COMMISSION OF THE EUROPEAN COMMUNITIES

COM(94) 69 final
Brussels, 20.09.1994

COMMUNICATION FROM THE COMMISSION TO THE COUNCIL
AND THE EUROPEAN PARLIAMENT

Final evaluation of the results of Eurotra: a specific programme concerning the preparation of the development of an operational Eurotra system for Machine Translation.

COMMMUNICATION OF THE COMMISSION

TO COUNCIL AND THE EUROPEAN PARLIAMENT

Subject: Final evaluation of the results of Eurotra: a specific programme concerning the preparation of the development of an operational Eurotra system for Machine Translation

I. INTRODUCTION

1. This communication concerns the evaluation of the results of the Eurotra research programme. The evaluation was carried out through a panel of independent experts, in accordance with Article 4 of the Council Decision 90/664/EEC of 26.11.90 concerning "the preparation of the development of an operational Eurotra system". The Decision also stipulates that the evaluation should be transmitted to Council and the European Parliament. The evaluation report entitled "Final Review Panel Report, February 1993" and the Opinion of the Eurotra Advisory Committee are annexed to this Communication.

2. This communication gives a short overview of the Eurotra programme, the main conclusions and recommendations of the final evaluation and the position of the Commission.

II. THE EUROTRA PROGRAMME

3. In November 1982, the Council decided to launch the Eurotra research and development programme *(Council decision 82/752/EEC of 4.11.82)*. The objective was to overcome language barriers: "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".

4. The programme was staged over five and one half years (1982-1987) at an estimated cost of 16 Mecus. It comprised three phases: preparatory actions, basic and applied linguistic research, and stabilisation of the linguistic models and evaluation of the results.

The expected result of the programme was an operational prototype for a machine translation system in a limited field and for limited categories of text dealing with all official languages of the Community. This prototype would provide the basis for development on an industrial scale in the period following the programme.

5. Following the enlargement of the Community, the Council decided in 1986 to add the Spanish and Portuguese languages to the system. At the same time, the budget was increased by 4.5 Mecus and the duration was prolonged for two and one half years to 1989.

6. In 1988, Eurotra was reviewed by an Evaluation Committee of independent experts headed by Dr. A.E. Pannenborg (ex- Vice-Chairman of Philips). Their Final Report was transmitted to Council and Parliament *(COM (88) 270 final)*.

7. The Evaluation Committee reached the following conclusions :

   Eurotra had succeeded to generate substantial cooperation between Member States, in a field of growing importance. Eurotra had contributed substantially to strengthen the human resource base for research in computational linguistics and for the emerging language industry. Substantial progress was made towards achieving the scientific and technical objectives of the programme. However, efforts should be made to involve industry and the area suffered clearly from a lack of long term policy at the Community level.

8. Following this evaluation, the Council decided the transition of the Eurotra Programme to the third phase (Council Decision 88/445/EEC of 25.7.88), and in 1989, the Council decided to accord another 7 Mecus for the completion of a machine translation system of advanced design (Eurotra). *(Council Decision 89/410/EEC)*. This programme aimed to implement a system prototype, improve the Eurotra software, linguistic specifications and training methods, prepare for the industrial development of Eurotra, and set out evaluation objectives and criteria.

9. In March 1990, a second evaluation on the Eurotra Programme was carried out by independent experts, chaired by A. Danzin (ex-Director IRIA and Vice-President of Thomson-CSF) and transmitted to Council and Parliament *(COM (90) 236 final)*.

10. The Danzin report confirmed the findings of the Pannenborg evaluation of 1988. The original expectations of the mid 1980s appeared to be over-ambitious. Genuine progress had been made in the project since the Pannenborg evaluation, however.

11. The Danzin report stressed that, by the very fact of its existence, Eurotra has laid the foundations for a Community achievement in the field of language technologies. The report strongly reiterated the need for a long term Community strategy in the field.

12. Based on the findings of the Danzin report the final phase of the Eurotra programme was approved by Council (Council Decision 90/664/EEC of 26.11.90). This phase was allocated 10 Mecus and aimed at "the development of a high-level scientific prototype in the field of automatic translation". This phase ran from 1990 to 1992. In line with recommendations from the evaluators, shared-cost projects were launched, notably for system development, testing and research. New avenues of work were opened up on advanced system architectures, lexical and terminology resources, standards, education and training.

III. THE FINAL REVIEW: AIMS AND SCOPE

13. When Eurotra came to an end in December 1992, a final evaluation of Eurotra was subsequently carried out by a panel of independent experts *(Final Review Panel Report, February 1993)*. The evaluation took place in the first quarter of 1993 in accordance with Article 4 of the Council Decision 90/664/EEC of 26 November 1990. The panel was led by Prof. Brian Oakley, former Director of the UK Alvey research programme and retired chairman of Logica Cambridge.

14. The final evaluation had two complementary aims:

" to appreciate the achievements of the Eurotra programme in the years 1991-1992, or more precisely, in the period since the last evaluation (i.e. 1990)", and

" to appreciate the outcome of the programme which was conceived in the late seventies and had lasted ten years".

Furthermore, the final evaluation should look at the way in which the recommendations of the previous evaluators have been taken up both in the 1990-1992 phase of the Eurotra programme and in follow-up programmes.

IV. CONCLUSIONS AND RECOMMENDATIONS OF THE FINAL REVIEW

15. The Final Review Panel Report examines three distinct aspects of the Eurotra programme:

- its main research phases between 1983-1990;

- the final phase preparing the development of an operational system during 1990-1992;

- and future options for Community action in the area of language technology.

16. Main conclusion

The main conclusion of the Final Review Panel report is that the language area is of highest importance. The report states that:

"The problems of language are amongst the largest challenges facing the European Community. We are divided by our different languages and the resulting communication failures...The cost, both in direct economic terms and in loss of cohesion generated, is very heavy, especially compared with our main competitors in the USA and Japan...With Eurotra...a base has been constructed on which future programmes can build, in the struggle to bring technology to bear on the language problem of the Community".

The report makes a series of detailed conclusions on different aspects of the programme, as set out below:

17. Initial Objectives:

The evaluators conclude that the difficulty and scope of the initial objective of the Eurotra programme (especially that of developing a system for handling all languages in parallel) have not been adequately recognised at the start of the programme. A longer-term plan would have been needed, with a more pragmatic aim and with strong involvement of industry in the mainstream development work.

18. Technical considerations:

It is noted by the report that the programme revised its original objectives in line with the previous evaluations, and worked towards a scientific prototype, rather than an operational system. The programme has now achieved the scientific basis for industrial developments in this area.

The scientific quality of the work is felt to be an achievement of the programme and may well turn out to be influential in future systems designs in Europe and elsewhere. The research work on semantics, for example, is considered of very great importance

The technical outputs include notably the production of a language specification for each of the official Community languages. The Eurotra Reference Manual, together with the specifications, is a remarkable record of the outcome, containing a description of the largest linguistic effort ever made at a multilingual level.

The objectives of the work on the software development platform (ALEP) and on re-usable lexical resources are considered by the evaluators to be eminently desirable and practical. Further development and exploitation by the research community make excellent sense

It is felt that more attention should have been given to dictionary development, to user interfaces, to interactivity and to producing demonstrators at the appropriate stages.

The creation of a very coherent community of computational linguists in every country of the European Community is a very considerable achievement, where the credit lies largely with the Eurotra programme and the supportive governments.

The shift from contracts of association with Member States to cost-shared projects with growing involvement of industry is to be welcomed. On the whole, however industry participation remains low. The report notes that this situation is being remedied in the follow up programmes to Eurotra.

19. Results and exploitation:

The Final Review Panel points to several results of the Eurotra programme. At the technical level, the Reference Manual, the language specifications, software systems developed centrally or in individual centres, are prime examples.

But probably the most important output from the programme is the manpower trained in computational linguistics and in machine translation. Over 400 experts have been trained at some time on the programme. It is believed that they since have been involved in virtually every industrial natural language project currently being pursued in Europe today.

20. Achievement relative to the original objectives:

The original approach of tackling the broad problem of developing a machine translation system for all Community languages is felt to have been over-ambitious and far beyond what is presently technically possible. On the other hand, a more industry led approach, with a less ambitious objective, would not have made the same strategic impact on the science and technology base in Europe for natural language processing and machine translation. The panel notes that in its later phases the programme was correctly adjusted to more industrial aims.

The main achievement of the programme is certainly the development of a strong computational linguistics community in almost every Member State.

21. Organisation and Management:

The Eurotra management should have been strengthened from an early stage with the appropriate technical and industrial skills. The Contracts of Association with Member States seemed to add complexity to the management process.

22. Panel recommendations for the future:

The panel stressed that the need to bring computational linguistics to bear on language barriers is more urgent than ever. Technology can contribute significantly in the area of translation, albeit with human revision (post-editing). It is recommended that the efforts should be focussed on machine assisted translation and on aids for translators. At the same time, longer term research to fully improve automatic translation needs to be continued.

The immediate need for the future is:

- exploitation of the results of the Eurotra programme
- maintaining the Eurotra human network
- continuing research, whilst widening the technological approach.

The longer term need is for a broad based language technology programme, including:

- technology assessment
- lexical resources development
- an applications programme, aimed at markets where natural language processing can be most effective
- a Commission own projects scheme to meet its own internal needs
- enabling research, at the academic level, based on a multi-disciplinary approach, but bringing in industry wherever possible.
- a training programme
- increased international cooperation.

V. COMMISSION POSITION

23. The Commission has analysed the report and appreciates its positive outlook and the importance which it attaches to the area of language technology. It considers that the report gives a fair and balanced overview of the achievement and shortcomings of the Eurotra programme.

The Commission notes that the conditions for completing a machine translation system for all Community languages were very difficult in the early stages of the programme. However, the research climate for computational linguistics has evolved considerably since then, with more involvement of industry actors.

The Commission agrees that Community research programmes in this area should have a long-term outlook, whilst at the same time adressing shorter term objectives. Generic research is needed to improve the scientific and technical foundation of this complex area. It is also needed to encourage a growing corps of trained language engineers to develop in Europe. Industrial development work is needed to help develop tools, methods and resources, which can be put to use in a variety of different areas, where speech and language technologies are essential components.

The Commission recognises the immediate need to exploit the results of Eurotra, whether technical or in terms of human resources and will take the necessary actions.

24. The Commission points out that many of the recommendations of the evaluation panel, e.g. with respect to involvement of industry, orientation towards applications, dictionaries and other language resources, have been and are being taken up by the Linguistic Research and Engineering (LRE) Programme. LRE is one of the area's of the Telematics programme (Council Decision 91/353/EEC of 7.6.1991). The panel has recognised this evolution clearly in their report.

25. With reference to the longer-term recommendations, the Commission is preparing a broader-based research initiative within the fourth Framework Programme in the area of language engineering. Language engineering aims to stimulate improvements in the way information and communications systems handle spoken and written language. Automated tools, methods and resources will be developed and applied in multi-sectoral areas: document creation and management; multilingual computer-assisted services; telematic translation services; computer-assisted language learning and training and technology-mediated person-to-person communications. Work will also include the constitution of electronic language resources (dictionaries, terminologies, corpora etc) and general research into computational linguistics and language technology.

    The expertise and know-how gained in Eurotra will contribute in a beneficial way to many of the activities proposed. The recommendations concerning research work on specific technical topics will be taken on board as far as possible in this context.

26. The Commission will examine the appropriateness of an Own Projects Scheme, as recommended by the panel. Here, it is necessary to distinguish clearly between the internal requirements of the Commission's services and the general industrial research goals of the language engineering initiative.

Annexes

1. EUROTRA Final Review Panel Report

2. Opinion of the Eurotra Advisory Committee of 11 March 1993

OPINION OF THE ADVISORY COMMITTEE OF 11 MARCH 1993 ON THE EUROTRA EVALUATION REPORT - "FINAL PANEL REVIEW"

1. The Committee welcomes the basic content of the evaluation report, which was introduced by Mr. Oakley and discussed and which, **by stressing both shortcomings and achievements**, appears suitable as a basis for the preparation for future R&D programmes in the field of computational linguistics.

2. The Committee requests the Commission that it takes into account in its future proposals the need of a mechanism to ensure the coherence and continuity which have been provided so far by the Eurotra programme **and insists that the principle of equal treatment of all the Community official languages shall be strictly observed.**

3. The Committee endorses the recommendations of the report, in particular that future programmes should widen the scope of topics covered by the Eurotra programme, that special attention should be paid to training of manpower and to the development of the methodology of system performance measurement, and that the insight obtained on the reusability of resources should be followed up in view of future standardisation.

4. The Committee fully supports the recommendation of the panel that research programmes should be matched by adequately funded exploitation support programmes.

5. The Committee agrees that the Commission should see that future programmes are adequately staffed **from the very beginning** and that it should give serious consideration to the creation of an Agency.

6. **The Committee acknowledges the importance of increasing the availability of language resources and the need to promote international cooperation in computational linguistics in future programmes.**

EUROTRA

FINAL REVIEW PANEL

REPORT

FEBRUARY 1993

EUROTRA FINAL REVIEW PANEL REPORT

INDEX

5. ET6/7/9 Projects

1) The Move to Individual Projects
2) The ALEP Projects
3) Lexical Resources Study
4) ET6/7/9 (C&R)

6. ET10 Projects

1) Organisation
2) Selected Proposals
3) LRE Scheme
4) Cost-Shared Approach (C&R)

7. Outputs and Exploitation of the Programme

1) Outputs
2) Reference Manual
3) Language Specifications
4) Exploitable Computational Linguistic Property
5) Software Systems
6) Individual Centre Developments
7) Eurolang
8) Trained Manpower
9) Assistance for Exploitation
10) Comparison with Original Objectives (C&R)

8. The Individual Centres

9. Organisation and Management (C&R)

1) EUROTRA Management
2) An Agency?
3) Relations with the Centres
4) Staff

10. Governments' Rôle

1) Contracts of Association Mode
2) The Start-Up Period
3) Management Committees
4) Governments' Influence
5) Conclusions (C&R)

11. The Future (C&R)

1) Strategy for the Future
2) Programme Organisation
3) The Immediate Need
4) A Broad Based Language Technology Programme
5) International Collaboration
6) Cooperation with the Commission's Translation Service

Appendices

1. Terms of Reference
2. Membership
3. Visits Schedule
4. Questionnaire Analysis
5. The EUROTRA System Design
6. The Eurotra Liaison Group's Proposals
7.1 CD 82/752 - Annex on Objectives and Programme of Work
7.2 CD 90/664 - Annex on Objectives etc, for Final Years
8. Pannenborg Report Recommendations
9. Danzin Report Recommendations
10. Advances in Computer Technology
11. EUROTRA : Key Events
12. Glossary
13. Bibliography

EUROTRA : Final Review Panel Report

Preface

The problems of language are amongst the largest challenges facing the European Community. We are divided by our different languages and the resulting communication failures; we all pay the price and some countries suffer a real penalty behind their minority language barriers. The cost, both in direct economic terms and in the loss of cohesion generated, is very heavy, especially compared to our major competitors in the USA and Japan who have no such internal communication problems. But our languages are of great importance to all of us, epitomising as they do our past, our history, and our culture. So in a world where much of our differences and individuality has to be surrendered to the greater good of the emerging new Europe, where we have to improve our ability to communicate with each other, it is more than ever important to hold on to and enhance our languages, to cling on to that reminder of our roots in an increasingly shared culture. Technology can help to resolve this paradox.

It was brave of the Commission, the Parliament and the Council to undertake the EUROTRA programme, for it can be seen as a symbol of the Community's determination to improve its internal communication ability, without destroying the diversity and richness of our individual language cultures. If the explicit objective was not reached, the implicit objective of strengthening our ability to tackle language technology was most certainly achieved.

It was a pleasure to study the work of the EUROTRA programme, and especially to visit the many Centres throughout Europe where the work was carried out. The enthusiasm for their work through the study of their languages, and indeed the very existence of these Centres of Excellence in all the official languages of the Community, is a tribute to the foresight of the founders and supporters of the EUROTRA programme. A base has been constructed on which future programmes can build, in the struggle to bring technology to bear on the language problem of the Community.

EUROTRA Final Review Panel
February, 1993

The EUROTRA Programme Final Review Report

1. Executive Summary

1.1 Introduction

The EUROTRA programme has been running for more than ten years, with coordinated work going on in every country of the Community. Some 16 Centres were established or enhanced, and at its peak there were some 200 research workers. largely computational linguists, working in those Centres to a common programme tackling the nine official languages and 72 language pairs of the Community. Until the final years of the programme all of these workers were academics, or working in non-profit institutions.

1.2 The Main Programme (1985 - 1990)

1.2.1 *The Key Objective* of the programme was to carry out a research and development programme which would prepare the way for the "creation of a European machine translation system of advanced design". In a simplistic, direct sense this was not achieved, for no prototype system emerged from the programme. But in an indirect sense the way was prepared to enable translation aids and systems of all types to be built, both through the training of some hundreds of experts who will form the basis on which industry and academia can build, and through the establishment of a body of knowledge about how to build a machine translation system, and especially about the grammars and language specifications of the nine official Community languages.

1.2.2 *The Problems.* Language is such a complex and still difficult subject to reduce to a simple set of rules which a computer can execute that unaided machine translation is not achievable, except in special, limited, circumstances. Until the problem of extending our computer comprehension of the semantics not just of a sentence but a whole paragraph or more is solved, there is no prospect of achieving good quality translation without human intervention. Though this was certainly recognised by the EUROTRA community, the programme chose to tackle the general problem. An alternative, which perhaps an industrially dominated team would have espoused, would have been to start from the human translator and to take a whole system approach to what was needed to achieve improved efficiency. Out of this would have emerged a Machine Assisted Human Translation system prototype, perhaps for a limited domain, which could readily have been passed on to industrial production. However, this far less ambitious objective would not have served so well to build up the computational linguistic community.

1.2.3 In terms of the approach adopted, two other problems were encountered:

1) Because the programme was dominated by academic linguists interested particularly in the problems of grammar, the work programme became unbalanced to the detriment of other aspects, such as the crucial dictionaries which received less attention than that aspect deserves.

2) Perhaps for the same reason, the running and testing of the system was somewhat neglected, with a software system becoming available too late and an architecture being chosen that was not efficient. Systematic testing and performance measurement at run time seems to have been given too little attention.

1.2.4 *The Achievements.* The EUROTRA System Reference Manual and, especially, the Language Specifications are a monument to the programme, and, if made widely available, will prove valuable to workers on Natural Language Processing systems, both in industry and the academic world, for many years to come, despite the penalty of being tied to a particular EUROTRA system architecture and formalism.

1.2.5 The trained manpower is perhaps the most lasting legacy of the programme. As well as the, at least, 380 people who worked on the programme, various courses were established as an indirect result of the work of the staff of the Centres. There is unlikely to be any significant project in the Natural Language Processing field in the Community nations for some years to come that does not employ people who received their advanced training as a result of the programme. This can already be seen in major programmes like Eurolang, where some of the EUROTRA Centres are directly involved, and many of the staff received their training on EUROTRA.

1.2.6 Perhaps in the long run, it will come to be seen that the most important legacy of the programme is that every country of the Community has been awakened to the importance of their language and the potential for language technology. EUROTRA has created a core expertise in computational linguistics in every official Community language, and has generated a human network of experts who work together as a single team across Europe. Europe has taken its place in the forefront of language technology, and the coherence of our community of experts is the "envy of less favoured lands".

1.3 The Final Phase (1990 - 1992)

1.3.1 The Panel was asked especially to examine the final few years of the programme, 1990 - 1992, when somewhat different methods of working were introduced, in parallel with the continuation, on a lesser scale, of the coordinated Centres approach. The objective was to open the subject up,

to introduce new blood and ideas, and to introduce a more competitive approach through cost-shared projects. New blood was certainly brought into the programme, though the number of industrialists involved remains disappointingly small. Many of the topics, the absence of which from the main programme has been criticised by this and the earlier Review Panels, were picked up in the cost-shared projects of ET10, and the subsequent LRE programme. However, these projects are too few in number, and too small in size and duration.

1.3.2 One legacy from the Final Phase is the ALEP system, a toolkit or software framework for Natural Language Processing research. This employs a fully declarative formalism, and certainly appears a significant advance on the EUROTRA work, both in design and run-time performance, as is to be expected for a system laid down some years after the EUROTRA design was frozen and in the light of the EUROTRA experience. Perhaps significantly, the work is centred in industry, though some EUROTRA Centres are involved. The final product is likely to be of lasting value to both the academic and industrial research community. It is unfortunate, though perhaps inevitable, that the new has tended to prematurely oust the old. By the knowledge of its development, ALEP has tended to inhibit exploitation of the EUROTRA system work.

1.4 The Future

1.4.1 It would be a tragedy if the expertise in Natural Language Processing that has been built up were to be thrown away. In 1991 the Danzin Committee, in a major study of the part the Commission should play in Language Engineering, made a set of recommendations for future work. Following that Report, in Chapter 11 of this Report the Panel makes various proposals for a future programme. One of the lessons to emerge from EUROTRA is that it is important to keep a balance between competitive individual projects and retaining cooperation and coherence in the, inevitably, thin and scattered community, in what is still a somewhat immature subject. Natural Language Processing is a subject where comparison of the languages can only serve to improve the work.

1.4.2 In the short term the EUROTRA work and experience should not be allowed to fade away. Projects are required to update and transfer the grammar and language specifications to the emerging ALEP system. There are various possibilities for exploitation in industry and commerce of the work of EUROTRA, and a scheme to assist this should be deployed without delay, for the usual exploitation gap in Europe will develop if assistance is not provided to take the work through to the stage where particular applications can be demonstrated.

1.4.3 For the future (see Chapter 11) the Panel propose a balanced and enlarging programme tackling in a multi-disciplinary way the real needs of the Community in the Natural Language Processing field. Steps should be

taken to involve industry as far as possible, and to strengthen the industrial NLP community, in the same way that EUROTRA did for the academic community. A major drive on creating dictionaries for use in a variety of Natural Language Processing applications is proposed. An Applications programme is proposed, aimed at very practical problems but in a way that will pull through to use the advanced technology. Applicable and Enabling Research should continue to be supported, as in EUROTRA, for without a strong basis in research, progress will falter in this difficult but all important field of language engineering.

1.4.4 *Relations with the Commission's Own Work.* One other issue deserves particular attention in the future. The Commission is a major user of MT, perhaps the major potential customer in the world. It is strange that the EUROTRA programme was conducted with virtually no interaction with the Commission's own work with Systran for their own translation. This is especially unfortunate, remembering that the translation service was a sponsor of the EUROTRA programme in the early years, and provided staff to help run it. A very different programme would have emerged had a careful study of the needs of the Commission's own translators been undertaken early in the planning of the programme. In view of the richness of the Systran dictionaries, it is unfortunate that there was so little cooperation over this aspect.

1.4.5 It is important that in any future programme of MT work, close cooperation with the translation service should be established. In particular, in the proposals for a major programme to build up lexical resources, the Commission should be a major partner and participant, in the light of the need to re-engineer the Systran system.

1.5 Conclusions and Recommendations - Summary

For ease of reference, the conclusions and recommendations scattered throughout the text of the report are gathered together below, with a reference to where the topic is treated in the main text.

INITIAL OBJECTIVES

1. Timescale

A much longer programme was required than the five and a half years originally planned. The programme ran for ten years without producing a complete system that could be fully tested and evaluated (3.9.1).

2. Had a ten year programme been authorised initially a much better programme might well have resulted (3.9.2).

R1: Where it is recognised that the difficulty and scope of a programme will require a long run, the Commission and Council should face up to this

from the initial decision, of course with suitable review and break-points built in (3.9.2).

3. *Work Plan*

It was unfortunate that no charted workplan could be drawn up and followed throughout the programme. Had a ten year programme been foreseen, one model for a workplan might have envisaged, say, seven years of research followed by three years of prototype development (3.9.1, 3.9.2).

4. *Languages*

The way in which the programme was conducted with work in parallel on all nine Community official languages and all 72 language pairs was wasteful and inefficient (3.9.6).

5. *Industrial Involvement*

It was unfortunate that there was no industrial involvement in the mainstream programme. Steps should have been taken to involve industrialists with the Centre teams (3.9.7, 4.2.1, 4.2.3, 4.2.4).

6. *Much of the prototype software should have been entrusted to and created by industry (3.9.7).*

7. *The Contracts of Association Approach*

On balance, the Contracts of Association approach to the programme was an unsatisfactory way of organising it. Normal "ESPRIT" type funding would have provided stronger central control, run in a centrally coordinated way to create the close-knit community which was certainly achieved by the approach adopted. (3.9.9, 3.9.10).

8. *The Move to Individual Projects*

This move, in 1989, opened up the programme to new participants and ideas and is to be welcomed but steps need to be taken to keep the cooperation and coordination of the projects and teams (4.5.6, 5.1.1, 5.4.1).

TECHNICAL CONSIDERATIONS

9. *Quantitative Measurement*

It is a failure of the programme that so little, virtually no quantitative measurement of performance was undertaken (4.2.5, 4.4.4).

10. Narrow Domain

Despite the encouragement to do so in the original Council Decision, little attempt was made in the programme to restrict the work to a narrow domain or market except for the dictionaries (4.2.7, 4.6.11 - 4.6.14).

11. Operational MT Prototype System

The programme failed to achieve an operational MT prototype system, but, in the words of the Danzin Panel, worked towards a "scientific prototype" (4.2.9).

12. The Computational Linguistic Community

The creation of a very coherent community of computational linguists with workers in every country of the Community, is a very considerable achievement where the credit lies very largely with the EUROTRA programme and the supportive governments (4.5.1, 4.5.2).

R2: The human network of computational linguists built up across Europe by the EUROTRA programme should not be allowed to decay (4.5.2).

13. Language Specifications

One of the achievements of the programme has been to produce a language specification for each of the nine official languages (4.5.4, 7.3).

14. Clamp on Publication

In retrospect it is easy to see that it was a mistake to introduce a clamp on publication in the early years of the programme (4.5.8).

15. The Reference Manual

The Reference Manual together with the Language Specifications, is a remarkable record of the outcome, containing a description of the largest linguistic effort ever made at a multilingual level (4.3.2, 4.6.1, 7.2).

R3: The Reference Manual and Language Specifications should be made widely available (4.6.1).

16. The E-Framework System

The E-Framework System architecture makes it difficult to relate the research to work elsewhere based on more conventional approaches (4.6.2).

17. *Dictionaries*

In view of the economic importance of the dictionaries in a practical system development it was unfortunate that so relatively little attention was paid to them in the balance of the programme (4.6.3).

18. *It is particularly unfortunate that so little attention was given to the portability aspects of dictionary design in mainstream EUROTRA (4.2.8, 4.6.3).*

R4: The task of building up machine tractable dictionaries for a multilingual community is one that requires encouragement and support from the Commission. After a suitable study phase, the Commission should launch a major project to create knowledge bases containing lexical, semantic and syntactic information usable in natural language processing systems for the European languages and language pairs (4.6.3).

19 *Semantics*

The importance of research on meaning, on the interlingua approach and the semantics needed, is very great indeed (4.6.5).

R5: Bringing to bear semantic information from a wide part of the text, the use of world knowledge, and intensified research on lexical meaning should be priority topics in future Commission programmes of IT research, both in natural language research and in wider IT research programmes such as ESPRIT. These programmes should also take into consideration the rôle of the language-independent knowledge bases and interlingua systems.

20 *The Statistical Method*

It is understandable that the approach does not feature in the mainstream programme, but it is good to see it feature in an ET10 project (4.6.6, 6.2.3).

R6: Statistical methods, as a complement to rule-based solutions and as a method for human aided knowledge retrieval from parallel corpora and monolingual corpora and, furthermore, neural network solutions should be priority topics in future programmes (4.6.6).

21 *The User Centred System Approach*

It is to be regretted that no real attention seems to have been paid to the user of the system that would ultimately result from the work, even considering that the prototype was conceived as a batch system (4.6.7).

R7: Whole system design and the User Centred Approach should be priority topics in future programmes. This implies efforts to bridge the gap between linguistics and computer science (4.6.7).

22 *Interaction*

It was a mistake, even if an understandable one, not to make the investigation of interaction a part of the programme (4.6.8).

R8: Interaction and learning, and automated inference systems making human interaction more effective and less repetitive, should be priority topics in future programmes (4.6.8).

R9: Future programmes should particularly embrace promising topics and techniques that have been under-represented in the EUROTRA programme (4.6.9, 4.6.10).

23 *Demonstrators*

It is unfortunate that the Centres were not always encouraged or prepared to produce demonstrators at all appropriate stages, and especially to produce demonstration systems at the end of the programme (4.6.15).

R10: The Commission should take continuing steps to develop the methodology and practice of MT system performance measurement (4.6.16, 4.6.17).

24 *Scientific Quality of the Work*

Some aspects of the mainstream EUROTRA work may well turn out to be influential in future systems designs in Europe and elsewhere (4.6.18).

25 *The evidence for a significant increase in the number of scientific papers stemming from European computational linguists involved in EUROTRA demonstrates the improved presence of European workers on the international scene. It is to be welcomed and is an achievement of the programme (4.6.19).*

THE ET6 AND ET7 STUDIES

26 *Fully Funded Studies*

The objectives of the ALEP and Reuse of Lexical Resources Studies seem eminently desirable and practical, though one must ask why these problems had not been addressed in the main programme in the preceding eight years (5.4.1).

27 *ALEP*

It is excellent that the production of a linguistic software development and testing environment was initiated in 1989 and is now proceeding (5.4.2).

28 *The impact of the ALEP work on the mainstream EUROTRA work has been*

unfortunate in the short run (5.4.2).

29 *The plans for the further development and use of ALEP by the research community make excellent sense (5.4.2 - 5.4.4).*

R11: The Commission should continue to develop and maintain the ALEP system, making it freely available for research purposes (5.4.2 - 5.4.4).

30 *The Reusability of Lexical and Terminological Resources Study*

This study is an important topic and is much to be welcomed. It is only to be regretted that it was not started early in the EUROTRA programme so that the lessons could have been applied to the work of the main programme. and the proposals followed up in that work (5.4.5).

R12: The Commission should follow up the ET7 Reusable Lexical Resources recommendations in its research programmes, standardisation and training activities (4.5.9, 11.5).

ET10 AND LRE COST-SHARED PROJECTS

31 *The shift to cost-shared projects is to be welcomed for research projects, though it would not be appropriate for large development projects. But for a subject like MT that requires a coherent attack on standards formalisms, interfaces, etc, it is desirable to take special measures to ensure that "continuity, completeness and coherence" is retained across the teams (6.4.1 - 6.4.4).*

32 *ET10 Projects*

Some of the projects fill holes in the scope of the mainstream EUROTRA work (6.2.1 - 6.2.2).

33 *The Statistical Complements project, led by IBM, is particularly to be welcomed (6.2.3).*

34 *The participation by industry remains disappointing in number (6.2.3).*

35 *The LRE EAGLES project*

This project, as a drive to coordinate the European language industry and research community, is to be welcomed. It is encouraging that so many of the major Language Engineering projects in Europe are represented on the Management Board (6.3.4 - 6.3.6).

OUTPUTS AND EXPLOITATION

36 *The Reference Manual*

The Reference Manual is very much a detailed working reference manual rather than a polished text book, but it is likely to be referred to throughout the world community of computational linguists and so is a lasting monument to the programme (4.3.2, 4.6.1, 7.2).

R13: The Commission should consider whether it is practical to prepare and issue an updated version of the Reference Manual, for this would certainly be desirable (7.2.3).

37 *The Language Specifications*

These are of outstanding value to any language technologist interested in a specific language, whether for monolingual or multilingual work (7.3).

38 *Exploitable Computational Linguistic Property*

Because the Reference Manual and Language Specifications are so intimately tied to a particular system architecture and formalism, they will date rapidly in some respects, and so are of little direct economic value, but are the main intellectual output from the programme and are valuable in that respect (7.4).

39 *Software Systems*

The ALEP1 tool set has the potential to be of value to research laboratories, and to industrial teams throughout the world, who might wish to use it to assist in their system developments (7.5).

40 *Individual Centres' Systems*

Where individual Centres have developed more run-time efficient systems, based on the EUROTRA work, these may form the basis for commercial products for niche markets. If this arises it will represent a very real exploitation of the work and expertise in the programme (7.6).

41 *Eurolang*

It is excellent to see the involvement of EUROTRA staff and some Centres in the work of Eurolang, and the use of the relevant Language Specifications. It demonstrates the value of the EUROTRA programme in developing the supply of the skills in this field in Europe (7.7.1 - 7.7.3).

42 *However, it is disappointing to see so little sign of Eurolang basing their work around the mainstream EUROTRA software, architecture, and formalism developments (7.7.3).*

43 *Trained Manpower*

Probably the most important output from the programme is the manpower trained in the techniques of computational linguists, and the particular problems of Machine Translation (7.8).

44 *It is believed that people who had worked in the EUROTRA Centres, or had been trained on their courses, have been involved in virtually every industrial Natural Language project current in Europe today (7.8.4).*

R14: It is highly desirable that the supply of trained manpower in computational linguistics in Europe be maintained and enhanced. In its future support for Linguistic Engineering, the Commission should take steps to monitor the supply of trained manpower, and to assist the training programmes should that prove necessary (7.8).

45 *Assistance for Exploitation*

Several of the EUROTRA Centres are struggling to find funding to assist in the problems of exploiting their EUROTRA work for specific market niches (7.9.1 - 7.9.2).

R15: The Commission should ensure that all its research programmes, like EUROTRA, are matched by exploitation support programmes with adequate funds (7.9.1 - 7.9.2).

THE ACHIEVEMENT RELATIVE TO THE ORIGINAL OBJECTIVES

46 *The programme has not achieved the key objective of an operational system prototype directed to the "creation of a machine translation systems of advanced design" (7.10.1).*

47 *The state of the art today, and probably for years to come, makes it a much more feasible proposition to design useful systems for limited domains (4.6.11 - 4.6.14, 7.10.2).*

48 *The programme ran for nearly twice as long as originally planned at three times the cost. However, these figures are misleading and the actual increases are not as significant as they appear at first sight (7.10.3).*

49 *The indirect objective of developing a stronger computational linguistic community in the European Community was certainly achieved (7.10.4).*

50 *Over the ten or more years of EUROTRA progress has been made in machine translation, inside and outside the programme. It would be desirable to set the work of EUROTRA into the wider perspective (7.10.5).*

R16: The Commission should establish a study to document what progress has

been achieved in MT worldwide over the period covered by EUROTRA (7.10.5).

ORGANISATION AND MANAGEMENT

51 *It is difficult for an individual to lead such a complex and technical programme without adequate experience and status. The Commission should have recruited a technical leader for the programme (9.1.6).*

R17: Where a programme is intended to lead to industrial exploitation, leadership should be placed with an individual with appropriate qualifications, reputation, and, if possible, industrial experience (9.1.7).

52 *There are strong arguments for creating an Agency for running such programmes, but there are also counter-arguments. Much depends upon the particular circumstances and nature of the programme (9.2).*

R18: Serious consideration should always be given to the creation of an Agency, whenever the establishment of a programme on the scale and with the complexity of EUROTRA is being planned (9.2).

53 *Initially, the Commission's staff were grossly overloaded (9.4).*

R19: In establishing a future programme of the cost and complexity of EUROTRA, the Commission should ensure that it is adequately staffed (9.4).

GOVERNMENTS' RÔLE

54 *Had the normal competitive cost-shared projects been established as the way of working from the beginning of the programme, the participation would probably have been concentrated in a few countries, and the Centres would not have been established in those countries where direct government intervention was required in order to get them set up. Thereafter, it was right to shift to a more open, more competitive approach (10.4, 10.5).*

55 *After the start-up phase, there seems little advantage in the added complexity of the Contracts of Association process (10.5.2).*

THE FUTURE

56 *The Community's Need*

It was imaginative and, indeed, brave of the Commission to propose the EUROTRA programme, and for the Council to approve it. It could be said that the need for the developments of computational linguistics to be brought to bear on the language barrier of the Community is more urgent than ever (11.1.1).

57 *Machine Translation Prospects*

Human post-editing will remain essential, for general text Machine Translation work, if good quality translation is required, for many years to come. However, it would be wrong to deduce from this that technology cannot contribute significantly to the language problems of the Community (11.1.4).

R20: In the field of Machine Translation the Commission should concentrate on Machine Assisted Human Translation, on aids to the translator, while continuing to support longer term research that will improve automatic translation (11.1.4).

58 *Programme Organisation*

The reversion to a carefully drawn up workplan after an open call for proposals, in the last phase of the EUROTRA programme and in the subsequent LRE programme, seemed entirely appropriate at that time and for much of any future programme. But it is essential to back this competitive approach by some mechanism to coordinate and pull together the industrial and academic community working in the field. ELSNET may form a basis for this, and certainly the EAGLES standardisation cooperation will also help (11.2.1).

59 *Programme Management*

An Agency to run a future programme should be considered. The important point is to place the leadership on someone who has the appropriate experience and motivation, backed if necessary by appropriate technical experts (11.2.2).

60 *Scale of a Future Programme*

The language problem is one of the most important facing the Community, both for economic and social reasons. The investment that the Community should be making in language technology should be commensurate to the impact it could make on the language challenge. But a programme should steadily ramp up in its investment, as the qualified resources to tackle it become available (11.1.7 - 11.1.8).

61 *Industrial Participation*

It must be an objective of the next phase of the Commission's Language Engineering Programme to repeat the success in stimulating work in the academic community, but now also directed to the industrial community (11.1.6).

62 *The Immediate Need (11.3)*

A continuation of the current Language Technology work with the following features:

1) Exploitation assistance for the EUROTRA work

2) Exploitation of the EUROTRA grammars, etc, via the ALEP system including updating the Reference Manual and Language Specifications

3) Maintaining the EUROTRA human network, and widening it as appropriate

4) Continuing research, widening the technological approach.

63 *A Broad Based Language Technology Programme (11.4)*

A broad based, inter-disciplinary approach, firmly based in practical solutions, should address:

1) A Technology Assessment Programme to assess opportunities for Language Engineering in industry and commerce, and to create awareness of them

2) Lexical Resources Development. A major attack on the task of building dictionaries and terminological database for all the official Community languages and language pairs

3) An Applications Programme, aimed at markets where NLP technology can be most effective.

4) A Commission Projects Scheme based on Language Engineering projects to meet the needs of the Commission in its own work

5) Enabling Research. A continuation of research in the academic world, based on a multi-disciplinary approach, but bringing in industry wherever possible

6) A Training Programme.

64 *International Collaboration*

There was little encouragement for interaction with the outside world until a late stage in the EUROTRA programme.

R21: International collaboration should be encouraged wherever that is appropriate, in particular with the centres of expertise in the field in the USA and Japan (4.5.9, 11.5).

65 *Cooperation with the Commission's Translation Service*

It is unfortunate that there was so little cooperation with the Commission's translation service, and especially with the work of building up the dictionaries (4.3.10, 4.6.3, 11.1.4, 11.6).

R22: In any future work in MT, the Commission should ensure that there is close cooperation with the actual work and needs of its own translation service. The opportunity will arise, due to the need to re-engineer its Systran system. The lexical resources programme, as proposed in this Report, is an ideal vehicle for close cooperation (11.6).

2. Introduction

2.1 Terms of Reference

2.1.1 The EUROTRA Final Evaluation Panel was established in response to the request in the Council Decision which authorised the final two years of the programme. The main objectives of the EUROTRA Evaluation, as called for in the Panel's Terms of Reference (see Appendix 1) are:

1) "to appreciate the achievements of the EUROTRA programme in the years 1991 - 1992, or more precisely, in the period since the last evaluation (ie 1990)"; and

2) "to appreciate the outcome of a programme (change of state) which was conceived in the late seventies and has lasted ten years. This applies both to the scientific and technical and to the policy aspects".

(Ref Council Decision 90/664, dated 26th November 1990.)

2.1.2 Furthermore the evaluation should "appreciate the way in which the recommendations of the 1987 Pannenborg and 1990 Danzin reports have been taken into account both in the 1991 - 1992 programme and in the follow-up programmes (LRE - Linguistics Research and Engineering - in Framework Programme 3, and in the preparation of a strategic programme in Framework Programme 4)".

2.1.3 The European Commission has been funding Machine Translation R&D work in a number of European Centres over the last 10 years within the EUROTRA framework. This programme is now complete and following the publishing of a Council Decision in the Official Journal to review the work, a Panel has been formed to do this. Their terms of reference are detailed in Appendix 1, but are interpreted briefly below:

2.1.4 *Impact.* EUROTRA has lasted ten years with an overall CEC budget of 37.5 million ecu. It has made some impact on policies and activities in computational linguistics both at Community and national level in the EC and outside. The evaluation should therefore compare the situation of Machine Translation (MT) and Natural Language Processing (NLP) related policies and activities of, say, 1980, and the end of 1992.

2.1.5 *Awareness.* The evaluation should assess to what extent discussions concerning EUROTRA have contributed to the increased awareness of the policy and decision makers, both at the Community and national level, of the importance of language and language engineering work and to the definition of the rôle of the EC in this field, especially with a view to the

future.

2.1.6 *Cohesion.* When EUROTRA started, the level of activity in MT was very low indeed in Europe, and activity in NLP in general was much lower than now and very unevenly distributed in the Member States. The evaluation should assess to what extent EUROTRA has contributed, in the computational linguistics field, to the general increase, balance of activities and expertise, and international cooperation within the EC. The evaluation should also assess how far the programme has created the possibility of starting new activities outside the EUROTRA context. In particular, the two aspects need to be addressed:

- at the academic level, the creation and continuation of institutes as a result of the EUROTRA work, their scientific status, reputation and ability to survive;

- at the industrial level, the influence of EUROTRA on industrially and commercially oriented projects.

2.1.7 *Scientific and Technical Impact.* The evaluation should examine to which extent research activities in EUROTRA and outside have interacted and influenced each other and the impact EUROTRA had on computational linguistics in general.

2.2 Membership

The Panel was established in the Summer of 1992. It consists of Dr Brian Oakley (Logica UK) as Chairman, together with Prof Sture Allén (Göteborg University, Permanent Secretary of Swedish Academy), Dr Alessandro Osnaghi (Olivetti SpA, Italy), Dr Jean Rohmer (Group Bull, France) and Professor Dr Hartwig Steusloff (ILTB Fraunhofer Institute, Germany). Prof Allén and Dr Steusloff served on the earlier Pannenborg and Danzin reviews of EUROTRA, thus providing an element of continuity. The Panel was supported by Dr Iain Rae of Logica, who also provided logistical support. (For outlines of members see Appendix 2.)

2.3 Mode of Working

2.3.1 The Panel met five times, receiving documents and presentations from the Commission's EUROTRA staff. Meetings were held with a group of experts involved with the programme, with the Director of DG XIII-E, and received demonstrations of the system. Individual members met with the EUROTRA Liaison Group. (For details of the visits and meetings see Appendix 12.)

2.3.2 Visits were made by two or more members of the Panel to virtually every EUROTRA team. Meetings were also held with industrial teams involved with the programme, IBM, PE International, BIM, SRI (Cambridge), and

with the Eurolang company of SITE. A meeting was held with a group of international experts in the field to examine the scientific quality and impact of the programme.

2.3.3 Questionnaires were sent to all the EUROTRA Centres, to firms in Europe to assess the industrial aspect of the work, and to independent scientists to assess the scientific impact.

2.3.4 Because this is the final evaluation of an important programme the Panel decided to examine the whole programme, though giving particular attention to the last few years and especially to the new mode of working introduced in that period.

3. History and Objectives of the Programme

3.1 Origins of the Programme

3.1.1 The European Commission has, perhaps, the most difficult and certainly the most extensive and costly translation workload in the world. By the second half of the 1970s there were six and then seven Community official languages, which means that documents might, indeed would, have to be translated between every one of 42 language pairs. Today the Commission employs over 1,000 professional translators, and the translation workload costs the European taxpayer over 150 Mecu each year. In 1976 the Commission acquired the machine translation system Systran, which had been developed in the 1960s in California by Peter Toma for the US Air Force. The decision to purchase a system of non-European origin caused dissention in the Community for there were at that time several research centres in Europe working on machine translation systems of more advanced design (notably in Grenoble and Saarbrücken). The development of further language pairs after the initial English to French provided by Toma was proving more difficult than had been expected. So in 1978 the Commission started preparatory work for a European R&D programme. A group of representatives from some thirty European universities and research centres was called together by the Commission. They named the programme EUROTRA and formed the EUROTRA Coordination Group. ISSCO at Geneva under Professor Maghi King was given a small contract to coordinate the work. This and other small preparatory contracts were funded under the Commission's Multi Lingual Action Plan programme.

3.1.2 It took the Commission five years before the approval of the European Council and Parliament was obtained for the programme. Finally it was authorised by the Council Decision 82/752 of 4th November 1982.

3.2 The Preparatory Phase of the Programme, 1979 - 1985

3.2.1 The Council Decision envisaged a first preparatory phase of two years, followed by a second main phase of "basic and applied linguistic research", with a final phase of 18 months for "stabilisation of the linguistic models and evaluation of results". In practice the programme is usually considered as having three phases, though of much longer duration. The first preparatory phase was concerned with setting up the programme and its organisation, determining the participating Centres with the national governments, and agreeing the Contracts of Association with them. In fact the first contract was signed in June 1984, but it was not until the autumn of 1985 that a sufficient number of contracts was signed to allow their implementation and the second main phase to start. Two were not signed until 1987.

3.2.2 In practice considerable work was carried out in this prolonged preparatory phase. The specifications and programme of work drawn up by the EUROTRA Coordination Group, covering both the linguistic and computational aspects, were distilled into the Reference Manual, of which the first version was released in 1979 at the first annual workshop. As well as the linguistic specifications drawn together in the Reference Manual, software specifications and a prototype implementation were drawn up during this phase. So, though the work was proceeding on temporary study contracts, more than purely preparatory work was carried out in this first phase. During this phase the proposal was made to base the EUROTRA development on the GETA formalism from Grenoble, under Professor B Vauquois. But in about 1980 the decision was taken to adopt a more modern basis for the programme, stemming from USA developments.

3.3 The Main Phases 1985 - 1988, 1988 - 1990

3.3.1 The programme funding was stepped up in 1985 as the Contracts of Association came into being, providing funding for the Centres from both the Commission and their national governments. In 1985 the CAT formalism was developed, based on the Term Unification, PATRII work from the West Coast of the USA. Then in 1987 the decision was taken to freeze the EUROTRA ETS formalism, though other formalisms were worked on as "sidelines" such as CLG (1990), MiMo (1990), MiMo 2 (1991) and CAT 2 (1991). And in the final phase of the programme the ALEP formalism was developed in parallel with ETS, so in a sense the programme consisted of a set of developments proceeding in parallel.

3.3.2 With the accession of Spain and Portugal to the Community on 1st January 1987, Council Decision 86/591 of 26th November 1986 extended EUROTRA to these two countries. It also increased the number of languages to be handled from seven to nine, the number of language pairs from 42 to 72. It extended the programme from five and a half years to seven, increased the funding, and the number of staff authorised to run the programme. So the programme was now authorised to run until the end of 1989, subject to review at the conclusion of each phase. The Second Framework Programme for Community R&D was authorised nine months later on 28th September 1987 (Council Decision 87/516). This made reference to the completion of the multilingual prototype machine translation system by 1990, and to the support to the industrial development of a machine translation system. Rather surprisingly, this language technology activity was classified not as a mainstream research and technological development work but as part of the activity to support "Dissemination and Utilisation of Science and Technology Research Results".

