governments of the Member States will have the assurance that the financial aid they grant to their specialized university research units will no longer be sunk in subsidizing purely experimental translation systems;

- As a supplier of Eurotra hardware and software, the European information processing industry will be high on the list of those benefiting indirectly from the programmes;
- The new translation system will be a natural adjunct of the data bases in the Euronet network, which will be upgraded and extended to European and non-European users as a result;
- Finally, the Community institutions will be able to adapt Eurotra to their own particular translation requirements.

5. Implementation

- To give maximum efficiency, Eurotra will be established in close cooperation between the Commission, the authorities and research institutes of the Member States, and the European information processing industry.
- The specifications cover the preparation of
 - . a complete analysis software for each language,
 - . a complete synthesis software for each language,
 - . a multilingual lexical data base
 - . a transfer mechanism for all the language pairs and
 - . a common modular basic software.
- The development and operation will be made the responsibility of a coordinating centre for each centre or language, which will distribute the work between specialist institutes and ensure compatibility and quality. To this end, the Commission will ensure that the objectives relating to portability and compliance with international standards are accurately reflected in the specifications.
- The working programme will be the subject of a series of contracts of association between the Commission and the specialist institutes of the Member States.

It will be carried out with the assistance of an Advisory Committee including representatives of all the bodies concerned;

- The work is to be financed jointly by the Member States and from the Community budget, and the Commission will be responsible for producing the information processing components and assembling the system;
- The Commission will be guided by the advice of the scientific and Technical Research Committee (CREST) and the Committee of Experts for the Transfer of Information between Community Languages (CETIL).

6. The proposed budget is :

	millions of EUA					
	1st year	2nd year	3rd year	4th year	5th year	Total
Contributions of the Member States	0,6	0,8	1,2	1,2	1,2	5,0
Community share of the expenditure of the national centres	0,3	0,4	0,6	0,6	0,6	2,5
Direct Community expenditure	1,3	1,5	0,8	0,8	0,8	5,2
Total	2,2	2,7	2,6	2,6	2,6	12,7

Proposal for a Council Decision on the adoption
of a European Economic Community research and
development programme for a machine translation
system of advanced design

The Council of the European Communities,

Having regard to the Treaty establishing the European Economic Community,
and in particular Article 235 thereof;

Having regard to the proposal from the Commission;

Having regard to the opinion of the European Parliament;

Having regard to the opinion of the Economic and Social Committee;

Whereas, under Article 2 of the Treaty, one of the Commission's tasks is to
promote throughout the Community a harmonious development of economic
activities, a continuous and balanced expansion and an increase in stability;

Whereas the multilingual nature of the European Community is of high
cultural value, but is also in practice an obstacle to closer ties between
the peoples of the Community, to communications and to the development of
the internal and external trade of the Community;

Whereas the benefits of teleinformatics and of communication and information
networks cannot be fully reaped at Community level unless the language barrier
is overcome;

Whereas the development of computational linguistics is likely to contribute
to the overcoming of this barrier;

Whereas considerable research has already been carried out on this subject
in the Member States;

Whereas this research, by its very nature, can only bear fruit provided that it is supported by a Community action which would, among other things, have a catalytic effect on work already carried out;

Whereas such Community action can, in particular, consist in the creation of a European machine translation system of advanced design;

Whereas such a system would have numerous industrial applications and be of direct benefit, inter alia, to the various industrial sectors and, in particular, to exporting industries;

Whereas the preliminary work already completed has demonstrated the technical feasibility of such a system;

Whereas a Community research and development programme on machine translation is therefore likely to make an effective contribution to the achievement of the above objectives of the Treaty;

Whereas the Treaty has not provided the necessary powers to this end;

Whereas the Scientific and Technical Research Committee (CREST) has delivered an opinion on the Commission's proposal;

HAS DECIDED AS FOLLOWS:

Article 1

A research and development programme for the creation of a machine translation system of advanced design is hereby adopted for a period of five years commencing on

The content of the programme is described in the Annex hereto.

Article 2

The expenditure required for the execution of the programme is estimated at 7.7 million EUA (1979 value); personnel requirements are estimated at 16 persons.

These figures are given for guidance purposes only; the funds needed to execute the programme shall be determined every year by the budgetary procedure.

Article 3

The Commission shall be responsible for the execution of the programme, in particular by means of research contracts. It shall be assisted by an Advisory Committee on Programme Management, the membership and terms of reference of which are set out in the Council Resolution of 18 July 1977 on advisory committees on research programme management⁽¹⁾.

The Commission shall inform the CREST at regular intervals on the progress of the work. Furthermore, the Commission shall submit an annual report to the Council and to the European Parliament concerning the execution of the programme.

Article 4

The dissemination of information resulting from the execution of the programme shall be effected in accordance with Council Regulation (EEC) No 2380/74 of 17 September 1974 adopting provisions for the dissemination of knowledge relating to research programmes for the European Economic Community⁽²⁾.

Article 5

In accordance with Article 228 of the Treaty, the Community may conclude agreements with third countries on their participation in the research programme which is the subject of the present Decision.

The Commission is hereby authorized to negotiate such agreements.

(1) OJ No C 192 of 11 August 1977, p. 1

(2) OJ No L 255 of 20 September 1974, p. 1

1. Objectives

The objective of the project is the creation of a computer-aided translation system capable of dealing with all official languages of the Community.

2. Programme of work

The programme of work includes the following elements:

- The development of an analysis module for each language, which will carry out morphological, syntactic and logico-semantic analysis of source text to the levels specified by the definition of the commonly agreed interface structure.
- The development of a generation module for each language, which will be able to accept an interface structure of the type commonly agreed and generate from it a surface representation in that language.
- The development of a transfer module for each language couple, which will take as input an interface structure of the type commonly agreed, derived from an analysis of the source text, and produce from it an interface structure of the same type, from which the target text may be generated.
- The compilation of monolingual dictionaries in two steps:
 - a) an initial test dictionary for each language, to be used both in analysis and in generation, comprising approximately 2 500 entries in a specific subject field;
 - b) an expanded monolingual dictionary, whose final size will be approximately 20,000 entries.
- The compilation of a multilingual lexical data base for use in the transfer phase, in two stages as for the monolingual dictionaries.
- The development of a basic software package, including support software for dictionary handling and updating as well as software capable of dealing with the analysis, transfer and generation modules described above. This will include interpreter compiler programmes for the specially defined high-level definition language in which grammars and dictionaries are to be written.
- The development of the software for the user-system interface, by which the different modules may be combined into integrated systems corresponding to the various user options defined.
- The Commission will ensure that the objectives relating to portability and compliance with international standards are accurately reflected in the specifications.