European Communities

# EUROPEAN PARLIAMENT

# Working Documents

## 1981 - 1982

25 May 1981

## Report

drawn up on behalf of the Committee on Energy and Research

**on the proposal from the Commission of the European Communities to the Council (Doc. 1-352/80) for a decision on the adoption of a European Economic Community research and development programme for a machine translation system of advanced design**

**Rapporteur: Mr G. ADAM**

By letter of 4 August 1980 the Council of the European Communities requested the European Parliament to deliver an opinion on the proposal from the Commission of the European Communities to the Council for a decision on the adoption of an EEC research and development programme for a machine translation system of advanced design.

The President of the European Parliament referred this proposal to the Committee on Energy and Research as the committee responsible, and to the Committee on Budgets for their opinion.

On 25 November 1980 the Committee on Energy and Research appointed Mr Adam rapporteur.

It held an initial exchange of views on 4 December 1980, considered the report at its meetings of 27 February and 24 April 1981, and adopted the motion for a resolution and explanatory statement at the latter meeting by 13 votes in favour with 3 abstentions.

Present: Mrs Walz, chairman; Mr Normanton, vice-chairman; Mr Adam, rapporteur; Mr Beazley, Mr Fuchs, Mr Herman (deputizing for Mr Rinsche), Mr Linkohr, Mrs Lizin, Mr Müller-Hermann, Mr Muntingh (deputizing for Mr Percheron), Mr Pisani, Mr Price, Mr Purvis, Mr Rogers (deputizing for Mr Linde), Mr Sassano and Mr Vandewiele.

The opinion of the Committee on Budgets is attached.

# C O N T E N T S

The Committee on Energy and Research hereby submits to the European Parliament the following motion for a resolution together with explanatory statement:

## MOTION FOR A RESOLUTION

embodying the opinion of the European Parliament on the proposal from the Commission of the European Communities to the Council for a decision on the adoption of an EEC research and development programme for a machine translation system of advanced design

The European Parliament,

- having regard to the proposal from the Commission of the European Communities to the Council [1],

- having been consulted by the Council (Doc 1-352/80),

- having regard to the report of the Committee on Energy and Research and the opinion of the Committee on Budgets (Doc, 1-193/81),

- recalling that a substantial proportion of the staff of the Community institutions is employed in translation services and that translation is expensive and a hindrance to the efficient working of the Community,

- aware that enlargement of the Community will put a disproportionate strain on the institutions' translation services,

- aware that steps will have to be taken to ease the pressure on the translation services long before the proposed machine translation system can be operational,

1. Recognises that Europe's cultural heritage is immeasurably enriched by its different languages;

2. Considers nevertheless that every effort should be made to reduce the delays and misunderstandings which arise from having different languages, to draw citizens together, and to ease trade and commerce, and that a machine translation system can help achieve these aims;

---

[1] OJ No. C 234, 12.9.1980, p. 2

PE 70.657/fin.

3.  Points out that the feasibility of machine translation has been
    demonstrated and that the programme provides an opportunity to
    develop expertise already acquired, to design a system of widespread
    applicability, and to use and adapt the system for its own transla-
    tion needs;

4.  Believes, however, that the programme should be divided into a
    preliminary phase and an implementation phase, each of two-to-
    three years' duration, and that the progress of the programme should
    be examined before embarking on the implementation phase, with
    consultation of Parliament on any amendment to the programme;

5.  Believes that the likelihood of producing a commercially usable
    system would be improved if its development in the implementation
    phase was not confined to university teams and Commission bodies;

6.  Urges prompt action following review of the organisation of the
    translation services in the Community institutions, and in particular
    the early adoption of word-processing technologies which are readily
    available;

7.  Approves the proposed programme, but urges that, in addition to
    the EUROTRA system, an urgent feasibility study be carried out on
    computer-based translation systems which would be suitable for the
    high proportion of repetitive translation in some Community
    institutions, and which would be available sooner;

8.  Demands that in negotiations for research contracts the Commission:

    - shall not deviate from the principles laid down in Document XII-105/76
      to the detriment of the Community,

    - ensures that the Community has a say in the exploitation of
      industrial property arising from the programme,

    - provides for royalties to be paid to the Community on the commercial
      exploitation of the system, so that the Community might recoup its
      contribution;

9.  Urges that provision be made for the inclusion of Greek, and, at the
    appropriate time, for Spanish and Portuguese in the system, and

10. Approves the proposal for an R. & D. programme for a machine trans-
    lation system of advanced design, subject to the incorporation in
    the Decision of the following amendments pursuant to Article 149(2)
    of the EEC Treaty:

RECITALS and PREAMBLE unchanged

Article 1

A research and development prog-
ramme 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 1

A research and development prog-
ramme 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.