3.3.3 Under the 1982 Council Decision the Advisory Committee on Programme Management established to monitor EUROTRA development was required to submit to the Commission and the Council a detailed report

at the end of each phase. Council Decision 86/591 amended this to require the Council to authorise the decision to proceed to the next phase on the basis of the report from the Commission. This led to the report prepared by a committee of independent experts under the chairmanship of Dr A E Pannenborg. This was carried out in the spring and summer of 1987 and delivered later that year. While critical of many aspects of the programme the review gave general encouragement to its continuation, and in particular to prevent any interruption of funding between phases two and three. The Pannenborg report led to the Council Decision 88/445 of 25th July 1988 authorising the EUROTRA programme to move on to the third phase on 1st July 1988. This Decision did not authorise any further funds since the programme was now seen as part of the Second Framework Programme. However a further Council Decision 89/410 of 20th June 1989 authorised increased funds for the third, two year, Final or Transition phase from July 1988 until 30th June 1990. The Decision called for a review by independent experts.

3.3.4 In 1987 the management and technical direction of the programme was placed firmly in the hands of the Commission's DG XIII staff, with the termination of the ISSCO contract to provide the technical leadership, apparently because of political pressure from some member states. The participation of Switzerland in EUROTRA was discussed but did not take place. While the linguistic specification work was very much decentralised to each national language group, some linguistic research of general interest, which constituted the basis for the work of the national teams, was carried out by the members of the national teams, but in a strongly coordinated way by the "central teams" through special clauses of the Contracts of Association. The linguistic specifications were frozen at the end of 1990 in Reference Manual 7.0. This decentralised approach was replaced for the software construction by work by and directly for the Commission's team in Luxembourg. The Commission's staff was augmented for this work by staff seconded from the Luxembourg team and by staff hired from software companies to work at the Commission.

3.3.5 As called for in the Council Decision of June 1989 a further assessment of the EUROTRA programme was begun in October 1989 by a panel of independent experts under Mr A Danzin. The final report of this assessment was delivered in March 1990. As well as assessing the quality and suitability for industrial development of the work to date, the Panel was asked to make proposals for a specific programme for 1990 and 1991, and outline a strategic programme for the field of language engineering for the 1990 - 1994 Third Framework Programme. The recommendations in this report led to the final fourth phase of EUROTRA from 1991 - 1992, as well as to the LRE programme.

3.4 The Final Phase, 1991 - 1992

3.4.1 *Mainstream EUROTRA.* The final phase of EUROTRA, sometimes called EUROTRA II, was a hybrid between traditional Community cost-shared contracts and the EUROTRA Contracts of Association. Now that EUROTRA was to be funded from the Second Framework Programme, rather than under a special Council Decision, there was pressure to conform with the normal cost-shared contract procedures of Community R&D. But, quite separately, the Danzin Panel had recommended that while the EUROTRA mechanism should be continued in the interim, albeit with revised objectives, the Panel also recommended that different mechanisms for funding language technology should be pioneered during what was seen as an interim two years before the new Framework Programme in 1993. The Danzin Report explicitly stated that the original objective of obtaining an operational prototype was unrealistic, and that the development stage was still far off. These recommendations led to a continuation of the EUROTRA Contracts of Association approach, but at half funding for the final two years 1991 - 1992. In parallel, cost-shared and fully funded projects were introduced. The mainstream work was largely concerned with the completion of the language specifications; further work on the dictionaries; the development of various practical versions of the ETS and the sideline formalisms, such as CAT 2; and the development of software for EUROTRA demonstration systems. Final activity reports were produced covering the main work in the period up to the end of 1990. Final "Implementation" reports are due in early 1993.

3.4.2 The Third Framework Programme was authorised by Council Decision 90/221 of 23rd April 1990 for the period 1990 to 1994. The Decision calls for a sustained effort in language research and engineering, and the encouragement of the development of operational systems linked to information and communication systems. This was followed by Council Decision 90/664 of 26th November 1990 concerned with the development of an operational EUROTRA system. This authorised a programme for two years from 26th November 1990. (Apparently, formally this Decision was adopted under the earlier Second Framework Programme.)

3.4.3 With the reduction in funding to the Centres, the growth of the cost-shared projects, and the need to secure the future felt by the teams, the last two years were largely a consolidation period, with various teams finding variants of the main ETS formalism to enable their work to be applied in practical systems. From January 1990 a P-E International team in Luxembourg were contracted to develop, maintain and distribute the EUROTRA software.

3.4.4 *The ET6/7 Projects.* Even before the final phase of EUROTRA the Commission had been making plans for preparatory work geared to post-EUROTRA NLP and MT activities. In April 1989 there was an invitation to show interest in fully funded studies, which were awarded in 1990. The

three ET6 studies were directed to the development of a "Comprehensive Linguistic Software Development and Testing Environment". The one ET7 study was a feasibility and project definition project on the "Reusability of Lexical and Terminological Resources in Computerised Applications". These studies ran for up to 18 months from January 1990 until mid 1991. The aim was both to produce a more up to date formalism than ETS, and to involve industry in the Eurotra work.

3.4.5 *ET9 Projects.* The ET6 studies were followed by a call for tenders in March 1991 for development work based on the formalism developed in the studies. There were eight bids and two fully funded contracts were awarded, one from the EUROTRA budget for the ALEP0 system, contractor P-E International; and one from the LRE budget for the ALEP1 system, main contractor BIM. These contracts were let in January 1992 and run for two years. A preliminary prototype ALEP0 system was available in the last quarter of 1992 for use in a number of ET10 and LRE projects. The first version of ALEP1 is due in March 1993, final implementation in mid 1994, with further support and development work being required thereafter.

3.4.6 *ET10 Cost-Shared Projects.* In the final phase of the EUROTRA programme the concept of cost-shared projects was introduced taking 30% of the budget. A call for proposals was issued on the conventional, ESPRIT-like, Commission pattern, in March 1991. From about 27 bids six projects were awarded in January 1992. They ranged in duration from 16 months to 2 years, in value from 162 thousand ecu up to 408 thousand ecu. Though largely drawn from academic teams, in particular the EUROTRA Centres, there were three industrial partners. Several projects were related to the ALEP formalism development, and can be seen as complementary to the main EUROTRA work, filling some of the gaps detected in the main programme.

3.5 LRE Projects

3.5.1 Though not strictly a part of the EUROTRA programme, the LRE programme, launched in 1991, can be seen as an extension of the move to cost-shared projects started with the ET10 projects. The call for proposals was issued in August 1991 and the decisions announced in January 1992. The nine projects last from 24 to 30 months duration, cost from 590 thousand ecu to 2.8 million ecu. The projects range across language and speech technology, though some have direct relevance to machine translation and ALEP. It is noticeable that the projects contain a significantly larger proportion of industrial partners than EUROTRA, with some 17 industrial partners to 30 academic partners. Most projects involve one or more EUROTRA Centres.

3.6 Future Language Engineering Plans

3.6.1 A second call for proposals for LRE cost-shared projects was issued in October 1992 and a further call is being planned for 1993 subject to the availability of funds. Plans are being drawn up for a Language Engineering programme for the Fourth Framework Programme due to start in 1994. The academic and industrial community has been consulted about the workplan in large panel meetings held in November 1991, January and May 1992.

3.7 The Council Decisions

3.7.1 *The Eight Decisions.* Since the programme was authorised by Council Decisions (CD) after approval by the European Parliaments it seems worth examining what was stated to be the objectives, timescales, costs, management arrangements, as described in these Decisions. There were eight CDs during the ten year life of the programme which make explicit or implicit reference to the programme. However two of these authorise the second and then the third Framework Programmes with, in each case, a following CD concerned with the specific programme.

	DATE	REF NUMBER	PURPOSE
1.	04.11.82	752	Launch of EUROTRA
2.	04.07.84	338	Change of advisory committee from ACPM to MCAC "Linguistic Problems" (CGC-12)
3.	26.11.86	591	Adds Spain and Portugal
4.	28.09.87	516	Second Framework Programme
5.	25.07.88	445	Authorised third phase to start on July 1st 1988
6.	20.06.89	410	Authorised funds for the third phase until 30th June 1990
7.	23.04.90	221	Third Framework Programme under which the LRE programme was authorised
8.	26.11.90	664	Authorised final two years 91/92 until November 1992

3.7.2 *The Objectives.* There were, essentially, two CDs that addressed the objectives of the programme, that of 4th November 1982 that launched the programme, and that of 26th November 1990 that authorised the final two

years with modified objectives, the first and the last of the CDs concerned with EUROTRA. The launching CD of November 1982 refers to the barriers that language differences create in the Community to communication and trade; to the likely impact of computational linguistics on this barrier; to the research work already carried out; and says that action by the Community can *create a European machine translation system of advanced design* to the direct benefit of European industry.

3.7.3 This initial CD calls for a "research and development programme for the creation of a machine translation system of advanced design". The annex states that this system should be *"capable of dealing with all official languages of the Community"*. On completion of the programme an *"operational system prototype"* should be available in *"a limited field and for limited categories of text"*. This prototype would *"provide the basis for development on an industrial scale"* in the period following the programme. The annex lists the work to be carried out in some detail, including basic and applied linguistic research; the construction of the basic software "by invitations to tender"; and the systematic testing and evaluation of the pre-operational prototypes. The annex refers to extending the lexical bases to cover "the chosen field as exhaustively as possible *(about 20,000 entries in all languages)*". This annex makes it clear that the industrial development will fall outside the programme, but requires the programme to prepare a proposal for the development of an operational system on an industrial scale for commercial exploitation. This annex is reproduced as Appendix 7.1 to this Report. In Annex 2 quite detailed instructions are given to the Advisory Committee on Programme Management and, amongst other matters, requires it to contribute to *"the clarification of the user requirements*, in particular in the field of information and documentation".

3.7.4 The Council Decision for the Second Framework Programme in September 1987 refers to the Language Engineering Programmes only in the section on the "Dissemination and Utilisation of S/T Research Results", but states as the objective of the programmes "to develop rapid and efficient computerised systems for translation and interpretation". The CD also refers to the activity covering "completion by 1990 of a first multilingual prototype machine translation system". So in 1987 that is still seen as the objective of the EUROTRA programme. The original phrase "completion of a machine translation system of advanced design" is used in the CD of 20th June 1989 with objectives unchanged.

3.7.5 The only reference to language engineering in the CD for the Third Framework Programme in April 1990 seems to be "Making services easier will require a sustained effort in language research and engineering. Following work already done as part of the EUROTRA programme, it is now necessary to encourage the development of operational systems linked to information and communication systems.". Perhaps it is a mistake to expect to deduce anything about the objectives of an individual programme

from the Framework Programme Decision, for this is for an authorisation of the LRE programme, rather than of the final stage of EUROTRA.

3.7.6 The Council Decision of 26th November 1990 implementing the Second Framework Programme is concerned with the "preparation of the development of an operational EUROTRA system" in the final two years of the programme from 26th November 1990. The overall objective continues to be the first step towards the "development of an operational machine translation system of advanced design, capable of dealing with all official Community languages". The three specific objectives are stated to be the "Creation of the conditions for the transition to an operational system; advancement of work on lexicography and terminology; and training and cooperative projects.". Priority action lines are listed as:

1) system development, testing and research environment (ET6/9 ALEP contracts);

2) language-specific research and development work (Eurotra Centres' work);

3) linguistic research of general interest (ET10);

4) research into advanced system architectures (ET10);

5) reusability of lexical and terminological resources (ET10);

6) standards for textual, lexical and terminological data (Text Processing Initiative EAGLES);

7) education and training (ET grants).

In view of the relevance of these objectives to the Panel's evaluation of the last stage of the programme this Annex to the CD of 26th November 1990 is reproduced in full in Appendix 7.2. The topics 3), 4) and 5) above were followed up by the Call for Proposals of 8th March 1991, asking for proposals by 8th May 1991. The training topic was implemented by a small scale post-doctorate grants scheme for the Centres costing 45 Kecu in total.

3.7.7 CD 91/353 of 7th June 1991 is concerned with the authorisation of the Telematic systems area of the Third Framework Programme, of which Area 6 is the Linguistic Research and Engineering, LRE, programme. There is an explicit reference to the work being based on the results and experience drawn from EUROTRA. This CD was followed up by a Call for Proposals for the LRE programme on 21st August 1991.

3.7.8 *Authorised Costs and Timescale.* The initial authorisation in CD 82/752 of 4th November 1982 refers to a five and a half year programme from 13th

November 1982 at a cost of "16 Mecu including expenditure on a staff of eight temporary agents", in three phases:

Phase 1) Preparatory phase : 2 years, 2 Mecu

Phase 2) Phase of basic and applied linguistic research : 2 years, 8.5 Mecu

Phase 3) Phase of stabilisation of the linguistic models and evaluation of results: 18 months, 5.5 Mecu

Total: 5½ years, 16 Mecu

3.7.9 The addition of Spain and Portugal resulted in CD 86/591 of 26th November 1986 authorising an increase to 20.5 Mecu and 14 temporary agents over seven years

Phase 1)	Unchanged (past?)	2 years, 2 Mecu
Phase 2)	Increased to	3 years, 13 Mecu
Phase 3)	Increased to	2 years, 5.5 Mecu
	Total :	7 years, 20.5 Mecu

3.7.10 CD 89/410 of 20th June 1989 authorising Phase 3, the completion of the programme, increased the cost of that phase from 5.5 Mecu to 12.5 Mecu, and named 30th June 1990 as the end of the authorisation. It broke down the use of the extra 7 Mecu as follows:

Community contribution to the national groups	4.3
Basic software	1.1
Linguistic specifications	0.2
Training, workshops, supplies, etc	0.3
Preparation for industrial development	1.0
Evaluation	0.1
	7.0 Mecu

3.7.11 The Final Phase 4) of the EUROTRA programme was authorised by CD 90/664 of 26th November 1990 for two years at a cost of 10 million ecu including the expenditure on five temporary staff. The indicative breakdown of this sum was:

System development environment	2.0
Community contribution to the national groups	4.0
Shared-cost research projects	3.0
Training, subsidies, evaluation	1.0
	10.0 Mecu

3.7.12 Thus the total cost authorised amounts to 37.5 Mecu over the ten years from November 1982 until November 1990. This includes the cost of temporary staff.

3.7.13 *Management and Evaluation.* The initial authorisation places the responsibility for the execution of the programme on the Commission, assisted by the Advisory Committee on Programme Management (ACPM). Each phase should include a review, but the Commission are not required to obtain Council authority to pass on to the next phase.

3.7.14 CD 84/338 replaced the ACPM by the Management and Coordination Advisory Committee "Linguistic Problems" (CGC-12), to bring the programme into line with the requirements of the Second Framework Programme. Then CD 86/591 lays down that the Council should authorise the move on to the next phases, on the basis of a report from the Commission and the opinion of CGC-12. Then in CD 90/446 of 26th November 1990, authorising the final two years, the "Committee of an advisory nature assisting the Commission" is stated to be "composed of representatives of Member States and chaired by the representative of the Commission".

3.7.15 CD 90/446 also calls for an evaluation of the results achieved through a panel of independent experts - this panel.

3.8 Cost of the Programme

3.8.1 *Budget and Expenditure.* The budget for the first two phases was revised in the Decision of November 1986 to be 15 Mecu, and the actual commitments and payments came in at that figure. The budget for the main third phase was increased to 12.5 Mecu in CD 89/410 of 20th June 1989, and the commitments and payments came in at that figure.

3.8.2 The budget for the final phase, 1991 and 1992, was set in the CD 90/664 of 26th November 1990 at 10 Mecu, and the commitments came in at that figure. It is too early to comment on the payments which may be somewhat lower as the EUROTRA teams ran down in numbers faster than might have been expected. At 5.6 Mecu the Community commitment to the Centres was larger than the, purely indicative, breakdown figure of 4.0 Mecu in the CD. The figure of 5.6 Mecu includes the cost of the Training Grants, and the special contracts to Luxembourg and Ireland for the general support function. The commitment on the ET10 cost-shared projects was lower at 2.0 Mecu than the expected 3.0 Mecu, because of the higher than expected cost of these payments to the Centres.

3.8.3 *National Contribution.* It should be remembered that the national governments were also contributing to the costs of the programme, in proportion to the Commission contribution. The agreed division of contribution was:

COUNTRY	NATIONAL CONTRIBUTION AS % OF TOTAL
Belgium	40
Denmark	20
Germany	75
Spain	40
France	75
Greece	20
Ireland	20
Italy	67
Luxembourg	15
Netherlands	40
Portugal	20
United Kingdom	70

3.9 Conclusions on Initial Objectives

3.9.1 *Timescale.* The original Council Decision of November 1982 envisaged a programme of five and a half years. *This was clearly a considerable underestimate.* However it must be remembered that the climate created by the adoption of the Systran system of US origin for use by the European Commission made it reasonable for some people to envisage the rapid construction of a European system. No doubt the programme proposal was designed to ensure the support of the Council of Ministers. Yet it seems inconceivable that experienced computational linguists could have believed that an operational machine translation system prototype of advanced design could have been built in that timescale. The US academic experience was hardly encouraging. However, it is clear that many, perhaps most of the "founding fathers" of the programme, especially the people with a software background, did believe that a useful operational system prototype could be built in a few years of work. The Programme of Work envisaged a two year preparatory stage, followed by two years of basic and applied linguistic research, followed by an eighteen month testing and evaluation stage. *In hindsight it is clear that a much longer programme was required, and indeed the EUROTRA programme ran for ten years, without producing a complete system that could be fully tested and evaluated.*

3.9.2 Subsequently to that original Council Decision there were several further Council Decisions revising and extending the EUROTRA Programme. *Conducting a programme of this nature by stages is not an efficient way to operate.* Of course the Council might not have been prepared to authorise a longer programme, even with the stage by stage reviews envisaged in the original Decision. *However had a ten year programme been authorised initially a much better programme might well have resulted.*

3.9.3 Though of course the culture in Europe is against commitments longer than five years, it is notable that ten year programmes can be very much more productive. Perhaps the best example is the ten years given to the Xerox Palo Alto Research Centre team in the 1970s which led to the

office automation revolution, and to numerous innovations. The Japanese often authorise their National Research programmes for periods of ten years, and some of these, such as the Semiconductor Programme of the 1960s and 1970s and the Supercomputer programme of the 1970s have been very successful. Of course, not all ten year programmes are as productive. For example, the Japanese Fifth Generation programme in the 1980s is generally considered not to have been a success.

R1 : Where it is recognised that the difficulty and scope of a programme will require a long run, the Commission and Council should face up to this from the initial decision, of course with suitable review and break points built in.

3.9.4 *Research and Development.* At the time EUROTRA was being planned there was no coherent Council policy or plan for programmes of R&D with industrial objectives. These did not really emerge until ESPRIT in 1982, leading on to the First Framework Programme in 1983. The EUROTRA programme was an uneasy mixture of Research and Development, running in parallel. In the context of this programme the word "Development" refers to the objective of achieving an "operational systems prototype", a pre-industrial prototype even though industrial participation in parts of the second phase of the programme seems to have been anticipated. Despite the two year preparatory stage envisaged in the initial Council Decision *there seems to have been no comprehensive and charted work plan that could be followed throughout the programme. This was unfortunate,* though of course in some ways understandable, and perhaps inevitable, for a programme that was authorised in steps. The decisions to freeze the formalisms were described by one senior "Eurotrian" as too early for Research, too late for Development. No doubt this is always the feeling if such decisions are not taken against a clear work plan. It is noticeable that to some of the research teams in the programme the word "Development" seems to have been used to refer to software system construction.

3.9.5 It is tempting to follow many of the Eurotrians in saying that the programme should have separated Research from Development. However, this is not necessarily the right approach. The original decision was taken in the belief that an "operational system prototype" would be the end product, leading on directly to a fully commercial product. A Research programme alone might well not have been authorised. *One model for a work plan might have envisaged, say, seven years of research followed by three years of prototype development.* To give focus and balance to the research a study team for the development phase would be established from the beginning of the programme, with research being planned and conducted to meet all the foreseen difficulties in the development, revised as new problems emerged. Of course the understanding of the problems of Machine Translation systems was pretty immature in the early years of the programme.

3.9.6 *Languages.* Though it does not seem to have been an explicit objective of the original Council Decision, one considerable achievement of the programme was the building up of significant computational linguistic capability in all the 12 nations of the Community. However, *the way in which the programme was conducted with work in parallel on all nine Community official languages and all 72 language pairs (until the final stage) was not efficient, judged in terms of achieving the systems prototype goal.* This approach tended to result in the work progressing at the speed of the slowest. It led to a pseudo democratic decision making process that made it difficult to reach sensible decisions in the view of many of those involved. Most of the technical decisions were made after discussion in the EUROTRA Liaison Group, consisting of the directors of the Centres, under the chairmanship of one of them. But in practice the ultimate decision power rested with the Commission's Programme Director, who on occasions overruled the Liaison Group of which he was always a member. It would have been possible to devise a much more efficient programme that achieved better research and development, at the same time as building up technology transfer and training for the less developed teams. It was, however, right to involve all nine languages, at more or less the same time, for valid political, cultural and scientific reasons. This issue is considered in more detail in 4.5.4 below.

3.9.7 *Too Academic a Programme?* Until the last stages of the programme *there was no industrial involvement which was clearly unfortunate* as a development programme was envisaged. The Commission's staff recognised from the beginning that there were few firms in Europe who could have taken a constructive part in the linguistic work of the programme. But the programme would have benefitted if industry had been involved from the planning stage onwards. Even in the last stage the industrial involvement was relatively small, though important. Had a study team for the development stage been established, as proposed in section 3.9.5 above, this should have been largely staffed by people from industry. *Much of the prototype software system should have been entrusted to industry,* who are used to the problems of maintaining and updating software. This indeed did happen after 1989. *Steps should have been taken to involve industrialists with the Research Centre teams,* if only by creating industrial advisory committees and "uncles". This would have brought an industrial view to bear on the work, and aided the exploitation of the work and the technology transfer to industry. But there is no doubt that industrialists alone could not have met the objectives of the programme, if only because of the lack of computational linguists in industry at the time.

3.9.8 It was also most unfortunate that there was so little involvement of the potential user of a MT system that might have stemmed from the programme. Industry might have been closely involved in the programme both as a product developer and potential exploiter, but also as a potential user of a resulting system. The influence of users from industry and commerce might have resulted in a more practical approach to what the

translator needs, to the integrated systems approach. Of course the Commission is itself the largest user of translation in Europe, probably in the world. And in the later stages of the programme the Commission's translation service, SdT, did provide a number of secondees to help with the running of the programme. By then the programme was too firmly established to turn to a user oriented approach. But perhaps their influence can be detected in the increased emphasis on the user interface and dictionaries in the final phase of the programme.

3.9.9 *Government Involvement.* To the extent that the involvement of governments in the original Contracts of Association process ensured that, at least, most of the 12 governments took an active interest in the programme, this procedure was sensible. This procedure was laid down in the original Council Decision. But it brought with it various undesirable consequences. It took a long time for some countries to nominate and fund their Centres; the process led to freezing of participation to the initial Centres, and the process made the financial control very difficult. On balance, *this was an unsatisfactory way of organising the programme.*

3.9.10 Partly because of the way the programme was organised, initially Centres took different interpretations of the objectives of the programme. As time went on the teams increasingly learnt to cooperate successfully, and a strong degree of coordination eventually emerged. But it was initially very difficult, and never easy for the central team in the Commission to exert control. This resulted in a weaker Central Management than was desirable, and indeed than was desired by some of the teams. The Central Management could not have established its authority without the power to redeploy its funds as the situation demanded easily and rapidly. *Normal "ESPRIT type" funding would have provided stronger central control, though that approach would have had to be coupled with coordination and "networking" of the projects and the project teams, to create the close-knit and integrated approach that was certainly achieved by the approach actually adopted.*

4. The Main EUROTRA Programme

4.1 Mode of Operation

4.1.1 The main EUROTRA programme was operated through Contracts of Association with each of the 12 national governments. Each nation nominated one or more Centres to participate in the programme. In total there were nineteen Centres, but sixteen main ones. All were non-profit making research centres, most being in universities. The exception were in Germany and Italy where the participating Institutes were semi-independent bodies also carrying out work on linguistics and translation outside the EUROTRA programme. There were no commercial or industrial bodies participating in the main part of the programme, though firms were contracted for some of the software environment work and the ALEP studies and implementation. In the final part of the programme the ET10 projects did bring in a few firms.

4.1.2 The Centres cooperated in the work, organised initially through the EUROTRA Coordination Group and small contracts, and then through the Commission and the EUROTRA Liaison Group. This consisted of the leaders of each Centre meeting with the Commission's project leader. Decisions in the Liaison Group required a two-thirds majority including the Commission. Over this was a Common Steering Committee, primarily for financial decisions. There were several ad hoc committees to organise aspects of the work, for example the Linguistic Specification Group. Most of the funding for each Centre came from the grant from the Commission together with that from their national funding agency. The Commission did award extra small contracts to enable certain tasks to be organised and implemented.

4.1.3 Much of the work on the linguistic specifications was carried out in parallel on each of the nine official languages in the designated Centres (Luxembourg and Ireland being given special tasks). This work was carried out by special contracts to individual researchers in the different teams, and was then distilled by the central team into the Reference Manual which brought together these linguistic specifications and grammars, in a common formalism.

4.1.4 Most of the work was carried out by linguists, and computational linguists who became computational linguists, often in language or linguistic departments of universities. There were relatively few software engineers in these Centres. After a stage of trying to develop the software system with mixed teams of linguists and software engineers in the Centres, much of the

software work was carried out in Luxembourg under the direct control of the Commission, at first with the assistance of people seconded from the Luxembourg Centre and then after 1986 by staff from firms, in particular P-E International, working in Luxembourg. This centralised software system work was far from satisfactory both because it, inevitably, became available only at a relatively late stage in the programme, and because of the comprehension gap between the software engineers in Luxembourg and the linguists in the Centres. Though certain of the Centres did receive small contracts for software work, the Centres were dependent on a software system from Luxembourg to test their linguistic work but it was a long time before a system was available, it was very slow, and all the Centres had to join in the debugging of early releases. However, it has to be remembered that there was a range of machines to be tackled with a variety of operating systems, so the logistic problem of providing software for the different Centres was formidable. Moreover, performance was always and remains a problem. No sooner was a faster computer system available than the complexity of the sentence to be parsed would expand, leading to the requirement for even more computing power.

4.1.5 The problems of controlling many diverse Centres, geographically widely separated, with skills differing in nature and quality, must have been formidable. These were compounded by the very limited size of the Commission team in the early years and their lack of authority in terms of total funding control and intellectual pre-eminence. It is hardly surprising that the programme seems to have proceeded more as a set of loosely coordinated parallel research projects than as a focused objective-led, directed, programme that must have been the primary objective. It is noteworthy that some of the Centres feel that the best work was carried out on the so-called official "sidelines".

4.2 <u>The Key Objective</u>

4.2.1 *1) Lack of Industrial Participation.* Though it may not have been intended by the Council in their 1982 Decision, the fact is that the programme was conducted throughout its main phases in an academic environment. Even the software work was conducted under the direct control of a Commission team lacking the industrial imperatives and experience. So it is not surprising if the key objective of the programme to provide an "operational system prototype of advanced design, capable of dealing with all official Community languages" preparatory to the "development of an operational system on an industrial scale" was not achieved.

4.2.2 Of course it does not follow that the fact that the work was conducted in an academic environment was a necessary reason for failure. The "large physics" community of Europe has demonstrated through CERN, through

the large telescope developments, through space research and through the fusion research programmes, that it is quite capable of organising itself to build on an industrial scale. However, there seem to be three main differences from the EUROTRA programme:

1) The "large physics" programmes are almost always conducted from a central facility or research facility, even if much of the preparatory work is carried out in a decentralised way.

2) The objective for these programmes is the pursuit of scientific rather than industrial objectives except perhaps for the fusion programme. The EUROTRA programme was clearly intended to lead on to an industrial objective, with all that that implies in terms of market considerations. So it is clear that the seeds of failure were implicit in the way that the programme was set up without industrial participation.

3) Compared with the physics community, the computational linguistic community is new and perhaps the study of the subject is at an early and still immature stage in its development.

4.2.3 It has been argued that it would have been very difficult to find European firms with the will and the capability to participate. The fact that major machine translation developments have been going on in Europe during the EUROTRA decade rather gives the lie to this. One thinks of the Gachot SA development of Systran, the Siemens or SNI development of Metal, the Philips development of ROSETTA, and more recently the SITE group's development of Eurolang. There have been other significant projects with industrial involvement, such as GRAAL and Genelex. It is true that most of these projects have received some funding from public sources, under the Eureka and ESPRIT programmes. And these projects have benefitted from the EUROTRA work and trained staff. But the drive has been commercial, even if commercial products will not necessarily emerge from all of them. However in the early days of EUROTRA it would have been difficult to find much professional competence in European industry.

4.2.4 Because the Commission is itself a major customer for machine translation systems there would have been justification in purely economic terms for Community funding beyond the normal 50% had that been deemed essential. The fact that the ALEP work, both in its ET6 study phase and in the subsequent ET9 development phase, has been conducted through fully funded contracts demonstrates that such an approach is recognised as appropriate in some cases. In the case of ALEP, the Commission wishes to fully control the property rights and so fully funds the work.

4.2.5 *2) Technical Considerations.* It was explicitly intended that the EUROTRA system should lead to a system of "advanced design". Nowhere was the meaning of this term explained, still less quantified. *(It is a notable feature of the whole programme that so little, virtually no quantitative, measurement of performance was undertaken* despite the explicit encouragement to evaluation in the various Council Decisions.) However, in view of the origins of the programme in the feeling that Europe could produce a better system than Systran, a system of US origin, it is reasonable to assume that "advanced design" implies a significantly better performance than Systran. The test would be in the eventual marketplace; would the system stemming from the EUROTRA programme seize the market? Had quantitative targets been established early in the programme this might have had the benefit of focusing the work on more commercial concerns, though it would have been difficult to establish satisfactory quantitative targets in a field where no commonly accepted measures of performance exist. It is encouraging to see that the Commission is now tackling this topic in the LRE programme.

4.2.6 Performance in terms of quality of translation must be the first consideration. However, even at the time of the inception of the programme in 1978 it must have been clear that there was absolutely no prospect of achieving a system that would be usable without human post-editing for normal translation purposes, and, indeed, the EUROTRA programme did envisage post-editing though it gave no attention to that aspect. The US National Science Foundation and National Academy of Science ALPAC report of 1966 had stated that high quality machine translation was not possible, and more to our point, that it would not be possible for many years to come. This report had had such an influence on the US Natural Language research community that it is inconceivable that it was not well known to the founding fathers of EUROTRA. However, that was twelve years later, though there was no evidence of any significant change in the situation then, just as it remains true today, despite the steady progress in the understanding of the complex linguistic problems and the very large progress in computer and system performance and human interface understanding and provision. So improvement in performance has to be measured in terms of the productivity of a system involving a human post-editor. While a good document handling working environment for a post-editor can bring about considerable productivity gains it takes a very considerable improvement in the quality of a machine translation to make a significant improvement, since the translator has to familiarise himself with the document.

4.2.7 One way in which machine translation systems might perform sufficiently well to avoid post-editing would be in a very narrow domain, where the document author is limited in the vocabulary and grammar he may employ. If the material, such as a technical manual, is drawn from a narrow enough domain the ambiguity it contains is reduced. The original Council Decision

does call for the system prototype to be available in a "limited field and for limited categories of text". The lexical work was concentrated on the telecommunications domain, and the satellite communications sub-set of it. But *there was virtually no attempt in the programme to restrict the work to that which might be appropriate for a narrow domain.* Perhaps there is little restriction on aspects like grammar that could have been applied in the research programme. In the "Coverage Descriptions" for the second and third phase of the programme the grammatical coverage does take account of the special needs of the telecommunications text.

4.2.8 From the viewpoint of the market one very important consideration is how robust, modular, extensible and maintainable the system is. In this respect it would not be difficult to produce a markedly better system than Systran, based as it is on the software technology of the Sixties. Since little attention was paid to the software system, until the last phase of the main EUROTRA programme, little attention was paid to the robustness, etc, of the system. However there is no question but that the ALEP tool-set will be well engineered compared with systems of the Systran generation. *It is regrettable that so little attention was paid to the engineering of the EUROTRA system, for example to the portability of the grammars and dictionaries,* despite the proposals on this from at least one Centre. However, this was studied in the ET7 study and is now being worked on in an LRE project.

4.2.9 The Danzin Panel, like the Pannenborg Panel, concluded that the EUROTRA programme will not lead to an operational machine translation system, but to what they called a "scientific prototype". This referred to "a sum of theoretical and experimental results, the reliability of which would be demonstrated and which could ultimately lead to an "industrial prototype". *After two further years' work it is clear that the judgement of the Danzin Panel was correct in the scientific prototype nature of the outcome, even if it was optimistic on the demonstration of the reliability of the results.*

4.3 The Scientific Achievement

4.3.1 Despite the success in pulling the disparate teams together, since the programme was more often in the nature of a set of coordinated academic research projects than a closely directed R&D programme, it seems appropriate to examine it in terms of its scientific achievement. However there are three factors that must mitigate against scientific output:

1) The inevitable clash between scientific research and the objective of producing a prototype system led to promising research lines being cut off too early. It is said that some of the best work was carried out in sidelines, whether official or unofficial.

2) The requirement to carry out work in parallel across the nine languages and 12 countries led to a loss of output through duplication, though this was offset to a limited extent by the benefits of being able to contrast the different environments and the intellectual creative tensions caused by the clash of cultures and backgrounds.

3) The partial ban on publication in the early years led to a loss of publications, and, to some degree, of intellectual contacts with peer workers outside the Community, though this ban was being lifted from 1982 onwards.

4.3.2 *The EUROTRA Reference Manual* is a remarkable piece of work, being perhaps the most extensive description and specification of an MT system that exists. It has detailed sections dealing with all the major aspects of a system such as the virtual machine approach, the linguistic theory, the grammar rules for the handling of Syntax, and of Semantics, Morphology, and the Dictionary. It enabled the widely separated workers on the nine languages to work to a common framework.

4.3.3 *The "E-Framework" System* adheres to the mainstream of current MT development, using a unification-based stratificational model approach. This is described in the "EUROTRA Linguistic Specifications" and in the "EUROTRA Formal Specifications", the first two volumes of the "Studies in Machine Translation and Natural Language Processing" published by the Commission, and fully defined in the Reference Manual. (See outline in Appendix 5.)

4.3.4 The *E-Framework formalism* is intended to be declarative and, though it falls within the unification approach, was developed and defined within the EUROTRA programme. Though a relatively minor facet of the work it does single the whole work out from that done elsewhere, rather as the use of a special computing language, say a variant of Prolog, distinguishes and separates a software system from others concerned with the same class. Inevitably this choice of formalism was a contentious issue, perhaps the most contentious in the whole technical development. Any formalism must be a compromise between the desire, on the one hand, to be as pure, as close to a fundamental set of logic rules and as independent of the particular hardware and implementation software, and, on the other hand, to run efficiently for a given generation of hardware and software. The larger the system, in respect of the complexity of the sentences it handles, the number of grammar rules, the size of the dictionary, the more computation time it takes - in some aspects rising exponentially with the complexity. So what may be perfectly efficient enough for a system to be used in a research environment,

may be quite unacceptable for a system of operational size. There were a variety of formalisms developed in the mainstream, and official and unofficial "sidelines" of the programme, such as CAT, CAT 2, MiMo 2, etc. The main ETS formalism had to be modified by those wishing to develop practical applications of the EUROTRA work, in order to provide a system capable of being run efficiently. Finally the ALEP formalism was developed in the ET6 study and implementation is proceeding in the ET9 projects. ALEP is both a much more fully declarative formalism and capable of being run very much more efficiently. However, it may be doubted if it can be used for an operational sized system without modification as distinct from a research tool system, though this remains to be established. Perhaps the topic of formalisms received over much attention in the programme due to the influence of academics interested in that aspect, to the detriment of other important aspects of the MT problem. The subject of the formalisms was discussed in Annex II to the Danzin Report.

4.3.5 *The Eurotra System Design* has the normal three main phases; analysis, transfer and synthesis, with stratification of the analysis and synthesis phase. Both the analysis and synthesis phase are completely monolingual, but the transfer phase depends on the specific pair of languages. So for the nine EUROTRA languages 72 transfer phases are needed. Evidently, the aim must be to keep the transfer phase as simple as possible for a system designed to handle so many languages. (It is accepted that the use of an Interlingua to reduce this power-law relationship is beyond the state of the art as errors and ambiguities would compound.) So the aim of the EUROTRA design was to simplify the transfer phase, essentially by the use of a bilingual dictionary to replace the source lexical units with the equivalent lexical units in the target language. One advantage of the EUROTRA architecture, for the distributed team approach adopted, was that much of the work could be carried out in a monolingual way. More details of the EUROTRA system and the E-Framework approach are given in Appendix 5.

4.3.6 *EUROTRA Dictionaries.* The EUROTRA work has been carried out very largely by grammarians, with relatively few lexicographers in the teams. So, inevitably, the EUROTRA framework design was dominated by considerations of syntax rather than dictionaries. The result is that a separate monolingual dictionary is required at each of the four appropriate levels, as well as the transfer dictionaries for each language pair in both directions. For nine languages that means 36 monolingual dictionaries, (together with minor transfer-between-levels dictionaries), as well as the 72 bilingual dictionaries. Though this is, in a sense, more a matter of database organisation of the lexical entries rather than multiplication of the material, it is a significant difference from the much simpler dictionary structure that

would arise if a stratification approach had not been adopted. Of course the structure adopted has the organisational advantage for EUROTRA that much of the lexical work is monolingual and so can be carried out in each country separately.

4.3.7 Except for the bilingual dictionaries of the transfer stage, entries in all the other monolingual dictionaries contain information determined almost exclusively by what the grammar needs. So the dictionaries may be stored in a modern, modular, relational database but still the contents are intimately related to the particular EUROTRA structure and grammar. (The EUROTRA dictionaries could be stored in a relational database, though many of the lexicographers preferred to stick to sequential files, because of ease of testing, etc.) The amount of work on these dictionaries was pretty much limited to the needs of a research test bed. The original Council Decision had blessed this by calling for work in a limited field and limited categories of text. The CD referred to a vocabulary of around 2,500 entries, until the third stabilisation phase when the lexical bases should be extended "to cover the chosen field as exhaustively as possible (about 20,000 entries in all languages)". The chosen field was telecommunications, and for the first few years it was based on a fifty page corpora, available in the nine languages, of Commission material relating to a proposal from the Commission to the Council proposing an ESPRIT programme in telecommunications. For the third phase (1988) this was widened to the ITU Handbook of Satellite Communications. At the time the Commission's translation service, SdT, was putting together the satellite communications section of Eurodicautom, the vast terminological database of the Commission. This was useful as a basis for the 72 bilingual transfer dictionaries which are essentially terminology databases in the sense that ambiguity should have been resolved by the earlier analysis stages.

4.3.8 In the final two years of EUROTRA the decision was taken to cut back on the language pairs, allowing each country to choose three out of their eight possible language pairs. English tended to get chosen by all the nine language groups for one of their pairs.

4.3.9 The Pannenborg Panel commented on the problem for dictionary generation caused by the chosen EUROTRA architecture and on the under-estimation of the resources required for dictionary compilation work in the programme. The work of making the entries in a dictionary is very different from the grammar related work of dealing with the rules for constructing the various dictionaries. Grammar comes first to many linguists because grammar is, essentially, a set of assertions about the combinatorial nature of words in a sentence. In recent years the lexical approach has come to be seen as more important, if not dominant, in modern formalisms like HPSG, but this was after the formative years of the EUROTRA structure. So perhaps it is

understandable that so relatively little attention was given to the dictionaries. It is doubtful if there is any property of any value in the actual dictionary entries, as distinct from the work on the way of specifying a particular dictionary entry. Unfortunately that work is a function of the EUROTRA system approach, with dictionary information conforming to the requirements of the Linguistic Specifications in the Reference Manual.

4.3.10 For a practical MT marketable system the dictionaries are everything, totally dominating the workload, the costs of development, and the system performance as things stand today. The Eurotra approach recognises the importance of modularity so that an improved grammar for a particular phase can be slipped in to replace an earlier version. Because of the dominating cost of the dictionaries it is particularly unfortunate that so relatively little attention was given in the design of the system to enable dictionaries to be ported easily from one system to another. This is a topic of major importance, which has been recognised in the attention that was belatedly given to the subject in the ET7 project and in an LRE project. It is particularly unfortunate that there was so little cooperation with the Commission's own Systran work in the dictionary field, in view of the richness of their dictionaries; but there is a need to re-engineer the dictionaries, along with the rest of the Systran system.

4.3.11 *Eurotra Software Implementation.* The ETS system, as implemented by the central Luxembourg team, runs under UNIX, and is written in Prolog and C. It comprises some 600 source code modules, and the whole system takes up about 100 Mbytes of disc storage. A UNIX machine of 3 Mips CPU power, with at least 8 Mbytes of core storage is required to provide a reasonable run time performance. However, it must be remembered that this represents a pretty minimal system in terms of dictionary size and grammar completeness. For short, simple structure sentences this system can respond in under a few seconds, but for longer and more complex sentences the parsing time can take many minutes on such a machine. The Pannenborg Panel commented on the inappropriate initial choice of software implementation methods and hardware of low performance.

4.3.12 The system is a scientific laboratory prototype system for the computational linguist user to develop, test and demonstrate grammars written in the formalism. The man-machine interface provides either a menu based or a command interface. Text handling input and output is based on the SGML standard for describing the text layout.

4.4 The Final Stage

4.4.1 *The Requirement.* In CD 90/664 of 24th November 1990 the final two years of the EUROTRA programme was authorised, running from November

1990 until December 1992. As well as authorising the ALEP work and the ET10 cost-shared projects, the Decision authorised certain work in the main EUROTRA Centres. These continued to be funded under the Contracts of Association, but the funding was reduced by half. The national governments continued to fund their percentage of their Centres' costs, except that the French CNRS, who were running the programme on behalf of the French government, negotiated for a reduction in the French contribution from 75% to 67%. Of course for some Centres where the staff and overheads were funded by virtue of their employment in a state funded institute, the government contribution could be somewhat notional. These final two years were seen as a transition programme for the Centres to enable them to convert, relatively painlessly, to the competitive environment of the normal cost-shared projects' approach. The reduction of funding at the start of these two years certainly was painful, with some of the Centres having to cut back on their staff dramatically. However, at the end of this period the Centres do not seem to have had quite such a painful experience, perhaps because it was foreseen from the start of this final phase, or perhaps because by then most of the Centres had experience of competing for, and sometimes winning, some of the ET10 or LRE projects. The staff of the Centres did decline in the final year as they sought employment elsewhere.

4.4.2 The Council Decision referred to the reviewing of the existing analysis modules, and the extension of the grammatical coverage to include additional text and discourse types. No large scale lexical development work was to be undertaken "pending the outcome of the research on the reusability of lexical and terminological resources". The work seems to be seen as rather tentative "..... is intended to gradually improve the linguistic performance of the system" and "It can be predicted that some progress will be made....., but additional effort must be foreseen for the future".

4.4.3 The "Programme of Work 1991 - 1992" prepared by the Commission makes considerable play with the recommendations of Pannenborg and Danzin. The objectives for this final phase were defined as "to revise in depth the existing implementations" and "to carry out applied contrastive research which includes the implementation, testing and evaluation of the results".

The organisational changes for this final phase were:

1) to reduce the numbers of language pairs, primarily because of the reduction in team size;

2) to carry out the contrastive research work in clusters of groups.

This would have the consequence that the coverage of all the modules in the system would not be equal, but it was argued that "the sum of the research

themes treated..... will be larger than in a fully synchronised approach". The success of the "clusters" approach that emerged as the way to get concentration of effort was acknowledged by the Centre teams.

4.4.4 *The Achievement.* The final progress reports covering the last two years are now being prepared but will not all be available for some months, so it is difficult to make any serious assessment of the work carried out by the depleted Centre teams on the mainstream EUROTRA work during the final phase. It seems to have been largely of a "tidying-up" nature. Many of the research workers were also involved in ET10 and LRE projects, which they probably treated as a priority. And of course they were looking to their future, seeking new positions, etc. Some of the teams were working to develop practical demonstrators in order to attract support for future applications work. So perhaps it would not be all that surprising if the final two years of mainstream EUROTRA work do not prove to have been as productive as the work during some of the earlier years. *It is to be regretted that so little performance measurement and evaluation seems to have been carried out in this, or any earlier, phase.*

4.5 Conclusions and Recommendations

Main Programme : Operation of the Programme

4.5.1 *The Computational Linguistics Community.* When the EUROTRA programme was initiated in the early 1980s, there were computational linguists in some of the countries of the Community but none in others. Today the picture is very different. In 1990 there were some 220 EUROTRA researchers, and there were a further 160 who worked in the programme before or after that date. This makes a considerable body of research workers now in the field, and moreover they are spread across the Community with teams in virtually every country, though many of them are still young and relatively immature. Judging by the technical journals the European research community seems to be at least comparable in strength to the equivalent US community. Moreover, the European research workers are now experienced at working together so they represent a very coherent community, certainly much more coherent than the US scene. *This is a very considerable achievement, where the credit for the growth of computational linguistics in Europe lies very largely with the EUROTRA programme and the supportive governments.* However, it must be remembered that it is Europe rather than the USA that has the multilingual problem within its boundaries. This problem will only get more demanding as the European Community is enlarged.

4.5.2 *The Human Network.* As a direct result of the way the programme was organised the Community computational linguists are now a tight-knit community. The programme had a liaison committee drawn from every team, and numerous standing and ad hoc committees. While this could be interpreted as an unusual way to run a research programme, it was an excellent way to build up the coherence of a community, and brought real scientific benefits. This was strengthened by the use of common software, standards, formalisms, etc, across the programme. *The coherence of the research community through this "Network" activity is a tribute to the EUROTRA programme.* However, there is a danger of this network dissolving with the end of the programme. This issue is dealt with in Chapter 6 below.

R2 : **The human network of computational linguists built up across Europe by the EUROTRA programme should not be allowed to decay with the ending of the programme.**

4.5.3 While the human network was a considerable achievement of the programme it did tend to leave outside those computational linguists in the Community who were not in the nominated Centres. They felt isolated, and resented what they felt was too much of the financial support going to the favoured

Centres. An alternative approach would have been to have created a European "Centre of Excellence" which could have acted as the hub for a dispersed spider's web of research workers, wherever they were located in the Community. Often this approach has been adopted in the Japanese national programmes, but it would have been difficult to staff the Centre of Excellence in the European environment of low mobility. Progress in this difficult subject will certainly require a multi-disciplinary approach, creating mixed teams from the different disciplines that are relevant. This was attempted in EUROTRA, but in practice was not always achieved.

