

# COMMISSION OF THE EUROPEAN COMMUNITIES

COM(76) 705 final.

Brussels, 23 December 1976.

## COMMUNICATION FROM THE COMMISSION TO THE COUNCIL

Plan of action  
for the improvement of the transfer of information  
between European languages

---

COM(76) 705 final.



Plan of action  
for the improvement of the transfer of information  
between the European languages

---

1. Purpose

1.1. General

The European Community comprises a multilingual group of peoples seeking integration, which is none the less continuous and progressive.

The multiplicity of languages, which is a permanent feature, represents a priceless heritage of civilization for it reflects the variety of cultural backgrounds of our nations.

But this linguistic diversity poses problems in daily life which hamper the ongoing process of integration and the smooth operation of the Community institutions. Immense efforts are needed - long before agreement is reached - simply to understand each other in these multilingual settings. One could never quantify the cost of communication and assess the inevitable wastage of time and intellectual effort in this environment, let alone the sum of frustration that ensues.

Despite endeavours to reduce the number of pages of documents and the number of languages into which translations are requested, the Commission is faced by a level of demand which is increasing at a rate of some 10 % per year. To cope with this demand, the Community institutions have had to establish extremely large translation services, probably the most highly staffed in the world.

Even so, in order to facilitate the translation process, particularly of repetitive texts which can be simply conveyed from one language to another, it seems highly desirable to make use of a wide range of working tools, some of which are available and some being developed.

Such tools range from the dictionary and multilingual thesauri to the terminology bank and computer-aided translation systems. They are expensive to develop and use. But the scale of the Community's present and foreseeable linguistic activities is such that the systematic use of multilingual tools cannot but lead to savings in the long term. In this respect the Community is in an exceptional situation.

## 1.2. Background

Dictionaries, card indexes and some automated instruments have been used in translation for a very long time. In the last thirty years, the computer has arrived to complete the range. In 1950-60 there was something of an obsession, particularly in the United States, for experimenting with automatic translation. At the end of that period, the authorities reduced their financial backing and research declined.

But the performance levels which have now made the computer an irreplaceable tool have continued to improve. Memories have become more capacious and the memory unit less expensive. Access has become more rapid and less costly. The information stored is better organized and data retrieval can be more sophisticated, so that it is now possible to "process" the most complex language structures.

Direct access has been refined to such a point that dialogue with data systems and data can replace consultation of reference books, sorting through index cards and a fair proportion of repetitive intellectual work.

Several research centres in the Community have tackled the problem of communication between different languages. The Commission itself has pursued certain lines of research on linguistic problems at the Ispra Establishment of the Joint Research Centre. The translation services have begun to build up a terminology bank and the scientific information services have set up a multilingual information retrieval system for metallurgy.

More recently, under the Community's first plan of action in the field of scientific and technical information and documentation adopted by the Council in 1975, the object of which is to set up a European information network under the title of "EURONET", the Commission has been studying the possibilities of applying various tools and methods to the transfer of scientific and technical information from one language to another. Some of these methods have already proved their worth in practice and several of the Community's advisory committees, particularly CIDST and CREST (\*), agree that the instruments designed to facilitate the transfer of scientific and technical information have been developed to such a point that they should now be turned to account in other fields.

## 2. Basic principles

Obviously the operation of the Community's institutions could not fail to be substantially improved, from the efficiency point of view, by the gradual application of computerized methods to the transfer of information between the Community languages. The Commission has therefore decided to initiate a three-year plan of action, to come into force as early as 1977, in view of the risk that present conditions in the institutions may be aggravated by the accession of new member countries.

---

(\*) CIDST: Committee for Scientific and Technical Information and Documentation.

CREST: Scientific and Technical Research Committee.

It must first be made clear that this plan of action does not concern true research and development, but a series of tests, pilot experiments, and trial applications using systems designed elsewhere and selected on the basis of their positive features and their development potential. There is no system nor technique capable of solving all problems at once. The present first generation of computerized instruments will undoubtedly be followed by a second generation of more sophisticated ones.

Consequently, while paving the way for the advent of the future systems, advantage should be taken of the facilities offered by systems already at the operational or preoperational stage. But for the reasons mentioned, it would be unwise to consider only the latter, or only one of them.