The programme will be divided
into two phases. Progress will be
reviewed after two years and the
programme revised accordingly. The
Advisory Committee and the European
Parliament shall be consulted on
these revisions.

Article 2

Unchanged

Article 3

Unchanged.

Article 4

Unchanged.

Article 5

Unchanged.

Article 6 (new)

(a) Insofar as the Community shall
promote the programme by means of
research contracts, provision shall
be made for repayment of part of the
contribution made by the Community
if the results of the programme are
used industrially or commercially
and/or lead to industrial property
rights and the subsequent award of
licences.

(b) The Commission shall negotiate
and conclude the necessary contracts.
For this purpose it shall draw up
a standard contract defining the
rights and obligations of each party,
including where appropriate condi-
tions and procedures for possible
royalty payments.

## EXPLANATORY STATEMENT

I. INTRODUCTION

1. That the different languages spoken in Europe contribute
to its cultural diversity, there is no doubt. But that they also
give rise to misunderstandings and delays is also clear; this is
not purely a cultural or political problem, but also one which
could increasingly hamper European industries as the
information revolution  gathers pace.

2. Within the Community institutions we sometimes forget how
much their working depends on the language services. In the
Parliament alone, there are upwards of 600 people (i.e. about one-third)working
directly in the translation service, with more than 1300 employed
at the Commission. Except, of course, when things go wrong,
the old English saying "out of sight, out of mind" might have
been coined for the Community's translation services. It is
perhaps unfortunate that, in the present context, an early machine
system transformed this saying into "invisible lunatic" (A. Calder-
Marshall, The Listener, 23.4.1964).

3. It is clear that every effort should be made to improve the
efficiency of the translation services, and a machine translation
system could obviously play a role in this. It is not however a
panacea. The aim is to do away with the banal and boring aspects
of a translator's work, and the gains in efficiency should arise
in speed and availability as well as reduced cost.

4. The volume of translation work increases disproportionately
with enlargement of the Community. With nine member states and
six languages, there are 30 language combinations to deal with.
The advent of Greece has raised that to 42, an increase of around
one-third. With a Community of 12, operating in nine languages,
the number of combinations rises to 72, more than double the
number over the 1980 figure. Clearly the problem requires
urgent and serious attention.

## II. THE EXISTING TRANSLATION PROCESS

5.   It is worth recalling that several stages are involved in translation, and that machine translation will assist only with one of them, although associated word-processing techniques can help with some of the other stages.

6.   Although the different roles of the various Community institutions mean that the types of translation they do, the numbers of people needed, and the procedures used, may differ somewhat, translation of a document in essence involves:

(a)   allocation of the document by a central administration to a translator;

(b)   translation (written or dictated);

(c)   typing of the translation;

(d)   scrutiny of the translation by a revisor; and

(e)   retyping as necessary.

In smaller organisations, of course, these stages are sometimes all done by a single person.

7.   In the Parliament, for example, each language section comprises around 100 people directly involved in translation, of whom  about a half are translators, about half a dozen secretariat and the remainder typing staff.  There are in addition terminologists and certain planning staffs.

## III. THE PROPOSED PROGRAMME

8.   At the end of 1976 the Commission approved a first Plan of Action concerning the transfer of information between European languages; this concentrated on developing thesauri and terminology, and on investigating automatic translation.  The rights to a US system were acquired (SYSTRAN).  A second Plan is now proposed, with a separate but allied proposal concerning an R & D programme for a machine translation system of advanced design.  Parliament has been consulted on this latter proposal.

9.   For machine translation, the Commission proposes the development of a new system (EUROTRA) rather than pursuing SYSTRAN. It suggests that the latter is rather inflexible, being essentially bilingual and therefore requiring separate setting-up for each language pair.  SYSTRAN appears to have commercial potential in some fields however, and proposals have been invited to develop this.

10. The proposed EUROTRA system, however, is modular, and the translation process comprises in essence the following steps:

- passage in language A analysed and broken down into its constituent parts such as words and phrases, using analysis software,

- parts compared with multilingual lexical data base and transformed into constituent parts of language B using the transfer mechanism relevant to that language pair,

- the constituent parts in language B are composed into a text using synthesis software.