4.5.4 *The Nine Languages.* For valid political and cultural reasons it was desirable that all nine community languages should figure in the programme. And there are some linguistic benefits in being able to compare and contrast the usage in one language from that in another. Nevertheless it was wasteful of resources and a significant factor in delaying progress to proceed with all languages and all language pairs being given even weight together. From a purely commercial viewpoint there are some language pairs that are required by the market more than others. This provides a reason for selecting a limited number to form the basis for the early work in the programme on grammar and dictionaries. A preferable alternative from a linguistic viewpoint would have been to have chosen one Romance language, one Germanic one. At a later stage when the initial problems have been sorted out and the systems architecture stabilised it would be time to extend to other language pairs. This would have been a much more efficient way to have worked. However, what is efficient may not have proved acceptable in the way the programme was organised. Moreover there are very tangible benefits stemming from having the nine languages addressed in the programme, both in the language specifications developed for all nine languages and from the experienced teams now available in all countries. The Danzin Panel recommended that the work should continue on a limited number of language pairs: "those where they feel they have achieved the most advanced, most illustrative or the most useful results". And, indeed, it is significant that the number of language pairs was reduced to three per language team in the final years of the programme. The teams were free to choose their own three pairs. *One of the achievements for the programme has been to produce a language specification for each of the nine languages.* This has been a useful and sensible activity, and these specifications are likely to be of continuing value to system developers.

4.5.5 The Pannenborg Panel reflected on the impact that tackling the nine languages in parallel would have on the risk involved in achieving the programme's objectives. The Danzin Panel commented on the tension between the need to take account of market forces, which give priority to a small number of languages, and the need to protect the cultural implications of all the languages. The Panel proposed that a study should be undertaken

of this issue. This does not appear to have been done. It should be noted that the Commission's overall policy is based on the principle of equal treatment of all official Community languages. In the case of SYSTRAN, which primarily aims at satisfying the operational needs of the Commission itself, budget and resource limitations led to the asymmetric development of the three most commonly used target languages (E, F, D) and nine source languages. This seems an eminently sensible policy and in retrospect could have been applied to help achieve an operational prototype of Eurotra. It was a pity that the policy was established too late to influence the EUROTRA programme.

4.5.6 *Exclusive Teams.* A further problem deriving from the way the programme was organised in the main phase was that *it was difficult to change the main Centres or bring in fresh blood to the programme.* Of course for some countries the Centres were created by EUROTRA and represented the only talent available in the early years. However, one improvement in the final phase was to bring in new teams through the cost-shared programmes. There is some evidence from these projects and elsewhere that good talent was excluded from the programme by the initial decisions on Centre selection that became frozen over the ten years; for example, Grenoble in France, Limerick in Ireland, Edinburgh and Cambridge in the UK.

4.5.7 In Germany BMFT organised and financed regular annual meetings of all the main players in the computational linguistic field. This provided an opportunity for a wider exchange of information than the exclusive EUROTRA teams. It would have been open to the Commission's EUROTRA team to have organised meetings on these lines, if only to expose the EUROTRA work to wider analysis and criticism by its peers. The evidence from the final phase is that it was possible to run the main EUROTRA process in parallel with cost-shared work organised competitively. This alternative will be analysed in Chapter 7 after the cost-shared projects are considered.

4.5.8 *Secrecy.* In the early years of the programme, the Commission's team and some of the national teams apparently held the view that the task laid down in the original Council Decision of 1982 would be achieved, leading to a system of real commercial value. It has to be remembered that the programme was born in the atmosphere in Europe of annoyance that Systran, a system of US origin, had been purchased for use by the Commission, easily the largest customer for Machine Translation systems in Europe and probably the world. *So a secrecy clamp was imposed on the work of the Centres, only being fully lifted by 1985. In retrospect it is easy to see that this was an unfortunate mistake.* While this publication ban was not total, it did discourage interaction with the work in the USA (for example at Stanford and Brigham Young Universities, MIT, and in various

industrial centres such as IBM and SRI) and Japan which might have been of real benefit to the programme. It was also contrary to the culture of open publication that is a desirable aspect of the University scene worldwide. The publication record of the programme has recovered in recent years. Of course where industry is involved in projects, caution has to be exerted in publication, but experience from programmes like ESPRIT suggests that firms rarely totally ban publication by their academic partners, though they may wish to delay the publication of certain details. The issue of publication is also dealt with in 4.6 below.

4.5.9 *It was one of the weaknesses of the programme that so relatively little interaction and connections to the outside world were established.* Systematic efforts to establish links with the rest of the worldwide NLP community, through publishing, demonstrations, invitation of external speakers to EUROTRA workshops, etc, did not start until relatively late in the programme.

4.6 Conclusions & Recommendations

Main Programme : Linguistics and other Technical Aspects

4.6.1 *The EUROTRA Reference Manual.* No doubt the large, multilingual, dispersed team made it inevitable that a reference document should be produced. *The resulting EUROTRA Reference Manual is a remarkable record of the outcome, containing a description of the largest linguistic effort ever made on a multilingual level.* Both the linguistic description and the virtual machine (the E-Framework) are thoroughly described and would be highly appreciated by the rest of the natural-language processing community. It is, in fact, the intention of the Commission to make the Reference Manual and the Language Specifications available to research workers everywhere without cost.

R3: The Reference Manual and Language Specifications should be made widely available.

4.6.2 *The E-Framework.* The theoretical choice of the EUROTRA research community led to both the adoption of the stratificational model and the heavy focus on syntax as opposed to semantics, lexicon, and system engineering. The linguists were given a formal language - the E-Framework - in which to encode their language description monolingually, step by step. Lexicon development was regarded, more or less, as a mere extension of data, and semantics was not used for disambiguation purposes until at a very late stage and at a very low level. Thus, it could be foreseen that it would not be possible to build a transfer link between a source language IS level in analysis and a target language IS level in generation. The E-Formalism was

constructed as a prototypical formal hierarchy that over-generated very much in contrast to human language processing. *The architecture makes it difficult to relate the research to work elsewhere based on more conventional approaches.*

4.6.3 *Dictionaries. In view of the economic importance of dictionaries in practical system development it was unfortunate - as pointed out by the previous panels - that so little attention was paid to the dictionaries in the balance of the programme.* What work that was done was tied to the particular formalism, though it is claimed that the EUROTRA dictionaries could be transferred almost automatically to the ALEP system. *It is particularly unfortunate that so little attention was given to the portability aspects of dictionary design.* EUROTRA could have set standards for future work on machine tractable dictionaries had their dictionaries been large enough and good enough to be used in new applications. However, attention has been given to these basic problems of MT dictionary design in the ET7 and LRE projects. It might be added that current methods for knowledge retrieval may contribute to making lexicons reusable, especially if they are corpus linked for further information retrieval. Ongoing work in this sub-field should give new guidelines.

In the ESPRIT programme there are projects for standards to organise electronic dictionaries (Acquilex, Multilex). A programme to construct dictionaries for a wide range of Natural Language systems is proposed for the future (see 11.4.3). The Commission itself should be a major participant and customer in this programme.

R4: The task of building up machine tractable dictionaries for a multilingual community is one that requires encouragement and support from the Commission. After a suitable study phase, the Commission should launch a major project to create knowledge bases containing lexical, semantic and syntactic information usable in natural-language processing systems for the European languages.

4.6.4 *Semantics.* At the time EUROTRA started, practically all work on machine translation was syntactic with semantics restricted to, at most, the sentence. Yet a human translator makes use of much wider information in resolving ambiguities of meaning and generating the appropriate translation. A translator will read all the technical magazines available to him in the field he is about to work in for the purpose of assembling world knowledge about the text he is to translate. This world knowledge is combined with the translator's language skill, and the quality of the translation is the sum of these two factors. Language skill includes knowledge of pragmatics and style as well as the rules of correctness in morphology, syntax, and semantics including lexical semantics. On the whole, EUROTRA restricted

itself to two of these parts, morphology and syntax. Current machine translation projects in the USA takes a much more complete approach, working on discourse analysis, interaction, interlingua systems and world models for machine translation.

4.6.5 *Artificial Intelligence and World Knowledge.* American research on machine translation is linked with Artificial Intelligence and research focused on natural-language interfaces and expert systems. Language is studied as a procedure that handles data, i.e. knowledge. This knowledge has to be formalized in order to be computable and the formalization of knowledge is referred to as Knowledge Representation. Several university projects use world models as a means to reach a language-independent level of lexical meaning on which to base meaning nodes in an interlingua knowledge representation. The importance of research on meaning, the interlingua approach and the semantics needed is very great indeed.

In the light of this, EUROTRA's consistent preoccupation with syntactic problems stands out as questionable if not seen in a purely historical context. The syntactic representations arrived at are, to a considerable extent, too ambiguous for practical applications. It would be worth while exploring whether a key to the narrowing down of the possible interpretations of a text could be found in the methods just outlined.

This is what is going on in the American MT community. A change of focus in European MT development - emphasizing the lexical and world-knowledge aspects - is called for.

R5: Bringing to bear semantic information from a wide part of the text, the use of world knowledge, and intensified research on lexical meaning should be priority topics in future Commission programmes of IT research, both in natural language research and in wider IT research programmes such as ESPRIT. These programmes should also take into consideration the role of language-independent knowledge bases and interlingua systems.

4.6.6 *Statistical Methods.* Though the technique of MT based on using a statistical probabilistic method was considered in the early days, even in the 1950's, it was too machine power intensive to be taken seriously until recently, even if quantitative techniques have always been employed by linguists. Stemming from work in IBM Yorktown Heights laboratory, directed in the first place not to MT but to speech recognition, the technique has received a renewed burst of attention in the last few years. Alignment programmes translating between English and French have been produced. These systems work on parallel corpora in the two languages and make assumptions based on statistics as to which translations are equivalents.

Performance remains unspectacular on those language pairs. It remains to be seen if the algorithms work on other language pairs. The tests show that rule-based systems like Systran still maintain a higher quality than the statistical solutions and that these should be used as a complement to rule-based solutions.

Since the Commission has vast corpora of parallel texts in the nine community languages it is in a particularly favourable position to conduct experiments using statistical methods as a component in translation or, more precisely, generating approximations of bilingual transfer dictionaries. It seems that statistical methods should be used along with qualitative linguistic analysis if the best results are to be achieved. *It is good to see the approach feature in one of the ET10 cost-shared projects.*

R6: Statistical methods as a complement to rule-based solutions and as a method for human-aided knowledge retrieval from parallel corpora and monolingual corpora and, furthermore, neural network and similar machine-learning solutions should be priority topics in future programmes.

4.6.7 *System Design. It is but little excusable that no real attention seems to have been paid to the EUROTRA user, even considering the fact that the prototype was conceived of as a batch system.* It is true that the decade of the programme has neatly spanned the period in which far more attention has been paid both to the user interface and to the User Centred Approach to systems. Work on projects such as the ESPRIT Translators Work Bench has demonstrated that considerable improvements in professional translator efficiency can be achieved by providing him with easy access to normal dictionaries, terminology data banks, etc., as well as to MT, all integrated into a conventional word processing environment with spelling checkers, etc. The EUROTRA programme did not examine the user interface in any detail.

The EUROTRA formalism is in reality a high level programming language where the researchers can describe grammar and lexicon. The integration of computer science and linguistics has been very low in the programme. This problem has to be dealt with in future research programmes aiming at European natural-language processing tools for the market. Current research in the USA and Japan regards the field of machine translation and natural-language processing as the next phase of everyday use of computers. Structuring requires system design. Current work in the USA is very user-oriented and regards machine translation as a field where there are several modules of language tools, monolingual and multilingual, that can be assembled according to the users requirements.

One reason for the failure of EUROTRA to address such important parts of a complete MT system was the lack of a comprehensive work plan driven by industrialists with an eye to the market, and so on what the users really require. No doubt another reason was that many, perhaps most of the teams were dominated by academic research workers, attacking the still unsolved language problems rather than working on a preliminary solution to the users MT problems.

R7: System design and User Centred Approach should be priority topics in future programmes. This implies efforts to bridge the gap between linguistics and computer science.

4.6.8 *Interaction.* When the programme began the available computers made it very expensive to experiment with real time interaction. Today the computer power available is perhaps two orders of magnitude more powerful, three orders of magnitude more cost effective, and much more effective computer networks are readily available. Moreover, it has now become realistic to plan MT systems using economic local computers such as widely available PCs, albeit powerful versions with large stores. In these circumstances it is natural to look at interaction as a part of any commercial system. *It was a mistake, even if an understandable one, not to make the investigation of interaction a part of the programme.*

However, interaction is an ambiguous concept since the user can be a developer or a post-editor and the system can be constructed to require interaction for disambiguation purposes during analysis or during selection in generation. The developer or expert has been well supported int he EUROTRA programme and since the aim was to make a batch system the only user modelling that needed to be done was that of the post-editor. This was never started since there was never any complete running system.

Human interaction during analysis and generation is still in a phase where a lot of research has to be done. Research has not yet solved the problems of learning and the repetitive questions asked by the system make users avoid it. Today every machine translation system under development has to take this problem into consideration.

R8: Interaction and Learning - automated inference systems making human interaction more effective and less repetitive - should be priority topics in future programmes.

4.6.9 *New Approaches.* The specific technical points of weakness in the main EUROTRA programme, discussed in section 4.6 above, are examples of problems that arose because the background of the EUROTRA teams was often too narrow, in particular being dominated by linguists particularly

interested in aspects of grammar. The Pannenborg Panel commented on this weakness. Had the teams more often been drawn from research workers with different backgrounds, such as lexicographers, computer scientists, human factors and human communication experts, a more balanced programme would probably have been achieved.

4.6.10 As the Danzin Panel pointed out, current cognitive science and artificial intelligence work is opening up new light on the traditional way in which linguists have attempted to solve the complex problem of describing a language. Since the programme started, new approaches to NPL have been developed; for example AI or knowledge based systems approach, notably at Carnegie Mellon University; and the neural network and other machine learning approaches, for example the work at San Diego, University of California. These are no more than examples of new approaches to MT that should be studied.

R9: Future programmes should particularly embrace promising topics and techniques that have been under-represented in the EUROTRA programme.

4.6.11 *Limited Vocabulary Markets.* On the principle that it is better to walk before running, it would have been better to have had limited market objectives for the prototype system. This seems to have been recognised by the authors of the initial Council Decision for the Annex to that document calls for a prototype for a limited field and limited categories of text. The programme did work to a limited vocabulary, but did not attempt to aim at a limited field, except for the Irish work in the final stage of the programme though the "Coverage Descriptions" did provide for some limitation of grammatical coverage. Yet there are clear advantages in aiming for a limited field. It is not simply that a system aimed at a limited, niche market, requires a much smaller vocabulary. More important may be the reduction in ambiguity that results from the system being directed to a limited market.

4.6.12 Extreme examples are fields like weather processing, knitting patterns, food recipes, police and customs communications. In these examples, some of which now employ machine translation systems very successfully and all of which need them, the vocabulary can be limited but also some aspects of the grammar. Of course such limited fields would limit the scope of the research, but would have enabled a practical operational system prototype to have been achieved. Many of the currently available systems on the market aim for the technical translation market, for manuals for maintenance purposes, etc. The market for machine translation for such systems is very large, and because the need to produce translations rapidly in many languages is usually part of the requirement this is a particularly suitable

field for machine translation work. Of course, such systems require a general vocabulary as well as the limited technical term vocabulary, but it is possible to restrain the input material to use a limited vocabulary and grammar, with automatic checking of the authors' source material. Within the work of the Commission there are many examples of suitable limited fields, where the Commission staff can be restrained in vocabulary and grammar in the interests of obtaining translations swiftly and cheaply. It is noticeable that the two commercial systems now under development under the Eureka programme (GRAAL and Eurolang) both aim at specific limited markets in the first place. *It was unfortunate that the EUROTRA programme did not aim for a limited market.*

4.6.13 It has to be admitted that there is something distinctly unpleasant about encouraging the use of limited grammars as this work may serve to do. However, the economic benefits, and the ability to achieve translation which might not otherwise be provided, this "formal" language approach may be justified when techniques like pre-editing interaction are inadequate. Luckily the human spirit is not likely to take readily to a restriction on his right to use and innovate with whatever grammar he chooses, except when the system demands the restriction!

4.6.14 It is interesting to note that the Irish Contract of Association, drawn up at the end of 1984, had envisaged the Irish team working on the relevance of "sub-language" for MT. Little work was done on this until the last two years of the programme, when that team built up expertise in the lexicon and grammar of a limited text field, and now expect to find commercial support for such limited systems.

4.6.15 *Demonstrators.* Though the Commission did take steps at the end of the programme to construct a useful demonstrator, through the work of its own staff in Luxembourg together with input from all the teams, *it is unfortunate that the various teams were not always encouraged or prepared to produce demonstrators at all appropriate stages and especially to produce operational systems at the end of the programme.* Demonstrators were, of course, produced by various teams, for example of the CAT sidelines. The fact that the main ETS formalism could not lead to a practical system without modification made it difficult to produce demonstrators without diverting from the main line of the work. Yet a practical demonstrator is vital if a potential exploiter is to be encouraged to support the work.

4.6.16 *Programme Measurement.* Since the programme was dedicated to the production of an operational system prototype it was perhaps understandable that a sideline like performance measurement did not get any serious attention. However the programme did develop a test suite of sentences, essentially to test various grammatical issues. Had the programme

developed a series of prototypes, as it might have done if it had been market led, then it would have been essential to have developed a process of measuring performance so that progress could be monitored. The Pannenborg Panel regretted the lack of practical test criteria for the end of phase two of the project. It is much to be regretted that so little performance measurement work was undertaken, despite and, indeed, because of the difficulty of establishing satisfactory measurement methods in this field.

4.6.17 Until system performance measurement is taken seriously it will be impossible to make comparative statements about the relative quality of systems, or how one system improves from issue to issue. The Commission has demonstrated that it understands this by giving the subject priority in the latest call for proposals for the LRE programme.

R10 : The Commission should take continuing steps to develop the methodology and practice of MT system measurement.

4.6.18 *Scientific Quality of the Work.* The comparison of the quality of research work is notoriously difficult, until the perspective of time sorts the wheat from the chaff. It is made particularly difficult in a programme that was intended to be a mixture of research and development. During the life time of the programme, throughout the 1980s, the main lead in Natural Language Processing probably lay in the West Coast of the USA with work at places like Stanford and SRI. Certainly the EUROTRA formalisms are derived from the PATR II formalisms from the USA. This is true for ETS and ALEP formalisms. The search for a fully declarative formalism stems from the parallel work going on in software engineering languages. It is excellent that EUROTRA chose to follow this emerging approach, avoiding all the prevalent error of innovating just to avoid following a lead from elsewhere - the "Not Invented Here" syndrome. Whether the variant adopted in the ETS formalism was so sensible is another matter! The EUROTRA Centres have produced a linguistic specification and grammar for every one of the nine official Community languages, but it is difficult to identify any other specific work in the EUROTRA programme that breaks new ground in any major way. The use of the stratification system architecture by EUROTRA makes it difficult to make comparisons with other MT systems' work. The bias towards syntax and away from lexical problems has been commented on above. *It may be that some aspects of the work will turn out to be influential in future systems designs in Europe and elsewhere.*

4.6.19 Due to the EUROTRA programme, there has been a significant increase in the number of European computational linguists whose papers get quoted and who are listened to with respect in international circles. *This evidence of the improved presence of the European workers on the international scene is much to be welcomed, and is an achievement of the programme.*

5. ET6/7/9 Projects

5.1 The Move to Individual Projects

5.1.1 In 1989 the Commission began to plan a different way of working for the last few years of the EUROTRA programme. In 1991 the direct funding of the Centres from both the Commission and the national governments was reduced to approximately half what it had been in the preceding phases. After the ET6 and 7 series of four studies, these cutbacks provided funds for the two ET9 ALEP development projects and the ET10 series of six cost-shared projects. *This move to open up the programme to new participants and ideas is to be welcomed.* The invitation to express interest in participating in the fully funded ET6/7 studies was issued in April 1989. There were some eight expressions of interest and the studies were awarded in January 1990. A small number of EUROTRA Centres (UMIST, IAI) took a part in the ET6 studies, and four Centres (Pisa, Paris, Saarbrücken and Stuttgart) took part in the ET7 study, along with new participants in EUROTRA (SRI, Siemens, SEMA, Oxford University Press, Van Dale Publishers, Hachette and the Universities of Oxford, Bochum, and Heidelberg).

5.1.2 The purpose of the ET6/7 studies was said to be to prepare the ground for the development of practical MT systems based on the EUROTRA system prototype, as well as for wider initiatives in the language field. But by going to external tender the Commission was able to form a window on work going on outside EUROTRA. They tackled two of the perceived problems of building a full system: 1) the absence of comprehensive linguistic software development and testing environments; and 2) tools and methods for the creation and storage of reusable lexical resources. It is interesting to note that the Pannenborg Panel recommended that a parallel stream of work should be set up, involving industrial firms and universities. The introduction of the ET6/7/9 projects implements this concept.

5.2 The ALEP Projects

5.2.1 *The ET6 Studies.* The aim of the ET6/1 study (main contractor: SRI with DFKI and the UMIST Eurotra Centre) was to draw up a detailed requirement specification for a flexible, state-of-the-art, virtual machine architecture and formalism required for grammar coding. It was to allow for an efficient implementation. Calling on the work of the US West Coast community through their Menlo Park laboratory, SRI (UK) were able to propose a fully declarative architecture that was both "purer" (ie avoided procedural features) and was able to operate very much more efficiently than the ETS mainstream EUROTRA formalism, which was, of course, some years older in time. The study seems to have been a success, leading on to provide the design for the ALEP (Advanced Language

Engineering Platform) system under development in the ET9 contracts. The Danzin Panel recommendations encouraged the development of tool-sets.

5.2.2 Related to the ET6/1 study for the architecture and formalism were two other studies. ET6/2 was to prepare proposals for the open, portable, software environment and related tools. The main contractor was the IAI EUROTRA Centre together with CAP and the SNI (Metal) team. The third study was ET6/3 and was for the text handling sub-system carried out by SEMA (Belgium) and an Oxford University Computing Service team. The ET6/2 study produced an outline software environment specification for a system using the object oriented approach. The ET6/2 study resulted in specifications for an SGML standard document interchange format. Together with the formalism and architecture from the ET6/1 study the three studies provide the basis for the development of an advanced toolkit for MT and NLP research purposes.

5.2.3 *The ET9 ALEP Contracts.* The ET6 studies were completed in mid 1991, but before that in March 1991 the call for tenders for the ALEP development projects was made. This led to two fully funded projects, one with P-E International (Luxembourg) for an interim ALEP0 system, and another with BIM (Brussels) for the main ALEP1 system. The contracts were awarded in January 1992 for two years. There is continuity with ET6 studies through SRI acting as consultant to the ET9 work. The ET9 projects will cost about 0.96 Mecu for ET9/1 and about 1.49 Mecu for ET9/2, representing about 25% of the EUROTRA final stage budget.

5.2.4 The P-E International (or rather their Westvries Dutch subsidiary) ET9/2 contract is conducted in Luxembourg in close contact with the Commission's software development team. The contract is for software development, support and consultancy services to the Commission, but is intended to be at rather more arm's length than previous contract support services to that team. The work is concerned both with the maintenance of the current EUROTRA (ETS) demonstration system and with the emerging ALEP system. The three man team has already distributed a very early ALEP0 prototype to some 25 of the EUROTRA Research Centres and projects contractors for use on ET10 and LRE projects.

5.2.5 The main ET9/1 ALEP1 development contract is placed with the BIM team in Brussels, who are known for a fast Prolog compiler development. The ALEP1 software environment is based on the following requirements:

1) a relatively conservative architecture in order to ensure an efficient implementation on mid-sized UNIX workstations;

2) as far as possible it is independent of linguistic formalisms;

3) it is modular and user reconfigurable;

4) the system is designed for further developments;

5) it is aimed to provide a multi-user environment for both Natural Language and Machine Translation development work;

6) it aims to be user friendly and robust;

7) it uses standards wherever possible (eg X-protocol, OSF/Motif widget set);

8) the system should be portable for use on different (POSIX and X/Open) UNIX platforms.

5.2.6 There should be a first release of the development model available in March 1993 followed by the main release in mid 1994, when the system will be tested by researchers working on LRE projects. Thereafter it is expected that the Commission will let further support and maintenance contracts. It is intended that the ALEP system will be made widely available for use by the research community, as an open, portable and reusable workbench for language engineering in a research context.

5.3 The Lexical Resources Study

5.3.1 The ET7 project was selected and funded at the same time as the three ET6 studies. The 18 month fully funded study contract was awarded in January 1990 for delivery in mid 1991 to a large consortium led by Stuttgart University (with Universities of Bochum, Heidelberg, Manchester IST, Pisa, Paris VII, Saarbrücken together with SEMA (Belgium), Oxford University Press, Van Dale Publishers and Hachette). The objective of the study was to provide guidelines aimed at developing standards to enable the reuse of lexical and terminological resources. The study investigated the feasibility of standardising monolingual and multilingual resources in such a way that they can be reused in different applications using different formalisms and system architectures.

5.3.2 The study resulted in a series of 11 monographs. A survey of lexical and terminological applications and resources was carried out. A feasibility study was made of possible architectures for reusable resources. Standardisation and R&D project proposals were made to the Commission.

5.3.3 Some of the proposals have been followed up in LRE I projects, such as the DELIS project, for the development of tools for dictionary building. Other proposals form the background to the Research and Resources part of the LRE II call leading to further projects. And the proposed standardisation actions are being implemented through the EAGLES initiative.

5.4 Conclusions and Recommendations : ET6/7/9 Projects

5.4.1 *The Introduction of Fully Funded Projects.*

The decision to introduce a different approach to the EUROTRA programme is much to be welcomed, though inevitably it raises the question of why the approach of fully funded studies was not adopted earlier in the programme. Maybe it was the main EUROTRA programme that created the computational linguistics community capable of taking part in the cost-shared projects. It is interesting that the Pannenborg Panel commented that a programme of this type would never have been undertaken as a commercial research proposition, and could only be undertaken with full public funding. One theme tackled one of the major weaknesses shown up by the main EUROTRA programme; the need for a comprehensive linguistic development and testing environment. And the other study directly attacks the weakness in the main EUROTRA programme in its failure to address reusable lexical resources. *So both these objectives seem eminently sensible and practical, though one must ask why these problems had not been addressed in the main programme in the preceding eight years.*

5.4.2 *ALEP*

It is excellent that the production of a linguistic software development and testing environment is now being tackled. The fact that the ET6/1 formalism is fully declarative and the whole system approach makes it easier for grammars developed under the ALEP architecture to be readily portable to other, similar, environments. But in practice, as with other formalisms, it is likely that procedural elements will have to be introduced if the system is to run efficiently, though no doubt fewer than with the mainstream EUROTRA ETS formalism. *The impact of the ALEP work on the mainstream EUROTRA work has been unfortunate.* By rejecting the ETS formalism round which the major part of the EUROTRA work was based, the impression has been created to the outside world that nothing of value is emerging from the main EUROTRA work. So far there has been only a limited interaction with mainstream EUROTRA work, though plans have been made to make use of the third call for the LRE programme to achieve the transfer of EUROTRA material to the ALEP formalism. Because of the procedural features in the ETS formalism the grammars are not automatically transferable to ALEP. Of course the ALEP formalism is much more up-to-date and run-time efficient than the ETS formalism. However, ALEP1 has not yet been tried and tested, compared with the ETS-based work which has had much work carried out round it. It is true that the ETS formalism cannot be run efficiently and without sometimes stopping, for example if a word cannot be found in the dictionary, unless it is modified. Perhaps the mistake lay in not giving serious attention to the development of a runnable system based on the mainstream EUROTRA work for immediate use, at the same time as the development of an ALEP more modern system for use as a tool for

research in a few years time.

5.4.3 The ET6/2 and ET6/3 were studies directed, respectively, to the architecture and specification for an open, portable, software environment and to the specification for an SGML interface between an input text stream and the internal linguistic system. *These studies seem worthy and sensible but it is not obvious how they are going to be used, exploited or followed up - except through the ALEP system as is the intention.*

5.4.4 *The plans for the development and use of ALEP by the research community make excellent sense.* However, other tool kits are being developed by firms, and the Commission will need to monitor and assist these developments when appropriate.

R11 : It is recommended that the Commission continues to develop and maintain the ALEP system as one alternative in the field, making it freely available for academic and industrial research purposes.

5.4.5 *Reusability of Lexical and Terminological Resources.* This ET7 study was carried out by a large consortia of eleven institutions led by a team from Stuttgart University, and including three publishers and one software firm. The objective was to develop standards for lexical and terminological resources, both monolingual and multilingual, so that they can be reused by various applications, including different formalisms and frameworks. The eleven reports stemming from the project include a study of a possible architecture for reusable resources. The Final Report makes proposals for Community action, including actions to create standards, proposals for R&D projects, and promotion and training activities. *This work is very important and much to be welcomed. It is only to be regretted that it was not started early in the EUROTRA programme so that the lessons could have been applied to the EUROTRA work, and the proposals followed up in the programme.* Some of the proposals are being followed up in LRE projects and in the work of the EAGLES standardisation initiative.

R12: The Commission should continue to follow up the ET7 Reusable Lexical Resource recommendations in its research programmes, standardisation and training activities.

6. ET10 Projects

6.1 Organisation

6.1.1 Following the Danzin Panel recommendations, the six ET10 projects were introduced to the programme by a call for proposals in March 1991, at about the same time as the call for tenders for the ET9 projects. But, unlike the fully funded ET9 contracts, the ET10 projects are cost-shared in the conventional Commission model, for example that adopted for ESPRIT. The total costs are shared roughly 50/50 between the partner in the team proposing the contract and the Commission. This is interpreted as the normal half the total costs including overheads, or for non profit-making bodies they can choose, if they prefer, to take all the costs excluding overheads. These terms can be quite attractive to academic bodies that have other sources for their overheads, such as their University funds. The total cost to the Commission is 2.84 Mecu, or an average 0.41 Mecu per project. Most projects are for 18 months, one for 16, one for 24 months, starting in January 1992.

6.1.2 Unlike the normal cost-shared projects, the evaluation of the ET10 bids was handled by the Commission's EUROTRA staff, and then the selection put for endorsement to the Advisory Committee. While it can be argued that the staff know the community well, and so can base selection on some wider knowledge than that contained in the written proposals, it is always undesirable when competitive bids are not judged by as impartial a peer review committee as can be put together. Justice has to be seen to be done. Of the 27 bids for ET10, six projects were retained.

6.2 The Selected Proposals

6.2.1 The projects have not been running long enough for a serious assessment of the quality of the work. But it is possible to make some response to the projects selected, especially in contrast to the mainstream EUROTRA work:

1) Semantic Analysis, using a Natural Language Dictionary.
Birmingham University (UK), Bochum University (FRG), Consorzio Pisa Ricerche (Italy, EUROTRA), CST Copenhagen (DK, EUROTRA).

2) Reusability of Grammars for ALEP Formalism.
Essex University (UK, EUROTRA), IAI (FRG, EUROTRA), FBG Barcelona University (E, EUROTRA), IMS Stuttgart University (FRG).

3) Formal Semantics for Discourse.
Leuven Katholieke University (Belgium, EUROTRA), Gruppo Dima, Torino (I, EUROTRA), Salford University (UK), Rijksuniversiteit Groningen (NL).

4) Statistical, Text-Corpora Based Complements for EUROTRA : Terminology, Lexicon and Preference.
IBM (F), Dublin City University (Ireland, EUROTRA), Instituto di Linguistica Computazionale, Pisa (Italy), C2V (F, Software House), Essex University (UK, EUROTRA), Lancaster University (UK).

5) Terminology and Extra Linguistics Knowledge.
Dublin City University (Ireland, EUROTRA), CRP-CU (Luxembourg, EUROTRA), ILTEC (Portugal), INLOM (FRG).

6) Collocations.
Stichting Taaltechnologie, Utrecht (NL), Essex University (UK, EUROTRA), Instituto di Linguistica Computazionale, Pisa, (Italy, EUROTRA), Swisstra, Geneva (Switzerland), Oxford University Press (UK).

6.2.2 It is notable that at least three of these six projects relate to the ALEP formalism and system. The centre of gravity of support has clearly shifted from mainstream EUROTRA work to the newer ALEP approach, which heightens the impression that the Commission has left mainstream EUROTRA behind. Some of these projects help to plug the evident holes in the EUROTRA programme, for example the "Reusability of Grammars" and "Statistical Text-Corpora Based Complements for EUROTRA" projects.

6.2.3 While some of the bids from the EUROTRA Centres were disappointingly unadventurous the EUROTRA teams feature in every project, which perhaps is a tribute to their competitive ability, despite the years cushioned by EUROTRA. Of the 27 partners in the six projects, 23 are academic or Institutes based on academic campuses. The representation of industry is disappointingly thin, being essentially confined to the IBM participation in the Statistical Complements project, where one might expect to find IBM since the company revived the interest in this approach from their work in Yorktown Heights. This project is much to be welcomed. The emphasis on Dictionaries and Terminology in three projects is also to be welcomed, in contrast to their relative neglect in mainstream EUROTRA.

6.3 LRE Scheme

6.3.1 Though it is not strictly a part of the EUROTRA programme it is interesting to look at the LRE programme because it is a natural development of the ET10 projects of EUROTRA. The Language Research and Engineering scheme is part of a broader programme

adopted by the Council in June 1991 (Telematic Systems in areas of General Interest). It was launched at virtually the same time as the ET10 projects, with a call for proposals in August 1991 with the first projects for LRE 1 announced in January 1992. A second call for proposals was made in October 1992, with the proposals due in mid January 1993. Some 81 bids were formally accepted, and of these nine were accepted. The larger number of bids compared favourably with the bids for ET10 projects but maybe the ET10 call was "reserved for the EUROTRA community", in a sense to create a bridge to the normal cost-shared approach. But it might have been due to the wide scope of the call, covering as it does Language Technology in general, and not just Machine Translation. And the publicity given to the LRE programme was considerably more extensive than for ET10.

6.3.2 *Objectives.* The total budget for the LRE programme, 1991 - 1994, is 22½ Mecu. The first call committed 6.5 Mecu, and LRE II in the Spring of 1993 will commit a further 9.5 Mecu. The LRE programme is organised round five themes:

1) research of general interest;

2) development of linguistic resources and related computational tools;

3) setting of standards and guidelines for the encoding and interchange of linguistic data;

4) pilot and demonstration projects;

5) supporting actions, especially training in computational linguistics, and the setting of common specifications and guidelines.

6.3.3 *Projects.* Once again, the projects are dominated by academics, but 15 out of the 47 partners are firms, though often small firms with strong academic links. One of the projects is worth 2.83 Mecu, total cost, 1.4 Mecu, from the Commission, far larger than the others which average about 1 Mecu total, 0.7 Mecu from the Commission. It is concerned with the pronunciation of up to 1,000,000 names for each of the nine Community languages. The COBALT project is concerned with the capture of factual knowledge from textual sources, which is an interesting project for the creation of the very large knowledge bases that will be required if the problem of background knowledge is to be tackled. The TRANSLEARN project is aimed at a toolbox for helping the human translator, for example to deal with repetitive work. It is interesting to see a very practical project related to the translators' real needs, so ignored in the work of EUROTRA. The DELIS project is concerned with methods and tools for the development of dictionaries, stemming from the ET7 project. The RGR project is aimed at the reuse of grammatical resources, and is essentially concerned with formalisms based round and extending beyond

ALEP.

6.3.4 *EAGLES.* The ninth project approved under the LRE 1 programme is very different from the others. EAGLES is the "Expert Advisory Group on Language Engineering Standards". It aims to establish a set of coordinated expert groups for pre-normative linguistic research. The Group of experts will be the driving force behind the development of common functional specifications for the description and representation of linguistic data. The Group will define, demonstrate, evaluate, validate, promote and disseminate these specifications. The Commission bears the costs of the meetings, but the participants bear their own labour costs.

6.3.5 The Group has a Management Board, with working groups and hosting organisations. The Management Board comprises the representatives of the European project consortia MULTILEX, PLUS, ACQUILEX, NERC, GENELEX, SAM-A, SUNDIAL, EUROLANG, TWB, ONOMASTICA and DELIS, together with the European bodies ESCA, ELSNET, FOLLI and the European chapter of ACL. Five working groups, each supported by a hosting organisation, are envisaged for: Text Corpora, Computational Lexica, Formalisms, Evaluation, and Spoken Language Resources and Methods.

6.3.6 It is clear that, potentially, EAGLES has a very important rôle to play in driving the coordination of the European language industry and research community. This is a long term endeavour, which should long outlive the LRE programme. *It is too early to comment on how it is working, but it is encouraging that so many of the major projects in Europe are represented on the Management Board,* though there are notable omissions. However, the Board is already quite large enough.

6.4 <u>The Cost-Shared Approach</u> (C&R)

6.4.1 The ET10 projects make a clear transition from the EUROTRA approach to the conventional cost-shared project approach. There are benefits and penalties in this approach. The main benefits are:

1) Provided there is a genuinely open call, and a properly constituted and conducted peer review body, this approach provides the best way of opening work to those best qualified to undertake the work.

2) The competitive approach may bring out the best in the bidders, stimulating them to respond well to the challenges of the work plan. The main EUROTRA approach lacked external competition, even if the in-fighting over technical issues provided some internal stimulation.

3) The relatively short timescale of a cost-shared project (never more than five years, typically three) allows the work plan to be adjusted

as the field develops worldwide.

6.4.2 However there are drawbacks:

1) There is a danger in lack of continuity as a project team builds up, and then has to disperse when the next contract is placed elsewhere. The main EUROTRA teams had the benefit of ten years of continuity, which was very important in providing a stable set of participants, who grew in stature, and experience of working as a distributed team.

2) The main EUROTRA programme was able to build up a community and coherence between the teams in the 12 countries. With competitive cost-shared projects it is far more difficult to create and maintain that cooperation.

3) EUROTRA was a programme, with the individual teams playing their part in a coherent whole. Though it is possible to ensure a group of cost-shared projects use the same standards and approaches (ie work to the ALEP formalism) it is far more difficult to build an integrated system. Indeed, it is unthinkable to achieve this through a group of projects and even the EUROTRA approach suggested it was almost impossible with a set of separated Centres, and the weak powers of the central leadership that the EUROTRA mechanism entailed.

4) The involvement of the governments in the programme does not arise in the cost-shared approach. Some, but not all, governments took an active interest in the EUROTRA Programme, as they do in Eureka projects but not in ESPRIT after the project selection has been made.

6.4.3 It is a tribute to the teams in the EUROTRA Centres that they welcome the move to cost-shared projects; one might expect they would prefer to retain their privileged, protected status. Yet in the Panels' visits and in the paper "How to combine the best of the ET and LRE schemes" (see Appendix 6) the Centres have shown that they see the benefits of the cost-shared approach, as well as the penalties. The arguments in that paper deserve careful study. They see benefits in a mixture of the main EUROTRA "Contract of Association" approach together with cost-shared projects, as has been in place during the last two years of the programme. *The Panel concludes that the shift to cost-shared projects is to be welcomed for research projects, though it would not be appropriate for large development projects. But for a subject that requires a coherent attack on standards, formalisms, interfaces, etc, it is desirable to take special measures to ensure that "continuity, completeness and coherence" is retained across the teams.* This is discussed further in Chapter 11.

6.4.4 *While welcoming the introduction of individual cost-shared projects, it is important to keep a balance between the competitive project approach, and the coordination of work across the Community that the subject demands. It is unfortunate that the current LRE projects are funded for such a short period. Longer and larger projects would be more satisfactory. The number of approved projects is dangerously small in relation to the demand. The high cost of preparing projects will cause industry and other bodies to abandon the attempt if the failure rate is known to be very high, due presumably to the relatively little funds available.*

7. Outputs and Exploitation of the Programme

7.1 Outputs

The outputs of the programme fall into two classes, the artifacts that remain to be used by workers in the Machine Translation field, such as the Reference Manuals and Language Specifications, and the trained manpower that has resulted from the programme. In most cases of exploitation it is likely that it will involve some of the EUROTRA trained staff, together with the use of some of the written material. But in the long run the main impact of the programme is likely to come from the trained manpower, some of whom are likely to participate in every major Natural Language project in Europe for years to come.

7.2 The Reference Manual

7.2.1 The Reference Manual is a detailed specification for the linguistics and architecture of the system, giving detailed specifications and guidelines to the far-flung EUROTRA workers on all aspects of the mainstream system design work that was undertaken. The chapters start with outline descriptions and then go into detail of design or rules under the heading "Legislation" and are followed by more rules under the heading "Pragmatics". The seventh and final edition of the Reference Manual was issued in 1990 and runs to about 1,000 pages of close typescript. Because it is all tied to a particular system design and formalism much of it is ephemeral. The ETS formalism was never very satisfactory, and is now certainly outdated even within EUROTRA where the ALEP formalism has superseded it. But the grammar rules, with illustrations drawn from a variety of the European Languages are of lasting value. Most chapters indicate who some of the key EUROTRA workers were in that particular field and conclude with a set of very valuable references.

7.2.2 There is no doubt that this remarkable document is of very considerable value to those in the computational linguistics field. Despite the ephemeral nature of much of the details, the whole work will be a detailed reference book for research workers and system designers for years to come. Quite rightly, the Commission plans to make it available to research workers everywhere, and this approach is much to be welcomed and encouraged. *Though very much a detailed working reference manual rather than a polished text book, it is likely to be referred to throughout the world community of computational linguists and so is a lasting monument to the programme.* Of course the work is unfinished - it never will be or would have been however long the programme had gone on - and is uneven in that it reflects the variable effort directed to the various aspects of the system, to the various aspects of linguistics. Work to transfer the grammars to the ALEP formalism has started under an ET10 and an LRE

project.

7.2.3 It is unfortunate that the last version (7.0) of the Reference Manual was issued in 1990. The Implementation Reports, now coming in early 1993 from the individual language groups, do complement and extend the Reference Manual.

R13: The Commission should consider whether it is practical to prepare and issue an updated version of the Manual, for this would certainly be desirable.

7.3 Language Specifications

As a form of extension to the Reference Manual, the Language Specifications add another, and perhaps most important, element to the documentation. There will be nine, one for each of the official languages, when they are complemented by the Implementation Reports early in 1993. They are also tied to the architecture and formalism, and because they date two years after the Reference Manual are a representation of what has been run on the EUROTRA system software. The Implementation Reports describe how the Reference Manual has been applied to implement each Language Specification in the grammar and dictionaries. *For any language technologist interested in a specific language, whether for monolingual or multilingual work, these Language Specifications are of outstanding value.*

7.4 *Exploitable Computational Linguistic Property.* The most important property stemming from the main EUROTRA programme is the Reference Manual and the nine Language Grammar Rule Specification sets. These are definitely useful to a commercial new system developer. But they are essentially academic documents from which it is difficult to obtain a significant financial return. There are now many computational linguists in Europe who could reproduce the Reference Manual, and linguists in the individual countries who could reproduce the Language Specifications. Because the main EUROTRA ETS formalism is out of date the Reference Manual may be rapidly losing its value, whereas the Language Specifications will form a basis that will grow over the years. *So, while there is little of direct economic value in the output from the main programme, it does have property of considerable intellectual value.*

7.5 Software Systems

7.5.1 The EUROTRA demonstration system software developed at Luxembourg provides a framework for the demonstration of the mainstream EUROTRA work, but is not developed to be of commercial value. Certain of the EUROTRA Centres have developed versions of the EUROTRA ETS formalism that provide more efficient runnable systems, and so provide a potential route to the demonstration of the system for

particular applications. But the main output of the programme in software system terms will be the ALEP system, now available in first prototype form, but to be available in ALEP1 first release form in March 1993, with the main release in mid 1994 (see chapter 5.2). It is intended that this should be developed over time, and that the EUROTRA grammar and language specifications should be steadily converted to run on the ALEP system over the next few years.

7.5.2 The ALEP1 formalism is, compared with ETS, a modern formalism with all the advantages of being fully declarative. *So the ALEP1 tool set is likely to be of value to research laboratories, and to industrial teams who might wish to use it to assist their system developments.* This is hardly likely to provide any large market in the Community, if only because there are few firms developing MT or Natural Language systems. (The Commission seem to believe that the number of firms in the field is growing fast, judging by the applicants in the recent second call for LRE projects.) But it is a useful contribution to assist academic research. And there are many research laboratories elsewhere in the world who might be customers for the system, especially in the USA and Japan. There are said to be 20 commercial suppliers of Natural Language processing systems in the USA who might be interested in the tool kit for development purposes.

7.6 Individual Centre Developments

7.6.1 Several of the EUROTRA Centres, notably Copenhagen, Group DIMA in Turin, and IAI in Saarbrücken, have adapted the ETS formalism to produce an efficient and runnable system. Copenhagen has a commercial partner for a niche system in the form of legal firms interested in the translation of patents. Turin and Saarbrücken are holding discussions with automobile manufacturers interested in systems to translate technical manuals. The path from research work to success in the marketplace is likely to be long and difficult. *If these systems develop into commercial products this will be a very real exploitation of the work of and expertise developed in the programme.* There may well be other Centres who achieve exploitation of their skills and perhaps of some of the material stemming from the programme, probably for rather narrow niche market applications.

7.6.2 There were other outputs from the programme that have received a warm welcome from the Panel; notably that stemming from the joint work of Leuven and Turin, ELISA. This product is currently demonstrating that voice output was not entirely neglected in the programme, but language developments are eagerly awaited in the next few years.

7.7 Eurolang

7.7.1 The Eurolang programme is an interesting example of a major MT industrial programme in Europe, where one might expect to find

exploitation of the programme. (GRAAL is another example). Eurolang is a subsidiary firm of the documentation and language translation company, SITE, who are owned by the CORA group in France. The objective of the Eurolang programme is stated to be the development of a second generation machine translation system for five language pairs, namely: French/English, German/English, French/German, Spanish/English and Italian/English. The project started at the end of 1991 and is intended to run until the end of 1994 at a cost stated by SITE-Eurolang to be 489 MFF (some 65 Mecu). It is a Eureka project and the participants may receive support from their governments. The SITE group certainly does receive support from the French government, as well as the backing of their parent company, the CORA-REVILLON group. Siemens-Nixdorf are major partners in the project, along with several minor partners including the Rank Xerox company, Cap Innovation and GETA.

7.7.2 There are several EUROTRA teams receiving some support from Eurolang for work directed to build up the system, often through their knowledge of the Language Specification of their particular language. And there are a considerable number of people in the 50-strong Eurolang central team in Paris who were working in or trained by EUROTRA. *This involvement is excellent and demonstrates the value of the programme in developing the skills in this field in Europe.* No doubt yet more will be involved before the programme is complete.

7.7.3 However *it is very disappointing that there is little sign of the EUROTRA work being adopted by Eurolang.* It is currently based on an uneasy mixture of ARIANE, stemming from GETA at Grenoble, and METAL. It is true that METAL has itself been influenced by the EUROTRA work. But one would have hoped that the mainstream EUROTRA work would have been adopted; perhaps it was felt in 1991 when the decisions were being made that the ETS formalism, like the other formalisms based on the unification approach, was too difficult to adapt to provide an efficient system. Maybe Eurolang would have taken a different path had they seen the various, loosely ETS based, systems that are now running. It is also disappointing to find that the ALEP formalism and work is not employed, but for the more understandable reason that it is seen to be too immature to base a major system development round it for the moment. However, it is known that Eurolang is interested in the EUROTRA Reference Manual and Language Specifications, so it is not only through the trained staff that they have benefitted from the EUROTRA work.