Consequently, the Commission intends combining the programme of applications with a continuous review of progress under way and active support for the development of new systems in the member countries, in order that such systems be geared to meet the needs of the Community in a coordinated manner.

### 3. Plan of action

The Commission will undertake a coherent plan of action on several fronts:

- automatic pre-translation of unprocessed texts drafted in natural language,
- automatic translation of texts drafted in limited syntax,
- terminology banks,
- multilingual thesauri,
- technical infrastructure,
- assessment of applied research,
- encouragement of multilingualism.

### 3.1. Automatic pretranslation of texts in natural language

The programme has been planned in the light of two considerations:

- (1) In the past the development of automatic translation systems for scientific texts has focused mainly on the Russian/English language pair, which is of no particular interest to the Community.
- (2) Disregarding the quality of the text supplied by the machine, no automatic translating system can operate properly in any sector without a substantial initial investment, equivalent to the establishment of a multilingual dictionary for the sector in question.

The Commission therefore intends to introduce an initial tried and tested system into its departments and bringing it first of all to the pilot-experiment stage and then to the operational stage for one pair of Community languages and one specialized field. If expectations are fulfilled, the system will be extended to include other specialized fields and other Community languages.

Thus, the initial system will be gradually converted into a European instrument belonging to the Community and will be capable of:

- rapid draft translations in certain fields, allowing, for instance, an engineer to grasp the substance of a document without necessarily understanding all the nuances;
- translations of sufficient quality to pass directly to checking or revision, as is currently the procedure for drafts prepared by translators.

The development work will be carried out mainly at the Commission's Computer Centre in Luxembourg, by Commission staff in collaboration with consultants.

Special attention will be paid to the cost/efficiency ratio and to the technical problems posed by the recording of texts and dictionaries.

Care will be taken to ensure that the dictionaries and other tools created for this purpose will be applicable later in the systems of subsequent generations, thus securing the best possible return on the resources invested.

### 3.2. Automatic translation of texts in limited syntax

There are systems which provide satisfactory automatic translations of texts drafted according to limited grammatical rules. Such systems are particularly useful for the translation of summaries in several languages for information retrieval purposes.

It is intended that an initial tried and tested system of this type be applied to scientific and para-scientific texts the drafting and translation of which are of interest to the Community, that the extension of this system to other specialized fields and to other Community languages be encouraged, that joint efforts be made to gradually lift most of the grammatical constraints which limit the use of the system, and if appropriate, that some staff be trained in the drafting of texts with reduced syntax.

### 3.3. Terminology banks

The compiling of multilingual terminology files in each country, each institution or even by each translator and interpreter separately for their own use represents a fantastic waste of



intellectual resources, which could be reduced by setting up terminology banks accessible to all.

The Commission is already in the throes of establishing, for its own translators, a terminology bank whose reliability is assured by a strict validity check. Subsequent development of this terminology bank, by extension of its content and improvement of accessibility, is provided for in the present plan of action, to be carried out in close liaison, including the exchange of available collections, with similar projects in hand elsewhere in the Member States and some non-member countries.

Once such a terminology bank has acquired the necessary volume and quality, its use by direct access will be far more economical than tracking down the same information in huge volumes of printed indexes, which are generally fairly inaccessible.

#### 3.4. Multilingual thesauri

These controlled multilingual vocabularies are essential tools for the transfer of information. Like terminology data banks, they are benefiting from the advances in computer techniques.

The best-known international information retrieval systems are based on multilingual thesauri. It is essential to use the most modern computerized methods to manage and use them for purposes of information retrieval.

The Commission has substantial experience in this field since it has helped to develop highly advanced methods of thesaurus compilation.

It seems advisable to compare and harmonise the methods developed and used in the Commission's departments with various others advocated by some international institutions so that common guidelines can be mapped out.

In particular, it is urgently necessary to draft recommendations for the establishment of multilingual thesauri by combining or dovetailing existing monolingual vocabularies in accordance with these guidelines.

### **3.5. Technical infrastructure**

A number of additional measures should be undertaken in order to situate the plan of action in a real operational context. The main purpose of such measures would be to improve the methods used for the acquisition, handling and retrieval of texts, both upstream and downstream of processing by computer.

They include

- improvement of conventional techniques for the acquisition of non-digital documents, particularly texts of all kinds;
- development of sophisticated tools permitting direct acquisition and transliteration of texts;
- use of video terminals for direct processing of texts;
- development of systems for returning translated texts on various media, in situ and at a distance.