11. This means that certain aspects of the system are common, whatever the language pair involved, and less investment is needed in dictionaries, because of their multilingual nature. Similarly individual modules can be updated as linguistic or software techniques improve; the system is better adapted to long term development than SYSTRAN. Similarly it is more easily usable on different computers, and is said to be more useful for training.

12. The total proposed budget is 12.7 million EUA (in 1979 prices), spread over five years. Community expenditure at 7.7 million EUA comprises around 60%, with member states contributing 5 million EUA. The Community will concentrate on developing the common parts of the system, and member states the modules for its own language (when both target and source language), via support for university research teams.

13. It is intended to achieve a workable system by 1985 or 1986, a system which will operate on unprepared text and produce an output requiring a 10% revision rate (i.e. one word in 10); the SYSTRAN revision rate is 25%. Manual translation also requires revision, of course. On a comparable basis it is expected that EUROTRA translations would cost 2 BF/word, compared with 3.5 for SYSTRAN and 5 for conventional translation. Speed and availability are often more important than cost, and this may have militated against more widespread use of SYSTRAN in Community institutions.

IV. OBSERVATIONS ON THE PROPOSAL

14. Making translation easier and more efficient is clearly an enterprise in which the Community can play a leading role, for reasons of self-interest and because of the importance of the subject to the European economy and its relevance to the computer industry. Experience indicates that machine translation is feasible, even if much remains to be done. The amounts of money involved (7.7 million EUA Community expenditure, rising to 16.45 million EUA if member states' contributions and the second Plan are

included, spread over five years) are modest compared with current expenditure on translation services in the Community institutions alone, which cannot be less than 50 million EUA per annum.

Objectives

15. Nevertheless your rapporteur considers that the objectives of the programme should be defined more clearly. This is, after all, a long-term project, running to 1985 or 1986 for developing the system and probably beyond before dictionaries are operational for some subjects, or indeed before certain languages are included.

16. Three general objectives exist:

(a) the development of a practical system which will give comprehensible results for straightforward documents, having in mind particularly industrial and trade use,

(b) the maintenance of progress in a field where Europe has built up a considerable body of knowledge,

(c) the improvement of efficiency in the translation services of the Community's own institutions, through the use of machine translation.

17. It is clear that governments of some third countries have developed a useful combination of support for research together with public procurement which has given their industries new markets. Machine translation provides just such an opportunity, if the programme is properly set up.

18. The development of a practical system (e.g. for industrial and trade use) would draw on and develop the existing body of academic knowledge. The potential market is wide and the industrial and commercial value considerable. Some aspects of the proposed system are novel (portability, multilingualism, modularity, extensibility) and need to be investigated before industrial exploitation can begin. The lack of manufacturer involvement is however unsettling if the objective is so market-oriented. Manufacturers could perhaps be represented in the management structure so as to keep this objective in view ; the development should not be an academic exercise, and should take account of changes in computer technology.

19. There is in principle no reason why the needs of general users and of the Community institutions should differ drastically, and the latter are clearly regarded as major potential users. The translation problems in the Community institutions require urgent solution, however; they cannot wait until the latter half of the decade.

## Needs of Community institutions

20. Your rapporteur doesn't wish to prejudge studies under way of translation procedures within the institutions; it is clear that improvements can be made. But it seems that little sustained effort has been made to introduce available word-processing techniques. The Parliament, for example, seems to have fewer than a dozen typewriters with any sort of memory capacity, never mind any more sophisticated apparatus. Yet these are in widespread use in industry and commerce. To make full use of the possibilities offered by word-processing would require changes in translators' working methods, combining some of the steps set out in paragraph 6 of this report; translators are unlikely to be displaced, economies arising more in the support staffs. Investigation of word-processing, possibly with a pilot programme, and preferably on a common basis between the institutions, is an urgent necessity.

21. Strictly speaking, the use of word-processing techniques in the translation services, and automatic translation itself, are separate topics. The former aims to make the secretarial aspects of translation more efficient, the latter the translation process itself. Of course word-processing is ideally suited to use in combination with machine translation. One might further envisage direct input of text via machines able to read handwritten text, although this is likely to be uneconomic for the time being.

22. It is clear, therefore, that from the point of view of the needs of the Community institutions, machine translation has a role to play but within a translation procedure which needs to be considered as a whole. The concept of EUROTRA is a sophisticated one, and one which deserves long-term support. It is an approach suggested by the research Directorate-General of the Commission, however, rather than the translation service and other machine-translation systems may be applicable alongside EUROTRA in meeting the special needs of the Community institutions.