7.8 Trained Manpower

7.8.1 *Probably the most important output from the EUROTRA programme is the manpower that has been trained in the techniques of computational linguistics and the particular problems of Machine Translation.* With a few exceptions the formal training courses were not undertaken directly by the EUROTRA Centres and were not provided under EUROTRA funding.

Nevertheless they were often very dependent on the EUROTRA Centres and staff, without whom they might not have been set up. If the training courses that have been established can be maintained, now that the Centre funding is ended, this will ensure a continuation of a supply of qualified manpower for the subject in Europe.

7.8.2 *At its peak in 1990, the programme was supporting 200 research workers in the 16 or more Centres spread across the Community, with at least some in every one of the 12 countries.* It is clear from the final reports that at least 380 people have worked within the 13 EUROTRA groups on EUROTRA contracts, excluding the administrative support staff. The majority of the 310 professional research workers were originally trained as linguists with a small number trained as computational linguists. Around 20% of the total had tenured positions in university or in associated institutes - the remainder being supported on renewable research contracts. Some still remain in computational linguistics in industry or universities, etc.

7.8.3 The undergraduate and postgraduate courses in computational linguistics at Leuven (KUL), at City University, Dublin, and at UMIST have been responsible for educating many students in computational linguistics. The content of these courses draws heavily upon the experience of the university staff who have worked on EUROTRA, and also utilises examples from the EUROTRA work to illustrate the various points. Many of the Centres have provided short courses, workshops, etc. For example, the 1990 European Summer School in Languages, Logic and Information, organised by Leuven, attracted 500 participants from 22 countries. Cross fertilisation programmes have taken place, eg EUROTRA-PT supported the 1989 Paris meeting on "The Portuguese Language and Translation". At the 1987 Copenhagen meeting of the ACL, members of the Greek EUROTRA team presented a morphological analysis of modern Greek developed with the Greek National Research Institute. In 1989 a meeting organised by EUROTRA-ES, jointly with the Energy and Education Ministries, was held to contact industries in Spain and brief them on new technologies in CL and MT. In Utrecht, throughout the programme, there have been close connections between the EUROTRA team and the ROSETTA team in Philips. Liege, Copenhagen and others are involved in student exchanges through the ERASMUS programme. Gruppo Dima has been involved with the Italian national computational linguistics programme. More generally, through conferences (eg Coling), workshops, Summer Schools, networking, personal contact, and publications (well over a thousand, of which a quarter are open refereed works), the knowledge of EUROTRA and its work has been diffused.

7.8.4 *It is known that people who had worked in the EUROTRA Centres, or been trained on their courses, have been involved in virtually every industrial Natural Language project current in Europe today.* The Siemens Metal project has employed EUROTRA people, as has Eurolang, GRAAL and GENELEX. Several of the senior scientists from the EUROTRA Centres

are now to be found in senior positions in the Universities and computational linguistic centres in the USA. While this represents a brain drain from Europe, it has, of course, been matched by an influx of workers from the USA, no doubt in part stimulated by the work going on in Europe under the EUROTRA programme. This exchange with the USA, and other countries like Japan, is to be welcomed and encouraged.

R14 : It is highly desirable that the supply of trained manpower in computational linguistics in Europe be maintained and enhanced. In its future support for Linguistic Engineering, the Commission should take steps to monitor the supply of trained manpower, and to assist the training programmes should that prove necessary.

7.9 Assistance for Exploitation

7.9.1 It is common wisdom that Europe is not good at exploiting the high quality research it carries out. There appears to be a tendency for countries in other continents to exploit the European research work first. So it is particularly unfortunate that the Commission programmes tend to cease, just at this key exploitation stage. It is true that the VALUE and SPRINT programmes exist to support the exploitation of research carried out under the Community's own programmes. But the scale of the funds available appears to be inadequate for the task, and in proportion to the size of the Community research budget. The Value programme is built up by a 1% "tax" on all Framework Programmes. This represents some 55 Mecu over the Third Framework period. But the funds are used to build up the infrastructure for technology transfer rather than to help projects directly: The SPRINT programme is also a technology transfer programme, outside the Framework Programme. It exists to help firms to adopt high technology, working through Chambers of Commerce and the like. Neither programme seems very appropriate for helping the exploitation of EUROTRA-based projects.

In any case, it is much better if the exploitation programme can be administered by those close to the original research work, rather than through some separate programme such as SPRINT.

7.9.2 *In the case of the EUROTRA programme several of the Centres are struggling with the problems of exploitation.* They have potential customers and firms interested in creating a product on the basis of the EUROTRA work and the Centres' expertise. But until they can see a demonstration of the work applied to their particular market interests, they hesitate to invest their own funds. The case of Group DIMA in Turin and a large automobile manufacturer is an example. The person responsible for the translation of the servicing manuals was sufficiently interested in exploiting the work of the Centre and the programme that he took the trouble to see the members of the Panel to explain his market interest. But, understandably, his firm concentrate their R&D investment in the field they are experts in,

namely automobile engineering. Until they have seen a demonstration of the Group DIMA system applied to the translation of automobile servicing manuals, the company hesitates to invest any funds. Other similar examples exist in other Centres. It is at this point that *Commission funds to stimulate exploitation are needed, but apparently are not available.* The Danzin Panel recommended that the Commission should encourage the search for industrial applications for the spin-off from the EUROTRA software environment, especially in the form of monolingual products.

R15 : The Commission should ensure that all its research programmes like EUROTRA are matched by exploitation support programmes with adequate funds.

7.10 Comparison with Original Objectives

7.10.1 In the Council Decision of November 1982 the programme was described as a "research and development programme for the creation of a machine translation system of advanced design". It was stated that "preliminary work already completed has demonstrated the technical feasibility of such a system". *The EUROTRA programme has not achieved this objective.* While it is difficult to say that it was wrong to claim that it was technically feasible to produce a system of advanced design, if only because the performance to be expected of such a system was not stated, the current evidence is that MT system performance remains dependent, above all, on the richness of the dictionaries. Indeed it is reasonable to expect that, had the EUROTRA programme led to a machine translation system which was equipped with dictionaries designed to match the system but of the number of entries of the Commission's Systran system, then the new system would have performed better than Systran due to the improvements to the grammar. However these improvements are not of a magnitude to make much improvement to the performance of the system, which will still be dominated by the quality of the dictionaries.

7.10.2 *The state of the art today, and probably for years to come, makes it a much more feasible proposition to design useful systems for limited domains where grammar, sentence complexity and dictionary size can be controlled. Had the EUROTRA Programme been aimed at such a system, making full use of interaction at the pre-editing stage to eliminate ambiguity, a system of more immediately exploitable value might have resulted.*

7.10.3 The Council Decision called for the programme to be carried out in five and a half years at a cost of 16 Mecu, including staff costs. In practice, the EUROTRA programme ran for ten years from the date of that Decision, and at a cost to the Commission of about 50 Mecu (formal budget 37.5 Mecu). At first sight the EUROTRA programme ran for nearly twice as long as originally planned at three times the cost. However, the Council cannot have expected that it would take three or more years to get the Contracts of Association agreed with the governments. Maybe it is fairer

to consider the programme starting from the end of 1985, and so lasting for seven years. Allowing for inflation, the cost to the Commission might reasonably be described as about 30 Mecu in 1982 terms. Moreover. the number of countries involved increased in 1986 from 10 to 12, the number of languages from seven to nine. So though the programme ran for longer than planned and cost more, *the actual increases are not as significant as they appear at first sight.*

7.10.4 Having made these critical statements, it is important to recognise that there have been other very significant benefits stemming from the programme, some of them described in the sections above, such as the trained manpower (7.8), the Reference Manual (7.2) and the Language Specifications (7.3) The work on the grammars has benefit for other applications in Natural Language processing work other than machine translation, which may have wide and more immediate applications. Some of the applications may be monolingual, some multilingual. In authorising the programme the Council explicitly refer to the likely impact of the programme in developing computational linguistics in the Community. *The objective of developing a stronger computational linguistic community in the European Community was certainly achieved.*

7.10.5 Over the ten or more years of work on EUROTRA progress has been made in machine translation. It would be desirable to set the work and lessons into perspective by a study of progress made over the period of the EUROTRA programme.

R16: The Commission should establish a study to document what progress has been achieved in MT over the period covered by EUROTRA work.

8. The Individual Centres

This chapter is based on the questionnaires filled in by the Centres for the Panel, and the interviews with Panel members. As such, the information in it is anecdotal in nature and has not been checked from other sources. The views expressed are compressed extracts from informal conversations and documents, and so distortions of the formal view of the Centres may have crept in. However, it is felt to be useful as providing some indication of the views of workers in the field, and of the issues in the programme and its management that worried them. It also serves to illustrate the way that the programme has built up teams and institutions, sometimes from nothing, in all the countries.

8.1 EUROTRA Leuven

8.1.1 *History.* Leuven has been involved since 1978. At this time none of the four major universities in the Dutch speaking part of Belgium had a programme in Computational Linguistics (CL) although they all had linguistics departments. There were no Belgian (public or private) initiatives in MT at that time. The Applied Linguistics Department within the Linguistics Department has taught CL since the end of the '70s. In 1984, during the study phase, Leuven was involved with the Coordination Group. In October 1984 the Katholieke Universiteit Leuven was awarded a CoA to work as part of the Dutch language group (with Utrecht): three researchers then; seven people at peak in 1988 - 1989. There was good cooperation with Utrecht. Leuven was awarded Addenda (to take part in Central Teams) to the CoA from early 1985 on.

8.1.2 *Leuven CCL.* The Centre for Computational Linguistics (CCL) was created in 1991 as an institution of the Katholieke Universiteit Leuven (KUL). The aim of the Centre is to promote research in the areas of computational and formal linguistics and applications of this research in language processing. It is currently involved in Computational Semantics (CS) in EUROTRA II (ET-10/61; coordinator), and LRE (LRE-62; consultant). This work directly utilises many of the discourse and semantic skills generated as part of the Leuven Dutch language activities on EUROTRA. In addition, the CCL is working on NLP projects for various funding agencies such as the Belgian National Fund for Scientific Research, and AIM. Leuven organised and ran the 1990 European Summer School in Language, Logic and Information. Many of the results of scientific research, especially semantics, have been reused in different systems.

8.1.3 *Staff and Related.* Recruitment at KUL was relatively easy to handle. The ERASMUS scheme created three positions for students (two semantics, one syntax, in Dutch). The Leuven CCL has developed directly from the EUROTRA (and others) teams in Leuven. Its scientific staff currently

comprises 30 persons, whose qualifications cover CL and related disciplines. Also in 1991 - 92 an additional six part time staff were employed on EUROTRA.

8.1.4 *Leuven's Views.* The main achievement of EUROTRA is the linguistic specifications (cf Reference Manual) and the application to nine languages. EUROTRA has the latest unification based formalism (the virtual machine), but has out-of-date implementation (Prolog). As for dictionaries, the aim was to describe 2,500 lexical items (one corpus for nine languages), supposed to be extended to 20,000 in 1991 - 1992. The decision was made to allow each language group to find its own language corpora - NL/B chose semi-popular text on telecommunications - Leuven say this approach never really worked. Teams in the Final Transition Phase knew their work would not be used which was demotivating, only research clusters looking at monolingual research were allowed to use the new ALEP formalism. The dictionaries can be converted to the new ALEP formalism, but not the grammars - yet this is the part which received the most attention in ET-10 and LRE I. Leuven say that the ET-10 selection was not in conformity with the CEC's Request for Proposals, eg research in morphology, syntax and semantics was required, but very different projects were chosen in the end, such as a statistical approach for dictionaries. As an alternative to the EUROTRA programme Leuven suggest that more realistic goals should have been set, such as the development of grammar or style checkers or MT for restricted sub-languages. Collaboration with colleagues in the rest of the world would have been valuable, LRE is not a long term programme. Exploitation will probably take place through the CCL, under Comett and COST schemes. Leuven's main achievement is the integration of model-theoretic semantics in MT, their ideas have been adapted by several other projects.

8.2 EUROTRA Liege

8.2.1 *Belgium-Liege.* The CoA for Liege was signed in April 1986. The signature of co-funding came into force between the Belgian State (Minister and Secretary of State for Scientific Research, and their department, SPSS) in October 1986. In the mid 1970s, the Liege team had pioneered work on machine readable dictionaries, in particular the Longman Dictionary of Contemporary English. Major publishers were interested in encouraging academic research on improving dictionaries, and, more broadly, on assessing the reusability of their lexical resources in MT and other fields. Liege retained this specialism throughout the EUROTRA years, but never succeeded in influencing the other EUROTRA Centres, nor the Liaison Group, to significantly examine the reusability of lexical resource issues. (This has in fact been taken up within LRE by other organisations.) Discussions took place with the Nancy group, and the Leuven group: the decision was made to fund Liege on two fronts - working on the French language monolingual aspects (receiving 8% ie 240 Kecu, of the French language funds) - and work on

computational lexicography, including work on terminology to be carried out in collaboration with the Irish and other teams (value 210 Kecu). Taking into account national government funding, the formula allowed for an annual budget of 150 Kecu, which provided for a small team of one head, three researchers and four half-time researchers through to 1992. Liege never seemed to be fully integrated into the EUROTRA network, and lost a major opportunity to increase their influence on EUROTRA when the Liaison Group turned down Liege's proposals for work on frames for terms (ie integration of terminology). In addition, the promised liaison with Dublin never seemed to take off. A consequence of this has been the growth of frustration and disillusion within EUROTRA, and now LRE, and EUROTRA's influence on Liege's future programmes will surely wane. Linguistic osmosis from the other Centres does not seem to have occurred in Liege, and it is hard to judge what effect EUROTRA has had on CL and NLP within Liege.

8.2.2 *EUROTRA Influence.* The team has carried out research in the field of lexicography and terminology. The team has liaised with the Irish on terminology, but perhaps due to the lack of precise assignment from the Liaison Group, this didn't work out as expected. In 1986, a new postgraduate programme within the "Faculté de Philosophie et Lettres" was created in MT and CL, and then in 1988 a new postgraduate course on "Lexical relations and databases" was created. Although posts were created through EUROTRA funding, the demise of funding means that these posts will disappear. The University of Liege does not appear to have regarded the work of the Liege team as an opportunity for growth in CL and NLP, and EUROTRA appears to have been viewed as an isolated project rather than an opportunity to grow the scope of the department. The team has now reduced to the original pair of University academics now in the EMIR project (see below).

8.2.3 *Liege's Views.* Liege believes the dictionaries were neglected throughout the EUROTRA work. The work that Liege wished to pursue, as apparently detailed in their CoA, was not done. Liege wished to examine the fundamental problems of portability in lexicography - addressing the question "How do you go about producing a dictionary for machine translation - in an innovative manner?". Liege felt that EUROTRA had too much of a tense aspect and too little of the drudgery and painstaking introduction of new dictionaries. They believe in future there will be an even stronger break between MT and lexicography. Liege had little contact with Eurodicautom. Via another project, Liege still has contact with ISSCO. Liege will apply for an LRE project with the University of Bonn (leading), with lexicographical work from British National Corpus, private companies, academics (Liege, Bonn, Copenhagen Universities, etc). Liege has been involved in the EMIR ESPRIT project headed by CEN (Saclay) dealing with research on NL front-ends for querying multilingual documents. Liege never got involved in pilot corpus studies within ESPRIT - it is only now that others have developed lexicographers'

workstations to deal with corpus work, to develop a dictionary from a large body of text. The COBUILD people in Birmingham have pioneered this kind of work - for MT and language processing the need is for more than just dictionaries - in translation, Liege's interests include the "dustbin" of publishers' dictionaries - some of the material they have to throw away because they have no space to enter the material.

8.3 EUROTRA Denmark

8.3.1 *History.* Denmark signed the CoA in October 1984. Denmark had participated in EUROTRA since 1978, and researchers at the University of Copenhagen had participated in study contracts. EUROTRA-DK had its offices in the University of Copenhagen, and all administration was done by the University administration. However, the unit was not an institute of the University but an independent research unit, managed by a Board drawn from various Danish organisations. Then in 1991 the Centre for Language and Technology (CST) was formed and this acted as an umbrella for the EUROTRA work. The funding of CST comes from national research funds, Nordic research funds, EC research contracts and increasingly, funds from commercial organisations. The EUROTRA work acted as a catalyst for work on the Danish language. Through the Liaison Group, chaired in recent years by Bente Maegaard, CST has exerted considerable influence over the work done in the various Centres. The major achievement of EUROTRA-DK has been a detailed formal linguistic description of the Danish language; this includes the running grammar and dictionary, and also the research that preceded it, in particular, valency theory for Danish, lexical semantics, morphology, the use of field grammar, description and the creation of a lemma dictionary. During the Transition phase concentration was made on the English, French, Italian to Danish language pairs, and it is the first of these which is being exploited in the PaTrans project. The success in training staff is indicated by the 34 research workers that have been involved in EUROTRA-DK. In addition, the influence on European CL work through the broad publications list must be significant.

8.3.2 *Copenhagen CST.* The Centre for Sprogteknologi was established in 1991 as a non-profit making public institution under the Danish Ministry for Research and Education. CST carries out research and development within the field of NLP both under national/international research programmes and as contractual work for private companies and public institutions. CST's staff currently comprises 17 persons, whose qualifications cover computer science/engineering, computational and theoretical linguistics, lexicography, knowledge representation, Danish and most other EC languages. CST has built strong relationships with a number of organisations including ISSCO, Geneva; HCRC, Edinburgh; SRI, US; the Prague School. A glance through the publications list for CST indicates that there is a good mix of internal CEC articles, conference reports, refereed publications in international (mainly English language,

some German language) journals, plus, notably, a number of popular articles raising awareness of the scope of CL/NLP/MT. At the moment CST is involved with: OFT "Translation of Technical Texts", (Danish National Fund); DIALOG (development of application-oriented dialogue systems with text and speech input), (DNF); EUROTRA, ET/10 "Semantic Analysis, Using a NL Dictionary"; ESPRIT "Network of Excellence 3701 in Language and Speech - NELS", (CEC); joint research project with HCRC, Edinburgh, and SRI Menlo Park, USA, on "Methodologies for Constructing Knowledge Bases for Natural Language Processing Systems". In addition EUROTRA-DK is involved with a Danish patent company for the PaTrans work - the first exploitation of the ETS formalism. Also CST has completed consultancy work with Canon Europa on aids for translation of manuals from English into the European languages.

8.3.3 *Copenhagen's Views.* Three demonstrations were shown to the Panel including: a demonstration of the treatment of modality in the EUROTRA system with special reference to epistemistic and deontic modalities; PaTrans work was also described and demonstrated. A prototype translating patents from English into Danish was run, as well as a system for the encoding of technical terms. The translation was based on an adapted version of the EUROTRA grammar augmented with guesses when a particular word wasn't found. Some of the points that were made include: the Engineering Framework was considered to have moved too fast into the ALEP system; CST is currently optimising ETS and the grammar; the Centre was of the opinion that the system could be generalised in the sense that it could be moved from one domain to another; an issue like ellipsis had been treated to a very limited extent; the treatment of optionality was mainly restricted to grammar; some work had been done on support verbs ("make an attack" etc) and on semantic features; there was an ongoing experiment on so-called relayed transfer involving English into Danish and further into French.

8.4 EUROTRA Spain

8.4.1 *History.* The EUROTRA-ES research unit in Spain comprises the Universidad de Barcelona (UB) Fundación Bosch Gimpera (FBG), and the Department of Logic and Linguistics at the Universidad Autónoma de Madrid (UAM). The CoA was signed on 27th December 1986. The establishment of the teams took considerable time, and involved two ministries (Education and Industry). FBG was a University Institute created to mediate between the UB and industry, and became the administrative manager and representative of the EUROTRA-ES group. In early 1986 the first team was established, comprising five researchers with linguistic background, but the operational start of EUROTRA-ES should be taken as August 1987 when payments were eventually received. In addition, at this time, two other projects on MT (METAL from Siemens and ATLAS-II from Fujitsu) began development on Spanish monolingual modules. The original two leaders of the EUROTRA team moved to join

Siemens and Fujitsu projects. In December 1987 the UAM team was created as a subcontractor to UB, for the development of lexical and terminological tasks and morphological modules. The aim was to spread the academic and scientific benefits of participation in EUROTRA as widely as possible in Spain.

8.4.2 *Barcelona GILCUB.* The GILCUB (Grupo de Investigatión en Lingúistica Computacional de la Universidad de Barcelona) was established in 1987 as a university institution to the University of Barcelona. Its administration is looked after by the Fundación Bosch Gimpera which was created as a group for administrating contracts between the University and industry. GILCUB has been carrying out research language processing under European and national research programmes and for private companies. Since its constitution, GILCUB has been involved in EUROTRA, ET-10/52, a contract with IBM Spain "Linguistic Specifications for the system MAT-IBM (90/91)", Integrated Spanish-British Actions 066 (with UMIST, and sponsored by the Spanish Ministry of Education and Science), Eureka Eurolang (EU676), LRE-1/029 LS-GRAM.

8.4.3 *Barcelona's Views.* The teams were developed completely from scratch. At peak they had 30 people in 1989 - now they are 14. They accomplished a good selection of grammar, and a reasonable dictionary. GILCUB is trying to start an institute of linguistics engineering. Members of EUROTRA-ES will be working for Eurolang. They believe ALEP is not suitable for Eurolang, but useful for research work. They said that EUROTRA is a translation system that, when it works, is better than others - however, when EUROTRA fails it fails badly. EUROTRA-ES are very enthusiastic about the contacts they have developed in Europe and the US. They believe they have done good monolingual work. All the staff came from a background of the rather pure approach that is taken in Spanish academia. They felt that the Liaison Group was too far removed from the workers and the right of veto of CEC was viewed by EUROTRA-ES as a negative aspect of management.

8.4.4 *Madrid University.* The School of Language Industry of the Fundación Duques de Soria and the Sociedad Estatal del Quinto Centenario was created in 1990, as a direct consequence of EUROTRA, and it appears the EUROTRA team has been encapsulated within this. Some 14 people have been trained by the Centre, of which two currently remain in the field of CL/NLP. The comment has been made by Madrid that LRE eliminates them for further work on EUROTRA related areas. Two Madrid staff have been developing linguistic specifications for IBM's MT project MAT. Another member participated in the evaluation of ATAMARI for the Junta de Extremadura.

8.4.5 *Madrid's Views.* Madrid has been part of the EUROTRA-ES team, as managed, at least commercially, but also it appears technically, by Barcelona. This has certainly led to some frustration within Madrid. However, achievements within the period 1990 - 1992 do seem to have been significant. The so-called "External Dictionary", a monolevel repository for words independent of the EUROTRA system, was built. A number of software tools have been implemented in order to produce EUROTRA dictionaries from the External Dictionary and vice versa. As a consequence of the exhaustive studies on Spanish derivational and compounding morphology, and inflectional models of the Spanish nominal and verbal paradigms, there exists a complete implementation of the Spanish inflectional morphology based on the Item-and-Arrangement theory. Madrid have also established criteria for the identification of terminological units relating to EIRETERM.

8.5 EUROTRA France

8.5.1 *History.* Before 1985 Professor B Vauquois, Director of GETA (Groupe d'Etudes sur la Traduction Automatique, a CNRS research team located in Grenoble) was one of the initiators of EUROTRA. The ARIANE prototype was considered as a basis for EUROTRA, but rejected around 1984. Meanwhile ARIANE was developed into a national project. The CoA was signed in 1985, when two teams were given the responsibility of working in EUROTRA: the Laboratoire de Linguistique Formelle (LLF) in Paris; the Centre d'Etudes sur la Langage et la Traduction Automatique (CELTA) in Nancy. These two centres ran into difficulties since they were centres of excellence in descriptive and theoretical linguistics, but not in CL. In 1987, following discussions between the CEC and CNRS, the Laboratoire d'Automatique Documentaire et Linguistique (CNRS LADL) in Paris, and GETA, were added to rectify the lack of CL skills. However, it was subsequently decided that the work should be focused in LADL and CELTA. This was the case from 1988 - 90, except that LADL was relocated in Paris and became the research group TALANA (Traitment Automatique du LAngage NAturel). The Paris team dealt with the analysis and generation of French, and transfers from the southern languages to French. The Nancy team dealt with the northern languages. Both teams worked in collaboration with EUROTRA-Liege who were responsible for terminology and lexicography. Despite the difficulties mentioned above, EUROTRA-France was going well at the end of 1990 and was well supported by the CNRS. An official demonstration organised in Paris in February 1991 attracted one hundred industrial and university specialists. The Contract of Association for 1991 - 1992 suffered some delay due to CEC administration and arrived for signature in May 1991. Changes in CNRS meant that the CoA was only signed in April 1992, ie 16 months after the beginning of the work and eight months before its end.

8.5.2 *TALANA's Views.* The consequences of contractual delays for EUROTRA-France were of course catastrophic. Several times, the team

envisaged having to stop working, however they kept on going as best they could and they even organised the 1991 annual workshop. Many members left, but luckily they easily found jobs in industry thanks to their EUROTRA experience. Not all of the work that was foreseen in the 1991 - 92 programme of work has been achieved but this can be said to be an exploit taking into account the work conditions. Again this bears witness to the loyalty of individual EUROTRA-France group members. The CNRS had signed the CoA only on the condition that the team would be dissolved in December 1992. Therefore, EUROTRA-France effectively disappears as such at the end of the year. The Nancy researchers will be integrated into another CNRS institute where they may not work on computational linguistics. The Paris team moves to TALANA. EUROTRA-France appreciated the research quality, the fact of working with European researchers from various linguistics schools, and the training they received. They wrote a substantial French grammar and are proud that French is part of the official EUROTRA demo.

8.6 EUROTRA-Germany

8.6.1 *Germany-Saarbrücken.* MT has a relatively long history in Germany: the University of the Saarland started a project in this field in the mid sixties. The efforts proceeded on the basis of a Special Research Unit "Electronic Language Research" (funded by the German Research Foundation DFG) which ended in 1986 and from which the SUSY system and all its descendants and variants originate. Other universities in West and East Germany also carried out research projects in MT. (ConText at the University of Heidelberg, maybe the best known and the most theoretically oriented one.) CL existed at the same time at several sites, eg in Hamburg, Bielfeld, Berlin and Stuttgart, mainly on the basis of personal interest of professors in Linguistics or Computer Science. On the industrial side Siemens started to sponsor the development of METAL at first carried out largely at Austin University, Texas. Thus at the time EUROTRA started, there was already a broad background for MT in FRG, although systematic research was restricted to the small unit in Saarbrücken where some tentative applications of SUSY derivations were carried out as small BMFT projects - this is where the EUROTRA-D Centre was set. Other universities were asked to offer subcontracts to Saarbrücken. In order to have a flexible administrative structure for EUROTRA-D a new institute IAI (Institute of the Society for the Promotion of Applied Information Science) was set up in Saarbrücken. The EUROTRA-D project was the major project of IAI at the beginning. During a short initial phase at the start of 1985 the people were hired, and IAI's infrastructure was created. The operational start of the EUROTRA-D project dates from June 1985 when seven people started work. The team structure soon looked like: 6-7 linguists and computational linguists, 4-5 translators, 1-2 computer scientists. Recruitment was from the Special Research Unit and general advertisements. Training was accomplished by attending the various EUROTRA beginners courses, summer schools.

Existing morphological analysers for German (for example in Systran, SUSY, etc) are mostly programmed directly in the lower level procedural programming languages (a more fully declarative approach was used for EUROTRA). Two versions of German morphology have been developed: one which treats only inflectional endings and is integrated fully into the EUROTRA prototype; and a second one which tries to perform full morphological analysis of prefixes, derivations and composita (used for experimental purposes). EUROTRA-D contributed significantly to the Liaison Group's Problem Office proposals for a system of semantic relations and for a network of semantic feature categories, and both are considered as topics for future activity. In addition, members of EUROTRA-D were active in the Dictionary Task Force. In 1986, EUROTRA-D were involved with speeding up the EUROTRA software and implemented a first prototype of EDB, the lexical database. These efforts were stopped in January 1987 as ETS was adopted. At IAI, work on the CAT formalism continued and led to CAT2. This formalism marks the change from a unification-based formalism to a constraint-based formalism. In 1990, a user-oriented interface with alphabets for nine languages and a lexical tool for the building and maintenance of larger lexicons, using the graphical tools available on Unix workstations, has been developed. Finally, IAI is active in LRE and other projects and remains a strong centre of CL activity despite the close of the EUROTRA programme.

8.6.2 *Saarbrücken IAI.* IAI was established in 1985 for the realisation of EUROTRA-D. It is a private institute, associated with the University of the Saarland, and is mainly active in R&D projects in the area of NLP (including MT), and in developing complex information systems. IAI is a subnode in the ESPRIT NELS (Network of Excellence in Language and Speech). IAI has been involved in the following nationally sponsored projects: EUREKA's Eurolang; Knowledge-Based MT; Verbmobil speech translation feasibility study. IAI is also involved with EUROTRA, ET-6/2, ET-7, ET-9, ET-10/52, ET-10/66, LRE61-029 LSGRAM. Cooperation in a burgeoning US MT programme has begun. IAI cooperates with a small company (STS) providing a translation service on the basis of post-editing for database materials (titles and abstracts). The government of the Saarland is funding smaller studies on aspects like knowledge-based MT and special problems in German-French translation. IAI intends to play a major rôle in MT, information retrieval and expert systems, and has focused on technological transfer between university and industry. IAI will continue to cooperate with the Universities of Stuttgart, Berlin and Hamburg, but there are also new links being forged in the former East Germany units which will help form partnerships in Eastern Europe. Whereas Stuttgart has the task of maintaining links with Japan and the US, Saarbrücken will remain the central German link for activities within the Community. Of the 100+ papers published by EUROTRA-D, about half are in English, many have been presented at Coling, many have been presented at various working parties across Europe, and about a fifth may

42

be considered to have been published in the open refereed literature. IAI staff members have consistently contributed to (on average) over half a dozen workshops every year since 1985.

8.6.3 *Saarbrücken's Views.* The EUROTRA-D group at Saarbrücken is one of several MT research centres in the FRG - the others (in Berlin and Stuttgart) are only funded by BMFT. Bonn University was subcontracted to Saarbrücken until 1990. Although there were different goals and priorities between the teams, there was a strong cross-fertilisation. The total value of funds from the CEC was 1.4 Mecu plus 4.45 Mecu from BMFT for Saarbrücken. The other projects were funded with an additional 3.5 Mecu by the BMFT. Saarbrücken would have preferred to concentrate on certain areas: to focus first on an efficient prototype, then to restrict on a limited number of languages and language pairs and finally extend to all Community languages and resulting pairs and large dictionaries. Saarbrücken's estimates were five years to extend the system beyond its current limitations. On the sideline CAT2, which was created as a consequence of the different priority views, there were some interesting issues: different kinds of linguistic approaches than the EUROTRA one, more user-friendliness, all with a view to taking it into industry. CAT2 has about 15 installations (five in the FRG) in universities and scientific organisations for research purposes. A pilot application project with a big software company is underway. Such projects and consultancy for commercial MT system manufacturers have brought in about 250 Kecu already. The main intellectual work in EUROTRA has been the contrastive NLP work with collaboration between the various participating groups, resulting in extensive documentation in the Reference Manual which is now used as a basis for major industrial development projects. As for the organisation of EUROTRA, Saarbrücken would have preferred a less "democratic" leadership, setting reasonable goals on well funded and well known bases. It was not a good decision to separate the software group in Luxembourg from the research teams in the Centres. The Luxembourg team was - at least during several years - too limited in computational linguistic skills. Saarbrücken's future priorities would include the further development of a comprehensive set of semantic features and rôles, the integration of conceptual knowledge and context as well as the integration of larger dictionaries.

8.7 EUROTRA-Greece

8.7.1 *History.* The CoA was signed for Greece in 1985. The first contract was for the creation of a specialised group, and this was established in the University of Crete, with assistance from Athens. The preparatory phase of the project was dedicated mainly to basic research, as the theoretical basis for the modelling of Greek was scarce and the implementation strategy was not yet decided. This phase ended in August 1985 and the team proceeded with the implementation of Greek formal grammars. In 1989 the whole activity of the project moved to Athens together and at the

same time a new director was nominated. The Greek team quickly managed to overcome the difficulties of lagging behind some other Centres in monolingual work, and modules for translation for all languages except Dutch and Danish have been developed. The monolingual dictionary includes 17,000 entries, and the bilingual dictionaries around 2,500 each. The Greek team has also developed a terminology databank of 7,000 telecommunications terms. One of the sidelines has been the statistical software package PROTIMISI which deals with overgeneration. Another is a dictionary construction package under MS-DOS named EUROLEXIS. The Greek EUROTRA team, together with the Speech Processing team of the National Technical University of Athens have been responsible for creating a new institute.

8.7.2 *Athens ILSP.* ILSP (Institute for Language and Speech Processing) was established in 1991, as an institution under the Greek Ministry of Industry, Energy and Technology, General Secretariat of Research and Technology. Its main purpose is to act for the development of speech and language technology in Greece as well as of the critical mass of human resources. It has an industrial orientation. ILSP's activities comprise research in CL and Machine Translation; Lexicology and Lexicography; Signal and Speech Processing/Synthesis/Recognition; development of tools (eg machine readable dictionaries, language checkers for Greek, etc); development of platforms in these areas. ILSP is coordinator of the nationally sponsored STRIDE framework project LOGOS and of the LRE 61-016 project TRANSLEARN. It also participates in the ESPRIT-FREETEL project (hands-free telecommunications devices) and in ET-10/63. It is a national node in the ELSNET network, participates in the NERC project, is starting its participation in the GRAAL project and is in close contact with the Text Encoding Initiative. Several demonstrations of the Greek grammars and dictionaries have been held in Greece and elsewhere. In June 1990, in Luxembourg, a demonstration of the Spanish-Greek module with a dictionary of 130 words was successfully given. Two Irish and one German students (scholarship holders) have participated in the EUROTRA-EL work.

8.8 EUROTRA-Ireland

8.8.1 *History.* Ireland joined the EUROTRA project in December 1984 when the CoA was signed by the National Board for Science and Technology (NBST). Initially, EUROTRA-IR was based at the NBST headquarters but relocated to the premises of the Institute for Industrial Research and Standards (IIRS) when IIRS and NBST merged to form EOLAS. In September 1988, the project relocated yet again, this time to Dublin City University in Glasnevin and responsibility for the project passed to the University. At this time, Ireland had little experience of CL, and there was no readily identifiable centre for CL - the original plan was that linguists would be seconded to work under NBST. The task allocated to Ireland at that time was more appropriate to people with a background in

translation and terminology. EUROTRA-IR became the terminology centre for the EUROTRA project - this led to early marginalisation of Dublin, until the importance of terminology was recognised by the other EUROTRA teams. The three areas in EUROTRA-IR's CoA were: terminology, sublanguage, text typology and classification. The work from 1985 - 1988 focused almost exclusively on extraction of terms and compilation of glossaries. In the third phase terminology policy was established by an external monitoring group, and Dublin's main function was to coordinate terminology work. In the past four years Dublin has been extremely active in the field of sublanguage. City University's significant contribution to EUROTRA was the design and compilation of a 10,000 English telecommunications terminology database (EIRETERM), with coverage to varying degrees for the other languages. This work was done in collaboration with Eurodicautom and the other Centres. City University are now looking for ways to exploit this facility.

8.8.2 *EUROTRA Impact.* The EUROTRA team has benefitted greatly by building relationships with the School of Computer Applications and the School of Applied Languages at Dublin City University. A group for MT has been established, bringing together people working in disciplines as disparate as languages, electronic engineering, psychology and computer applications. Furthermore, as a direct result of EUROTRA, a new undergraduate degree in Applied Computational Linguistics has been established. (In addition some research is being carried out on the reusability of lexical resources at the University of Limerick, and on lexical issues and the Irish language at Queens University, Belfast.) The group has submitted a proposal for LRE II terminology, sublanguage and CALL funding. Dublin is a centre for software localisation, and the EUROTRA team has been in regular contact with Microsoft, Lotus - future work may well follow. There are plans to make the EIRETERM database available to students through the library, and perhaps to the public through on-line access. Discussions are ongoing with Coiste Teirmiochta, the terminology committee of the Irish language who have a substantial database of Irish-English pairs to explore how EUROTRA-IR can become the national centre for terminology storage. In addition, there is ongoing work in the sublanguage area of knitting patterns. The EUROTRA team have expressed considerable regret that an opportunity for them to become involved in the linguistic aspect of EUROTRA, through analysis of the Irish language, has been missed. At this time METAL is being considered as a suitable translation tool, and discussions are ongoing with Siemens Nixdorf about the development of Irish dictionaries. The EUROTRA The EUROTRA work in Dublin City University has acted as a catalyst for further NLP work in Ireland as a whole, and there will be greater contact with the other centre for linguistics in Limerick in due course. The creation of undergraduate and proposed postgraduate courses in CL has begun to attract students from overseas. An international terminology seminar for terminologists and telecommunications engineers was organised in 1989 for representatives from all EC

countries. EUROTRA-IR has established themselves as a Centre for sublanguage research in Europe. They have also worked with DG XXI on multilingual harmonisation of customs tariffs - they designed the thesaurus. LRE II proposals have been submitted, but there is a funding gap - through which Dublin City University will have to cross to maintain continuity of its EUROTRA team.

8.9 EUROTRA-Italy

8.9.1 *History.* Gruppo DIMA, University of Pisa and ILC (Instituto di Linguistica Computazionale, Pisa) were involved in preparations for EUROTRA throughout the period 1978 - 1985. Professor Zampolli as head of ILC was the official head of the Italian group, but he worked closely with the head of Gruppo DIMA, Cesare Oitana. Gruppo DIMA is an association of applied research in the field of CL. It started in 1975 but was officially established in 1979. Its work was initially mainly concerned with comparative lexicography, but from 1977 its programs were directed towards syntax and semantics. In 1984/85 the Group built a PC-based analysis module for the Italian language for Olivetti SpA. Gruppo DIMA, as well as being involved in EUROTRA has been working on the national research programme for CL in 1987 - 1990. In September 1989, Gruppo DIMA decided to optimise the official EUROTRA framework and produced the sideline E-Star. Collaboration with the University of Pisa continued and from the practical point of view Gruppo DIMA and the University team are seamless. Most of the University members are on contract to the ILC, which is an institute of the CNR (National Research Council). From the start, linguistic research has been the main activity of EUROTRA-Italy: morphology, syntax, terminology, lexicography and semantics.

8.9.2 EUROTRA-Pisa, as a task force of ILC, will exploit the know-how and experiences acquired within EUROTRA, by participating in lexicography work in national and international projects. Gruppo DIMA will contribute to the promotion of CL by designing and implementing applications projects for public institutions and industries. Besides MT, the main applications are expected to be syntax checkers, training and learning systems based on natural language interfaces, automatic extraction, storage and retrieval of information, CALL, etc. Pisa is involved with both LRE I and LRE II activities.

8.9.3 *DIMA's Views.* The DIMA Group continues to see whether they can spin off companies post EUROTRA. There has been no direct support from the Italian government. DIMA regretted the shift to the new ALEP formalism, when they could have focused on exploitation of a version of ETS. The group faces extinction now. They felt it was a scandal that the CEC did not have the funds to exploit the EUROTRA work. A potential user said that they would put money into the exploitation if the team could show that the work would yield useful results.

8.10 EUROTRA-Luxembourg

8.10.1 *History.* In January 1984 discussions started between Luxembourg and the Commission, and in June 1984 the CoA was signed. CRETA (see below) was created for EUROTRA by the European Institute for Information Management (IEGI) in August 1984. The team grew to four full-time staff in 1986. In 1989 the IEGI president was appointed to head of EUROTRA-Luxembourg on IEGI's closure, and the deputy head of the EUROTRA team became head of CRETA's research unit in 1990. Since then the University has assumed responsibility for the six CRETA staff. CRETA's early work was on classification of the EUROTRA documents, with a view to facilitating their archiving, retrieval and dissemination. An on-line documentation database (Basio on MicroVAX II) was set up in 1988. The literature database (ETIN - EUROTRA Internal) then comprised 2,900 full text or bibliographically analysed and abstracted EUROTRA documents. CRETA assumed all the tasks linked to the acquisition and distribution of the EUROTRA software from June 1987. A help desk was provided for the EUROTRA teams. In the transition phase CRIS (CRETA Information Services) with three on-line databases was made available: ETIN contained 10,000 full text internal EUROTRA documents, external documents related to NLP or references to those; COLI (conferences database); ETUS (EUROTRA contacts). CRETA took on additional activities in testing and software clearing.

8.10.2 *Luxembourg CRETA.* CRETA (Centre de Recherches et D'Etudes et Traduction Automatique) was created in 1984 and is legally integrated in the CRP-CP (Centre de Recherche Pubil - Centre Universitaire de Luxembourg). Its purpose is the organisation of R&D in the field of technical science in the public sector, technology transfer and the technical cooperation between the private and the public sector. Its principle activities are serving as a documentation centre, clearing house for software and linguistic data and as a test and reference centre. CRETA participated in EUROTRA and ET-10. Outside EUROTRA there are one of two examples of use of EUROTRA material (Upsaala - outcrop of Denmark's work, Paris - Japanese French). The CRETA institute was dissolved by the Luxembourg government at the end of 1992. ET-9/2 software maintenance (being done by PE) was not available to the Luxembourg group because it was not research. The Association for Information Translation Services was started in July 1992 to promote information and user exploitation of MT and related work. It is looking for partners to form projects, to give courses and seminars. Two proposals for the VALUE programme have been prepared - one of these is for an exhibition booth at the Hanover event in 1993 to show aspects of machine translation - the other is an information server for language industries (more than 10,000 entries are stored on the database - articles on MT, etc) as an extension of the EUROTRA work on this which is coming to an end.

8.10.3 *Luxembourg's Views.* Several team members originated in Saarbrücken, and on their move to Luxembourg, they set up the document collection and dissemination centre. Abstracts were written and added to the database. Some other work included evaluation of software - grammars and dictionaries were sent to the Centre and evaluated, and later packaged and distributed to other Centres. Also the team worked on methods for AI of machine translation, eg Dublin liaison, and other ET-10 proposals - ET10/66. (EUROTRA itself had some AI but it is difficult to identify and extract this work.) In July 1988 software development started. The team was integrated with the CEC's software development group. There was also a software development group in Saarbrücken, and before that in ISSCO. It became clear from 1984 that Unix was an appropriate operating system. This became more complicated as the range of Unix architectures developed: eg Netherlands with DEC, Denmark with HP, etc; the DEC stations were faster but could not run YAP. This variety of platforms did lead to problems when the Prolog compiler was obtained. Luxembourg handled the licensing, and developed the user interfaces. In July 1987 Luxembourg became involved in software support. It would have been a good idea to have had some linguistic work in the Luxembourg EUROTRA team - but there are no Universities in Luxembourg, and this would have been difficult to put into practice. The Luxembourg team did however liaise with the other Centres and Universities (eg University of Saarbrücken).

8.11 EUROTRA-Netherlands

8.11.1 *History.* When the Netherlands became involved in EUROTRA around 1980, CL was already established within Dutch Universities; most literary faculties had regular courses in programming and linguistic computing, and MT was already under active exploration (Rosetta - Philips Research Labs, Eindhoven) or in preparation (DLT - BSO, Utrecht). In 1980, the founding members of EUROTRA approached researchers from the Technical University of Delft and the University of Utrecht - until that moment the interests of the Dutch language had been taken care of by the researchers from KUL Leuven. From 1981 - 1984, work on the study contracts concerning the Dutch language was done on a collaborative basis between Leuven, Delft and Utrecht. Participation in topics or design-oriented contract work took place on a personal basis, and staff from Utrecht took part in semantic research, linguistic specifications and framework design. Throughout this period efforts were made to establish a joint EUROTRA Centre for Belgium and Holland - this failed. After 1984 when Belgium signed their CoA, and the language specific study contracts were at an end, Delft left the project. In the course of 1986, STT (see below - the Foundation of Language Technology) was created to act as the EUROTRA agent, and the CoA for the Netherlands was signed in September 1986. The STT was not fully staffed until 1989.

8.11.2 *Utrecht's Views.* At the start of EUROTRA, Utrecht were not really involved in MT. Utrecht believe it a mistake to make MT behave like a human translator - the ultimate goal should be redefined as trying to overcome the translation problem, not to simulate humans. EUROTRA chose not to go for Machine Assisted Translation Systems (eg dictionaries) but to go for human replacement. The French (ARIANE) and the Germans' (SUSY) seemed to believe that it could be done. In the early 1980s it was too early to bring industry in to do something that had not been invented yet. The first EUROTRA workshop in France (Harry Somers, Maghi King, etc) did not involve Utrecht, who first became involved at the second EUROTRA workshop in Bangor in 1980, and started active participation in 1981. Much intellectual work took place 1980 - 1986. People from many Centres were contracted by the CEC (especially UK, ISSCO, Netherlands) to form specifications - the so-called Central Team (ca 10 - 12 people full time active). Their main task was to draw up the formal and linguistic external publications which were exciting at this time - MT was seen as a linguistic problem. There were some CL people around - ISSCO imported Prolog into the project. Many people were not really linguists.

In January 1987 the CEC said what had been produced was not implementable - developers of the specifications told the Liaison Group that another six months was required to make it efficient. CEC said go for a sufficiently implementable approach in a few months (ETS) countering the declarative proposals by the central team. This cut off the CAT formalism proposed by the Central Team - subsequently developed into MiMo almost fully declarative system. MiMo2 was based on HPSG like ALEP and so were very close relatives. Both MiMos were funded by CEC - at the same time as ETS. MiMo was perceived as more "sexy" (five on this), and between 25 - 30 people on "boring" ETS work. Since the ETS software was not usable, people did not really use it though the team fulfilled the plan. The Liaison Group should have been firm and insisted on keeping the CAT framework. This would have come against the CEC veto. There was no peer reviewing at the time of change to ETS - only political committees.

1991 saw the post-Pannenborg change to ET6 formalism from consortia outside the Eurotrian world (which was not encouraged to use ALEP at that time). It appeared to Eurotrians that ETS had been abandoned. Utrecht commented that to use ETS in the future requires grammars adapted to its peculiar properties. ALEP or MiMo2 were more mainstream than ETS. For the 1991 - 1992 programme it was agreed to use the first six months to consolidate the monolingual and bilingual components for a reduced number of language pairs, and to use these modules to evaluate the results of research work going on in parallel. It was felt to be an important improvement that there was no longer an obligation to let all research results converge into one single prototype system.