### **3.6. Assessment and orientation of applied research**

The measures described above must be periodically reoriented in the light of current research and development on multilingualism.

Particularly in automatic translation, there are so many persons involved and projects under way that new ideas and promising initiatives may emerge at any time; these must be channelled along certain lines which correspond with the action programme and take into account the evolution of the Community's requirements.

The Commission will therefore see that a limited number of incentives and encouraging measures be adopted and implemented in collaboration with the JRC and specialized institutions in the Member States.

### 3.7. Promotional measures

A large-scale plan of action requires the participation of all who are directly or indirectly concerned by multilingualism, including teachers, publishers, documentalists, computer specialists, linguists, translators, interpreters and terminologists, and even the users of scientific or other information.

To secure their participation, they must first be identified, approached, informed and interested. This could be achieved through specialized publications and the organization of courses, seminars and demonstration sessions.

A starting point in this direction will be provided by the organization of the Third European Congress on information systems and networks, entirely devoted to multilingual problems, in May 1977.

### 4. Staffing and implementation

For the planning and implementation of this work, the Commission intends to secure the assistance of a "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 experience of technical problems and requirements at European level.

Under this three-year plan of action, the work will be carried out with the assistance of consultants and the staff of the European institutions, in particular the Commission's Computer Centre. The Commission intends to form a single team of linguists and computer specialists, drawn from its own departments, possibly other institutions and on secondment, from other specialized organizations, in particular, specialized organizations in the member countries.

### 5. Budget

The expenditure involved in implementing the three-year plan of action is of the order of 2,750,000 UA, broken down as follows (in units of account, at 1976 prices)\*):

	(1977)	(1978)	(1979)
Automatic pre-translation of of natural language	180.000	360.000	460.000
Translation of texts written in limited syntax	110.000	240.000	280.000
Terminology banks	70.000	120.000	140.000
Multilingual thesauri	50.000	80.000	120.000
Technical infrastructure	40.000	70.000	100.000
Applied research	30.000	80.000	100.000
Promotion	20.000	50.000	50.000
<b>TOTAL :</b>	<b>500.000</b>	<b>1.000.000</b>	<b>1.250.000</b>

\* A financial statement is attached as annex 3.

FINANCIAL STATEMENT for the 1977 budget

Plan of action  
for the improvement of the transfer of information  
between the European languages

A. PART I: INTERVENTION APPROPRIATIONS

1. Relevant budget heading code

Item 3621

2. Title of budget heading

Projects supplementary to the three-year plan of action in the field of scientific and technical information and documentation.

3. Legal basis

Action directed to the improved effectiveness of the Commission's staff.

4. Description, objective(s) and justification of the project

The aim of the project is to develop computer-aided methods for carrying out pre-translations of texts of all kinds in the six official languages of the Community.

5. Appropriations

In 1977 :	500.000 UA
In 1978 :	1.000.000 UA
In 1979 :	1.250.000 UA

For the first year of the plan, 1977, it is likely that the commitments will be mainly concentrated in the second six months; for the following years, the commitments will be spread out evenly over the year.

The basis for calculation of the above figures was the number of computer instructions and the number of programmes which would have to be completed, taking into account past experience in a particular field, food industry, and for a particular pair of languages, English-French.

6. Type of control envisaged

Tests of quality of the service ; moreover, operations carried out under contract will be the subject of reports to be examined not only by the staff of the Commission but also by an advisory committee which it is proposed to set up; this committee will be composed of experts of the Member States as well as of the Commission.

B. PART II: COMPLEMENTARY INFORMATION CONCERNING A NEW ACTION

7. Total cost of the action during the period foreseen

2,750,000 UA

8. Information on the personnel and the administrative appropriations needed to carry out the action

The management of a plan of action of such scope will require the reinforcement of the staff presently available by a senior administrator A5/A4, an administrator A7/A6, an assistant B3/B2 and two secretaries C3/C2; the assistant, administrator and senior administrator need to have some knowledge in the areas both of computers and linguistics.

9. Source of finance

The appropriations will need to be included in the future budget; they have been asked for by the Commission in its budget proposals for 1977 and the Parliament has included 500,000 UA of payment appropriations for this purpose on the new budget line 3621 accepted by the Council.