23.  Among those special needs is the repeated translation of very
similar documents such as minutes, tenders and regulations;  these
amount to approximately half the translation work in the Council and
substantial proportions elsewhere.  This raises the possibility of a
machine translation system which, rather than analysing each passage
anew, relies on large memory capacity to store similar passages
translated in the past.  This possibility is discussed in more detail
below, but seems to your rapporteur to be a proposal worthy of study as
an adjunct to the EUROTRA proposal.

Technical risk

24.  Considerable progress has been made with machine translation and
experience gained during the 1st Plan of Action.  Early systems were
dictionary-based;  dictionaries remain very important although analysis
is now considered crucial.  Translation systems are necessarily big, so
the decision to opt for a modular system (with sub-modules also) is
sensible.  In any system, there is a distinction between the data being
manipulated on the one hand and the procedure for manipulating them on
the other.  This distinction is often blurred, and while this can be
done for machine translation between a single language pair, is
unacceptable in a multilingual system.  EUROTRA is the first system to
be set up from the start as a multilingual system.

25.  Of course, even in EUROTRA, the transfer module is unique to each
language pair but it is desirable to keep it as small as possible.  This
in turn implies that a deeper level of analysis of the source text is
necessary.  This, and the need for the analysis to be suitable for
translation into any of the other languages, means that the depth of
analysis being sought is very ambitious.  Similarly, clear rules have to
be developed to cope with ambiguities, and it is also intended to rely on
a sophisticated generation module to improve the readability of the
translation.

26.  For these and other reasons, first versions - to be ready in about
five years - will not deal with all possible sentence structures, nor
with all languages, nor with all subject areas.

27.  The basic concept of EUROTRA is a linguistics one, but the
translation problem can also be regarded as one of data processing.  This
is a conceptually more simple approach, and has been proposed as being
particularly suitable for certain repetitive translations in the
Community's institutions.  It would require substantial memory capacity
and draws more heavily on the raw processing power of computers.  Fed with

an increasing selection of passages, sentences or phrases already translated, the system would search these for something matching the passage to be translated. Modern techniques allow the search to be efficiently concentrated and also, more interestingly, to cope with a certain degree of mis-match (e.g. singular instead of plural). The output could show up those phrases translated in one type face for example, words translated using the Eurodicantom dictionary in another, and untranslated parts in a third. The eventual quality of translation will depend on the similarity of the passage to what has gone before, size of memory etc., and is not inherently limited (cf SYSTRAN).

28. This seems to your rapporteur to be an approach worthy of study, alongside the EUROTRA proposal. In the long run, the two systems could run side-by-side, and the simpler system could be available sooner, meeting the more immediate problems and providing some insurance in the event of delays to EUROTRA. The technical risk appears to be less, and the system would be susceptible to progressive improvement.

Languages_included

29. The multilingual and modular nature of the proposed structure of EUROTRA means that certain parts of the system are common whatever the language pair involved, while the analysis and synthesis parts are specific to a particular language. Thus while the needs of each language pair need to be borne in mind when designing the central parts, not all languages have to be brought on stream together. Development should therefore concentrate on those language pairs in greatest demand in translation of Community documents, and those pairs which are commercially significant.

30. Greek is a special case because of the totally different script employed. Results are awaited from international work aimed at defining appropriate standards for display and transliteration, but this should not rule out laying the foundations for extension to Greek at an early stage. Similarly, the foundations for eventual extension to Spanish and Portuguese must be laid at the appropriate time.

## Ownership

31. The Commission envisages that the commercial value of a working system will generate royalties on a world-wide scale. While it is probably too soon to negotiate in detail how the system might be exploited industrially, some clear guidelines ought to be set if income is to be divided not only according to financial contribution but also the know-how contributed to the project.

32. The Commission considers that its SYSTRAN system has some commercial value. It is true that this and other systems are such as to make substantial extension and improvement difficult. But it does permit rough translations and allow sufficient understanding of a document to know whether full translation is necessary; this is adequate in some fields. Proposals have been invited for the setting up of SYSTRAN on a commercial basis, and it is expected that Euronet host operators would be interested : there is now a lot of information available on data bases linked to Euronet, but these are monolingual and rough translation of even titles and perhaps abstracts is attractive. Translation of complete documents would probably be ordered off-line.