8.11.3 *Utrecht STT.* The actual project teams are embedded in the Research Institute for Research and Speech (OTS) which is a research organisation of the faculty of Humanities of the University of Utrecht. This institute grew out of a long Utrecht tradition in formal linguistics and phonetics, and is a platform for all research in these areas at the University. Since its inception in 1986, STT has been involved in EUROTRA, MiMo Sidelines, LEXIC (reusability of lexical resources with Philips and Van Dale), ROSETTA (with Philips Research Labs), and GRAMMAR (reusable grammars - with Tilburg University). Current activities include: ET10/75 (Collocations), DYANA 2 (ESPRIT Basic Research), LRE 61/61 (reusable grammar), LRE 61/62 (Discourse), EAGLES, CLASK (robustness study - combining linguistics and statistical knowledge; with CWARC/Montreal, SITE/Paris and funding from DG XIII International Collaboration). The future? - five applications for LRE II, one application for the National Information Technology Programme. Meanwhile, the CL part of the Research Institute brings in about 1 million hfl per year.

8.12 EUROTRA-Portugal

8.12.1 *History.* In May 1987, Portugal (Junta Nacional de Investigacao Cientifica e Tecnologica - JNICT) signed a contract with the EC to take part in EUROTRA - at that time there was no research in MT in Portugal. Furthermore there were no courses in CL in any of the Portuguese Universities. Computer scientists had attempted to conduct NLP work on Portuguese language but the linguistic work was very limited. At that time there were two Centres studying this latter aspect - the Universities of Lisboa and Porto. The linguistic department of the Universidade of Lisboa was the main source of linguistic skills and became the contract manager for the EUROTRA-PT work. In addition to the Universidade Nova de Lisboa, the Universidade de Coimbra has contributed effort. Communications between these latter Centres and Lisboa were not as good as they should have been. (The effort ratios for Lisboa, Porto and Coimbra appear 8:2:1.) The EUROTRA programme has acted, not just as a catalyst, but as a springboard for activity in CL within Portugal.

8.12.2 *Lisboa, Porto, Coimbra, et al.* ILTEC. In order to reinforce the importance of CL in Portugal, and to generate projects in Portuguese CL, the Instituto de Linguistica Teorica e Computacional (ILTEC), a non-profit making institution, was founded by the association of some Portuguese universities and cultural institutions. ILTEC integrates the EUROTRA-PT group, and is a direct consequence of the EUROTRA work. A proposal to create a postgraduate course in CL at the Faculdade de Letre de Lisboa is a direct result of the influence of EUROTRA, and the importance Portugal places on EUROTRA related activities. The EUROTRA-PT team has been in existence for only six years. The first two years were spent on "catching up", the next two on consolidation and raising awareness of their work within the CL community in Europe as well as Portugal, and the Transition years have seen the basis for future

work being laid, as the teams discuss collaboration and exploitation opportunities. ILTEC is currently involved in Ilterm terminological databanks (cooperation with industrial partners). Sócrates (CALL - Portuguese government), Gramático (syntax checker - Portuguese government), Eureka GENELEX, Eureka GRAAL, EUROTRA, ET-10, LRE and also European Social Fund postgraduate courses. Since 1987 the Portuguese group has been an active member of the CL community. There was an ILTEC stand at the Expolingua exhibition in 1989 and 1990, and a demonstration of MT (German-Portuguese) was shown. Portugal is highly motivated to identify consortia partners for projects such as LRE, and has had extensive discussions with a multi-national company on English-Portuguese MT.

8.13 EUROTRA-UK

8.13.1 *History.* At the time of the official launch of the EUROTRA programme, there was already a flourishing CL community in the UK. In the early 1980s the main national impetus to CL research was the Alvey programme which funded a number of projects in Natural Language Processing. UMIST and Essex both had established reputations in the field of CL and MT and were obvious candidates for the UK EUROTRA work. The UK CoA was signed in December 1985, but the involvement of Essex and UMIST in EUROTRA dates back to the very origins of the programme in the late 1970s. Indeed, both Centres supplied members to the coordination group which was set up in 1978, and which developed the proposal on which the EUROTRA programme was based (cf Council Decision 82/752/EEC, November 1982). Members of UMIST and Essex also conducted EUROTRA study contract work. Thanks to early support from the UK DTI, EUROTRA-UK comprised 14 members (six at Essex, eight at UMIST) by 1st January 1986: over half these original members remained with the team through to 1990. In the early years, a great deal of effort was put in to ensure cohesion between these two Centres. The management of EUROTRA-UK has always been a shared function, with representation on the Liaison Group being a shared function.

8.13.2 *Essex CL/MT.* The Essex group forms the core of the Essex CL and MT groups, an informal collection of about 15 researchers with interests in NLP. Although the group is also involved in other work (eg Eurolang, LRE, speech research) by far its largest project has been EUROTRA.

8.13.3 *Essex's Views.* 1980 - 1983 there was an explosion of interest in CL in attribute value structures - led by Xerox, SRI, CLSI Stanford. This work gave a standard focus for CL, as described in the standard model PATR-II (by SRI) which was being publicised in Coling 1984. The West Coast of the US was leading mainstream development. EUROTRA was first discussed in 1978/79 in the GETA tradition: the 5th generation Japanese work gave emphasis to toning and Prolog; a return of staff from the US led to the development of GPSG (eg Alvey tools) - an early unification,

coincident with finite state morphology. Over the last four to five years computational lexicographers and computational linguists have come together, and have recognised the need to pool resources. Speech and language people will interact in the next five years. MT is now a subset of NLP - morphological analysis, dictionary construction, generation analysis are all involved in MT. At the beginning of EUROTRA differences in scientific training and background was a problem - too many people, too diverse backgrounds. This is no longer a problem.

Various UK people were involved in the years 1978 - 1982. In 1984 an attempt to bring in new ideas was made by the steering committee. There was discussion about SUSY and GETA - GETA had been working for 15 years - they dominated the work intellectually. There was a pull between: "go for the best available" (UK wanted PATR-II *de facto* standard - not taken); "go for future" - the post-GETA, CAT framework was developed. However, the UK and Utrecht continued with their intellectual views and developed sidelines. The move to the ETS formalism was an engineering initiative - a better, more reliable implementation.

The Liaison Group fixed specifications late. The project was well planned - the work was moving at the forefront of technology - the main problem was the lack of suitable tools and resources to simplify the work, and allow development to be completed in a manner more likely to yield a successful outcome for the original aim. MiMo has served its purpose for the UK. CAT2 is exploitable - Saarbrücken put together a VALUE proposal. It is an open question how far the ETS grammars are reusable.

It doesn't matter whether it is a transfer-based approach or an interlingua approach with unification-based formalisms - the big divide is between the knowledge based approach and the statistical approach (cf ET-10). A rule based system with a transfer based approach could be built, but the lexical transfer approach would provide alternatives, and these would be decided upon using the existing rules developed through prior usage. The pure empirical approach can be bettered by far using lexical approaches.

8.13.4 *UMIST CCL.* The UMIST group is located within the Centre for Computational Linguistics (CCL) which is a separate research and teaching unit in the Department of Language and Linguistics. CCL was created in 1979 and how has some 25 teaching research staff, 60 undergraduate students and over 30 postgraduate students. The Centre was the first in Europe to provide an undergraduate course in CL, and its recently established MSc course in MT is well known. Professor Jun-ichi Tsujii from Kyoto University joined CCL in 1988 and since then CCL has developed strong research links with Japan, and is taking part in a number of collaborative projects on Japanese-English MT and lexicography. It also participates in ESPRIT and other projects in CL, MT, sublanguages and terminology, CALL, information processing and text linguistics.

8.13.5 *UMIST's Views.* Parsers, lexicons etc at UMIST would not have been developed without EUROTRA. On the linguistic side a vast amount of formal monolinguistic research would also not have been done without EUROTRA. The large English lexicon is reusable, but the grammar is more difficult (each entry takes hours to complete). UMIST viewed EUROTRA as essentially a research project. If it had been aimed at development it would have been spread over fewer sites, with less complex management, shorter timescales, etc. Communication was very cumbersome.

The approach to development amongst the centres was pragmatic. EUROTRA transformed theoretical linguists into CL workers - very few Centres had worked with computers. The competitive element between the countries was beneficial. UMIST's Japanese connections would not have arisen without EUROTRA. MT was not acceptable in the UK in 1977. Alvey (1984) had already funded UMIST, and gave UMIST the confidence CL was worth pursuing.

9. Organisation and Management

9.1 The EUROTRA Management.

9.1.1 Initially EUROTRA was run from DG XIII in Luxembourg with a very limited team. Much of the technical planning work was carried out under contract by Professor M King and her team at ISSCO in Geneva until 1985. The offer to take the technical leadership by Professor Bernard Vauquois of the GETA team at Grenoble was turned down, because the work of his team was not seen as a suitable basis for the EUROTRA development. So both the managerial and ultimate technical policy responsibility fell to the Commission's team in DG XIII.

9.1.2 The origins of EUROTRA lie in a meeting of experts in February 1978. The Council authorised the programme in November 1982, at which time the Commission's team consisted of two people. It is interesting to note that the output of the study work that had been going on resulted in the first version of the Reference Manual being released in 1979 at the first annual workshop. This was three years before the programme was formally authorised, and the work was supported by small study contracts from the Multi-Lingual Action Plan programme budget.

9.1.3 The Commission's management team consisted of essentially one person until 1981, two thereafter. The first Contract of Association was not signed until June 1984, though it was not until Autumn 1985 that enough were signed for the programme to be properly implemented. The Italian and Dutch contracts were not signed until 1987. Meanwhile the programme went forward on study contracts. No extra staff were made available until 1986 when the DG XIII team was increased to six. In addition, SdT, the translation service of the Commission who were co-proposers of the programme, did agree to provide a number of translators to the team. These eventually rose to 12, but only towards the end of the programme after SdT had made a special drive to recruit computational linguists. Much of the team has now moved on to work on the LRE programme.

9.1.4 The running of the programme, in both management and scientific sense, lay with the staff in DG XIII and in particular the team leader. Initially there was a programme management advisory committee (ACPM) set up in 1983 consisting of the government representatives, but they met infrequently, and probably had little influence. The ambiguity in their rôle is apparent in their title, containing the words "management" and "advisory". It was replaced in 1985 by a Comité de Gestion et de Coordination (CGC) with a broader remit across the field of linguistics, and a Common Steering Committee (CSC) consisting of representatives of the government signatories of the Contracts of Association. These

bodies had little influence over the technical direction of the programme

9.1.5 More influential over the technical conduct of the programme was the Liaison Group, consisting of the Directors of the Centres together with the Commission and normally chaired by one of the Directors. There were also numerous standing and ad hoc committees comprising staff from the Centres and Commission as the central funding body. On occasion, the Commission's team leader used that power of decision against the views of some of the Directors. The procedure was that a two-thirds majority was required in the Liaison Group provided that included the Commission's representative, who essentially had a veto. In important cases the decision could be referred up to the CSC. In practice the governments sometimes nominated one of their Centre Directors as their representatives on the CSC, so much the same argument - with the same outcome - would tend to be repeated in the CSC.

9.1.6 *Dual Rôle of the Director.* The Pannenborg Panel commented on the insufficiently distinguished executive and principal rôles in the management of the project. The burden on the Director of the programme would have been very considerable, even had he had adequate support staff, which he didn't have in the early years of the programme. The Director had both a managerial leadership rôle and a technical leadership rôle. It would have been feasible to split these rôles, the ultimate leadership falling to the managerial leader. (One is reminded of the way General Groves worked with Professor Oppenheimer in the Manhattan project.) Unless some entirely different organisation to run the Commission's language technology programme is envisaged (as discussed in para 9.2 below), it seems inevitable that the main management responsibility must fall to the Commission. Management power lies where the money is and though this was not as unambiguous as it should have been because of the complex arrangements with the national authorities it was always clear who held the ultimate power. But it would have been feasible to have brought in such a technical leader on some secondment basis or other, even though it might have been difficult to find a person with the right skills and reputation in Europe at the time. For such a prestigious programme the technical leader must attract the support and respect of those technical people who will work for him in the various Centres. It would be sheer luck and highly improbable chance if amongst the Commission's staff was a suitably qualified man. In this case the Director certainly won the affection and indeed respect of many of those in the programme, for his extremely hard work and devotion to the programme, but he did not have the international technical reputation to win their immediate technical regard. This made his rôle extremely difficult. There is much experience in industry of how to provide managerial and technical leadership in complex projects, by putting together the appropriate individuals supported by management boards as appropriate. Such leadership has the power to lead constructively including the important power to stop work when that is needed. *The*

Commission should have recruited a technical director for the programme who would have gained automatic technical respect and leadership.

9.1.7 *An Industrialist as Director?* The programme was clearly intended to lead on to an industrial product, even if it was foreseen that a stage of research would be required before the prototype was fit for industrial exploitation. One might therefore consider that it would have required someone with appreciation and experience of industrial objectives and market outlook. The conflict between the research ambiance of the Centres and the needs of the development aspects of the programme was always a problem. However, in practice the expectation of a directly exploitable product receded with time, so perhaps, by that time, a Director drawn from industry would have been inappropriate. Such a man might well have been in conflict with the research aspirations of the Centres. However, the Commission should bear in mind the need for the Director of such a major programme to have appropriate qualifications and experience. In particular:

R17: Where a programme is expected and intended to lead to industrial exploitation, leadership should be placed with an individual with appropriate qualifications, reputation and, if possible, industrial experience.

9.2 An Agency?

9.2.1 The concept of an Agency outside the Commission to run a programme like EUROTRA, and perhaps all of their programmes in the language engineering field, was proposed by the Danzin Panel, and followed up in a study by a panel under the chairmanship of Dr Coltoff. There are obvious advantages:

1) The Agency would be much freer to recruit staff as appropriate, unconstrained by the inevitably bureaucratic conditions of service of the Commission.

2) An Agency could move faster over contractual matters, etc.

3) There would be somewhat less need to balance the conflicting interests, because the Agency would provide a buffer from the national and regional concerns.

4) A good leader might more readily be attracted to the relative freedom of an Agency.

However there are disadvantages:

1) The Commission might find that it could not delegate certain of its powers. (It is possible to imagine the secondment of an

appropriate Commission officer to work in the Agency to overcome this problem.)

2) Programmes need the support of the staff of the Commission if they are to win the continuing approval of politicians in the Commission, Council and Parliament. There is a danger that, without direct and continuing involvement, the Commission's staff would lose interest and detailed knowledge about programmes they were sponsoring through the remote hand of an Agency.

3) The selection of the body to run or become the Agency is a highly politically sensitive operation, which can delay the start of a programme for an inordinately long time. The compromise that might arise, in satisfying the conflicting political interests, may lead to a weak or constrained Agency.

9.2.2 It is difficult to advise for or against an Agency without considering the specific situation of a particular programme. However, there are such practical advantages that the Panel does recommend:

R18 Serious consideration should always be given to the creation or employment of an Agency whenever the establishment of a programme on the scale and complexity of EUROTRA is in mind.

9.3 Relations with the Centres

Considering the complexities of the EUROTRA programme, *the Commission and especially its leader, Dr S Perschke, achieved excellent relations with the Centres.* There were inevitable conflicts from time to time, especially where matters of technical leadership were contested.

9.4 Staff

9.4.1 Initially, the Commission's staff were grossly overloaded. It is a waste of resources to fund such a programme and not provide the number and quality of staff to run it adequately. Delays in initiating the programme were excessive, at least in part due to staff shortages in the appropriate part of the Commission, and the problems of recruiting staff with the required skills. Conversely, towards the end of the programme there seems to have been an excess of monitoring staff, who were able to add little and occasionally created some resentment. These staff were also involved in building up the LRE programme.

R19: In establishing a programme of the cost and complexity of the EUROTRA programme the Commission should ensure that it is adequately staffed, especially in the difficult early days.

9.4.2 In fairness to the Council and those involved in that decision, the creation

of eight temporary staff to run the programme was explicitly authorised in the initial EUROTRA Council Decision of November 1982. It seems to have been an administrative error that led to the delay in staffing the programme appropriately.

10. Governments' Rôle

10.1 The Eurotra *"Contracts of Association"* mode of working is an unusual way for the Commission of cooperating with the national governments. The precedent lies in the way the European Fusion programme was organised. It is claimed that the advantage of creating a partnership with the governments, in the way that the Contracts of Association does, is that the governments take a detailed interest in the programme if they are contributing directly to the local costs of the programme. The disadvantage lies in the complexity of dealing separately and in detail with 12 governments. It can also lead to some loss of authority, to those running the programme in the Commission, if they are only providing a part, perhaps a small part, of the funding.

10.2 <u>The Start-Up Period</u>

How did it work in EUROTRA practice? After some five years of preparatory work the Council authorised the programme in November 1982. Yet it took until June 1984 until the first Contract of Association was signed, until the Autumn of 1985 before sufficient Contracts of Association could be signed to launch the main programme. A large part of this delay must have been due to the problems of negotiating with the governments separately. The legal departments in both the Commission and the governments were said to have been the cause of much of the delay. Thereafter, with a few exceptions, the relations with the governments worked well. In some cases individual governments found it hard to find the appropriate mechanism to act as their national agency; in Germany it was the science ministry, BMFT; in the UK it was the industry ministry, DTI; in France it was the research agency, CNRS. It was in France that the greatest problems emerged, perhaps because the French government has itself been consistently a significant supporter of similar work with its own funds. In some countries, like Ireland, it took some time for the government to find the right mechanism for organising its own contribution, and Ireland finally solved the problem by passing the responsibility to the Dublin City University who handled it very sensibly and effectively. Perhaps it was the German government who took the most active interest in the programme, organising and funding work for the annual meeting of all the Natural Language teams in Germany including the EUROTRA teams. This created an information exchange within the country, and served to lessen the tensions in the teams outside EUROTRA. In Scandinavia there are regular meetings of the NLP and MT research workers in which of course the Danish EUROTRA Centre participates. But these meetings stem from the academic research workers themselves.

10.3 Management Committees

During the running of the programme the governments formally interacted with the Commission through the CGC committee, formally the Management and Coordination Advisory Committee (CGC-12), or through the lower level Common Steering Committee. In the early years the programme was an independent entity outside the Research and Technical Development Framework Programme. It was sponsored by both DG XIII and SdT, the Commission's translation service. Then in September 1987 the second Framework Programme was authorised, including the EUROTRA programme. As with all other programmes under the Framework, individual Council Decisions are required for each programme, but they adhere to a common pattern over matters like management committees, etc. The Contracts of Association procedure remained in place for EUROTRA, but probably the governments came to see the programme more in terms of the standard mechanisms for Framework Programmes.

10.4 Governments' Influence

There were delays between one phase of the programme and the next when authorisation ran out. The Commission had to find ways of keeping the Centres going over the interim, with the cooperation of the governments. But, basically, the programme ran smoothly as far as the involvement of the governments was concerned. One can see their influence, along with the Pannenborg and Danzin reports, in the shift to open up the programme to new participants, and in the move to cost-shared projects. The tightening of the control over the authorising of the moving on to successive stages, and the external reviewing of the work, is also noticeable.

10.5 Conclusions

10.5.1 *Had the normal cost-shared projects been established as the way of working the participation in the programme would probably have been concentrated in a few countries, and the Centres would not have been established in those countries where direct government intervention was required in order to get them set up. Thereafter it was right to shift to a more open, more competitive approach.*

10.5.2 In the early stages the governments were certainly much more directly involved than in the more normal way of Commission programmes. But thereafter there does not seem to be much indication that the governments took a greater interest in the programme than they do in other programmes under the Framework Programme. Some governments take an active interest in cost-shared programmes, in order to encourage local participation and coordination with national programmes. *So, after the start-up phase, there seems little advantage in the added complexity of the Contracts of Association process.*

11. The Future

11.1 Strategy for the Future (C&R)

11.1.1 *EUROTRA Achievements.* There are lessons to be learnt from EUROTRA for the future nature and organisation of Language Engineering support by the Commission. *It was very imaginative and, indeed, brave of the Commission to propose the EUROTRA programme and of the Council to approve it.* If it did not achieve its ostensible objective, it most assuredly had a great success in stimulating computational linguistics in every nation of the Community; in bringing the participants together in an outstanding example of cooperation between the nations; and in providing material such as the nine Language Specifications that will underpin both academic and industrial work for many years to come.

11.1.2 *The Community's Need.* The original Council Decision of 1982 rightly recognised that the multilingual nature of the European Community is both of "high cultural value", and at the same time, "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". This is certainly as true today as it was ten years ago. Indeed, with the likelihood that the Community will be enlarged to contain a further five or more languages before the current decade is out, *it could be said that the need for the developments of language technologies to be brought to bear on the language barrier of the Community is more urgent than ever.*

11.1.3 *The Commission's Need.* The size of the language problem facing the Commission itself is immense. The cost of translation to that body probably exceeds 150 Mecu per annum. The hidden cost, in the failures of full communication, and the delays inherent in a system where translation is required but only available in due course, dwarfs the direct costs. Yet there can be no doubt that technology can serve to reduce these problems, not by glamorous total automation, but by attention to machine assistance for the human translator and the provision of an integrated document handling system, where language aids are provided wherever they can be of help to the users of documents, whether they be translators or other Commission staff.

11.1.4 *Machine Translation.* The failure of the EUROTRA programme to create a "machine translation system of advanced design" must lead to the question whether further work today would have any better prospect of success? While the accumulation of knowledge and experience is steadily improving the performance of machine translation systems, it would be repeating the error of the original Council Decision to assume that progress has reached the point where even a "system of advanced design", built with today's technology, would make much improvement to general

machine translation performance. Human post-editing will remain essential if good quality general text translation is required. Until, at some time in the future, systems are capable of handling semantics drawn from much more than the single sentence without human assistance, Machine Translation will be incapable of producing good translation. However, *it would be to follow error with error to deduce from this that technology cannot contribute significantly to the language problems of the Community.* Even in the extremely difficult machine translation field, technology in the form of specialised integrated document processing equipment, "Translators Workbench", will achieve significant improvement in efficiency. If the doubling of output that has been achieved in well attested applications of such workbench technology in the USA were to be achieved by the Commission's translation service the consequent savings would be worth some 75 million ecu per annum in translator time, and probably far more in the value of a faster document turn-round. Moreover there are many opportunities in the Community where niche markets and sub-languages can be successfully tackled by machine translation technology. When the EUROTRA programme was being planned the Commission's Systran system had not come into serious use by the translation service. But it was most unfortunate that there was so little cooperation with the translation service for it might have had a beneficial influence over the whole direction of the EUROTRA work.

R20: The Commission should concentrate on Machine Assisted Human Translation, on aids to the translator, while continuing to support longer term research that will improve automatic translation.

11.1.5 *Wider Fields of Language Engineering Application.* If the Council was wrong, in 1982, to concentrate on the creation of a machine translation system instead of a system to improve the efficiency of the translator, it was not wrong to invest in computational linguistics and Natural Language Processing. There are many other applications of language technology, beside machine translation, most easier to tackle and some addressing large markets. Monolingual as well as multilingual topics should be addressed. By far the largest is thought to be for the retrieval of information, where the spread of databases available over networks is creating a market for natural language interfaces, making it easy to obtain specific information in a natural way, in the language of one's choice. Text editing tools, and the creation of precis of text, are seen as other potentially large markets. In conjunction with speech processing there is a developing market for adaptive-dialogue database access systems, and eventually, when the technology improves, for systems that enable the user to talk and dictate to his computer.

11.1.6 *Industrial Participation.* The participation of industry in EUROTRA, even in the last phase, was disappointing. To encourage exploitation and to stimulate interest in language technology in industry, it is desirable to devise programmes that will be attractive to industrial participation. *It*

must be an objective of the next phase of the Commission's Language Engineering Programme to repeat the success in stimulating work in the academic community, but now also directed to the industrial community. This will inevitably require projects where the market application is clearly in view and where user participation can be encouraged.

11.1.7 *Scale of the Programme.* The language problem is one of the most important facing the Community, both for economic and for social reasons. The very large cost to the Commission is a measure of the wider cost to the Community. EUROTRA has demonstrated the importance that every nation, whether large or small, attaches to its language, for language lies at the heart of a nation's culture and past. The investment that the Community should be making in language technology should be commensurate to the impact that it could make to the language challenge. Yet the investment being made by industry in developing and using the technology is still relatively small, perhaps because the problem is perceived as being too difficult. Yet the evidence is that technology could have a major impact. The responsibility lies with the Commission to step up its investment, both in aiding the development of the technology and in making direct use of it in its own offices.

11.1.8 A programme of investment in the stimulation of technology should not move too far beyond the Community's capability to provide qualified human resources to tackle it. This means that the Commission should be planning to steadily ramp up its investment, aimed especially at increasing the competence of industry to work in this field, while maintaining the academic capability that has been developed. Attention should also be given to stimulating the use of language technology in industry, commerce and government.

11.1.9 *Advances in Computing Technology.* Before outlining work for the future it is worth remembering that computing technology moves on apace, in a way that impinges on the work of NLP and MT. In Appendix 10 Professor Dr Hartwig Steusloff suggests how current advances in parallel processing hardware, in AI techniques, and in the Object-Oriented approach to computing, should be taken into account in future MT work.

11.2 Programme Organisation

11.2.1 *Competition and Cooperation.* One of the achievements of the EUROTRA programme was to create a strong network of computational linguists across the Community. There is a danger that changing to a competitive individual project basis will cause this network to decay. This would be very regrettable because the subject is peculiarly one in which one team can benefit from interaction with another, one language base learn from the understanding of another. Steps should be taken to ensure the network is maintained, and indeed extended to embrace all the language engineering experts in the academic world and industry, whether they are

involved in Community supported projects or not. The ELSNET may form a basis for this, but perhaps a mechanism more specifically directed to language engineering is required. The EAGLES standardisation cooperation will also help. This is discussed in Chapter 6. What is required is a balance between competitive cost-shared projects of the ET10 or LRE type, but complimented by actions to focus and coordinate the work. Projects grouped in "focused clusters" round technology and language themes proved a successful approach in the final stage of EUROTRA.

11.2.2 *Programme Management.* The arguments for creating an Agency to run the language engineering programme were rehearsed in Chapter 9 above and will not be repeated here. The important point is to place the leadership on one who has the appropriate experience and motivation to keep the market always in mind, backed if necessary by appropriate technical experts with the intellectual prestige to give ready leadership.

11.3 The Immediate Need

11.3.1 *Exploitation of EUROTRA work.* Having built up a considerable body of material and expertise it would be folly to allow it to disappear without exploitation. It is very unfortunate that some of the EUROTRA Centres have potential users of their technology and skills, but need some support to get the EUROTRA work applied to the users' field of interest, at least to the point where those users can assess the exploitation potential. The current Commission schemes, such as SPRINT and VALUE, do not seem appropriate or large enough to provide support. There should be an "exploitation scheme" specifically linked to the Language Technology field to help with the follow-up of EUROTRA work but also LRE, etc. In practice this might be a sub-set of the limited vocabulary market applications programme outlined above in 4.6.11 - 4.6.14, or the wider applications programme proposed in 11.4.4. below. It should be a priority to set up a study of the exploitation of all the EUROTRA property and work to establish where help is needed and can be most effective.

11.3.2 *Exploitation via ALEP.* Work should continue to transfer the grammar and dictionaries of EUROTRA to the ALEP system. If appropriate, a new revised and updated version of the Reference Manual and the Language Specifications should be issued.

11.3.3 *Maintain the Network.* Action needs to be taken to continue the collaboration of the EUROTRA community, widening it as appropriate.

11.3.4 *Continuing Research.* Research should continue in the academic community, both to continue the language rule development pioneered in EUROTRA, and to widen the technological approach to solving the problems of language engineering. Some of the priority topics for research are discussed in Chapter 4.6 above.

11.4 A Broad Based Language Technology Programme

11.4.1 *Inter-disciplinary Attack on NLP.* A broad based programme is required, which will need an inter-disciplinary approach, firmly based on practical solutions that will make an input on the users' real problems. The Panel welcomes the Commission's activities to consult widely on the technical content and applications for a future programme. The elements of such a programme are outlined below:

11.4.2 *A Technology Assessment Programme.* This is a field where a team of technology assessors, experienced in what used to be known as operational research techniques, should be established to explore potential market opportunities in the Community. They would be tasked with exploring potential opportunities, assessing what technology development and training is required, and giving wide publicity to the resulting reports. This would serve to stimulate the market as well as directing research to the practical needs.

11.4.3 *Lexical Resources.* An attack is needed on the task of building up machine-based dictionaries and terminological databases for all the official Community languages and language pairs, to provide lexical resources for many of the NLP and MT projects now in Europe and for years to come. This major programme will need to be preceded by a careful study of the structure and specifications to ensure portability and wide applicability to different system architectures. This is a major, and probably long lasting programme, where the cooperation and participation of many of the on-going NLP commercial projects should be sought. Indeed a major customer and cooperator in this would be the Commission for its own requirements. The programme will require the building up of the lexicographic expertise required in many of the languages. A distributed workforce but centrally coordinated, on the EUROTRA pattern, will be required.

11.4.4 *An Applications Programme.* Rather than tackling head-on the unconstrained machine translation market, an applications programme should be established aimed at markets where NLP technology can be most effective. The systems approach must always be adopted, addressing the ultimate users' real problems and needs in a practical way. Markets, narrow in scope but not necessarily in magnitude, where the advantages of restricted grammars and dictionaries can provide effective solutions, should be tackled as a priority.

11.4.5 Such applications should be very practical in approach, but may serve to pull through enabling technology that needs R&D to enable or improve the application work.

11.4.6 *Commission Projects Scheme.* It is particularly appropriate to base applications and R&D projects on the Commission's own needs. There

should be a small internal "Operational Research" team actively seeking market opportunities in the Language Engineering field within the Commission's own work. The promising opportunities should be followed up with development projects, where it would be appropriate to provide full funding at least for the study phases. Within the translation field a priority should be to equip the Commission's translators with a fully integrated document handling environment and "Translators' Workbench". The use of corpora matching techniques, especially when attacking the sort of repetitive tasks often required within the Commission, seem especially suitable to the Commission environment.

11.4.7 *Enabling Research.* Having built up a competent academic computational linguistic research community it would be the height of folly to let it decay away. The field of computational linguistics is still relatively young, and there is much applicable and enabling research to be carried out. A multi-disciplinary approach will be essential, and links to the international research community should be encouraged. While such a programme will inevitably be largely based in the academic and associated non-profit institutions of the Community, such as the EUROTRA Centres, it is important to bring in industrial participation wherever possible, if only to provide monitoring of the work from the industrial viewpoint.

11.4.8 *Training Programme.* EUROTRA has been very successful, seen as a training and technology transfer programme. There is a continuing need for a training scheme, specifically oriented to students who may already be in industry or commerce, potential users of Language Technology as well as suppliers of the technology.

11.5 International Collaboration

The subject of language engineering is difficult and universal in its scope. There is everything to be gained by cooperation across national frontiers wherever skilled resources are available to make cooperation of mutual benefit. Of its very nature, linguistics is a subject that benefits from a wide perspective, of languages and linguistic usage; and of disciplines ranging from fundamental logic through computer science to linguistics and philosophy. In supporting research work, the Commission should ensure that the applicants demonstrate an appreciation of the international state of progress in their field.

R21: International cooperation should be encouraged, in particular with centres of expertise in the USA and Japan. Topics like system performance measurement and standards are particularly appropriate for initial steps in collaboration.

11.6 Cooperation with the Commission's Translation Service

As has been pointed out in various places in this Report, the Commission is a major user of MT and it is most unfortunate that there was so little cooperation with the Commission's translation service, especially over the dictionaries. This must not be allowed to happen in the future.

R22: In any future work in MT, the Commission should ensure there is close cooperation with the actual work and needs of its own translation service. The opportunity will arise, due to the need to re-engineer its Systran system. The lexical resources programme proposed above is an ideal vehicle for close cooperation.

APPENDIX

A

1

FINAL EVALUATION OF THE EUROTRA PROGRAMME

TERMS OF REFERENCE

---

1. BACKGROUND

Article 4 of the Council Decision 90/664/EEC of 26 November 1990 adopting a specific programme concerning the preparation of the development of an operational Eurotra system stipulates that :

> "1. During the first quarter of 1993, the Commission shall conduct through a panel of independent experts an evaluation of the results achieved and shall send the panel's report and its comments on this subject to the European Parliament and the Council.
>
> 2. This report shall be established having regard to the objectives set out in Annex I and in accordance with Article 2 (2) of Decision 87/516/Euratom, EEC."

The above-mentioned Council Decision is the last of a series of five :

Eurotra was initially adopted in November 1982 (82/752/EEC). In November 1986 (86/591/EEC) Council approved the extension of the programme to Spain and Portugal ; in June 1988 (88/445/EEC) the transition to the third phase was decided and in November 1989 (89/410/EEC) the integration of Eurotra in the 2nd Framework Programme.

In addition, Eurotra has been the subject of four reports of the European Parliament : Adam (1981), Pinto (1986), Desama (1989), Desama (1990).

The programme has been evaluated twice by panels of independent experts : 1987 by the Pannenborg Committee and 1990 by the Danzin Committee.

The foreseen evaluation should fulfil two complementary functions :

- it should appreciate the achievements of the programme in the years 1991-1992 or, more precisely, in the period after the last evaluation, i.e. 1990;
- it should appreciate the outcome of a programme (change of state) which was conceived in the late seventies and has lasted ten years. This applies both to the scientific and technical and to the policy aspects.

Furthermore it should appreciate the way in which the recommendations of the Pannenborg and Danzin reports have been taken into account both in the 1991-1992 programme and in the follow-up programmes (LRE in FP-3 and the preparation of a strategic programme in FP-4).

2. THE FOCUSSED EVALUATION OF THE SPECIFIC PROGRAMME

The Council Decision formulates in its preamble quite precisely and realistically the expectatio of a programme with a duration of two years and a budget of 10 million ECU in two whereases

> "whereas this programme should lead to the development of a high-level scientific prot type in the field of automatic translation"
>
> "whereas additional efforts will have to be made by the Community, the Member Stat and the European industry to reach this objective" [i.e. an operational Eurotra system]

It is, therefore, important that the evaluation takes into account the objectives set out in Annex of the Council Decision and in the programme of work agreed upon with the Eurotra adviso committee, rather than taking some hypothetical Eurotra system ready for practical applicatio and commercialisation.

The elements of the programme of work subject to evaluation are articulated under three hea ings:

(a) work carried out by the national research teams under the contracts of association whi includes:

- the monolingual and contrastive research (including publications, working p pers etc.) and its impact on the qualitative performance of the Eurotra prot type;
- the multilingual MT prototype, in particular its qualitative performance compared to the prototype which was available to the Danzin Panel in 1990;
- the use and the impact of the funds earmarked for grants;
- the contribution of the two national teams which did not directly participate the implementation of the prototype: Ireland (terminology) and Luxembo (documentation and software test and reference centre and clearing house).

(b) the shared-cost projects :

ET-10/52 (migration of Eurotra grammars to the new formalism)

ET-10/75 (collocations)

ET-10/51 (general vocabulary definitions)

ET-10/66 (terminology definitions)

ET-10/63 (probabilistic methods)

ET-10/61 (formal semantics for discourse)

(c) the work fully financed from EC funds in preparation of the follow-up programme, e cially in view of creating a common platform (methods, tools, resources, standards). concerns in particular :

- the definition study for a new formalism and software environment an prototype implementation (ET-6, ET 9 projects);
- the feasibility study concerning the reusability of lexical and terminologic sources (ET-7) and its impact on the activities aiming at standardization.

In general, this part of the evaluation should investigate :

- how far recommendation from the two previous evaluation reports have been implemented ;
- the progress made during the reference period both with regard to the S&T content and the management of the programme ;
- how far the starting conditions for follow-up programmes have been improved.

2. THE GENERAL EVALUATION OF THE PROGRAMME

(a) Impact

On the whole, Eurotra has lasted ten years with an over-all budget of 37.5 million ECU. Although this is a relatively modest amount (if compared with other IT related programmes), in the field of linguistics it is the first EC funded programme and the biggest single project ever in Europe.

The existence of Eurotra has undoubtedly had some impact on policies and activities both at Community and national level in the EC and outside, especially in the USA and Japan.

The evaluation should therefore compare the situation of MT and NLP related policies and activities of, say, 1980, when Eurotra was first presented to Council and European Parliament, and the end of 1992.

(b) Awareness

One of the reasons for the considerable delay in the adoption of the Eurotra programme (from June 1980 when the proposal was submitted to Council and Parliament to November 1982) was the lack of awareness - at the policy and decision making level - of the importance of language problems for the Community, and also of the understanding of the role, the EC should play in this domain.

The delays in the initial decision, and the subsequent decisions and scrutiny by the European Parliament and external evaluators (five Council Decisions, four Parliament reports, two external evaluations) are certainly disproportionate to the size of the programme, but they may have had some positive side-effect

The evaluation should assess to which extent the discussions concerning Eurotra have contributed to the increased awareness of the policy and decision makers both at Community and national level, and to the definition of the role of the EC, especially with a view to the future.

(c) Level of activity - Cohesion

When Eurotra started, the level of activities, both in MT and in NLP in general was relatively low in all of Europe and very unevenly distributed in the Member States.

The evaluation should assess how Eurotra has contributed to the general increase and balance of activities and expertise and international cooperation in the EC, and created the possibility of starting new activities outside the Eurotra context

In particular, the evaluation should consider two aspects:

-at the academic level, the creation of institutes especially for the participation in Eurotra, thei scientific status and reputation and ability to survive after the end of the programme and th impact of Eurotra on the creation of other institutes, jobs and the adaptation of university cur ricula

-at the industrial level the influence of Eurotra on industrially and commercially oriented project such as METAL or EUROLANG

(d) Scientific and technical impact

If one counts the preparatory period which started in 1978, Eurotra has existed 14 years. Durin this period, especially research activities have made considerable progress outside the project.

The evaluation should examine to which extent research activities in Eurotra (which has been relatively closed community) and outside have interacted and influenced each other and the im pact Eurotra had on computational linguistics in general.

2

Membership of Review Panel

Sture ALLÉN

1928	Born Göteborg
1965	Fil. Dr. Scandinavian Languages, University of Göteborg
1972	Professor of Computational Linguistics, Swedish Research Council and University of Göteborg
1980	One of the Eighteen in the Swedish Academy
1980 - 1986	Vice-Rector, University of Göteborg
1986	Permanent Secretary of the Swedish Academy

Publications (Selection)

Graphemic Analysis as a Basis for Text Editing 1 - 2, Diss. 1965
Natural Language Processing (Ed.) 1970
Frequency Dictionary of Present-Day Swedish 1 - 4 (with others) 1970 - 1980
Text Processing, Proceedings of a Nobel Symposium (Ed.) 1980
A Dictionary of Swedish (with others) 1986
Possible Worlds in Arts and Sciences, Proceedings of a Nobel Symposium (Ed.) 1986

Brian OAKLEY

1927	Born London
1950	MA in Physics, University of Oxford
1950 - 1969	Research in microwave devices, real-time and control systems, TRE, RRE, RSRE, Malvern, UK
1969 - 1979	Department of Trade and Industry. Computer Industry. Research Policy
1979 - 1983	Secretary, Science and Engineering Research Council
1983 - 1987	Director of the Alvey Programme
1987 -	Director, Logica Cambridge

Brian Oakley worked on computer systems from the mid 1950s. He has been much involved with academic/industrial cooperation. He was a member of the ESPRIT Management Committee from 1983 and then of the ESPRIT Advisory Committee until 1992.

Alessandro OSNAGHI

1940	Born Milano
1965	Politecnico of Milano, MS Nuclear Engineering
1966 - 1968	University of Bologna : Assistant Professor of Nuclear Electronics. Research activities on Fast Signal Processing
1968 - 1974	Montedel, Milano : Manager Software Department. Responsible for the development of the system software for the first Italian minicomputer
1974 - 1984	Olivetti, Ivrea : Director Software Technology Division. Responsible for the architecture and system software development of the Olivetti minicomputer line of products
1985 - 1986	Unix Europe Ltd, London, UK : Managing Director
1986 - 1988	Olivetti, Ivrea : Corporate Development and Product Strategy, Director Software Strategy
1988 - 1989	Olivetti Information Services (OIS), Milano : Director Software Technology
1989 - 1991	Olivetti Systems & Networks, Ivrea : Director Education Centre
1991 - present	University of Pavia : Associated Professor of Computer Assisted Engineering
	Independent consultant in the field of Information Technology.

Alessandro Osnaghi worked on system software development from 1968, mainly in the area of operating systems. He has also been involved in developing business strategies in the software technology area. He has covered senior executive positions in a major international company. He has been teaching Operating Systems from 1972 to 1989 at the Computer Science Department of the University of Milano.

Iain RAE

1953	Born Yorkshire
1977	Mathematics degree from Leicester University, PhD in Magnetohydrodynamics with Pilkingtons Ltd from Keele University
1977 - 1980	Mathematics teacher, Loughborough Independent School
1980 - 1982	Research Fellow, Solar Physics, University of St Andrews

1982 - 1983	British Aerospace, Hatfield, radar, millimetric and infra-red imaging analysis, and missile aerodynamics modelling
1983 - 1988	Scicon Ltd. Imaging consultant, manager environmental modelling group, software coordinator for UK Department of Environment radioactive waste management modelling work
1988 -	Logica. Business development and project manager in process industry, and later open systems group, also working with speech and language systems division

Jean ROHMER

1948	Born France
1970	Ingenieur in Applied Mathematics, Institut Polytechnique de Grenoble
1976	Docteur Ingenieur, University of Grenoble
1980	Docteur d'Etat es Sciences, University of Grenoble

From 1970 to 1980 Jean Rohmer was a researcher in public laboratories : IMAG Grenoble, and INRIA Paris. He worked and published about multiprocessors, database machines and text retrieval machines.

In 1980 he joined Groupe Bull, where he created the Artifical Intelligence research department, conducting work on logic programming, knowledge representation and natural language. He personally contributed to the fields of deductive databases and symbolic parallel processors.

Since 1987, Jean Rohmer has managed CEDIAG, the Bull business unit in AI, developing commercial products and services in various fields of AI.

From 1982 Jean Rohmer was instrumental in the early phases of preparation of the ESPRIT project.

Hartwig U STEUSLOFF

1937	Born in Gelsenkirchen, FRG
1977	Doctorate in computer science at the Technical University of Karlsruhe, FRG
1987	Professor at the Department of Computer Science of Karlsruhe Technical University

H U Steusloff is director at the Fraunhofer Institute of Information and Data Processing, Karlsruhe, FRG, and Institute of Applied Research in Computer Science for applications of computer systems in production. Main working areas are hardware, operating software and languages for real-time computer systems, including data bases and artificial intelligence.

His scientific and teaching activities are concerned with computer architectures for parallel processing and the supporting software components such as communications systems and systems engineering.

3

Visit Schedule

From the outset the Panel set itself the task of understanding the work of EUROTRA and the views of the EUROTRA community. It was agreed that a series of visits should be held to the EUROTRA Centres, to the Commission, to selected EUREKA, ET-10 and LRE organisations.

In addition, views were sought from peers in the field, from governments, and from colleagues.

The various visits undertaken within the period of the EUROTRA evaluation are summarised here, together with a brief listing of meeting inputs/outputs and attendees.

Files of all material generated during the study have been retained in Logica.

Visit Schedule Table (This table summarises the visits made by the Panel during the EUROTRA Review period.)

Location	Date	Attendees	Papers input / discussed	Meeting output
CENTRES				
Essex	21.10.92	Louisa Sadler, Doug Arnold BO, JR, ICR	Q response	Visit report
UMIST	2.11.92	Juan Sager, Harold Somers, Paul Bennet, Bill Black, Tsujii Jun-ichi, John McNaught BO, ICR	Q response	Visit report
Dublin	6.11.92	Jennifer Pearson, Lesley Davis, and colleagues (Dublin City University), Denis Toomey (BOLAS) BO, ICR	Q response	Visit report
Saarbrücken	16.11.92	Johann Haller and Jörg Schütz - IAI, Jürgen Wedekind - Univ Stuttgart, Birte Schmitt and Wolfgang Weißenberg - Univ Berlin, Heinrich Billing BMFT, Darmstadt SA, HS	Q response	Visit report
Leuven	17.11.92	Ineke Schurman, Frank van Eynde BO, ICR	Q response Leuven papers	Visit report
København	17.11.92	Bente Maegaard, Ann June Sielemann, Uffe Sonne Svendsen, Anna Braasch, Niels Jaeger, Annelise Bech, Lina Henriksen (CST) and Viggo Hansen (PaTrans); SA, HS	Q response	Visit report
Barcelona/ Madrid	17.11.92	Nuria Bel JR	Barcelona Q response Madrid Q response	Visit report
Paris TALANA	20.11.92	Laurence Danlos JR	Talana 2-page summary in lieu Q	Visit report
Torino Dima / Pisa	25.11.92	Cesare Oitana (Turin), Antonio Zampolli (Pisa), Luca Dini, Marco Buscaglione (Fiat) AO, BO	G.Dima Q response	Visit report
Luxembourg	27.11.92	Tom Gerhardt BO, ICR	Q response	Visit report
Utrecht	4.11.92	Steven Krauwer BO, ICR	Q response	Visit report
Liege	4.1.93	Jacques Nöel BO, ICR	Q response Liege papers	Visit report
Athens	22.01.93	Prof G Carayiannis and colleagues HS, BO	Q response	-
Lisboa / Porto	18.01.93	Maria Helena Mateus and colleagues BO, ICR	Q response	-

OTHER				
Coling	24.7.92	See below + BO	Papers from BO, NV	
Meeting 1 Luxembourg	25 9.92	Panel, CEC [Ref EWP1] Agenda : 1) ToRs, 2) Introduction to Eurotra, 3) Pannenborg / Danzin reports, 4) Panel's Work Programme, 5) Questionnaires, 6) Visit Schedule, 7) Meeting Schedule	List of LG Members List of EAC Members Eurotra Summary SP	Meeting 1 Minutes
Serge Perschke	12.10.92	SP, JG ICR	Questionnaire for SP	Report input to Meeting 2
Meeting 2 Luxembourg	19.10.92	Panel, CEC, [Ref EWP2] Agenda : 1) Minutes, 2) Panel activities, 3) Shared Cost projects, 4) Eurotra Costs, 5) LRE, 6) Danzin/Pannenborg, 6) Mondorf	Pannenborg/Danzin Summary Summary ET-10 projects Nino Varile Coling Summary Consolidated Costs Eurotra	Meeting 2 Minutes
Liaison Group Luxembourg	27.10.92	LG (see below) BO, ICR	-	Minutes
PE Luxembourg	28.10.92	Neil Simpkins, Gordon Cruikshank, Marino Groenendijk; BO, ICR	-	Visit report
Belgium Science Policy Office, Brussels	17.11.92	Monique Meert BO, ICR	-	Visit report
BIM, Brussels	17.11.92	David Sedlock BO, ICR	-	Report from BO/ICR
UK DTI, London	19.11.92	Gerry Gavigan, Peter Rothwell BO, ICR	Questionnaire	Notes from BO/ICR
IBM France, Paris	23.11.92	Jean-Marc Langé BO, ICR	-	Notes from BO/ICR
SITE Paris	23.11.92	Bernard Sitei BO, ICR, JR	-	Notes from BO/ICR
Meeting 3 Luxembourg / Mondorf	27-29.11.92	Panel, CEC, Central Team, Experts [Ref EWP3] Theme : analysis with input from peers Agenda : *(27.11)* 1) Meet with Lux Team, 2) ALEP, 3) ET-10, 4) DEMO *(28.11)* 5)Visit Reports, 6) Q Analysis, 7) Org & Mgmt, 8) Experts, *(29.11)* 9) Panel Discussion	DEMO and handouts LRE Summary ET-10 Summary Central Team Discussion	Minutes for Tom Gerhardt ET-10 minutes Visits minutes Frans de Bruïne minutes Experts minutes Panel closed session minutes
Meeting 4 Luxembourg	14.12.92	Panel, CEC [Ref EWP4] Review of material to date Review of Draft Report	Conclusions & Recommend's ELSNET contacts ET-10 work LRE programme ALEP Papers	Meeting 4 Minutes
Serge Perschke	5.1.93	SP (+ CEC) BO, ICR	Draft Report	Amendments to Draft
Meeting 5 Brussels	25.01.93	Panel, [Ref EWP5] Frans de Bruïne	Draft Report	Amendments to Draft

Serge Perschke	29.01.93	SP (+ CEC)	Draft Report	Amendments to Draft

PANEL
Brian Oakley (Chairman), Sture Allén, Alessandro Osnaghi, Iain Rae (Secretary), Jean Rohmer, Hartwig Steusloff

MEETINGS 1-5
Scheduled for the Panel, with presentations from variuos CEC personnel.