## V. CONCLUSIONS

33. The proposal has much to commend it. The feasibility of machine translation has been demonstrated, and the project is one well-suited to imaginative support for R+D and central procurement, building on existing expertise and opening up new markets both within and outside the Community institutions.

34. The markets exist, and cheap translation will increase the flow of information to the benefit of trade and industrial performance. Within the Community institutions, not only is cost important but also speed, and although sensitive texts will require further editing, machine versions should be adequate for a wide range of day-to-day documents.

35. The concept of the system is elegant, with its multilingual
and modular approach, and capacity for continuous improvement;
it seems worthwhile concentrating further development on this new
system, rather than SYSTRAN, for this reason. The sums of money
involved, compared with the potential benefits, appear to be
reasonable. Existing expertise in this subject will be developed,
expertise of increasing relevance in computer languages and the
man-machine interface themselves. Two items require further
thought however.

36. Promotion of new technology, through support of R+D and
central procurement, appears to be successful when close attention
is paid to eventual production and marketing. The wisdom of
confining the project to university teams and the Commission
therefore seems debatable. The difficulties of involving manufac-
turers at an early stage of the project are recognised, but prospects
for commercial success would be enhanced if they were.

37. The heavy load on translation services in the Community
institutions is well known. It is a load which will increase
disproportionately with enlargement. This is not a problem which
can await the arrival of EUROTRA. It requires:

(a)  prompt action on management and methods following the studies
     completed or in hand;

(b)  a concerted effort to make effective use of available technology
     such as word-processors; and

(c)  urgent consideration, as an adjunct to this machine translation
     programme, of computer-aided translation techniques to deal
     with common types of documents. Conceptually simpler than
     EUROTRA, these could be in service earlier and provide a back-
     up for the latter.

## OPINION OF THE COMMITTEE ON BUDGETS

Letter from the chairman of the committee to Mrs WALZ, chairman of the
Committee on Energy and Research

Subject: Commission proposal for a decision on the adoption of a European
Economic Community research and development programme for a machine
translation system of advanced design (Doc. 1-352/80)

Dear Mrs Walz,

The Committee on Budgets considered this Commission proposal at its
meeting of 24 April 1981. It noted the financial statement submitted by
the Commission, indicating that the total cost for the five-year programme
would be 12.7 m EUA, of which the Community share would be about 7.7 m EUA,
60%. This expenditure includes the creation of 16 new posts.

However, other documents submitted by the Commission make it clear that
the figures in the Commission proposal and the financial statement are based
on 1979 values and do not allow for inflation. If these figures are updated
the expenditure would amount to 15.1 m EUA at 1981 prices, of which the
Community would have to find 9.1 m EUA.

The 1981 budget contains no funds for this programme, as the Council
deleted the appropriations entered by Parliament for the purpose.

The working document also made available by the Commission made a number
of points of interest to the Committee on Budgets which led it to the
following conclusions:

1. It considers the additional 16 posts, for which no justification is given,
   to be unacceptable, and calls for staff to be limited to 5 grade A and
   3 grade B temporary posts for the duration of the programme. Personnel
   required for technical and administrative support for the project must
   be recruited from the Commission's existing staff. If staff taken on
   are limited to those required for research alone the Commission, on the
   basis of its own figures, could keep costs within its original estimates.

2. The contracts concluded between the Commission and national authorities
   must include a clause to the effect that aid granted for this research
   will be repaid if the research produces results suitable for industrial
   or commercial application and/or generating royalties or proceeds of
   sales.

3. The programme must therefore be reviewed during its third year,
i.e. after its preliminary phase, and possibly revised after
consultation of the advisory programme committee. The European
Parliament must be provided with an assessment of the proposals for
such a revision of the programme, and deliver its opinion on the
matter.

The Committee on Budgets requests the Committee on Energy and Research
to take these recommendations into account in its report and to put them
in the form of specific amendments to the Commission's proposal for a
decision.

On the subject matter itself, warnings were expressed from various
quarters urging extreme caution in respect of technical progress of this
kind which might turn out to have a dehumanizing effect.

With these reservations, the programme proposed by the Commission
was endorsed by 8 votes to none with 4 abstentions.

Yours sincerely,


Erwin Lange


Present: Mr Lange, chairman; Mr Notenboom and Mr Spinelli, vice-
chairmen; Mr Adonnino, Mr Balfour, Mr Bonde, Mr Forth, Mr Georgiadis,
Mr Gouthier, Mr Jackson, Mr Langes, Mr Newton Dunn and Mr Schön.