CEC DGXIII-E Eurotra management (at time of review)
Serge Perschke (Eurotra 1978-), Frans de Bruïne (Director), Jose Gasset, Nino Varile, Roberto Cencioni, Franco Mastroddi (advice).

CENTRAL TEAM (at time of review)
Lidia Pola, Anna Becker, Karsten Strørup, Poul Andersen, Dominique Groenez, Achim Blatt, Erwin Valentine

LIAISON GROUP (at time of review)
Cesare Oitana, Italy (It); Paul Schmidt, Germany (D); Laurence Danlos, France (Fr); Jennifer Pearson, Ireland (Eire); Tom Gerhardt, Luxembourg (L); Maria Gavrilidou, Greece (H); Paul Bennett (UK); Steven Krauwer, Nederland (Nl); Paulo Alberto, Portugal (P); Ineke Schurmann (Leuven), Archie Michiels (Liege), Belgium (B); Bente Maegaard (Chairman), Annelise Bech, Danmark (Dk); Nuria Bel, Spain (E) + Serge Perscke (SP), Jose Gasset, Nino Varile, Poul Andersen, Karsten Stronip (Commission)

A.3.4

COLING 92, NANTES

Experts
Prof Eva Hajicová, University of Prague, Deputy president International Comittee of CL
Pierec Isabelle, Dept of Communication, Canadian Government
Gregor Thumair, Manager, Siemens METAL team
Prof Alan Melby, Brigham Young University CL expert

Eurotra observers
Nuria Bel, Barcelona University; Bente Maegaard, CST Copenhagen; Louisa Sadler, Essex; Prof Antonio Zampolli, Pisa University; Nino Varile, CEC

MONDORE-LES-BAINS

Experts

Antonio Zampolli, Pisa.	Not directly involved in Euirotra - interest is CL in general. In particular last year the feasibility of tools and natural resources;
Steve Pulman, SRI.	Director of SRI International in Cambridge. Work in CL and formal methods. Also has position in Cambridge Univesity. Also the designer of Alep and working with BIM;
Bente Maegaard, København	Director CST. Doing Eurotra work and othe NLP work. Head of Danish Eurotra team and Chair LG since 1986;
Pierre-Yves Raccah, CNRS.	No connection with Eurotra. Working in AI. Interests include logic programming;
Maghi King, ISSCO, Geneva.	Specialising in NLP. Involved with E. 1978-87. Semantics and MT, nowadays, running the Institute;
Jean-Louis Binot, BIM	Advanced information processing group head;
Tom Gerhardt, CRETA.	With Eurotra from the beginning. Was member of Univ of Saarbrucken. In 1989 came to Lux as PM of the Lux team . Director of new institute in Luxembourg;
Frank van Eynde, Leuven.	Involved since 1979 coordinating the Eurotra team at Leuven. Founder of centre for CL at Leuven.;

4

EUROTRA Questionnaire Responses

As part of the Panel review process, questionnaires were developed and circulated to all Centres involved in the EUROTRA work. The completed questionnaires have now been received from most Centres.

From the ten questions asked we have created tables of replies. Each table reproduces the question in full.

Clearly, to copy each response for inclusion here would introduce a voluminous Appendix, swamped by detail. On the other hand, filtering out comments in a selective manner from particular questionnaire responses whilst introducing brevity, would neglect some Centre's comments.

We have chosen the middle ground. A Centre's comments on each question have been examined, filtered (using our judgement alone), and included. However, for brevity, each entry is a precis based on what we believe is relevant or important. We have included footnotes of explanatory remarks.

Finally, the purpose of this Appendix is to show points raised by the Centres in response to our specific queries. The points must be read in the context of EUROTRA, and our analysis elsewhere, and should not be quoted out of context. Our interpretation of points' meaning may not coincide with the intention of a Centre. Where comments have not been received, we have delved into the Final Reports, and extracted quotations as appropriate. The full questionnaire responses remain on file for future reference.

EUROTRA Questionnaire Analysis

Question 1 Involvement by Team

1) Date start work by the Centre

2) Average amount of professional time, in man months, put into Eurotra by workers at the Centre over the period of involvement a) senior researchers, b) postgraduate, c) others

[Editor's Note. Of all the questions, this one has been the most difficult to quantify. Many responses have given total man months, which seem easier to estimate. *Italics* are our estimates. Jan (Oct) work started before funds awarded.]

3) Total value of funds received for Eurotra-related work from: a) CEC, b) National Government, c) Other

	Start	Time	M/mths		Funds	K ECU	
Centre		senior	postgrad	others	CEC	National	Other
Leuven	Oct 1984(1)	46.5 (2) total	346(2)	111(2)	1159(3)	747(3)	-
Liege	Apr 1986(1)	45 total (2)	107 (2)	74 (2)	558 (3)	422 (3)	0
Københavns	Oct 1984 (1)	1 person	13 person	1.5 person	4032	1008	-
Barcelona	Jan (Oct)1987	*70*	*400*	*80*	2040(1)	872(1)	0
Madrid	Jan (Dec)1987	37 total?	234 (2) total?	97 (2) total?	-	-	-
France	1988-(1)	*172* (2)	*470* (2)	*50*+(3)	*1112* (4)	*3336* (4)	255 (4)
Saarbrücken	1985 (1)	360 total?	1200 total?	200 total?	1390	9488	-
Athens	May 1985	480(1) total	360(1) total	120(1) total	1640	248	-
Dublin	Dec 1984	192 total?	66 total?	72 total?	765(1)	180	0
Torino Dima	1986(1)	*150 total* (2)	*420 total* (2)	*36 total* (2)	1241	0	6.5
Pisa	1986	*30 total* (1)	*300 total* (1)	*36 total* (1)	*200* (1)	-	-
Luxembourg	July 1984	109 total	233 total	90 total	3107	345	(1)
Utrecht	Sept 1986(1)	12av (60 tot)	124av (748 tot)	19 av (116 tot)	1737	1098	0
Lisboa/Porto	May 1987 (1)	32.5 pa?	97.2 pa?	26 pa?	1920 (2)	72 (2)	-
Essex	Oct 1985(1)	1 f/t throughout	6-9 f/t RAs	1 h/t (f/t 91/92)	1327.5	1257.5	0
UMIST	Oct 1985(1)	*120* (2) total	*360 RAs* (2) tot		690 (3)	1323 (3)	0
Totals		2000+	8000+	2200+	22918	20396	261.5

Editor's note.
The figures given here do not fully correspond with those calculated using Commission information, and used earlier in this Report. No attempt has been made to validate the figures above.

Footnotes to Questionnaire Analysis Table 1

Leuven
(1) Also, advice during 1978-79; ET-4 (1980); ET-7 (1981); ET-10 & ET-10sem (1982); ETL-1, ETL-3 and ETL-4 (1983); ETL-5, ETL-8 (1984)
(2) 1984-90 (a 46.5, b 268, c 93.2) 1991-92 (a 0, b 78, c 18)
(3) 1984-90 (952.7 CEC, 639.3 National), 1991-92 (162, 108). Excludes 1992-93 45K received from CEC for training. Additionally Leuven received 300K ECU for contributions to the development of software and linguistic specifications of general interest (1987-90), and 62K ECU for a project in computational morphology and lexicography 91-92.

Liege
(1) Talks in September 1985 between Nancy and Liege
(2) Taking Prof J Noël (av 5%)/ Prof A Moulin (av 5%)/ Dr Archie Michiels (av 50%) as senior. Taking computer officer as 'other'.
(3) Financial provisions Eurotra-I 1986/04/01 to 1990/12/31 750K ECU; Eurotra-II 1991/01/01 to 1992/12/31 230K ECU; Total 980K ECU. Comprising 558K ECU CEC, 422K ECU national government.

Kobenhavn
(1) Study work 1981-1983 not included.

Barcelona / Madrid
(1) This is the combined España figure through to 1992, and includes addendum contracts. Both the Universidad de Barcelona and Universidad Autonoma de Madrid Eurotra teams are managed by Nuria Bel.

France
(1) The work is currently centred in Paris under Professor L Danlos. This situation was arrived at in 1988 when LADL and CELTA were moved to TALANA (treatment Automatique du LAnguage NAturel) in Paris Apart from L Danlos full-time, we have estimated research effort.
(2) Estimate only - based on Final Report for 1990. CNRS LADL (Laboratoire Automatique Documentaire et Linguistique) under Professor M Gross were involved in 1987. We have assumed 2 research students and 1 full time professor. CNRS LLF (Laboratoire de Linguistique) under Professor A Culioli in Paris was involved from 1985-1987- we have guessed the research effort. CNRS CELTA (Centre d'Etudes sur le Langage et la Traduction Automatique under Professor G Bourquin in Nancy have been involved since 1985. Professor B Vauquois Director of GETA developed the MT prototype "Ariane" on which the first specifications for Eurotra were based. The involvement of GETA ended in 1987. We have made assumptions about man effort in all these.
(3) Additional funding over the period 1988/89 51 man months for individual consultants.
(4) This is the total for France since July 1985 through December 1990. From Final Report.

Saarbrücken
(1) The University of Saarland started an MT project in the 1960s : this became the SFB 100 project funded by BMFT, CL work at this early time took place in Stuttgart, Berlin, Hamburg and Bielefeld. The SFB 100 team had been involved in Eurotra since the first meetings held in 1978. In 1984 the decision was taken to place the Eurotra team in Saarbrücken. The other Universities offered subcontractors to Saarbrücken (eg Bonn for terminology). The other Universities set up allied research groups. In order to effectively manage the Eurotra work, the IAI (Institute for the Society for the Promotion of Applied Information Science) was founded.

Athens / Crete
(1) At the start of the Eurotra work there were two centres - Crete and Athens. Since 1989 the work has centred in Athens. These figures are for Athens.
(2) There is a discrepancy between the Final Report and the Questionnaire. The latter states 2019 ECU from the CEC

Dublin
(1) The Eurotra Ireland team was initially to the National Board for Science and Technology. In mid-1988, Dublin City University took over the work, and the new team was recruited in January 1989. Comprising 45K ECUs for training, 720K ECU CoA.

Torino - Gruppo Dima
(1) Gruppo Dima has been actively involved in preparations of Eurotra since 1978, through its participation in the workshops, writing of papers, etc., and completing CEC 'study contracts' 1980-1984.
(2) Estimates taken from the Final Report.

Pisa
(1) Eurotra Pisa did not exist before the signature on the contracts of Association. However, Professor Antonio Zampolli actively participated in the preparation of the work prior to this date. Also ET-4 (1980); ET-7 (1981); ET-10 (1982); ETL-4 (1983); ETL-7 (1984). All figures are estimates - no questionnaire response was received.

Luxembourg
(1) The national government paid for the operation of the hosting institute " Institut Européan pour la Gestion de l'Information"

Utrecht
(1) In 1980, the founding members of Eurotra approached researchers from the Technical University of Delft (A G Sciarone), and the University of Utrecht (S Krauwer). Up to this time the Catholic University of Leuven had looked after the interests of the Dutch language. From 1981 through 1984 the work on Dutch was done collaboratively on study contracts between Leuven, Delft and Utrecht. The Netherlands gave advice on Eurotra 1978-80. Study contracts included : ET-7 (1981); ET-10, ET-10sem, ETS-1 (1982); ETL-3, ETL-4, ETS-3 (1983); ETL-7, ETS-6 (1984); ETL-7, ETS-6 (1985); ETS-9, ETL-9 (1986).

Lisboa / Porto
(1) The staff of Eurotra Portugal included linguists from the Universidad de Lisboa, Universidade Nova de Lisboa, and Universidade de Coimbra.
(2) To 1990, extracted from Final Report.

UMIST
(1) Both centres applied to be members of the Eurotra coordination group set up in 1978. By 1986, there were 8 staff in UMIST and 6 in Essex. Advice on Eurotra was given during1978-82. Study contracts were : ETL-1, ETL-3, ETL-4, ETS-3 (1983); ETL-5, ETS-6, ETS-8 (1984); ETL-8, ETS-9, ET-PR-1 (1985); ETS-12-UK (1986)
(2) Estimated from Final Report and questionnaire response.
(2) Obtained from the University accounting centre.

Essex
(1) Both centres applied to be members of the Eurotra coordination group set up in 1978. By 1986, there were 8 staff in UMIST and 6 in Essex. Advice on Eurotra was given during1978-82. Study contracts were : ETL-1, ETL-3, ETL-4, ETS-3 (1983); ETL-5, ETS-6, ETS-8 (1984); ETL-8, ETS-9, ET-PR-1 (1985); ETS-12-UK (1986)
(2) CEC : 690 KECU CoA (85-90); 495 KECU Other (83-90); 142 KECU CoA (91-92). Nat Govmt : DTI £1020K CoA (85-90); £237.5K (91-92).

Studies for Eurotra
ET - * [Study for Eurotra]; ET-*sem [Semantic research]; ETS - * [Software specifications]; ETL- * [Linguistic specifications]

EUROTRA Questionnaire Analysis

Question 2 Quality of Research

1) Key achievements. List with brief description, up to three of the key intellectual achievements of your Centre or national team

2) List, with brief description, the key intellectual achievements of the whole Eurotra programme?

3) Relative achievements i) How do these Eurotra achievements compare with others made in the MT and NL fields during the Eurotra involvement period? ii) What were the most important achievements, whether outside or inside the Eurotra programme, made during this period worldwide in the MT and NL fields during the Eurotra time period?

4) What were the technical and administrative weaknesses (if any!) of the Eurotra programme a) in your centre, b) Eurotra as a whole?

Centre	Key Achievements-Centre	Key Achievements - Eurotra	Relative Achievements	Relative Weaknesses
Leuven	a) Linguistics - esp Tense & Aspect(4) b) Eng-Du, Ger-Du modules in DEMO & monolingual Du c) MT evaluation	Unification based formalism m/stream CL - but slow, needs small dictionaries. Good linguistic specs used by METAL, Eurolang. Formal description in 9 langs. Grew NLP, MT in Europe.	i)Large size, decentralised, multi-lingual, internat'l, open, project. Transfer based interlingual subparts, sentence driven, rule governed (not example, corpus based). (5)	a) EU-LE found it easy to recruit CL, but difficult for computer scientists. b) CoAs awarded different times for Leuven & Utrecht- extra work for LE on Dutch. Cheap, poor performance computers.
Liege	Sidelines. Design & Devpt of Horatio; proposal for 'frames of terms' - rejected by I.G.	Raising of NLP awareness, rather than results.	Better picture for Eurotra results when sidelines considered.	Mixture of scientific with day-day mgmt. Focus on academia, image Suffered from NIH syndrome.
København	a) Formal linguistic theory of Danish (valency, lexical semantics, morphology, etc) b) running transfer pairs in Eng/Fr /It c) 1987-90 E-f/wk & preference mechanism d) 1987 coverage of description and systematic testing strategies	1st program for producing prototype multilingual MT - linguistic description of transfer-based. Multilingual MT system. Unique European MT network. Raised awareness.	i) Compares well with research elsewhere. Eurotra is state-of-the-art MT. ii) a) Unification formalisms introduced, b) Knowledge base MT explored, c) Statistical methods came up again in ET-10	a)- b)i) Only mgmt tool in CEC was financial withholding (not used) ii) Some Nat Gov'm't created obstacles. iii) CEC slow in preparing CoAs iv) Not all CEC staff able to scientifically monitor
Barcelona	To catch up with rest of Eurotra in early years (with formal development of Spanish grammar).	Linguistic theory, unification like for Eurotra(2). Formal descriptions for 9 languages. Prolog usage.	ia) Eurotra Spanish grammar unique. ib) It influenced METAL, showed viability declarative form. ii) unification grammars, linguistic approaches (eg Lilog,Rosetta), knowledge based represent'n, discourse analysis, expressivity etc.	a) Lack of precedents : the group had to solve problems and overcome difficulties mainly due to lack of expertise (work & contacts), b) Admin control not realistic, technical goal (pre-ind prototype) not achieved.

Madrid	Devpt of a theory of the lexicon. Study of Spanish nominal & verbal paradigms. Criteria for identification of terminological units established.	Def'n. of 'linguistic' theory of translation, & IS. Creation of dict of semantic readings. Modularity. User language for formalism.	Eurotra is only MT system based on CL advances - IBM"s MAT & MENTOR are based on ad-hoc Prolog. Eurotra's huge bibliography.	Bad flow of information between centres in Spain. No Madrid infra-structure - unequal distrib'n of res-ources. Computer skills input denied. Deficient s/w tools.
Paris TALANA	Substantial French grammar. Verb construction for Ref Man	-	-	French teams were excellent in linguistics but not in CL - this led to initial difficulties.
Saarbrücken	a) Use of Unification Grammars b) Systemization of Semantic Labelling c) CAT2 sideline	a) International science cooperation (eg prod of Ref Manual V.7 (800 pages) b) Feasibility of multilingual MT	i)IAI/ET-D was peer reviewed at yearly Statusssseminars-favourable Compared well with other NLP systems ii)New grammar form's. Creation of lexical & grammatic'l resources for new languages in same format as rest.	a) None b) Decentralised democratic structure not ideal for R&D. CEC team resource limited - lack of coordination. Delays through interdependency of Centres' results.
Athens	'Basic linguistics research' in Greek - this is the only existing formal grammar. Devp't of preference mechanism for overgeneration problems. 1991 foundation of ILSP.	Proved multilingual approach to MT is feasible. Supported comparative / contrastive research on all EEC languages. Formal description based on a unification formalism - all levels of linguistic analysis (text process through to semantic represent.)	Greece: active CL did not exist. Funds for this were non-existent before Eurotra It provided staff opportunity for seeing what was going on worldwide in CL. Recognition of necessity to combine linguistics, KBS, AI, statistical methods in MT.	a) PTT communications delays. Greek marginalised. Centre inexperienced in infrastructure. b) Successive modifications of linguistic theory & implementa-tion strategies - necessary, but repetitive.
Dublin	(1) i) Sublanguage research ii) Terminology research iii) Terminology resource	Showed multiling' MT feasible. METAL/Eurolang used IS struct' Enhanced European langtech/CL Comprehensive grammar/lexical resources in 9 languages.	a) Dublin only sublanguage centre, apart from Irish gov'm't pubs office b) Work based on philosophy of Infoterm in Vienna.	CL expertise grown in Centres - some didn't need it, others did (eg Dublin, Lux) but didn't get it.On balance Eurotra trained people to start work in CL/MT.
Torino Dima/ Pisa	Coindexation tool for dealing with unbounded dependencies A preference mechanism for linguistic objects. Lexical semantic theory (Italian)	Eurotra has improved upon the transfer scheme by the stratific-ational architecture based upon declarative approach. Common IS tested for 9 langs, belonging to 3 families (Rom/Germ/Other).	-	-

Luxembourg	Group not a research team: acted as a service team in the field of clearing house functions & documentation. Scope from 1989 includes AI R&D, & PR services for the Eurotra teams.	Intensified R&D in MT all over the world & motivated European CL staff towards MT. Unification grammar / formalism Purely declarative approach, Mathematical translation model.	a) No NLP research in Lux before 1989 - afterwards speech project. METAL used ET-6 results to improve their system. Raised European visibility in NLP (eg Euro Assoc'n for MT) b)MT is an application of CL	a)Frequent change of mgmt & staff. Preferred to have a CL task. Lack of future for CRETA. Dispersed Eurokom document stores. b) CRETA & CEC relations sometimes perceived indistinct.
Utrecht	MIMO, MIMO-2. Eurotra showed MT can be based on mainstream CL. Generic lexical resources.	Intellectual - not products or major contributions (except ALEP, Ref Manual) to MT. Linguistic MT approach ; Europe-wide MT	ia) BSO's DLT; Philip's Rosetta (better than Eurotra). ib)Eurotra not involved in mainstream CL ic)commercial desktop systems ii) transition from procedural to declarative; from rule to include statistical.	a)Utrecht(1) started late. Fast growth of too large team. Unreal CEC planning. Recruitment difficult. Uncertainty. Bad s/w performance b)Slow CEC procedures. Manager also science head. Poor LG working & CEC support.
Lisboa/Porto	Formal represent'n of Portugese grammars (analysis & synthesis) Creation of mono/ bilingual electronic dictionaries & terminologies. Devp't of basic NLP research.	Devp't in CL, MT, NLP. Influence of multilingual transfer approach on other MT systems Creation of an alternative MT approach - focus on linguistics.	ia) Previously nothing done in CL in Portugal. Now ILTEC work on electronic dictionaries / modular grammars / spell & syntactic checkers, etc b/c) MT with 72 language pairs	Demo work neglected. Decentralised, interdependent work - lack of integrated testing & breadth of testing (new problems will arise with free input). Transfer v synthesis unequal weighting.
Essex	Linguistic contibutions to Ref Man, MT evaluation, MiMo with Utrecht, 'constraint-based' MT.	Ref Man - resource for NLP work in CEC langs; ETS approach to 'transfer' and 'synthesis' relation; Put MT on the map in Europe. CAT2, MiMo, ETS ideas key. (8)	No other project has Ref Man equivalent. Poor cost-benefit in Eurotra cf other work. Linguistic & formal f/work in Eurotra not as mainstream as it could be. (8)	a) no serious problems, locally or at nat'l level. b) sheer size; lack of early infrastructure (later solved using Eurokom) - Eurotra pioneered - miracle it worked at all!
UMIST	Results of linguistic research (compounds, lexicology, control relations, clefting & topicalisa'n). S//w prototype experience. (6)	Explored idea that a 'translation relation' relates to linguistic properties of texts done in a multilingual way. "IS legislation & contrastive research reports".	i) Not interactive. Not looked at disambiguation - problem area. ii) Ignored lexical tools. Scaling up will introduce ambiguity. Other work- statistics & lexical based	a) Pointless implement'n tasks (eg targets for transfer of dicts, & changes), b) Easy access to structural information failed. Late integration of morphology.

Footnotes are collected on the following pages. They have been selected to expand the various points above.

Footnotes for Questionnaire Analysis Table 2.

(1) Reproduced from **Utrecht** response, because it exemplifies the concerns from several centres. "A last problem which we would like to mention is the problem of having to serve too many masters at once. First of all there were the contractual obligations, but since the tasks could most of the time not be classified as basic or applied research tasks, we had a major problems in justifying our activities to the colleagues and management of the Research Institute for Language and Speech of our University, in which our activities were embedded. Traditionally the output of research institutes is measured in terms of publications per researcher, and the project deliverables could hardly be counted as publications. In addition, we felt that for our employees their employment for Eurotra should constitute a step in their academic career, which meant that eg just letting someone write 5000 dictionary entries would be too meagre an output for one year of academic activity, and would hardly contribute to a better post afterwards.

(2) Reproduced from **Barcelona** response since it deals with a key component of Eurotra. "At the Interface Structure , as a level of representation, most of the surface discrepancies among individual languages have to disappear resulting in a common language of representation which will allow the systematicity of the relation between the languages. At the IS two kinds of knowledge are needed : knowledge for the mapping between monolingual motivated descriptions; knowledge for the mapping between ISs of different languages. The IS approach is economical only if the system intends to deal with more than one pair of languages, because decreasing transfer complexity with an IS implies increasing the labour of monolingual modules : analysis and generation. IS can also be seen as a level of representation where generalisations from a monolingual point of view can be stated. In that respect IS will act as an abstraction level of linguistic objects, ie Deep structure, which should be useful for other purposes than MT."

(3) From the **Dublin** response on terminology and sublanguage. "i) Sublanguage research. Until recently the focus of linguistic research ...in MT was to provide a description of language as a whole...as broad a base as possible. It is now generally recognised that MT is more likely to be feasible with systems which are designed to handle a subset of general languages. Eurotra Ireland has invested a considerable amount of time in research into sublanguage and text types with a view to i) identifying characteristics of sublanguage, ii) providing a list of criteria for assessing the suitability of material for MT.These criteria have been applied to select an appropriate sublanguage/text type (dress-making patterns). The results...presented at 5th Irish Conference on AI and Cognitive Science September 1992. ii) Terminology Research. As with sublanguage, the importance of terminology within the context of MT is now being recognised. Multi-word terms, in particular, do not necessarily have the same internal structure as multi-word units in general language texts and are therefore not amenable to the parsing strategies for general language texts. Eurotra Ireland, in collaboration with Eurotra Greece and Eurotra Portugal, has proposed a number of solutions to the problem based on the current prototype formalism. The results...[applicable to] any MT system. iii) Terminology Resource. Development of a medium scale multilingual terminological resource in the field of telecommunications. Ripe for exploitation under the aegis of LRE-style initiatives or industry initiatives documented in a number of published papers. iv) Terminology and Extra Linguistic Knowledge (ET10/66). Eurotra Ireland is the coordinator of a research consortium which is building an ontology using the extra linguistic knowledge associated with terms with a view to facilitating disambiguation in MT."

(4) From Questionnaire for **Leuven** because it illustrates some linguistic points. "Leuven's main achievements are in the field of linguistics research, especially in the areas of semantics and morphology : Tense and Aspect, Mood and Modality, Determination and Quantification, Aktionsart, Comparison, Transconstructionals and Compounding. The expertise of the Leuven team in these areas is clear from the fact that the Reference Manual contributions on these topics were all (co-) authored by members of Eurotra-Leuven. The work on Tense and Aspect has been of particular importance, in that it has resulted in an interlingual and computationally tractable treatment of phenomena which are notoriously difficult in translation and which other systems can only handle in a transfer based language pair specific way. In the meantime, other systems have imported various aspects of the new treatment."

(5) From Questionnaire for **Leuven**. " ii) In general, the most important achievements in MT/NLP were... The appearance and relatively quick acceptance of unification based grammars and Prolog based formalisms. The rapproachment between properly formalised theoretical linguistics and computational linguistics, especially in syntax (cf NLP on the basis of LFG, GPSG, HPSG, Categorical Grammar), but also in semantics (cf the growing influence of model theoretic

semantics). The first attempts to integrate real world and domain specific knowledge in NLP systems (cf knowledge based MT). • The attempts to make use of large corpora and statistical data in NLP (cf example based MT). The availability of more powerful, faster and - at the same time - cheaper machines.

(6) From UMIST Visit (*editor's notes on Japan& UMIST*). The Japanese connection came from discussions over lunch (Doug Arnold, Pete Whitelock, Rod Johnson) and a visit to Japan. Doug went to Japan for 15 months - to NTT (where Nomura San was), and Toshiba (fellowship), and twice yearly Japanese delegations would come to UMIST. There were 2 Eurotra Japanese meetings (Geneva, ET/1 contract + another). Good relations were built - UMIST funded a Chair, and Fuji San was given the position in 1988. He has brought and maintained contracts - UMIST now has visitors, students, and teaches Japanese.

(7) From Lisboa/Porto *"Evaluation of the Eurotra Project"* paper sent to the CEC, 16.07.92. " It seems to us that thinking about...whether main goals have been achieved...[but] whether things were set up in such a way that eg information flew and views were exchanged as frequently as it was supposed to....In our opinion,...namely that Eurotra represents a "relative failure" with respect to objective of building a "significant prototype", and the success in building expertise, correspond to what we could consider the visible and the hidden parts of the project. *Visible part*. It consists of i) software, ii) grammars, iii) clusters, iv) DEMO. Everybody agrees the software (not just the machines) is far too slow and had a damaging effect on the speed of the grammars. On the other hand it is clear that the grammar and theoretical work were the core of the project. Evaluation of his part is not very controversial...[Pannenborg and Danzin reports]. [The weakest point] is due to the fact that ...groups were never able to test their grammars in a systematic way. with what we could call reasonable accuracy, and Eurotra was never really confronted with as wide a range of translational problems as it could have. In other words we believe completely new problems will arise when Eurotra will be able to accept free input and provide a translation in less than a minute....Eurotra suffered from an unequal distribution of work carried out in analysis and synthesis, on one hand, and in transfer on the other - partly due to lack of CL expertise in some countries. Probably due to unclear guidelines, DEMO has been neglected. *Hidden part*. All agree this is the positive aspect. PT agrees with DK that the creation of an institution for language technology would have been impossible without Eurotra.

(8) From Essex response. "It should be remembered that before Eurotra there were essentially no computationally usable descriptions of many European languages. The ETS formalism provides an interesting attempt to solve some general translational problems eg the relation between 'transfer' (ie bilingual knowledge) and 'synthesis (monolingual knowledge). Several of the ideas embedded in CAT2, MiMo, ETS have been important: compositionality, the role of explicit descriptions of different levels of representation. Relative achievements : the quality of some of the descriptions of English is comparable to other large scale projects such as the Alvey tools in the UK, or the work at Hewlett-Packard in Palo Alto in the USA - though the result is not as interesting, because these projects were based on more 'mainstream' formalisms. PATR2 developed in the early 1980s, is the simplest of the generally accepted 'standard' models of linguistic representation (the 'feature structure', typed or untyped) and of a standard operation, namely unification. What was significant about this development was that it lead to formalisms in which it was possible to produce computationally usable, and linguistically well-motivated descriptions, which has lead to a convergence of linguistic and computational research, within formalisms / theories like GPSG, LFG, HPSG, etc. Something similar happened in morphology, where a model involving Finite State Machinery has evolved as standard. Though the distance between Eurotra and these other theories / formalisms is often overstated, it would have ben better if Eurotra had been more mainstream - eg by adopting one of these theories. This would have made the work that has been done of wider interest and value. Various moves in this direction were proposed early on but rejected, and it should be admitted that there are still difficulties with these formalisms / theories in the form of open research questions, & none of the theories provided a theory of MT as such - in this respect CAT & its successors are of interest."

EUROTRA Questionnaire Analysis

Question 3 Work Programme

1) Was the right work programme selected? In retrospect, would you have dropped some parts, built up others?

2) Was the Running prototype an appropriate part of the programme in view of the state of research? Was it right to mix the Development work associated with the Running Prototype in with the research programme?

Centre	Work Programme	Running Prototype
Leuven	Better to have a more gradual transition from preparatory research to large scale implementation : first test prototype for 1/2 language pairs, then 9; focus on better s/w rather than formal properties of virtual machines. Delays from interdependencies (esp 2/3 phase). It was de-motivating in	...phase 3 that ETS contrastive work not used beyond '92. Drop translation ALL pairs from start. Build up s/w, problem office (PO). Publish early. In 90/91 - reduced dependencies *right*, abandon Ref Man (PO) update scheme *wrong* - led to fragmentation of research & de-motivation. Q2. "Yes"
Liege	The transition phase work for Liege was rejected; continued French monolingual work was imposed. (2)	-
København	In retrospect, too ambitious. Over optimistic time planning (lack of efficient tools for implementation, no serious reuse of existing lexical resources). However, early discussion on dictionary size should have happened - impact on reusability. Lack of publishing early on - later change of attitude.	Entire running system (with comprehensive Danish coverage) can be attributed to the prototype. It has been used to test research results immediately. Adds an important element of concreteness to the research work. But it has been time consuming, & has constrained some research work.
Barcelona	[Better] balance between R&D. Waste of effort on large dict's recoding as grammar changed - better treat them as last step. Also late delivery of lexical tools (1991). Poor planning on	..developing modules, and experimentation - results neither robust nor exhaustive. However, 1991/92 was planned well, but the effort was used for redundant ETS system (pre-ALEP).
Madrid	Error to stress dictionaries before stable grammars - this had negative impact on dictionaries' development because constant revisions were needed. Lack of attention for s/w tools - they are rudimentary.	Not all research done has been incorporated. Strongly recommend mix prototype devp't work & research. Also recommend mix implementation with research *if* restricted lexical coverage to avoid major dictionary changes.
Paris TAUANA	Paris joined at start of new formalism - Nancy was already established and had to rewrite grammar. All was well through to 1990. Thereafter catastrophic consequences with..	..CNRS deliberations over CoAs, and a condition that the team be broken up Dec '92. Overall the team appreciated the research quality of Eurotra, & working with Eurotra centres.
Saarbrücken	Programme in principle right. Sometimes 9 lang strategy led to avoidable multiplicity. Wrong to form s/w group at CEC- should be close to grammar & dictionary writers. Need to have terms of contract longer than 1 year for staff.	The running prototype was appropriate because it is useless to work only on theoretical grammars. The right mix was not always chosen.
Athens	Given the options, the work programme selected was one of a few successful alternatives. More effort was needed for testing and basic research.	It received feedback from research throughout the programme. Mixing R&D was the only thing that could be done given demands/constraints, although for the Greek linguistic analysis this was not ideal.

Dublin	The change in emphasis over 1987-1991/92 towards terminology & sublanguage has allowed EU-IR to become more integrated. More effort on sublanguage /text types.	Conceived as Community initiative - difficult not to include tangible objective. In retrospect, optimism was premature given relative inexperience in European NLP.
Torino Dima/ Pisa	-	-
Luxembourg	A programme with one multi-national project with effects like awareness, basic research, development, s/w develop't, working groups & knowledge transfer, was a sensation itself. Eurotra's aim was addressed from initially poor linguistic structures - these have now been developed.	Problematic to conduct parallel R&D, but the ongoing confrontation was fruitful for the work. In the field of NLP it may be necessary to conduct work this way. The Saarbrücken EDP 100 work was conducted in this manner too.
Utrecht	Monolithic & incremental approach not for research project. Prefer multi-stranded approach : various theoretical explor'ns, methods, fewer language pairs, tools for MT.	Not a bad thing if oriented correctly. Large scale validation needs correct prototype. "Devpt activities (suggested by the Council Decision) should have been totally out of question"
Lisboa/Porto	The project had several reformulations & had as a consequence a slow dev'p't of the construction of grammar. The overall orientation of the program followed the formal generative paradigm, emphasizing too much the syntactic component of the language description, & neglecting...	..morphology & semantics. This had a bad consequence on the research and implementation in these fields, where no satisfactory results have been achieved. The Portugese team feel it would have been fruitful to have been integrated from the very beginning.
Essex	Wrong programme selected. R v D an unhappy mix. Goals were not met - needed to change them. Original programme too short, but accession Spain / Portugal meant programme too long without re-evaluation of goals and resourcing / timescales, etc.	Mix of R&D never right. Amount of development and time pressure meant always using immature tools - essentially research prototypes (actually were often only executable specs). On other hand research inhibited - modifications would have meant throwing away dev'pt results.
UMIST	Useful contrastive, too much emphasis on monolingual. Transfer module approach should have obviated this. Runn'g prototype used to test research results but lack of large dicts and coverage failed the aim 'pre-industrial' - this aim should	have been dropped. Team structure was not appropriate for developing the large computer system. Constraints of running prototype may have limited research on IS specs, and achievement of best possible solution.

(1) From France Final Report : "Professor B Vauquois, Director of GETA is one of the initiators of project Eurotra. He and his colleagues developed an MT prototype called "Ariane". Professor Vauquois' idea was to extend Ariane to the official languages of the Community. Nevertheless, Eurotra has been developed independently from Ariane, although the first specifications were largely based on Ariane. Ariane has been developed in a National project PNTAO."

(2) From Liege response. "As early as 1978 ie long before IBM's Yorktown Heights had published any important work in computational lexicography, the Liege team had pioneered work on MRDs, in particular the LDCE. Research contracts between Liege University and major publishing houses interested in encouraging academic research on improving dictionaries, and more broadly, on assessing the reusability of their lexical resources in MT and other fields. Hence Horatio sideline."

EUROTRA Questionnaire Analysis

Question 4 Output (also see Section 8)

1) Publications : List main publications stemming from Eurotra work of the Central team.

2) Exploitation : a) List main examples of the exploitation of the Centre's Eurotra work, b) Rough value, in financial terms, of exploitation.

3) Consultancy : How has the Eurotra work led to the Centre providing Consultancy? If so, please name the bodies to whom Consultancy has been provided, and the scale of the income generated.

4) Build up of relationships : To what extent has the Eurotra work helped to create relationships?

Centre	Publications	Exploitation	Consultancy	Relationships
Leuven	30 in journals '85->. 12 working papers (Utrecht/Leuven distrib'n w/wide) , in CEC Studies Books, 9 during Study contracts 80-84, 10 during CoAs 84-90, 9 during PO 86-90, 10 Du Utrecht/Leuven	implementation reports, 9 MA theses. a) AIM-project MENELAS b) ? Contact with DECIDE	No income: reviewing research proposals for KU Leuven Research Council & Dutch NWO. Referee/Editorial board. Expert for CGC-12, AC and TMC. Income from: LRE-1/62.	Univ Leuven contacts-> in Sept 1991 new CL Centre - 32 staff engaged in LRE, Erasmus, Tempus, Comett, COST, DELTA (esp HCM) - FoLLI contacts.
Liege	About a dozen Eurotra dictionary papers. Ref Man. Horatio sideline will appear in Eurotra Studies.	See Questionnaire table §4.3	Collins, Longman, OUP & Le Robert (Fr) signed contracts for mono & bilingual file research	With individuals rather than groups.
København	Over 60 since 1986.	PaTrans(1). Also project with Danish Research Council - KBS MT for car manuals.	To CEC in MLAP. To Industry. To Canon Europa - translation aids for manuals Eng-> European.	a) Unique society of researchers in NLP. Contacts with partners eg SRI, Prague School. b) Industry knew where to look for expertise
Barcelona	(87-90/91-92) 31/1 syntax, 19/10 semantics, 19/1 morphology, 13/3 lexicon/computational semantics, 9/1 formalisms, 6/7 contrastive studies. Total 128	IBM Spain (1 m/yr) 90/91 MAT linguistic specifications. Eurolang (1 m/yr) Spanish analysis & generation grammars for 93/94.	Consultants to National Administration Dept's, several Catalonian Dept's. Industry : GSI-Teesidel, EUDI & others. Always held 'increasing awareness' goal.	a) Contacts with Eurotra teams has led to ET/10 & LRE programmes. b) Participation in Eurotra was viewed as "guarantee" *(approval)* for industry.
Madrid	9 by Francisco Marcos-Marin 4 others by group	2 staff involved with IBM's MAT, managed by Barcelona	Staff evaluated ATAMIRI.	Created relationships in Europe. Madrid too zealous about IPR - limited exchange from Madrid. IBM more open, & trained staff.
Paris TALANA	39 publications, 31 communications since 1987	-	-	Participation in the Pragmatics Group, Dictionary Task Force, Experimentation group.
Saarbrücken	Over 100 since 1985	a)METAL MIR (Interface Representation) draws on Eurotra. SNI Muenchen. EUROLANG. b)not easily quantifiable - many	IAI is currently able to finance a third of its budget by consultancy bringing in over 200K ECU pa	With other European Centres, and NLP centres around the world. Cooperation agreements, exchange of staff. Industry partners

Athens	39 largely between 1989 & 1991	Staff involved with LRE, STRIDE, ESPRIT, ET-10; new collaborations with industry / public sector.	None	a)Built relationships with Univ of Athens, Nat Tech Univ of Athens b) Industrial partners in STRIDE, LRE, ESPRIT, etc.
Dublin	10 since 1991	a) Terminology database. b) Sublanguage : design of Irish language module.	Exploratory meetings with industry : advice on terminology management (Lotus,Softrans,ITP, Idoc). Siemens - Irish METAL.	a) Invaluable for broadening research base, sharing results. b) EU-IR fostered awareness of MT through Irish Trans Assoc
Torino Dima/ Pisa	Torino / Pisa : 6 MT, NLP; 6 contributions to working papers, 18 contributions to Ref Man; 4 PO & 26 Eurotra reports	Discussion with car manufacturer. Own sideline E-Star.	-	-
Luxembourg	Various internal papers. Editor of 3 journals and a newsletter.	a) Foundation of IT&S a.s.b.l; doc & cultural information server; MT translation agency b) No idea	Initial steps towards industry have been taken - no revenue to date.	MT Users' Group - also a pilot COSINE project. EAMT newsletter. CRP-CU ET10 66 contract employs CRETA staff.
Utrecht	Over 75 since 1985	No exploitation has taken place.	None	a) Academic contact Eur, US, Japan. -Utrecht recognised in MT. b) no exploitation contacts
Lisboa/Porto	15 between 1990 & 1991 *Exploitation* Spell checker,syntactic & stylistics checker, CALL	Building of Terminological Databases, Term Dictionaries. LRE - ET-10/TRANSLEARN EUREKA - GENELEX/GRAAL	1989/92 postgrad course in new technologies applied to translation. Also mailing algorithm.	ab)Created relationships with publishers and Portugese & foreign s/w companies. Worked with IAI on CAT2 project.
Essex	237 during 1983-90, 26 in 91-92	Eurolang will exploit expertise of Eurotra-Essex (worth several millions of ECUs). essex can bid for ET-10 and LRE projects.	2 weeks consultancy CIT Ltd - design of multilingual databases. Also UK DRA (4 m/m'ths, £22K) role of NL in req'mt specs.	Area where Eurotra had greatest success. Several groups now work closely. New relationships eg SITE.
UMIST	Over 237 since 1981.	Re Japanese connections and development	Rank-Xerox, BT, SERC. University of Technology, Kuala Lumpur - Eng/Malay grammar for ETS	Good contrastive research cluster relations & Dutch-English bilateral & Malaysia collab'n.

(1) Quotation from CST **København** : "Being developed by CST Denmark. The goal is to make a customised translation system for a private company for patents from English into Danish - it reuses and further develops the implemented Eurotra grammars and lexica for the two languages covered. We also built on the Eurotra software, which is enhanced and optimised, so as to comply with the requirements of a production system. This work was launched in 1992."

EUROTRA Questionnaire Analysis

Question 5 Trained Staff

1) How many research workers has your Centre trained as a consequence of Eurotra funding? Where are they now (eg industry, academic posts, etc)? Are they using MT expertise?

2) Has your Centre run any training courses, summer schools, etc., as a consequence of the Eurotra work?

Centre	Trained staff	Eurotra related training
Leuven	20 trained (4 to Univ, 5 to industry) 3 trainees on Eurotra grants 92/93. 8 trainees in Erasmus & Tempus programmes 88/92.	1985 organised 7th annual Eurotra workshop, and 2 3-day beginner courses. 1988 Pennsylvania/Leuven Summer School -> 3 week MT course. 1990 organised 2nd European FoLLI Summer School (50 courses, 9 workshops, 500 attendees. Also Groningen & Saarbrücken Summer Schools. New Centre for CL at Leuven.
Liege	Jacques Jansen looking for post. Mergeai is with Reuters, Luxembourg; Liebecq with EUROCARD. Gerardy works on Esprit in same department (EMIR multilingual information retrieval). None in	.... NLP as such. Fontenelle is assistant in English. Van Caille is with the MET Ministry. Delcourt is in corpus research in Liege University.
København	34 research workers, plus 7 students (2 of which are permanently employed at CST after graduation). Typically returned to Universities but also in industry (3-4).	EU-DK host for Nordic Seminar on MT 1986.(30 attendees). Participated in Gothenburg Summer School 1991. Host for "Sprogteknologisk Forum" 1992 (180 attendees).
Barcelona	40 trained since 1987. Leavers still in MT : ATLAS-II Fujitsu Spain, 6 other Spanish Universities, 1 New York.	Different PhD programmes and also summer schools.
Madrid	14 in Centre : 2 in industry, 1 NLP in US, 3 academics (1 Germany) , 3 lang teachers, 1 CL training in Essex - no CL course in Madrid.	School of Language & Industry (1990,91,92).
Paris TALANA	19 staff (Paris/Nancy). Continuation of NLP work in TALANA	Teams received considerable training from Eurotra work.
Saarbrücken	About 60. Several Univ professors, assistants, etc., involved in MT	One course pa in MT and MAT. 6-50 attendees. Participation in Saarbrücken Summer School Lang Ling & Logic 1991 (800 attend)
Athens	About 40. 5 now in USA/UK, 1 on Systran. 3 team members now Univ professors.	2 courses (91/92) 50 attendees. On NLP and CL. Also give lectures on C Lexicography & MT in other institutions.
Dublin	13 staff trained (4 left to University, 4 postgrads, 5 Eurotrians)	Telecomms terminology seminar (1989, 2 days, 30 attendees). 1992 Infotech MATS seminar. Demo EIRETERM Young Scientists Exhib 1992. University courses and seminars.
Torino Dima	*[20 staff trained]*	*Will try and integrate the two main trends in Italy (lexicography & AI) with the syntactic and semantic treatment of texts. Only Gruppo Dima seems to be focussing on MT.*
Pisa	*[12 staff trained]*	*Pisa is likely to exploit the know-how and experiences from Gruppo-Dima. A new faculty for communication science is being established.*
Luxembourg	8 trained. 4 now in consultancy. 1 private. 2 industry. 1 s/w company	Use of CRIS databases (Eurotra files) 20 attendees. HPSG & ET6 course 37 attendees. All for Eurotrians only.
Utrecht	Since 1986, 40 researchers. 13 (5 in industry) still involved.	Considerable influence on Faculty of Humanities teaching. Many researchers have taught both introductory & advance courses in CL.

Lisboa/ Porto	Most of the staff involved in EU-PT remain in ILTEC - some in Faculty of Letters in Portugese Universities.	
Essex	20 research officers during Eurotra period - most of these were linguists - no training in CL. 9 still with group, 3 academic teaching, 3 further research, 1 industry, 1 other, 1 teaching. Use of 9!/92 ...	.. resources for 1 research student. CL and MT expanded at Essex - new courses. Organised 4th European Summer School in Language, Logic & Information 1992 - 550 participants Europe, USA, Japan.
UMIST	17 staff involved since 1985. UMIST lecturers still in MT. Arnold in Essex. Johnson in Keil s/w house. Whitelock in Sharp Labs. Maxwell in Essex. Other research associates left & not in MT.	Hosted 1992 summer workshop. Difficult to separate Eurotra project work from general MT work, for which UMIST is UK's largest research and teaching group.

EUROTRA Questionnaire Analysis

Question 6 Infrastructure / Artifacts

1) Has the Centre been involved in any software tool, etc., development? Give examples and state whether this is seen as valuable work.

2) To what extent has the support of the central, Luxembourg, team been employed and of the value of this work?

Centre	Software Tools Development, etc	Use of Luxembourg Team
Leuven	Linguistic specs - Central team 87-90. Problem Office 88-90. Pragmatics group 88. Devpt s/w specs 87-90. NL/B team (UT/LE) developed a front-end (in place of EMS module) -used in AIM MENELAS	The CEC team was not really involved. The Central Team on Linguistic Specs was established exactly because CEC lacked the know-how. CEC lacked experience - translators rather than CLs used
Liege	A small parser generator in Prolog for testing frame proposals (system ANAL) also has Eng-Fr translation module.	-
København	S/w & Formalism - E-f/w design/devpt; SGML front-end; preference mechanism; parts of Ref Manual. Linguistics - Ref Man (modality, support verbs, dictionaries). Testing - systematic testing, test suites.	S/w - cooperation was good during E-f/w and front end development, and also in creating lemma dictionary via database tool. Taking over from Maghi King, the Lux team coordinated Ref Man satisfactorily
Barcelona	Members of the Spanish group have participated in central activities such as copy operators development, pragmatics group, linguistic specifications group, Ref Manual, experimental implementation	The Commission team has augmented its activities along the life of the project from offering just clerical support to collaborating and participating in technical topics.
Madrid	No involvement.	-
Paris TAUANA	-	
Saarbrücken	Several study contracts 1983-86 on s/w and formalism design and spec. Starting point for CAT2 sideline. Latest ET-6/7 design work in direction of mainstream ling. Same for ET-9 industrialisation.	Worked only satisfactorily during last three years where the official prototype reached a better performance.
Athens	15 simple utilities for the lexicon (eg word frequency count, implementation checking). Lexicon DOS environment (integrates texts, sentences, words, forms & lexicon files). Used for EU-EL dict's.	Always responded quickly and accurately to requests, helping solve software problems, and helping organise demonstrations.
Dublin	Designed / developed terminology database for Eurotra.	-
Torino /Pisa	*6000 entry dictionary for Italian, plus 100 feature lexical semantic*	*system. Bilingual: Dan-It moduleE-Star facility.*
Luxembourg	CRETA became a major source of European s/w and Eurotra support.	Liaison with Saarbrücken from the beginning. Staff transfer to Lux.
Utrecht	Prep'n and prod'n of the first prototype Eurotra framework - later MiMo. Lexicographer's aid placed in public domain. Ref Manual initiated by Utrecht & first 6 issues edited by Utrecht & ISSCO.	Have fulfilled a number of useful tasks, eg distribution of material. Role of CRETA v CEC never clearly defined.
Lisboa/ Porto	EU-PT has been involved in aspects of the Ref Manual and in clusters that developed fundamental and applied research, and also EU demos.	The support of the Luxembourg team has been essential to the development of the whole project.
Essex	With Utrecht developed MiMo. useful demo system. Collaborated with IAI on CAT2. Involvement with Ref Man.	Assume CEC meant. Admin part excellent - though overworked. Lack of CEC staff for scientific admin problematic early days - took time to achieve sufficient expertise. Lately - CEC team too large and has interfered too much with the work - also problem for ET-10/LRE
UMIST	-	-

EUROTRA Questionnaire Analysis

Question 7 Management Issues

1) How well have the external management arrangements for the Eurotra programme worked? How might they have been improved?

2) Have the network and liaison arrangements between the Centres proved satisfactory and reliable?

3) Have the funding aspects worked satisfactorily?

4) What has been the National Government involvement in funding and support of the work? Has this worked satisfactorily?

Centre	External management Arrangements	Network and Liaison	Funding Aspects	National Government involvement
Leuven	Worked fine most of time. Skew award of UT/LE CoAs. Delay in Add.4 award in '88. Poor CEC mgmt of telecom expert for LE	Good coop Utrecht, also Saarbr./ Essex/UMIST (Ger-Du, Eng-Du). Good UK coop 1990-> monoling modules lexic/grammar. LG good	see Questionnaire response table §7.1	Provided < 40% funds, plus additional transition phase funds.
Liege	*OK?*	*OK?*	Not enough to guarantee critical mass for Liege team.	*OK?*
København	No problems wrt CEC & Nat. Govmt. Nor with EAC, CSC(1)	EuroKom excellent. CEC & centre cooperation on demo good.	Worked well. Good support from University of Copenhagen.	Always funded, gave support, inc extra 5 year professorial grant
Barcelona/ Madrid	With CEC, Nat. Gov'm't, have worked satisfactorily for Eurotra	Fruitful interaction as demonstrated by joint proposals	2nd addenda not signed due to CEC - led to team reduction	CICYT has fully supported the group financially.
Madrid	Spanish representatives let Madrid down - formal complaint made.	N/work & liaison between Madrid and rest poor.	Funding managed from Barcelona.	Spanish representation NOT based solely in Barcelona!!
Paris TALANA	-	-	CNRS not fully supportive	Policy towards the Eurotra work unclear. Frustration.
Saarbrücken	Mgmt very complicated - research mgmt by 'pseudo-democracy'. Better in later years-team worked together. Relations with Nat Gov less complicated. Some opinion conflicts on work. Better define as research project (academ v indust)	Groups linked by Eurokom - very efficient, good for communication (eg 'Problem Office', 'research clusters'). Used for ET-ES (CAT2 & ETS demo; ET-GB (CAT2 & ET-10) workgroups.	Funding usually worked satisfactorily.	Worked satisfactorily.
Athens	Research mgmt faced problem that theory evolved in parallel with implementation - effort diverted to coord. Administration was appropriate for 12 countries.	Fruitful collaboration in clusters (Greece was in 5). 9 proposals submitted for LRE I using the Eurotra network.	Yes	Supplied 20% of the programme's budget. Excellent relations with government.
Dublin	CEC - excellent. Irish Govmt - complicated by mixed responsib-ilities EOLAS main contact.	Arrangements valuable & helped form consortia with other groups in bids for TE10 & LRE work.	Yes	No direct involvement or support from Govmt -no problem.Govmt not aware langtech opportunities

Torino Dima/ Pisa	-	-	-	-
Luxembourg	Visiting scientists:'a miracle this kind of mgmt approach works'.	Eurokom stable but expensive. Used heavily for cooperation.	"Funding was fun"	No problem 1989/92. None beyond.
Utrecht	Now satisfactory relations with Nat Govmt - but their disinterest.	Use of clusters good. Network excellent. LG questionable.	No funding problems	No problem, but lack of interest in LE technology at high levels
Lisboa/ Porto	External management arrangements OK	As good as they could be under the circumstances (12 groups).	The Portugese government funded University positions only: but	recognised the importance of Eurotra. CEC payments timely.
Essex	Decentralised org difficult. CoAs often signed post-work - speed up process. Funds-better CEC or DTI.	Network / Liaison excellent (Essex - 27 proposals, 11 centres in 91/92)	DTI helpful - no problems with funding apart from initial delays	DTI involvement xcellent.
UMIST	-	Fr-Eng collaboration on contrastive research was slowed down by CNRS.	-	-

(1) From the CST København Questionnaire response. "The Liaison Group consists of the heads of national research teams plus the project head from the CEC. This way all relevant actors are assembled together in a decision making body for the day-to-day management of the work. Normally, this has worked well; but in a few cases it has been difficult to reach a decision because of the different roles of the heads of national groups compared with the role of the head of the project: the CEC has the executive power for the programme. However, in the 1991-92 period the voting rules were changed, which has led to easier decision making."

EUROTRA Questionnaire Analysis

Question 8 Opportunity Cost

1) If your team had received a similar amount of funding for MT work, but with no restrictions on the nature of the work, the way it was carried out, etc., how would you have spent it?

2) Compare the LRE programme with Eurotra. is LRE an improvement? Which features are better, which worse?

Centre	Opportunity Cost - funding	LRE v Eurotra
Leuven	Would have divided budget across several action lines.1) Applications (R&D feasible objective for academia/industry) 2) Research (basic CL) 3) Grants (for overseas visiting researchers) 4) Teaching/ education (CL curricula - establishment of chairs).	Better : more realistic goals & work programmes; opens Eurotra community - widens accessibility of results. Worse : fragmentation; stimulates rivalry; short term views/planning; too little funding.
Liege	Programme of work rejected by the LG (1)	No experience - wait and see.
København	Would have sought cooperation with other European teams, fewer partners, concentrated on fewer research tasks.	Eurotra closed '84-90, but 91/92 improvement. LRE - competition & new teams, but content of LRE seems more arbitrary, communications meagre, bias against less favoured languages.
Barcelona	Conditions helpful to group to overcome lack of experience. Would now have preferred to diversify work from linguistics alone, to tools.	Difficult to compare. Good to have industry involved, but poor funding aspects for good results. Too early for Spanish industry.
Madrid	-	Favours most developed centres - forgets less favoured countries - eliminates existence of Spanish CL scholars.
Paris TALANA	-	-
Saarbrücken	Concentrating first on development of an efficient s/w prototype Not working on all languages at the same time (first 3 then rest) More focus on practical needs : link to terminological databases, text handling, robustness; less basic research on long-term topics	LRE improvement insofar introduces more objective criteria: proposals are evaluated. More influence of mainstream research in CL. Inconveniences are : no common infrastructure, less cooperation, political influences. Industrial prototype goal dissolves : quasi-Esprit
Athens	In all probability do things the same way. At the start no MT or NLP expertise in Greece. Eurotra trained, provided ideas exchange in parallel with comparative / contrastive work, and monlingual analysis.	LRE is evolution of Eurotra.It takes up where Eurotra stops & furthers its achievements. It aims at providing the framework for the development of theoretical / applied NLP. Without Eurotra, the less favoured languages would not have arrived at LRE competitive levels.
Dublin	EU-IR unusual status - coordinators of terminology collection & sublanguage research - no grammar. Politically Irish not official EEC working language. Slow start-up succesfully recovered 1991/92.	LRE likely to produce marketable results. Basic research will suffer from emphasis on applications? Danger- produce products based on 'old' technology - need NEW parallel research/production (cf Japan).
Torino Dima/ Pisa	-	-
Luxembourg	-	-
Utrecht	Ideal programme small size (5 staff for 10/12 years) flexible: i) basic research ("tomorrows"); ii) applied research (validate "yesterday's" speculations); iii)product view (industrial partner, end-user aids/tools).	Improvement in that not monolithic, no interdependencies, open to all. No guarantee of coherence, too small size, low chance of win, too short, no continuity, gaps between Call & startup.

Lisboa/ Porto	With no prior experience EU-PT would be compelled to spend most of the money on acquisition of know-how.	LRE improvement on Eurotra as it widens the scope of research. However negative point is (smaller) regrouping of Centres.for work.
Essex	Involve fewer partners. Wider range of research. Feasibility of concept demonstrator for 2/3 languages - leave rest at research stage.	LRE improvement-clear R v D. Worse i)lacks continuity -funding .. gaps ii) short-term iii) large admin iv) small funds. CEC sdopts customer role, not research commissioner eg reports format.
UMIST	10 years ago - same way. Now, build a transfer-based commercial system 3 languages. Eurotra work reasonable & neccessary. Use corpora for real life lexicons, grammars. New goal, procedure.	No. Funding piecemeal. Effort of preparing applications enormous. Scepticism about selection process.

(1) Edited extracts from unsuccesful Liege proposal to LG, and response to Question 8. It relates to MRDs. The main thrust of Liege's work has been i) processing existing lexical resources into relational database format and other formats to ease access to information, for ii) a) automatic detection of parent field, b) providing lexical entries for a lexicon-driven parser/generator of English, written in Prolog and using a dcg grammar, c) CALL, based on LDOCE and COBUILD. NLP lexicons tend to be restricted in scope, idiosyncratic and fragile (lack of extensibility and transportability). Dictionaries used in MT projects that are not AI-oriented are considerably bigger, but tend to include few aspects of semantics. Liege believe that the building of lexicons for NLP projects from scratch is a waste of resources. MRDs such as LDOCE - representing hundreds of man-years of work - should be utilised in the NLP context.

EUROTRA Questionnaire Analysis

Question 9 The Future

Now that Eurotra funding is nearly ended, what do you see as the future of your Centre and team? From where do you expect to get your funding? Would you have liked to see Eurotra continuing?

Centre	The Future
Leuven	The team has devoted a lot of energy to the spreading of knowledge about MT and NLP through the teaching of courses in CL at Leuven University and elsewhere, publications, lectures, demonstrations, collaboration on journals, setting up CL Centre at Leuven University. Leuven started preparing for post-Eurotra in 1990 : Centre for the Study of Language & Computation at KU, then in 1991 Centre for CL at Leuven University. The latter has 30 staff with projects in computational semantics, MT, document handling, computer aided language learning, corpora & lexica. Funds obtained from EC, Belgian Government, Flemish Community, industry, etc. Eurotra-like work should continue BUT it should use the ALEP formalism *and* good support from CEC/Central, good contacts industry, cluster-based, central training.
Liege	*[The EU-Liege team came from primarily an English department, and this led them to feel it was undesirable to work solely on French terminology.]* The direction taken by Eurotra has diverged from Liege's interests in dictionaries. Liege will bid for LRE work and will continue the contract with OUP, and also the small Esprit contract for multilingual access to information databases. Eurotra has been good for Liege, and they would like to have continued with more corpus based research and to have had stronger liaison with Dublin. Unhappy with Eurotra direction. *Liege's unsuccessful proposal was for i) monolingual lexicography (exploitation of dictionary definitions for the retrieval of semantic relations; comparison of definition styles - using the two machine readable monolingual dictionaries, namely LDOCE and COBUILD; exploiting examples to retrieve collocational behaviour) ii) multilingual lexicography (reversability of bilingual dictionaries; homography and polysemy; feasibility of establishing links between two monolingual dictionaries) iii) terminology (research on how to integrate domain-specific knowledge in a linguistic framework; automatic determination of text type and subject fields).*
København	In Denmark, the Centre for Language Technology was created in 1991. Eurotra-DK had the Eurotra programme as its only responsibility, whereas the Centre has a much broader scope of language activities. The Centre has been performing Eurotra work under the CoAs in 1991/92, and although Eurotra has been an important source of funding, they do not see a problem in changing to other sources. The funding of the Centre comes from national research funds, Nordic research funds, EC research contracts and to a large extent and increasingly from commercial contracts in Denmark and abroad. They would like a Eurotra-like programme to continue : for the continuity of the work, the production of modules that fit together, the guarantee that all languages are covered to a reasonable extent. One of the possibilities for future administration of linguistic programmes is the creation of a European Agency. This idea has been brought forward several times. The role of an Agency could be more easily definable than the current role of the CEC, and it would be possible to hire staff with exactly the right qualifications. If a linguistic agency is created it should not just be an administrative unit - it should do research itself. Staffed by a small permanent team and supplemented by visiting research secondees from all nationalities. Extension of 3rd Framework: CST is worried not to see the heading "Linguistics" appearing in the Telematics section. The Commission should have made mention of this. 4th Framework: CST do agree that linguistic resources are important, but this term should cover grammars, semantics, knowledge bases, etc., as well as the research which is necessary to produce these.
Barcelona	Future is unclear. Although Barcelona is in procedure of building up an institute for collaboration with the Department of Linguistics of the Universidad de Barcelona, the Department of Filología Espanola of the Universidad Autónama de Barcelona and with the ICE of the Universidad Politécnica de Cataluna, they are facing some problems getting funds which will allow them to keep all the personnel. In Madrid no clear strategies have been devised for maintaining the Group. It is a pity to discontinue Eurotra - especially for Spain - momentum lost.

Madrid	Future uncertain - hardware & software will be preserved thanks to Universidad Autonoma de Madrid. Further work on Spanish corpus. No more space for MT - unthinkable in Spain that work could carry on without CEC support. Varied views about continuation of Eurotra.
Paris TAUANA	-
Saarbrücken	IAI has acquired considerable know-how in MT and related techniques - useful for future CEC and industrial work. relations with some Centres will be maintained. Recommend continue Eurotra CoAs in small scale (say, 2 persons / country); management by a European Agency (independent from CEC, small permanent staff, guest researchers, etc). National governments should take over responsibility for administering & perfectioning of the created grammatical and lexical resources. CEC to participate small scale at least in Eurolang, & other. The Athens team would welcome concerted special actions of the Commission for the support of the less favoured languages and the promotion of language technology within Europe.
Athens	The question of Eurotra's end and the consequent problem of the loss of the expertise acquired and the resources created, occupied the team a long time ago. The creation of the Institute for Language and Speech Processing was the best answer to this question. ISLP was founded in 1991 and functions under the auspices of the Ministry of Industry, Energy and Technology. It participates in EEC and national projects (currently in 5); its funding comes from these projects, and it also gets additional funding from the Greek government. Continuation of Eurotra : Athens would like the infrastructure / network to continue - both for contact and for all European language support (unique in MT).
Dublin	*Dublin seeking to exploit their reputation in terminology and sublanguage.*
Torino Dima/ Pisa	*Dima : Balance the research activities with the commercial applications, which should provide the necessary fundings for the future. Eurotra : no need to undertake any specific action, as it is closely connected with the researches and activities of ILC and the University. The group is expected to participate in NLP and CI programmes. (Paraphrased Final Report.)*
Pisa	Eurotra-Pisa does not need to undertake any specific action, as it is closely connected with the researches and activities of ILC and the University. The group is expected to participate in NLP and CI programmes. (Final Report.)
Luxembourg	The non-profit organisation IT&S was founded in July 1992, but it does not dispose of capital nor of R&D contracts. The CEC toned down the problem of Luxembourg and now there is no more money from Eurotra nor is there any LRE to do something for Luxembourg. The director of the CRP-CU gave us, the employees of the CRETA, his notice as of December 31st 1992. A number of project ideas were developed, but it is pretty late for contracts to start at the beginning of 1993. Luxembourg have expressed the need for an Information Server for the Language Industry.
Utrecht	The Utrecht Centre lies within the Research Institute for Language and Speech, and as such it will continue to exist. Some team members will work on new projects, others will lose their jobs. Since hardly any funding can be expected from regular sources (university, NSF) we will continue to participate in EC programmes, and they will try to establish links with industry. Eurotra has outlived itself, and it is good that it will now come to an end. One of their major concerns is the continuity problem, and it is desirable that EC and/or national authorities committed to a number of longer term actions. Furthermore, there is some concern that the MT network built up in Europe, with Eurotra serving as the backbone, will now gradually fall apart.
Lisboa/Porto	ILTEC has some small projects that will allow it to survive, but there is no regular Government funding. The work should not be lost - development projects for industry should be CEC funded. Portugal is keen to work within future coordination structures. It doesn't matter whether it is a transfer-based approach or an interlingua approach with unification-based formalisms - the big divide is between the knowledge based approach and the statistical approach. (Cf ET-10).
Essex	Eurotra as an intellectual or scientific programme has reached the end of its useful life. As administrative or human ensemble it still has potential. Need to distinguish between R & D. Will participate in Eurolang (D). Will continue (R) with LRE, Esprit, UK research councils.
UMIST	*Continuing their successful work in CL*

5

The EUROTRA System Design

The "E-Framework" system uses a unification-based stratificational model approach. This uses a number of representation levels for each language, linked by simple transfer components. All the system design is modular so that the pieces can be developed in different places and times, yet assembled into a coherent system. This approach demands clear and strong specifications for the various modules and their interfaces, leading to the Reference Manual. Three basic principles underlie the EUROTRA approach:

1) A stratification approach; ie the translation process is broken into smaller steps by defining a number of representation languages and mappings between them.

2) Independent definition of representation languages; ie each representation language is described fully by a grammar and a feature dictionary.

3) Simple mapping between levels of representation; ie ideally the mapping can be stated compositionally.

The virtual machine consists of two components:

1) the generator, which interprets the grammar by applying rules;

2) the translator device which interprets the mapping between two adjacent representational levels.

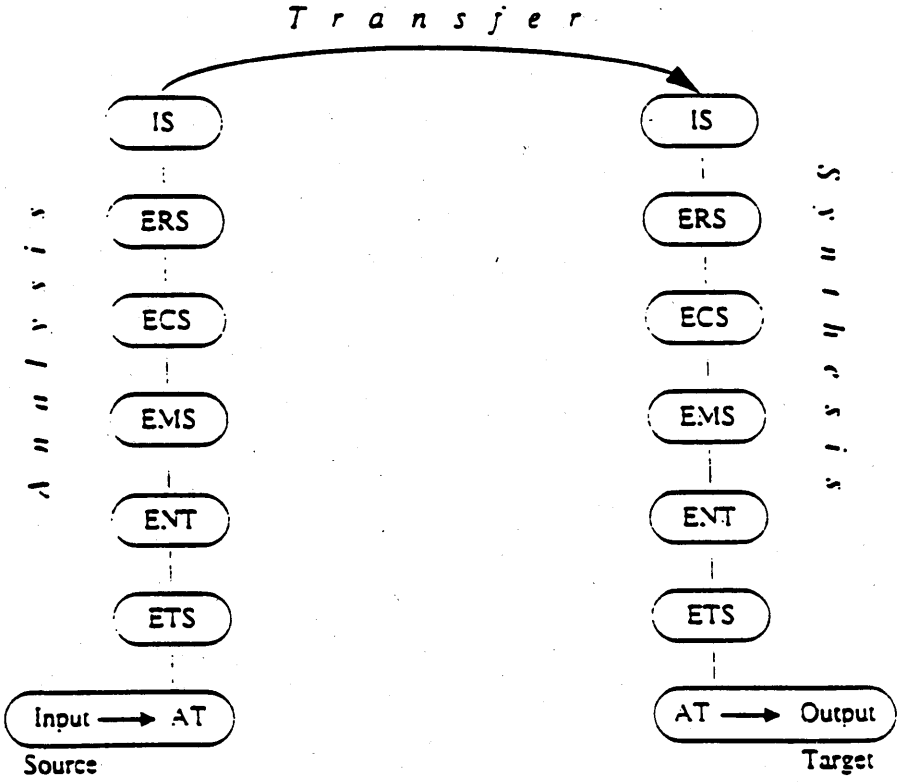
The EUROTRA system design has the normal three main phases: analysis, transfer and synthesis, with stratification of the analysis and synthesis phases. There were six strata in both the analysis and synthesis phases, with different steps of analysis or synthesis carried out tidily in the appropriate strata:

AT : Actual Text	as written possibly in a word processor format
ETS : EUROTRA Text Structure	separates the text from the structure of the document using SGML analysis and reverses the process
ENT : E Normalised Text	words are decomposed into word morphemes, such as prefixes, suffixes, stems
EMS : E Morphological Structure	combinations of morphemes are analysed to produce a feature bundle with a reference to the root lexical

	unit, together with features indicating how this unit is modified by the associated morphemes. At this stratum invalid decompositions are discarded
ECS : E Configurational Structure	identification of phrases and components within the sentence
ERS : E Relational Structure	handles relations between items, such as subject-verb
IS : Interface Structure	the deep syntactic representation which incorporates interlingual descriptions for sub systems

Most linguistics work was carried out on the ECS, ERS and IS levels. For detailed information, see the first two volumes of "Studies in Machine Translation and Natural Language Processing", published by the Commission in 1992. A summary is given below.

Diagram of the Eurotra Model



The standard E-framework hypothesis is that there are 3 intermediate representation levels between text and IS for each language. The EMS builds representations of the morpho-syntactic structure of word-forms by means of general morphological rules. The ECS is a level of phrase structure closely related to the level of c-structure in Lexical-Functional Grammar (LFG), deals with categories such as noun, verb, etc., and coordinator, quantifier, etc. The next two levels are ERS and IS: typical ERS treatments are subject-verb relation, and long-distance dependency (using the 'coindexing tool'). The IS is the most abstract level in EUROTRA: it deals with formal semantic analyses of phenomena such as tense and aspect, mood, quantification and negation.

The penalty in ETS from minimising the gaps between IS of different languages (simple transfer), is that the gaps between text and IS become large. Only by decomposing analysis and synthesis into a series of primitive translations, between intermediate levels of representation, can the task become more manageable. Each representation level is a formal language, comprising simple objects (called feature bundles). Sets of simple objects can be formed into connected trees, showing linkages, dominance, and precedence, and these are called structured objects.

A generator, based on grammar rules, can be applied to sets of structured objects to test hypotheses about grammar construction. The three basic rule types applied to objects are: structure building rules (immediate consolidation of the objects - straightforward unification, parsing, insertion); feature rules (test condition, apply action - eg add dictionary information); filter rules (for checking well-formedness).

Translators are 'one shot' devices in that the output of a source generator becomes the input to a target generator without creating any intermediate representations within the translator. They include a feature theory, a default translation mechanism, and a set of user-defined translation rules.

The generator and translator components i.e. the core of the system, are written in Prolog. The mechanism for applying these rules is the 'virtual machine'. It is a unification-based machine, non-deterministic, and offers rapid prototyping. Surrounding the core, but still written in Prolog, are a number of tools to aid linguists in writing correct grammar and translator rules, including: a debugger; a pretty-printer; a command interpreter to manipulate objects. Rules are written in a formalism (i.e. the user language) different from the virtual machine's Prolog. There is an interface to a Unify relational database system where a large number of dictionary items for each representation level of a language can be entered, stored and updated.

Finally, there is a top-level interface allowing the user access to all components of the system and to the Unix toolset.

6

How can you combine the best of the EUROTRA and LRE schemes

EUROTRA Liaison Group, January 1993

0. Preamble

When Brian Oakley met with the Liaison Group October 27 a discussion was started on the problem given by the title of this paper. Brian Oakley asked the Liaison Group to prepare a proposal for the EUROTRA Final Evaluation Panel.

In the present paper we describe the advantages and disadvantages of each of the two ways of organising a research programme, and make some proposals for the future.

1. EUROTRA and LRE.

1.0. Introduction

By the end of 1992 the part of the EUROTRA programme which is managed by contracts of association (CofA) ceased to exist. Only some minor activities most of them tasks funded by grants will last until spring 1993. NLP activities will continue for a while in the ET-10 series and will then gradually be taken over by the LRE programme. In the following we will try to analyse if LRE in the present shape will preserve the positive achievements of EUROTRA.

1.1. Achievements and Deficiencies

Before we start the analysis we would like to stress that fact that EUROTRA has a special structure in these years 1991-92.

In the years 1984-90 most of the research in the EUROTRA programme was managed by CofA. Some research work was managed by special contracts between the CEC and a EUROTRA research institute (e.g. production of RM).

In the 1991-92 programme the CofA have been cut to half their size, and an equivalent amount of money is used in the research programme ET-10 which works through calls for proposals. Finally, a sum is used for software production.

In LRE no CofA exist, so the research part of the programme works through calls for proposals, like the ET-10 part of EUROTRA.

Our general view is that LRE which is the follow-up programme to EUROTRA to a large extent is an improvement. The 1984-90 EUROTRA programme had some shortcomings related to the monolithic structure which sometimes created almost complete dependency of results from the groups involved.

It is also our conviction that the quality of scientific results will improve compared to original EUROTRA because the programme concentrates on more specific questions and is less ambitious. It avoids being overambitious in not trying to build a full MT system requiring to address more or less all basic research problems of NLP such as formalisms, morphology, syntax, semantics, transfer, analysis vs. synthesis, and all this under the multilingual perspective including 9 languages.

We also agree with the general orientation of LRE to build a linguistic technology, creating resources that can be used for all kinds of applications, among them MT. We agree that this is the better approch than building linguistic resources for a monolithic MT system.

Though it seems as if LRE has taken the right consequences we miss the following points that are considered essential if one wishes to preserve the positive achievements of EUROTRA. The 2 major points that are to be mentioned here are 'involvement' and 'coherence'. Additionally, we have some comments to the 'cost' the research institutions have to pay for the new structure.

(1) Involvement:
One key achievement of EUROTRA was that this programme induced an involvement of all countries with numerous positive effects:
A very modern principle, namely that of subsidiarity was enforced in the area of linguistics by EUROTRA. The national governments were forced by EUROTRA to participate by taking care of the national language.This had the known positive effects:

(i) Computational linguistics (even modern formal linguistics) were established which did not exist before and which would not exist nowadays.
(ii) Personnel in the area of NLP was trained and exists now in these countries which would not be the case if EUROTRA had not existed.
(iii) Major catalyst effects for national efforts in the field of NLP can be observed in most of the countries.

(2) Coherence:
Though the monolithicity of the EUROTRA project created complications, delays etc. it had some positive effects on the other hand: Working towards a common goal in all countries created a common understanding of problems, a common scientific background and approach, even a common scientific language, a European (!) collaboration and also common resources, the most valuable of them being computational grammars and lexicons for all European languages, morphologies, small-scale MT systems, a multilingual demonstrator that includes all (!) languages and common linguistic specifications in form of an voluminous reference manual.

It is these two areas where we feel that the EUROTRA approach had a positive impact and where LRE may have to be supplemented by additional measures. LRE will not produce a reference manual and there is certainly no guarantee that the results from the different quite disparate projects will fit together. We therefore propose additional measures in section 2.

(3) Cost:
The LRE scheme has added a very considerable overhead stemming from the elaboration of research proposals. The amount of person months that go into the production of such proposals is very high. In the EUROTRA organisation this type of largely wasted effort did not take place. When cost-effectiveness is considered, this point should also be taken into account.

2. Proposals

The proposals concern the organisation of LRE or its follow-ups, incl. the idea of a European language technology agency, and the creation of an MT network.

2.1. Organisation

As mentioned the LRE programme has advantages over the original EUROTRA setup by building on competition, thereby not being restricted to a closed group of research centres. This advantage should be kept in the future.

The disadvantages are
1) no necessary commitment from national authorities to support their own language,
2) no special commitment from the Community to support the less favoured languages, an investment which is too heavy for the countries concerned,
3) no continuity, completeness and coherence in the modules produced by the various projects, and thereby no guarantee that exploitation projects needing the combination of different modules (e.g. MT) can be easily made,
4) no special focus on MT.

Actually, we see the programme organisation of EUROTRA 1991-92, i.e. a mixture of CofA work and competitive research work (without commenting on the distribution between the two) as a possible way of combining the two programme schemes, and in particular of catering for the disadvantages 1)-3) above. Maybe other schemes can be devised.

If it is not possible for the Commission and the Member States to join forces as in the CofAs, we alternatively see the Agency as having an important task in taking on the responsibility of securing the continuity, completeness and coherence of the research done. This can be done by strongly monitoring a certain part of the contracts to ensure the production of the resources needed.

These measures can only be taken in the long term. For the short term we have consequently been seeking possible ways to maintain the current coherence of modules. The LRE programme does not have much to offer, but we do see two possibilities:

1) Ensure that all EUROTRA grammars are migrated to the new Alep formalism, (this is already part of the LRE programme of work)
2) LRE has a branch for application programmes. Make sure to use the existing resources when building applications.

Finally, we should like to add that a few EUROTRA centres have been dealing with themes that are not language specific, but yet highly relevant to the project and its coherence, in particular Ireland and Luxembourg. Plans for the future should take these countries and the expertise they can offer, into consideration.

2.2. MT Network

2.2.1. The situation

One of the major achievements of EUROTRA is that it has created a network of MT specialists spread all over Europe.

The size of this network is considerable. The most recent list of Eurotrians (prepared by CRETA in April 1991) contains some 220 names of people working for EUROTRA in 1990, and the number of people who left EUROTRA before that time, and those who joined after that date may amount to another 100 or even more.

Many of those people are still active in the field, most probably in academia, but a fair number have moved to private industry.

This network has been very effective within EUROTRA (both in CofA work and in ET-10), but has also shown its strength outside (cf. the partnerships in LRE and other programmes), and has led to a number of other joint actions both in research and in educational activities.

During the execution of the EUROTRA programme this network could rely on EUROTRA

as its stable backbone, not only for those actually working for EUROTRA, but for many others as well.

By mid 1993 the EUROTRA community as such (CofA institutes) will have ceased to exist, and there is no other body or organisation that will naturally take over the backbone function.

This means that the existing European MT network will have to be based on personal or working relations, and it is to be feared that the result will be that the existing overall network will gradually fade away and in part be replaced by other network structures, and in part just disappear.

This will create a situation where the current massive work force, with experience and expertise in MT, will slowly fall apart.

2.2.2. The proposal for an MT Network

We propose to aim at the creation of a European MT network including representatives of all relevant actors (funders, researchers, developers, vendors, end users, etc).

The main long term objective of this network would be to promote those research and educational activities which may eventually lead to the design and development of MT systems, and the shorter term objective would be to (a) investigate with regular intervals how current knowledge and technology can be applied in order to overcome the language barriers in Europe by means of translation facilities, aids or systems, and stimulate the implementation of projects oriented towards this goal;
(b) identify with regular intervals the direction which research activities should take in order to generate the knowledge and technology for the next generation of feasible facilities, aids and systems, and stimulate the implementation of research and training activities aiming at these goals.

The network would undertake various types of actions, such as:
(i) coordination of postgraduate and postdoctoral programmes,
(ii) provide connectivity with other networks in related areas,
(iii) help coordinate the creation and dissemination of resources,
(iv) increase flow of information between academic and industrial research groups and (potential) individual or corporate end users,
(v) establish a coordinated and representative source of expertise for consultation by national and EC organisations.

The network would be based on a modest infrastructure (a small coordination point), with communication facilities (mail, phone, fax, email), and resources for the organisation of meetings, seminars, workshops.

The possibility of joining forces with an existing network of European researchers, namely ELSNET (European Network in Language and Speech) under ESPRIT Basic Research, has been put forward.

As stressed above, we find it important that a visible structure is created for the MT area. This may be possible to achieve within ELSNET, but it would require a change of the structure and shift of the main focus of ELSNET, the main purpose of ELSNET being the integration of NL and Speech research.

Preliminary discussions with the ELSNET Executive Board to illuminate these topics will be taking place late January.

7

Council Decision 82/752 of 4th November 1982

ANNEX I

1. Objectives

The objective of the programme is the creation of a machine translation system of advanced design (EUROTRA) capable of dealing with all official languages of the Community.

On completion of the programme an operation system prototype should be available in a limited field and for limited categories of text, which would provide the basis for development on an industrial scale in the period following the current programme.

2. Programme of Work

The programme is divided into three phases:

(a) Preparatory phase (two years, 2 million ECU)

In this phase the following work would be carried out in parallel.

1. First:

- setting up of the ACPM;
- definition of the project and its organisation and of the responsibilities of the participating countries and centres;
- definition of the methodology of the work;
- preparation of a detailed programme of linguistic work to be carried out by the participating centres, and of the sectors and categories of texts covered by the research;
- definition of the allocation of intellectual property rights and definition of the arrangements for disseminating the results of the work in accordance with the actual combination of each participant;
- examination of the value to the Community of participation by third countries and, where appropriate, definition of the conditions for such participation.

2. Second:

- preparation of more detailed specifications of the linguistic models and strategies for the various components of the process (analysis, transfer, generation);

- preparations of detailed and binding specifications for the EUROTRA basic software and the data processing programmes capable of carrying out the various processes: analysis, transfer, generation, monitoring functions and text management;

- preparation of more detailed specifications for the lexical database;

- preparation of the contracts of association including financial and other contributions to be made by the associated parties.

The Commission will ensure that the objectives of portability and compliance with international standards are correctly reflected in the specifications referred to in the first three indents above.

At the end of this phase the opinion of the ACPM must be obtained on the above specification in order that the linguistic work can progress quickly and so that the widest possible invitation to tender for construction of the software can be issued as soon as possible (see point 2 (b) below).

(b) Phase of basic and applied linguistic research (two years, 8.5 million ECU)

On the successful completion of the first phase, and after consultation with the ACPM and Crest, the second phase will be divided into two parts:

1. *Basic linguistic research*

This part will consist of the following work[1]:

- the development of initial linguistic models for the analysis and generation of each of the official Community languages and for transfer between these languages. This work will be based on a corpus and vocabulary in a limited field, estimated at around 2,500 entries;

- preparation of the lexical data base, for the above mentioned vocabulary, which will serve both for the analysis and for the

---

[1] Some of this work could continue in the following phase.

generation of each of the languages and for the transfer between these languages;

- a study of the linguistic strategies best suited to machine execution of the various processes.

2. *Construction of the basic software for EUROTRA*

This part comprises the following work:

- issuing of invitations to tender, the specifications for which will have been defined during the first phase;

- scrutiny by the Commission of the replies to the invitation to tender and selection, after consultation of the ACPM, of a body to construct the EUROTRA basic software, within as short a time as possible;

- development of the basic software by the body selected, including:

  - the high level language for describing the linguistic data and strategies;

  - the high level language for interaction between the user and the system, which will make it possible to introduce the various modules into integrated systems corresponding to the different utilisation options;

  - the utility software for compiling the high level languages, for tests and for management of the data bases.

This initial version of the software is intended to enable the development and machine testing of the linguistic models defined by the participating centres when they are sufficiently advanced. Its development is consequently a prerequisite for validating the linguistic work under this programme.

The industrial development of the EUROTRA system, including adaptation of the software to the performance and reliability requirements for producing translations under commercial conditions, will not be put in hand until this programme has been completed.

(c) Phase of stabilisation of the linguistic models and evaluation of results (18 months, 5.5 million ECU)

After options have been received from the ACPM, Crest, CIDST and

Cetil at the end of the second phase, ie when it is possible to carry out systematic testing of the initial linguistic models, comprising complete language pairs and consisting of analysis, transfer and generation, the objective of the work will be concentrated on the following aspects:

- adapting the linguistic models, in order to produce linguistic modules which are as reliable as possible. The modules will then be fit for pre-operational use;

- progressively extending the basis of the text corpus, the linguistic models and the vocabulary for a specific field, and on texts of increasing complexity;

- revising and progressively extending the lexical bases to cover the chosen field as exhaustively as possible (about 20,000 entries in all the languages);

- evaluating the technical and economic performance of the system;

- preparing a proposal for the development of an operational system on an industrial scale and proceeding to the stage of commercial exploitation.

Council Decision 90/664 of 26th November 1990

ANNEX I

1. Objectives, Evaluation and Content of the Programme

1.1 Overall Objectives

This programme constitutes the first step towards the development of an operational machine translation system of advanced design, capable of dealing with all official Community languages. The specific objectives of the programme are:

(a) *Creation of the conditions for the transition to an operational system*

- implementation of a development, testing and research environment capable of supporting large scale systems;
- extension of the linguistic coverage and large scale testing of the analysis and synthesis modules for all languages covered by EUROTRA;
- the definition of common methods for large scale development for machine translation and other applications involving natural language;
- experimentation and evaluation of relayed transfer, using an interface structure as pivot;
- research, prototype implementation and evaluation of new linguistic models aimed at the improvement of the interlinguality of the interface structure and control of overgeneration;
- research, prototypical implementation and evaluation of methods for the use of subject-field and text-type specific knowledge for translation and other applications.

(b) *Advancement of Work on Lexicography and Terminology*

- definition of common methods and tools for the integration of existing lexical and terminological collections;
- participation in the definition of international standards for textual, lexical and terminological data;
- close cooperation with research bodies in the Member States with a view to harmonising lexical and terminological resources and to making existing systems compatible.

A7.2.2

(c) *Training and Cooperative Projects*

- Training of researchers and engineers through a grant scheme;

- setting up, experimentation and evaluation of cooperation schemes between research institutes and industry.

1.2 Evaluation

At the end of the programme the results will be evaluated by independent experts against these objectives.

2. Priority Action Lines and Scientific and Technical Content

2.1 System Development, Testing and Research Environment

On the basis of the critical review of the prototype implementations and the specifications to be produced by the end of June 1990 the implementation of a system development, testing and research environment will be commissioned to the European software industry on a turn-key basis.

This environment should have the following characteristics:

- a powerful and user friendly formalism for describing linguistic facts (to encode dictionaries and grammars);

- a special purpose data management system for the creation and maintenance of large scale dictionaries and grammars with special user services for the addition, inspection and modification of the linguistic data;

- an efficient rule interpreter capable of dealing with large dictionaries, grammars and texts;

- a set of testing tools with special attention to interactive run-time testing and correction of dictionaries and grammars.

The system development environment will pay special attention to modularity to ensure the reusability of the linguistic resources created through the possibility of combining the various modules in different ways to fulfil special tasks and of interfacing them with external applications.

ANNEX II

Implementation of the Programme, Rates of the Community's Financial Participation and Indicative Allocation of Funds

1. Modalities of Execution

The various action lines pursue different objectives which demand different forms of organisational, contractual and financial schemes.

The participants may be universities, research organisations and industrial companies, including small and medium sized enterprises, individuals, or any combination thereof established in the Community.

1.1 *Service Contracts*

The implementation of the system development, testing and research environment (action line 2.1) which will provide all participating parties with a common set of tools, will be entrusted to industrial contractors on the basis of calls for tenders. It will be financed fully from the Community budget.

1.2 *National Research Teams*

The work concerning the different languages (action line 2.2) will be carried out by national research teams in the Member States, and co-financed by the Community and the Member States.

1.3 *Shared-Cost Projects*

The linguistic research of general interest (action line 2.3), research and development into advanced system architectures (action line 2.4) and reusability of lexical and terminological resources (action line 2.5) will be carried out as cooperative ventures between industries, research centres and EUROTRA teams. Shared-cost research projects should as a general rule be carried out by independent participants from at least two Member States.

The contracts for shared-cost research projects shall, as a general rule, be awarded following a selection procedure based on calls for proposals published in the *Official Journal of the European Communities*.

For shared-cost contracts, the Community participation will as a general rule be up to 50% of the total expenditure, but this percentage may be varied according to the nature and the stage of the development of the research. Alternatively, universities and research institutes may, for each project they carry out under this programme, opt either for 50% funding of total expenditure or 100% funding of the additional marginal costs.

1.4 *Grants*

Grants will be awarded to qualified postgraduate students.

1.5 *Subsidies*

The Commission will award subsidies to professional associations and standards organisations for action line 2.6.

2. Indicative Allocation of Funds

The indicative breakdown of the amount of ECU 10 million deemed necessary for the execution of the programme is as follows (in thousands of ecus):

(a)	System development environment	2000
(b)	Community contribution to the national research terms	4000
(c)	Shared-cost research projects	3000
(d)	Training, subsidies, evaluation	1000
	Total	10000

2.2 *Language Specific Research and Development Work*

2.2.1 Reuse and extension of the existing implementations

For each of the nine languages covered, the analysis modules produced by the EUROTRA programme will be thoroughly reviewed to ensure their generality and adjusted to the features of the revised formalism.

On the basis of these revised implementations the grammatical coverage will be gradually extended to include additional text and discourse types. No large scale lexical and terminological development work is planned for this phase pending the outcome of the research on the reusability of lexical and terminological resources (see point 2.3).

2.2.2 Released transfer

The research and experimentation are to determine the feasibility of this approach and the optimal strategy for its implementation.

2.3 *Linguistic Research of General Interest*

This action line is intended to gradually improve the linguistic performance of the system and the quality of translation. It will be organised along three main directions:

- general linguistic research to increase the interlinguality of the interface structure and to reduce overgeneration;

- use of subject-field specific knowledge (terminologies, classification schemes, paradigmatic relations, domain models, knowledge bases, etc);

- use of text and discourse type-specific constraints to reduce overgeneration.

It can be predicted that some progress will be made in the course of this programme, but additional efforts must be foreseen for the future.

2.4 *Research into Advanced System Architectures*

To create a potential for innovation and keep pace with the fast advancement of hardware and software technologies, continuous research into new formalisms, software and hardware architectures is to be foreseen, which will lead in selected cases to experimentation and prototype implementation (eg parallel system architectures).

2.5 *Reusability of Lexical and Terminological Resources*

The details of this action line will be defined through project definition studies to be carried out in 1990.

It is expected to have two main components:

- development of methods and tools for the conversion of the formalised parts of existing dictionaries which cover mostly orthographic, phonological, morphological and syntactic information;

- research into the utilisation of non-formalised portions of dictionaries which concern mostly subject classification, discourse types, definitions and examples or citations. This is an advanced research topic whose outcomes cannot be predicted now.

2.6 *Standards for Textual, Lexical and Terminological Data*

This activity is closely related to the reusability of linguistic resources in the future. The Commission will support and stimulate international activities in this field in close cooperation with professional associations and national and international standards organisations.

2.7 *Education and Training*

A number of research grants will be awarded to postgraduate students to participate in the research and development work in the projects outlined hereabove.

8

Pannenborg Report - October 1987

7. CONCLUSIONS

Scale of Project

105 Machine Translation is carried out by a system. The architecture of systems generally evolves in small steps (mainly through improved sub-systems) and occasionally by a radical change of concepts. The latter is risky. Accordingly the new concepts should be tested on as small a scale as possible.

106 EUROTRA has chosen to base itself on a larger volume of fundamental research than any existing translation system. In order to test the validity of the outcome of research, one would normally prefer to test it on as small a scale as possible. The political decision that was made for EUROTRA has overridden this approach and required the project to proceed with research and implementation of the nine languages in parallel. The magnitude of the risk involved has thus been greatly increased, while reducing the likely achievability of a practical translation system.

107 EUROTRA at present is based on the assumption that all CEC translation work will be done centrally. The Panel query this assumption, and would expect some translation to be done in a decentralised manner within member states.

108 It is apparent that this type of project would never have been undertaken as a commercial research proposition and could only be undertaken with full public funding.

EUROTRA Principles

109 Fundamental progress has to be made in several critical areas: the level of abstraction to be used in the processing of source languages, interfaces between one language and another, understanding of context, computer techniques. Development will emerge from the deeper knowledge, inevitably enriched by a high degree of empiricism, that will be derived from fundamental linguistics, from the science and architecture of information processing systems (textual, non-mathematical) and to a certain extent from artificial intelligence.

110 Although most of the participating national teams are university based, some of the teams have a more independent status, with a stronger practical orientation. It is to be noted that the latter group has not turned away from the EUROTRA approach despite its language research orientation. This implies that experts with a stronger link to practice than the average university scientists also believe in the ultimate utility of the EUROTRA project.

Reference Manual

111 The reference manual has been trying to fulfil two rôles, namely those of standards and of regulation, as well as serving as a method of communication between research groups. This has caused a certain amount of confusion as the two rôles have not been sufficiently obvious within the text of the manuals.

Management

112 The central organisation and direction of the project have been hampered by long delays in provision of staff and resources. Although the liaison group appears to work well at this stage, a greater central management burden than necessary has been put upon them because of the lack of central resources. As a mechanism for running a distributed research project it has been effective, but is not seen as an efficient way of managing any future development project.

113 The executive and principal rôles have not been sufficiently distinguished within the project. The project appears as an integral part of an administrative department of the Commission, which is not an effective project management scenario.

114 There has been no attempt to establish practical test criteria for the end of Phase 2 of the project. This would appear to be partly due to the lack of sufficient central expertise to evaluate and integrate the results of the various research components. One of the criteria has to be based on a comparison with human translation.

115 In such a costly and ambitious project the results of the work must be demonstrated in the form of applications, or the stages of progress should be marked with practical results. This is why it is so desirable that there should be a permanent association between the research and the candidate organisations for the creation of a language industry. Only industrial firms can identify commercially exploitable objectives that are compatible with the state of the art.

116 The areas of application for computational linguistics are very numerous and a (non-exhaustive) list is given in Chapter 3.

Finance

117 There have been problems with the lack of central financial resources. This has meant that the central personnel have not had the budgets necessary for close liaison with some of the national groups.

118 For many reasons, and in many cases, the Commission funding has taken an excessively long time to reach the national groups. In some cases national funding has been available to fill in the gap. In others this has caused unacceptable delays in the project.

Language Rules

119 The choice of a declarative method for encoding the rules of language seems plausible but is unproven in its efficacy. The plausibility rests on two arguments. The first one, which seems to have guided the choice, rests on the better opportunities for modularity and also on more easy coordination between various languages. The second argument is derived from increased attention to declarative languages in computer technology in general.

System Design

120 Machine translation deals with languages. In the light of the nature of the project it is therefore not illogical to direct the major research efforts at language analyses and syntheses. The more mechanical work of translation has to be carried out by computers. These have enormous capabilities of speed in data handling, but also have their limits. It would be logical for more account to be taken of the possibilities of computer hardware and software. The present almost exclusive emphasis on the linguistic side of the system prevents this desirable interaction and again increases the risk of not achieving the ultimate goal of a practical system.

121 The project philosophy ignores any potential interaction between the translator and the system. The Panel questions the wisdom of this approach. In view of the many developments in computer software which have been advancing mechanisms and sophistication of pre-editing, some cautious steps in this direction could be considered.

122 It would also be reasonable to include some element of post-editing, which has been widely agreed on among those active in the project and is to be looked upon as quite natural.

Computer Systems

123 The software bottleneck of EUROTRA seems to be one of the most pressing problems needing a solution. It seems likely that a solution will only be found within an acceptable timescale if appreciable talent in the field of software architecture and engineering is contracted at short notice from third parties.

124 If the software cannot be improved, there is a possibility that neither the grammar nor the dictionaries can be appropriately tested at the end of Phase 2.

125 The hardware and architecture requirements for overall system design are also causing a bottleneck; this will become worse as the software is improved.

Dictionary Development

126 In the development of practical translation systems and their subsequent continual updating, the bulk of the work and the cost comes from composing and extending

the dictionaries. Accordingly it is customary to make use, as much as possible, of existing electronic dictionaries.

127 The EUROTRA approach leads to the compilation of dictionaries which are not directly compatible with existing ones. Only when the ultimate results have been proven to be vastly superior to anything else will this justify the appreciable additional cost.

128 Insufficient forward planning seems to exist in the present EUROTRA structure with regard to the task of compiling the dictionaries. It is clear to the Panel that to a certain extent this work requires a different kind of people and a different kind of organisation than is needed for the applied research on the rules of language.

129 In addition it would appear that the resources required for dictionary compilation work have been seriously underestimated in the project.

Promotion

130 The EUROTRA project has certainly achieved its goal of promoting computational linguistics in the member states. It has increased awareness of the subject in general and has encouraged work on those languages which are less well developed in the field of linguistic research.

Importance of Project

131 The complexity of the problem of automatic language translation has only been realised gradually during the course of the project by the authorities. The linguists look at the project as a real challenge.

132 A research and development team represents a high level of expertise. If funding is interrupted, there will be no knowledge transfer from Phase 2 to Phase 3 and the primary goals will definitely not be reached.

Summary

133 It is impossible to judge at this stage whether the project has fulfilled all its goals. The general conclusion of the Panel is that EUROTRA has so far fulfilled its political, education and training goals, and has partly achieved its scientific and technical goals. The economic goals do not appear to have been considered at this stage. The Panel's recommendations aim to rectify this situation.

8. RECOMMENDATIONS

134 There are three parts to the recommendations which may be considered as separate entities:

I THE FUTURE OF EUROTRA

135 The importance of the project to the European Community as a whole has to be emphasised as well as its enormous cost saving potential. It has implications for all the Community Institutions - the Council, European Parliament, European Court, European Patent Office, as well as cultural implications. Having instigated such a project and mobilised the expertise, it would be a retrograde step for the Commission to abandon it.

136 The funding for the project should not be interrupted, in the Panel's opinion, in particular some national groups should not be made to wait for others. Of necessity there will be a staggered development from research to development. The transfer from Phase 2 to 3 is already staggered in time for various participants, this is borne out by the Contracts of Association (see Appendix A).

137 There should be more realistic deadlines for Phase 2, and a modified basis for Phase 3. The suggested deadline for completion of Phase 2 for all participants is the end of 1988.

138 Work on the implementation of language pairs should not be stopped because other pairs need to "catch up".

Organisational Form

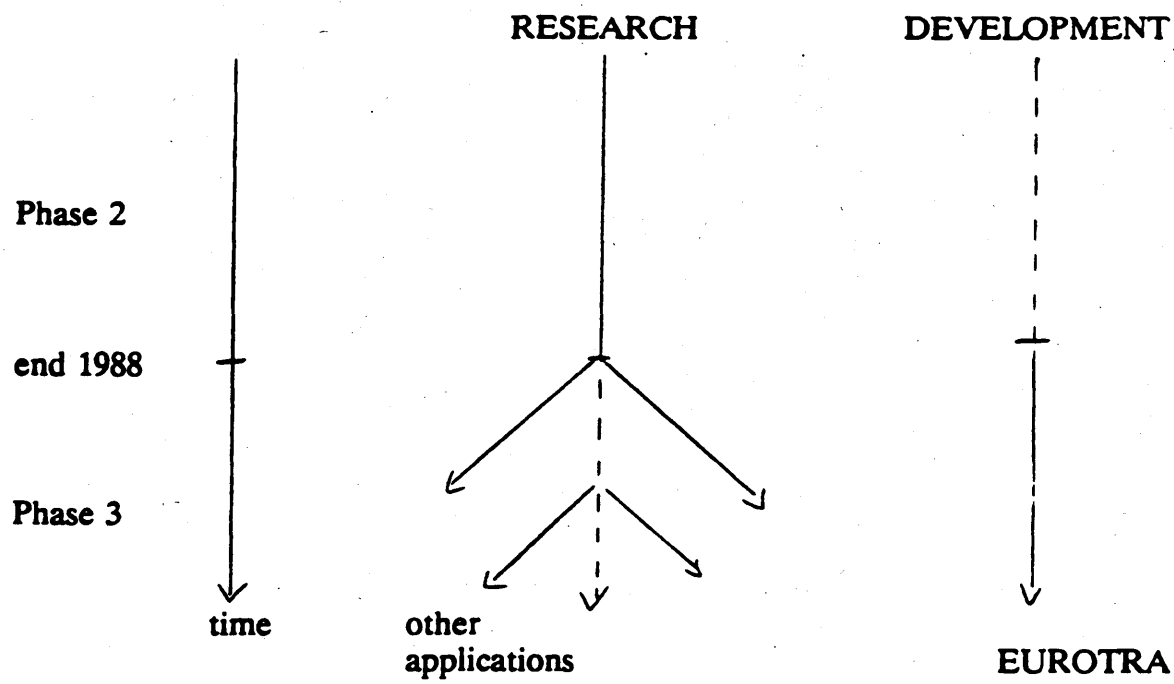
139 The management organisation of the project should be reviewed immediately for the remainder of Phase 2. The central management in particular should have a specific review of resources required. For Phase 3 a detailed management plan is required. This should be looked at in the near future, because of the necessary adaptation required to cope with the involvement of third parties. This also applies to the recommendations in III.

140 The Panel opinion is that the essence of Phase 3, as described in section 2, should be that the research and development work be separated, and that development should be transferred to an industrial footing. EUROTRA should then proceed on two parallel tracks, each with its own clear sets of objectives:

(1) research on linguistic aspects, building on the progress achieved in Phase 2;

(2) development of practical applications of the results of research, leading towards the production of a fully operational automatic translation system.

141 The following Figure gives a summary of the two parallel tracks:



II RESEARCH

Content

142 It is evident that research in computational linguistics should continue for a long time to come. Research in this area is important for the creation of "language industries" related to information services, which the panel sees as fundamental to the emerging new economy.

143 With regard to the present EUROTRA programme, in the linguistic research more attention should be paid to the following areas: the crucial dictionary component of the system; the contrastive aspects of language translation, since only a minority of the 72 pairs of languages have been studied from this point of view earlier; and the semantic problems involved (semantic features and relations), which are very important for the interface structure, require a great additional effort.

144 There should also be more attention paid to: standard computer architecture issues, better use of the capabilities of the existing architecture, and better use of existing software tools.

145 There is a need for stimulation of research into advanced computer architectures, in particular parallel and associative architectures. These could have major implications upon the future language industry.

146 The Panel recommends that CGC12 concerns itself with the way in which research, academic or pre-competitive, could be carried out in parallel with

EUROTRA. Stimulus would be provided by having, in mutual competition, a small number of European teams of workers with similar ideas, working in parallel with mainstream Phase 3 development. These teams would be constituted preferably by association between universities and industrial firms, R&D proposals could address either clearly defined practical problems or more fundamental questions. Additional research funding could be available from projects such as ESPRIT, or from the Framework Programme.

III DEVELOPMENT

147 The Panel is of the opinion that the original definition of the Phases of EUROTRA are not realistic. This applies especially to the transition from the present mode of operation to industrial development, which is a process spread out over time. The Panel believes that the suggested involvement of industry only after the end of Phase 3 does not fulfil the essential goal of EUROTRA.

148 The EUROTRA programme could result in products with potential for exploitation outside the Commission and the Community. However, this type of project would never have been undertaken as commercial research, and the Community institutions are likely to be the only customers for a system with these particular 72 language pairs. It must be recognised, however, that it is the wide spread of language coverage which puts EUROTRA in a class of its own outside other MT systems. Therefore if such a comprehensive system is to be realised then total public funding is required.

149 The amount of money needed for the development of a practical system for use by the Community institutions by an (in essence) industrial consortium cannot be estimated at the present time. It is certain, however, that the funds for EUROTRA, committed and earmarked now, will not be sufficient for that purpose. If an attempt was made to squeeze this development project into the present budgets, it would have a doubly negative consequence. It would kill the continuation of the research effort, and would lead to a very imperfect system, which could not be expected to improve on alternative, existing systems in performance.

150 The Panel recommends that a third party is commissioned to carry out a study about the definition and cost of the development of a practical EUROTRA system, based on the present and shortly expected research results.

Organisation

151 As stated above, further discussion and study is needed on possible realistic targets for a development project for the machine translation system. How advanced a system should it be? For what customers? As with all products, the more limited the objectives, the better the chances of success.

152 The Commission should pay more attention to the organisational requirements needed to execute the next phase, for example the work on dictionary

compilation. In the Panel's opinion, most of the present groups do not have adequate resources to cope with large scale dictionary work.

153 Before the end of Phase 2, private enterprise should be involved, both to help in the achievement of targets and definition of the final product. Specific areas requiring external input are (a) dictionary compilation work, and (b) specifying software needed for Phase 3.

Planning and Execution

154 The Commission should ensure that steps are taken to bring about the formation of a (multinational) industrial consortium to take on the development work suggested above. Early involvement of industrial organisations during Phase 2 will facilitate the formation of the consortium.

155 More thought is needed on how such a consortium would interact with the EUROTRA organisation; participants would have to consult with EUROTRA research groups to evaluate the applicability of their research results There is no need to postpone this consultation until Phase 3. The expertise of the existing National Groups should be used in the preparation work for the development phase. Several of the Groups are already concerned with external contract work.

156 It is suggested that proposals for Phase 3 research projects should be invited from the present research groups. Some of these might be in conjunction with industrial partners.

157 The work of the industrial consortium should be backed up by the activities of Central Operations (performed for instance by the IEGI as the Luxembourg National Group) in testing, maintenance and distribution of the EUROTRA product on behalf of the European Commission.

9

Danzin Report - March 1990

CONCLUSIONS AND RECOMMENDATIONS

55. The assessment panel is aware that these conclusions and recommendations may go somewhat beyond the terms of reference, which were to assess EUROTRA as a programme designed to provide a tool for the automatic translation of the Community languages. In our conclusions and recommendations, we have shifted the emphasis to the language technologies as a whole and we propose a policy and structure for coping with the impact of the new information technologies on natural languages.

The shift of this emphasis is in keeping with the internal dynamics of the programme and the events which have added to the corpus of knowledge in the field since EUROTRA was originally launched.

56. EUROTRA will not lead to an operational machine translation system but merely to what we have agreed to call a "scientific prototype"[1], which will moreover be imperfect and incomplete.

However, by the very fact of its existence, EUROTRA has laid the foundations for a Community achievement in the field of language technologies, and this is very important since it corresponds to a need which has become clear in the course of the past decade. We have endeavoured in our report to describe what is at stake and how we may meet the challenges.

Recommendations

57. Our main recommendations to the Commission are set out below. They fall into three categories, the first concerning the main developments of the project, in terms of objectives and organisation, and the other two concerning more technical aspects in the fields of linguistics and of computer environment.

58. The main developments proposed are based on the observation that the original ambition - ie that the third phase would already yield an operational prototype functioning in a given field on certain types of texts with a vocabulary of approximately 20,000 entries - was unrealistic. What we are more likely to obtain is a prototype usable exclusively by the researchers for their subsequent work. Thus, the development stage is still far off, which is understandable in view of the great difficulty of the objective. However, genuine progress has been made in the project since the last evaluation and it is out of the question that it should be

---

[1] The assessment panel agreed on this concept of "scientific prototype", to refer to a sum of theoretical and experimental results, the visibility of which would be demonstrated and which could ultimately lead to an "industrial prototype".

abandoned. Thus we propose arrangements which will enable research to be continued and to tackle development at a level which is more realistic and more in keeping with market requirements. Our conclusions are summarised in the following recommendations:

- *Recommendation No 1* : Given that EUROTRA provides the only incentive to theoretical and computational linguistics research for certain European languages and since work in these fields would be reduced or discontinued if the programme were called to a halt, EUROTRA should be protected, whatever its shortcomings. Thus, the efforts should be continued, albeit with revised objectives.

- *Recommendation No 2* : The new objectives should be such as to maintain or enhance the already positive spin-offs from the programme. In other words, the benefits in terms of basic research and specialist training should no longer be regarded as mere by-products of the project, but must become a formal objective.

- *Recommendation No 3* : When exploring new objectives, account should be taken of the fact that EUROTRA is still nowhere near being able to generate industrial products in the field of machine translation. It can, however, contribute to several monolingual applications, as mentioned in this report and the previous one, for which there is a market. Thus, the project should take a new direction and work toward the development of tools for monolingual applications.

- *Recommendation No 4* : These monolingual tools should be designed and implemented in close cooperation with industry.

- *Recommendation No 5* : If research and training work is to be carried out in parallel with pre-industrial development, changes must be made to the organisational structure, which is at present based on a single type of objective. In addition, the fact that the EUROTRA organisation can act both as awarding authority and project supervisor means that it has too much authority over its own affairs.

  Thus it is suggested that over the next two years, the research work should be accompanied by the study and implementation of a new organisational structure. We propose setting up a European language technology agency with functions and procedures as described in Chapter IV of the report. Particular attention would need to be paid in this study to synergy with the ESPRIT programme, and between the EUROTRA and SYSTRAN projects. Systems of financing which would be more suitable for all concerned should also be proposed.

- *Recommendation No 6* : The EUROTRA research teams should be given sufficient freedom to continue their work on a limited number of language pairs - ie those where they feel they have achieved the most advanced,

most illustrative or the most useful results - so that the validity of their work can be demonstrated in a number of cases before covering all 72 pairs.

The work of the Assessment Panel was also complicated by the tension between the need to take account of market forces - which give priority to a small number of languages - and the need to protect the cultural implications of all the languages. This question should form the subject of a specific study which would enable the relevant decision makers to gain a better understanding of all the aspects involved.

59. In the linguistic field, the recommendations in the Pannenborg report generally concerned semantics, dictionaries and the contrastive aspects. These recommendations have been followed, albeit to a limited degree. The existence of certain faults, such as overgeneration, was confirmed as the project progressed. Thus, the following recommendations reflect the opinions of the Committee on these problems:

- *Recommendation No 7* : EUROTRA's stratified approach is based on the traditional way in which linguists have attempted to solve the complex problem of describing a language. Current research in cognitive science, artificial intelligence and linguistic corpora shed new light on this question however. No serious consideration appears to have been given to this in the EUROTRA project. Thus alternatives must be sought to the stratified approach, the shortcomings of which are described in annex 2.

- *Recommendation No 8* : Certain progress has been made on semantics in certain groups. However it has not been nearly enough, particularly as regards the interface structure. Thus, improvements must be made to semantic representation if the resolution of ambiguity is to be improved and the correct choices made at the transfer stage.

- *Recommendation No 9* : An interactive approach would be a useful aid to the resolution of ambiguities, at least at the prototype stage.

- *Recommendation No 10* : Work on terminology and dictionaries, both monolingual and for transfer, is totally inadequate. Even if the definitive structure of the dictionaries cannot be fixed until the grammar is fixed - which it is not in numerous respects - considerably more attention must nevertheless be paid to this vital aspect of the project. Thought could be given to certain fundamental questions - in particular the use of knowledge bases for the representation of dictionaries.

- *Recommendation No 11* : Ultimately, limiting the examination of context to the sentence under construction will be a serious drawback. Thus, the idea of studying context beyond the limits of the sentences must also be studied.

60. There have been substantial improvements in the software environment since the last assessment. The aim of the following recommendations is to consolidate this improvement and promote its application.

- *Recommendation No 12* : The means of assessing and validating tools, particularly formalisms, should be defined, since while proposals exist for an assessment procedure for the project as a whole, there is apparently nothing permitting assessment of deliveries by one team to another, or for validating a tool. Benchmarks to assess the functionalities and performance of a module must be established along similar lines to those used for conventional software.

- *Recommendation No 13* : The resources at the disposal of the team responsible for the software environment should be increased so that the team can give more efficient support to the research and pre-development work being carried out in the fields of natural language processing (NLP) and computer assisted translation (CAT). In particular, the software environment should enable the national teams to cooperate in distributed activities via a computer network, by developing appropriate software and connections systems etc. In short, a EUROTRA network should be set up.

- *Recommendation No 14* : Dissemination and use of the software should be promoted for all the formalisms used in the project, with a view to testing it more effectively and reinforcing the EUROTRA community.

- *Recommendation No 15* : Work should continue on finding industrial applications for the spin-offs of the EUROTRA software environment in the form of monolingual products. In order to define these spin-offs more precisely, the project should include market research and a study of the technical aspects of rival products, including in the United States and Japan.

10

The Influence of Advances in Computer Science and Computer Technology on Machine Translation

H Steusloff

Machine Translation (MT) is, above all, still a linguistic problem. Dealing with the transformation of different natural languages into other natural languages means dealing with the complexity of human life and national cultures as expressed in those languages. This complexity, among others, requires the consideration of semantics and of enlarged contexts which both are still in a very early state of application to MT. In addition, natural language translation needs to be regarded as part of a complete process starting with document creation and ending with the availability and the use of a document in different languages. MT is just one of several steps in this process which needs computer support in most or, perhaps, all of its phases.

This Appendix will deal with recent developments in computer science and computer technology which could support MT.

Processing Hardware

The current development of processing hardware can be characterised by an annual increase of computational power by a factor of approximately 1.4 and triannual major changes of processor architectures. However, requirements of computational power for MT are immense and such hardware trends will not substantially ease the task or improve the efficiency of MT in the near future. The implications of considering a wider context in MT, (ie considering a context of more than one sentence for the translation of each sentence), calls for parallel processing systems which will be able to translate several sentences at the same time and then exchange context information, for example for disambiguation.

Since the size and price of distributed computer systems are decreasing at a similar rate as their increase in power, it would be advantageous to employ such distributed or networked computer systems with the same basic software systems and formalisms as are currently available, and to introduce information exchange between such MT systems working in parallel.

Consideration should also be directed to the idea of transforming the currently available framework software to make it run efficiently on a parallel processor system (ie analysis and synthesis of one sentence). Again, the availability of economical distributed computer systems and multiprocessor systems should improve the efficiency of MT, through the introduction of parallel processing into language translation.

Progress in Artificial Intelligence (AI)

AI techniques for the manipulation of symbolic information have reached a stage of applicability that would justify a detailed investigation of their applicability to MT. Since

the inclusion and the treatment of semantic information is essential for any substantial progress in MT, AI techniques could contribute to a breakthrough in semantic driven natural language processing. The successful use of Prolog in EUROTRA, for example for the implementation of the new ALEP formalism, is an encouraging example for a beneficial application of AI tools to implement current MT formalisms. In addition to the use of such AI driven implementation tools, the extensive introduction of AI techniques to the treatment of the very complex and, depending on ongoing culture-based changes of natural languages, rather specific semantics in different natural languages could be useful.

Object-Orientation

One of the major advances in designing and understanding information processing systems is the introduction of the object-oriented paradigm. Object-orientation supports the "natural" matching of real-world objects of all kinds to computer system structures and procedures. Object-orientation also provides means for an efficient and less error prone implementation of software systems through the concepts of classes and inheritance. Objects combine data structures and procedures and communicate with other objects supporting the parallel execution of the procedures within such objects.

The advent of object-oriented data base systems could be another stimulus for investigating the applicability and the advantages of object-orientation for MT. Object-orientation may be a satisfactory way of dealing with semantics due to the combination of data structures and procedures in a well controlled and systematic manner.

11

EUROTRA : Key Events

• 1976	Acquisition by Commission of English to French Systran.
• 23.12.76	First Multilingual Action Plan authorised, under which EUROTRA preparatory costs were funded, 1979/81.
• 1978	Formation of EUROTRA Coordination Group to prepare programme.
• 1979	Reference Manual, first release. First Annual Conference.
• 04.11.82	CD 82/752 authorising EUROTRA programme.
• 1984	Leuven workshop makes decision to follow the PATR II developments, rather than the Grenoble GETA formalism.
• 1984	CD 84/238 replaces Advisory Committee on Programme Management with the Management and Coordination Advisory Committee : "Linguistic Problems" (CGC-12).
• June 1984	First Contract of Association signed (Luxembourg).
• Autumn 1985	Sufficient Contracts of Association signed to allow programme to proceed.
• 1985	CAT formalism developed.
• 26.11.86	CD 86/591 adding Spain and Portugal.
• 1987	Decision to freeze formalism development on ETS.
• 28.09.87	CD 87/516 authorising Second Framework Programme.
• Autumn 1987	Pannenborg Report delivered.
• 25.07.88	CD 88/445 authorised programme to move to third phase on 1st July 1988.
• April 1989	Invitation to express interest in ET6/7 fully funded studies.
• 20.06.89	CD 89/410 authorised completion of EUROTRA to 30th June 1990.
• January 1990	ET6/7 studies awarded.

- March 1990 — Danzin Report delivered.
- April 1990 — CD 90/221 authorised Third Framework Programme.
- 26.11.90 — CD 90/664 authorised final two years of EUROTRA, 1991/92
- End 1990 — Issue of final linguistic specifications, Reference Manual.
- 08.03.91 — Call for proposals for ET10 cost-shared projects.
- March 1991 — Call for tenders for fully funded ET9 ALEP projects.
- Mid 1991 — Completion of ET6/7 studies.
- 21.08.91 — Call for proposals for LRE cost-shared projects.
- January 1992 — ET9 ALEP contracts awarded for two years.
- January 1992 — ET10 projects awarded.
- December 1992 — LRE1 projects awarded.
- December 1992 — Final stage of EUROTRA completed.
- End 1993 — ET10 completed.
- Jan - July 1994 — LRE1 projects completed.

12

Glossary

(This Appendix includes a description or definition of many terms and acronyms referenced in the text and in the tables of this Report.)

ALPAC US Automatic Language Processing Advisory Committee concluded in 1966 that human translating was faster, more accurate, and less expensive than MT, and that no further support should be given. As a consequence, only a minimal amount of M research was carried out in the immediate following years.

ATAMIR A multilingual system created by mathematician Iváan Guzmáan de Rojas using Aymara as pivot language. An evaluation was made by Madrid for the programme Extremadura Enclave 92 of the Junta de Extremadura.

Ariane MT system developed by Professor Vauquois in Grenoble GETA)

anaphora A feature of grammatical structure referring to something already expressed. "When Mary saw John *she* waved"

applied linguistics The application of theory, method of linguistics to practical problems.

aspect The duration or type of temporal activity denoted by a verb eg completion or non-completion of an action.

'Basic Linguistics Research' Eurotra. Research on morphology (inflection and derivation), syntax (NP-structure, anaphora, infinitives and relatices), semantics (tense and aspect systems) and computational lexica.

Basic English Sublanguage. Simplified natural language developed by Charles Kay Ogden in 1930. British American Scientific International Commercial consists of 850 words selected to cover everyday needs. This is supplemented by scientific words. Of historical interest but the first example of a sublanguage.

CALL Computer Aided Language Learning

CAT2 Efficient and simple sideline. Presented for the first time in 1987, and then again in several conferences (eg MT Summit, COLING) - it showed the possibility of building pre-industrial prototypes based on the linguistic concepts of Eurotra.

CoA At the basis of the programme is a series of bilateral *Contracts of Association* between Member States and the Commission, and about half the overall budget is directly contributed by the National funding authorities (the precise proportions differ between countries). The same regime operates for the core activity by 'language groups' in the Transitional Programme which involves researchers from all member states, while the CEC provides an additional ECU 6m for funding 'shared cost' research, training and industrial participation. Within the CoA structure, Central Contracts, either special study contracts with the Commission devoted to special problems, or special paragraphs within the Addenda of the CoAs were supported.

Chomsky Noam Chomsky wrote (1957) that grammar is a 'device of some sort for producing the sentences of the language under analysis'. Chomsky subsumes all aspects of sentence patterning, including phonology and semantics and introduces the term 'syntax' as the more specific notion, ie grammar = phonology + syntax + semantics. A more traditional approach is language structure = phonology + grammar + semantics. Chomsky developed the system of rules and symbols that provides a formal description of the underlying syntactic, semantic, and phonological structure of sentences. In recent years new approaches not based on Chomsky's generative grammar have been developed.

COMET CEC's programme for higher education in information technology

CSC The Common Steering Committee dealt exclusively with the CoAs and intellectual property rights : it comprised CEC personnel, or nominees.

'Coindexation tool' For dealing with unbounded dependencies within the Eurotra framework. A first component of the tool was designed and implemented by the Eurotra-Turin team at Gruppo DIMA in 1988. The recursion markers were subsequently designed and implemented by the Eurotra-DK team. The results of this collaboration have been published.

collocation The habitual co-occurance of lexical items "peanut butter"

comparative linguistics A branch of linguistics that relates the characteristics of different languages or varieties.

computational linguistics The application of the concepts and techniques of computer science to the analysis of language.

concordance An ordered list of words used in a particular text or corpus.
constituent analysis A process of analysing a construction into its major components, each component being analysed until a set of irriducable elements is left.
context The linguistic environment of an element.
contrast Any formal difference that serves to distinguish meanings in a language.
contrastive analysis The identification of structural differences between languages.
corpus A collection of language data brought together for linguistic analysis

DLT Developed by BSO in Netherlands. Part funded by National Government.
DECIDE Within Comett-programme, Leuven is involved in DECIDE. "Development of European Course on Information and Datacom Engineering". This project aims at the development of courses for SMEs. In Leuven 2 courses were written, one on NLP (together with the University of Groningen) and one on Electronic Dictionaries (together with the University of Uppsala). The contact with this project are not with the SMEs themselves, but mainly with the Chambers of Commerce and the like.
declarative grammar A grammatical construction used in expressing a statement "the linguist spoke"
declarative programming Prolog permits a very simple, direct implementation of augmented context-free grammar. Context free analysis is available as a special case of the general control structure provided by Prolog. Prolog is an example of a declarative programming language - the word order of a sentence can be analysed independent of the execution of Prolog commands of a procedural language such as Fortran.
deep grammar / structure An underlying level of grammatical organisation that specifies how sentences should be interpreted.
derivation The set of analytical steps required to generate a sentence.
determiner An item that co-occurs with a noun expressing such things as number of quantity "some books"
dictionary A reference book listing words or terms and giving information about a particular subject or activity.
discourse A continuous stretch of language larger than a sentence.

EAC Eurotra Advisory Committee. Chaired by the CEC, and its members comprised representatives from the various national governments
E-Star Since 1989 Gruppo Dima has carried out the design and implementation of a sideline E-Star. : a new Prolog lingware and s/w prototype for expressing and applying Eurotra-type linguistics for multilingual MT. The lingware formalism is based on Eurotra and retains data structures, unification, subsumption, etc. Whereas Eurotra is purely an experimental MT demonstration system, E-Star is conceived as an operational translation tool suitable for batch as well as interactive applications. New devices (not in Eurotra) allow 'fail-soft' translation in case of trouble, and 'revocable preferences' are relevant to fully automated batch translation. They make it possible for single choice of translation. Also a user-friendly MMI for human aided MT is available.
EUROLANG From SITE. See Section 7.7
EUROTRA Though originally envisaged as being of 44 months duration, the accession of Spain and Portugal meant that the project was extended until the end of 1990: it comprised phase 1 (1983-84, preparation), Phase 2 (1985-88, basic and applied linguistic research), Phase 3 (1989-90, development). It was succeeded by the Transitional Programme for Eurotra (1991-92), the transition in question being that from a pre-industrial prototype to an operational one. Further work is now under way (January 1993-95) as part of LRE. The total budget for the 'pre-industrial' phase of Eurotra (1985-90) was about ECU 44m of which around half was contributed by the CEC. The CEC contribution to the Transitional Programme is about ECU 10m ECU, and its contribution to the LRE programme is forseen as about ECU 22m. (In addition, prior to the advent of the CoA structure in 1985, various study and consultative contracts were awarded.)
E-framework (ETS formalism) For ETS to avoid making the analysis component of the MT system target language dependent, the **transfer** approach was chosen, based on the following principles: a) transfer should be as simple as possible - preferably limited to the replacement of lexical material, preserving structure and features (the notion of **simple transfer**); b) analysis and synthesis should be strictly monolingual - ie not devised with one or more target languages in mind [this allowed 9 language groups to work on the official 9 languages]; c) abstract representations,

called **interface structures** (IS), should act as the vehicles for delivery of analysis and transfer to synthesis, and the receipt from synthesis to transfer and analysis - they take the form of dependency structures, enriched with semantic information; d) the mapping of sentences onto interface structures (and vice versa) is not one-shot, but is performed by a number of intermediate representations (the principle of **stratification**).

ET-6 The ET-6 studies in the Transition Phase were intended to assess the strengths and weaknesses of the current prototypes with respect to the state of the art in CL and NLP and propose an improved framework. A number of high level requirements were placed on the formalism redesign, amongst which that the design had to be totally mainstream and extensible as new phenomena and capabilities can be added. The first of these developed specifications for a new formalism (the ET-6 Formalism, ET-6/1), and the second led to specifications of a user and grammar development environment (ET-6/2), and the third (ET-6/3) dealt with issues of low-level text encoding and handling (including some morphological analysis).

ET-10 a) Collocations and the lexicalisation of semantic operations - collocational restrictions (not idioms) (eg 'rancid butter' v 'sour milk'); b) Terminology - the definition of internal representation of terminological definitions and their use in analysis and generation, the parsing of definitions, and the output of such parsing in analysis and generation; c) Knowledge Bases - this involves the feasibility and effectiveness of the (semi-) automatic parsing of dictionary definitions (from COBUILD) as a form of knowledge acquisition for ET-6, with wider relevance for other natural language systems; d) implementation of probabalistic and Corpus-based methods in Eurotra within the ET-6 architecture; e) the Reusability of Grammars for ET-6 - involves research on the migration of grammars to the new ET-6 formalism.

ellipsis The omission of part of a sentence (eg for economy, emphasis), where the missing element is understood from the context. "Where is the book? On the table"

expression Any string of elements treated as a unit for analysis eg a sentence, a idiom.

FoLLI The European Foundation of Logic, Language and Information

finite state grammar A simple kind of generative device that is able to process only a very limited range of sentences.

formal logic The study of systems of deductive argument in which symbols are used to represent precisely defined categories of expressions.

frame A specific structural context within which a class of items can be used.

formalisms The mathematical or logical structure of a scientific argument as distinct from its subject matter.

formalisms for EUROTRA To 1985, the accepted processing model was essentially inherited from SUSY and GETA, namely a Controlled Production System, involving the successive transformation of structures by means of pattern-matching rules, which could be organised into sub-grammars under various ordering regimes. Typical US west coast formalisms at that time were PATR-II, LFG, the GPSG family. However, by 1985, it was felt that this was too unconstrained a model to be effective in the highly distributed setting of Eurotra. It was also rather isolated from what was then clearly emerging as the mainstream of NLP, involving *unification* based formalisms. This dissatisfaction led, in early 1985, to the CAT Framework. With a few variations and additions the basic ideas of this framework persist in both the 'mainstream' Eurotra formalism (E-framework or ETS), as well as the 'sideline' prototypes that were produced exploring alternatives, namely CAT2 (1989-92), MiMo (1986-88), MiMo2 (1989-90). Subsequently, in 1991, the ET-6 'new formaiism' studies involved PATR-II (ELU, ISSCO, LTAG) redesign and the adoption of the ALEP formalism.

GENELEX Eureka project. Building conceptual models for electronic dictionaries. (IBM France, GSI Erlic, SEMA, ILTEC-PT, LADL-Fr)

GRAAL Eureka project. Toolset to help with NL, MT and knowledge extraction.

"Generic lexical resources" The Utrecht group has taken a special interest in reusable grammars and dictionaries.

generative grammar A description of a language in terms of explicit rules that ideally generate all and only the grammatical sentences of a language.

GPSG Generalised phrase structure grammar. This theory does not recognise the role of transformations in a generative grammar. Instead it focusses on developing the phrase structure dimension to grammatical analysis.

genitive An inflection that expresses such meanings as possession or origin "the database's content"

grammar The study of sentence structure, especially with respect to syntax and morphology, often presented as a textbook or manual. 2. A systematic account of the rules governing language in general, or specific languages, including semantics, phonology, and often pragmatics.

Horatio Sideline from Liege. A parser for a subset of English, with focus on the treatment of multi-word units and the importation of material from a machine-readable dictionary, namely LDCE.

hierarchy A classification of linguistic units into a series of successively subordinate levels, especially an analysis of sentences into clauses, phrases, words and morphemes.

Infoterm Infoterm in Vienna has pioneered terminology research in Europe and the Eurotra Ireland research is rooted in the Infoterm philosophy. Infoterm, originally a terminology centre for the translator or documentalist has, in recent years, started exploring the possibility of applying terminological knowledge structures to the field of NLP (text retrieval, hypertext) and MT. Eurotra Ireland is the only centre carrying out research into sublanguage and terminology within the context of MT, but some research is being carried out on the reusability of lexical resources at the University of Limerick and on lexical issues and the Irish language at Queen's University in Ulster.

idiom A sequence of words that is a unit of meaning eg "kick the bucket" (= die).

LDOCE Longman Dictionary of Contemporary English (MRD) - represents the work of about 100 man-years of people that are specialists in the field of lexicon design and maintenance.

LILOG German NLP programme

LINGUA CEC language learning programme.

LRE The post-Eurotra LRE (Linguistic Research and Engineering) programme - within the 3rd Framework Telematics programme. LRE is entirely funded by the shared cost scheme. LRE is intended to promote a range of R&D initiatives, not just in MT, but in NLP in general, and in various types of 'spin-off' applications. Work under LRE is grouped into five main headings : a) Research of General Interest : ways of increasing the interlinguality of linguistic representations of text / discourse; the use of domain specific knowledge (eg terminological, 'real-world' specialist, and 'heuristic' knowledge); interfacing NLP and speech technology; advanced computing; b) Common Tools and Resources : development of generic software tools, grammars, dictionaries, terminological collections, and text corpora, which can be re-used for a variety of applications and purposes. Eg integrated testing and development environments, tools for dictionary construction, workbenches, etc; c) Linguistic Standards - definition of commonly agreed data encoding schemes and formats for linguistic resources (eg dictionaries, grammars, corpora) - the EAGLES expert group; d) Applications - the aim here is to support pilot and demonstrator projects in areas such as : MT; automatic document abstracting and indexing; aids for mono- and multi-lingual document generation, storage and retrieval; MMI; computer aided instruction; construction of knowledge bases from natural language text; e) Supporting Actions - this covers training, initiatives to raise awareness, gather, synthesize, and disseminate information about NLP, with special emphasis on the economic and social impact of the technology, and legal problems that may act as barriers to its emergence. Eg via the VALUE programme.

level 1. A kind of representation recognised within the derivation of a sentence eg deep vs surface grammar. 2. One of a series of structural layers within a sentence (clause, phrase, word, etc)

lexical item (lexeme) Smallest contrastive unit in a semantic system "switch on".

lexicography The art and science of dictionary making.

lexicon (lexis) 1. The vocabulary of a language, especially in dictionary form. 2. A list of terms relating to a particular subject.

'lexical semantic theory' Eurotra context. Has been designed and implemented since many years in both the Italian dictionary and grammar. DIMA's improved version is about to be finalised and implemented in E-Star.

linguist 1. Someone who is proficient in several languages. 2. A practitioner of the subject of linguistics.

linguistics The scientific study of language.

MAT Machine Assisted Translation

MAHT Machine Assisted Human Translation

MENELAS An Access System for Medical Records using Natural Language). A front-end developed by Leuven.

METAL From Siemens Nixdorf. Distributed by Sietec. See section 7.7

MIMO systems Small experimental prototype MT systems, translating between English, Dutch and Spanish in all directions. Designed and built by Utrecht in close collaboration with Essex and ISSCO. MIMO-2 was developed in Utrecht. MIMO demonstrated the formal framework designed for Eurotra during the period 1985-87 (based on notion of compositionability). Run times reasonable (<1 min per sentence on 1MIP machines) for grammars with fair coverage and small (ca 300 words) dictionaries. MIMO-2 (produced 1988-90) based on mainstream computational linguistics (unification, HPSG) and had as one of its main principles, reversability. Performance like MIMO. Main point demonstrated by MIMO-2 was that it was possible to base a design for an experimental MT system on current mainstream CL.

MLAP Multilingual Action Plan (DGXIII)

MRD Machine Readable Dictionary

machine translation (MT) Use of a computer to carry out the task of translation.

modal A verb that signals contrasts in speaker attitude (mood) eg *may, can.*

modality The system of modal expression.

modification The structural dependence of one element (the modifier) on another.

mood Attitudes of fact, wish, possibility, etc., conveyed by a verb (a modal) or clause, eg indicative, subjunctive.

morphemes The smallest contrastive unit of grammar (eg bound forms de-, -tion, -s, etc)

morphology The study of word structure, especially in terms of morphemes.

'NLP Research' In the Eurotra context, has been referred to as including tense and aspect, determination, negation and quantification, morphology.

natural language A language with native speakers cf auxiliary language - a language adopted by different speech communities for the purpose of communication; cf artificial language - an invented language to facilitate international communication

noun phrase A phrase with a noun as the head "the tall man in a hat".

number The grammatical category that expresses such contrasts as singular, plural, dual (a grammatical contrast in some languages referring to "two of").

object language A language that is the object of analysis (using a metalanguage).

onomastics The study of etymology (the study of the history of origin and meaning of words) and use of proper names. (Re Onomastica project in LRE.)

PaTrans Being developed by CST Denmark. The goal is to make a customised translation system for a private company for patents from English into Danish - it reuses and further develops the implemented Eurotra grammars and lexica for the two languages covered. Built on the Eurotra software, which is enhanced and optimised, so as to comply with the requirements of a production system. Launched in 1992 after a feasibility study to investigate the possibility of transforming the Eurotra research prototype to a real-life system. The study concluded that it was feasible to build the desired system and that it would produce fairly high quality translations due to the strong linguistic approach inherent in the Eurotra model. The 18 month PaTrans project is the first attempt to use the Eurotra results commercially. The client has expressed interest in having similar system built for different source languages.

Problem Office In the period 1986-1990 the linguistic research in the project was organised by the Problem Office. This PO issued calls for tender, processed the tenders, defined the work programmes, and kept the project informed of the progress in the various research groups. Such groups typically consisted of linguists from different Eurotra teams. They usually worked together for a period of 6-10 months and summarised the results of the research in a final report. Eg *Interlevel Syntax* (1990, 116 pages) UMIST/Torino/Leuven/ Utrecht/Paris. *Word Structure* (1990, 220pages UMIST/Luxembourg/Barcelona/Utrecht/Saarbrücken/Salford/Leuven/Athens/Nancy/Lisbon.

Prolog was chosen for Eurotra, because its predicate calculus approach allows simple definition and implementation of special purpose tools - it allow rules or implications to be stated. An example of such a tool would be a formalism geared towards a specific task, eg coding a dictionary or grammar rules. In this way it is user-friendly for linguists and lexicographers, who have to formalize and code their grammars and dictionaries, although the penalty for this is poor performance.

'Preference Mechanism' Involving rules that can be optionally written for any level of representation of the Eurotra system in order to compare linguistic objects at that level, and select only those which fulfil the preferences. The mechanism resulted from the collaboration between Gruppo DIMA (which carried out the implementation) and other Eurotra Centres (especially DK and EL). A paper has been published.

'Preference mechanism for overgeneration problems'. The software and linguistic engineers in the Athens team developed this. The mechanism was later integrated into a general reference package, product of the collaboration of Italian, Danish and Greek researchers, which was presented at the ACL Conference in 1991.

parsing Analysing and labelling the grammatical elements of a sentence. Also diagramming, clause analysis.

phonology The study of the sound systems of languages.

phrase A group of words smaller than a clause forming a grammatical unit "in a box".

phrase marker A structural representation of a sentence in a generative grammar, usually in the form of a tree diagram.

phrase-structure grammar A generative grammar that provides an analysis of sentences into constituent elements. Taking the sentence (S) "The girl chased the dog". The first division produces a 'noun phrase' (NP) the girl and a 'verb phrase' (VP) *chased the dog*. The second division recognises a 'verb' (V) *chased* and another noun phrase *the dog*. The next divisions would produce combinations of 'determiner' (DET) and 'noun' (N) *the girl, the dog*. This is the phrase structure of the sentence and it ca be displayed in a tree structure, or as labelled sets of brackets.

pre-editing The human preparation of text for input into an MT system - usually by a translator, or technical specialist.

post-editing The proofing, editing and partial writing of translated text output from an MT system.

Reference Manual As a result of Eurotra's attempts to push the idea of linguistics based MT (as opposed to, eg AI approaches) to its limits, the Reference Manual can be seen as an enormous repository of linguistic information, described within a common framework, and with comparable coverage for all 9 languages.

'Research Clusters' Make common reports to the Liaison Group - collections of Centres / staff tasked with examining specific topics.

Rosetta Developed by Philips Netherlands. Part funded by National Government. Sound linguistic basis. Multilingual. Good coverage and performance. Not based on mainstream CL.

reduction The lack of one or more of the normal constituents in a construction "gone to town" cf ellipsis.

regular Said of a linguistic form that conforms to the rules of a language.

restricted language A highly reduced linguistic system found in narrowly defined settings, eg heraldry, weather reporting.

rule A generalization about linguistic structure. The rules of a generative grammar are objective descriptions of the grammatical patterns that occur. A prescriptive grammatical rule is a statement that indicates whether it is right or wrong to use a particular construction.

Semantic Labelling Study ET-D developed systems for labelling semantic relations and lexical semantic categories (1988, published in Steiner/Schmidt/Zelinsky)

SUSY MT system developed in Saarbrücken.

'Statussseminar' The method whereby German MT groups meet yearly and national experts peer review the work.

sentence The largest structural unit that displays stateable grammatical relationships, not dependent on any other structure.

source language A language from which a word or text is taken.

statistical linguistics The study of statistical properties of language.

stratification A model of language as a system of related layers, or strata.

structural semantics The study of the sense relations between words.

sublanguage Subset of natural language. Examples are the knitting pattern work of Dublin, and vocabularies based on the telecommunications terminology database.

surface structure / grammar A syntactic representation of a sentence that comes closest to how the sentence is actually pronounced.

syntax The study of word combinations. The study of sentence structure.

TRANSLEARN LRE I MAHT development.

TMC Telematics Management Committee

'Transition Phase' In the Transition Phase the following four activities were pursued:a) continuation of the E-framework R&D - especially contrastive research on linguistic topics - by the same teams as in Eurotra I, and on the same funding basis (CEC plus National Government funding, total ECU 8m) b) implementation of an enhanced development and research system (formalism, development environment, etc) along the ET-6 study recommendations (directly funded by the CEC ECU 2m) c) shared cost research involving industry (CEC ECU 2m) d) training, mainly in participating centres (CEC ECU 0.5m)

target language The language into which a translation is made.

tense A change in the form of a verb to mark the time at which an action takes place (past, present, etc).

term Name, expression, or word used for some particular thing.

terminology The body of specialized words relating to a particular subject. The study of terms.

terminology database A database of terms.

text A stretch of spoken or written language with a definable communicative function (news report, poem, road sign, etc).

textlinguistics The study of the linguistic structure of texts.

thesaurus A book of words grouped on the basis of heir meaning.

transformation A formal linguistic operation (a transformational rule) that shows a correspondence between two structures.

transformational grammar A grammar that uses transformational rules.

transformational rule In Generative Grammar, a rule that converts one phrase marker into another. Taken together, these rules convert the deep structures of sentences into their surface structures.

tree diagram A diagram used in generative grammar to show the heirarchical structure of a sentence.

Unification (Formalism) Grammars MT systems before Eurotra were mainly procedural. Unification is an operation that combines information from two objects (eg representations or descriptions), providing it is not contradictory.

universal grammar A grammar specifying the possible form a language's grammar can take.

verb phrase In generative grammar the whole of a sentence apart from the first noun phrase.

word The smallest unit of grammar that can stand alone as a complete utterance, separated by spaces in a written language.

Bibliography

The following references have been used to develop the content of this Report:

1. EUROTRA CSC/159/92 Final Activity Reports for CSC. (Available from the Commission.)

2. The EUROTRA Linguistic Specifications, in Studies in MT and NLP, Eds. C Copeland, J Durand, S Krauwer, B Maegaard, CEC 1991. (Available from the Commission.)

3. An Introduction to the EUROTRA Machine Translation System, Anthony Raw et al. in Working Papers in NLP Ed. Frank van Eynde (Leuven), Pius ten Hacken (Utrecht) 1990. (Available from Katholieke University of Leuven.)

4. EUROTRA : An Assessment of the Current State of the EC's MT Programme. Doug Arnold and Louisa Sadler in Working Papers in Language Processing, December 1991. (Available from University of Essex.)

5. LRE Call for Proposals 1992 : Technical Background Document. (Available from the Commission.)

6. The Cambridge Encyclopaedia of Language. David Crystal, 1987. Cambridge University Press.

7. Questionnaire Responses. (The completed questionnaires from the EUROTRA Centres are held on file in Logica and the Commission.)

8. Visit Reports. (The reports of members of the Panel during the Review period. Held on file in Logica.)

9. EUROTRA Reference Manual. (Available from the Commission.)

10. EUROTRA Assessment Report : Dr A Pannenborg (October 1987). (Available from the Commission.)

    EUROTRA Assessment Report : M A Danzin (March 1990). (Available from the Commission.)

11. "Vers une Infrastructure Linguistique Europeene". M A Danzin (March 1992). (Available from the Commission.)

ISSN 0254-1475

COM(94) 69 final

DOCUMENTS

EN **16**

Catalogue number : CB-CO-94-409-EN-C

ISBN 92-77-80585-4

Office for Official Publications of the European Communities
L-2985 Luxembourg

A / 3 - 3