European Commission
Directorate General XIII

MULTILINGUAL ACTION PLAN (MLAP)

Linguistic Research and Engineering

Overview of the Actions launched in 1994 with a view to preparing Language Engineering RTD in the Fourth Framework Programme

October 1994

❸

PREPARATORY ACTIONS IN THE

AREA OF LANGUAGE ENGINEERING RTD

(MLAP 1994)

*****

1. GENERAL

A call for proposals was published in the EC OJ on 15 March 1994, with a view to i) fostering active user involvement in RTD projects, ii) profiling specific usage situations and application scenarios within the broad domains covered by the Language Engineering (LE) Sector of the Telematics Applications Programme 1994-1998, and iii) laying down the foundations of future LE project clusters.

A total of 95 project proposals were submitted, which underwent technical assessment in early June. Some basic data concerning the proposals are provided here below:

Participants: 505

*	IT industry, including SMEs:	162 (32%)
*	Private sector other than IT industry:	46 (9%)
*	Universities and non-profit research centres:	228 (45%)
*	Others, including public authorities and utilities:	69 (14%)

The distribution of resources (project partners and EC) across categories of participants was as follows:

*	Universities and research centres (100% additional costs):	39%
*	Private sector and user organisations (50% actual costs):	61%

In spite of its limited scope, the call was very successful in attracting the interest of the private sector, including major IT companies, service/network operators, and SMEs specialised in speech technology and natural language processing, as well as user organisations including but not limited to administrations and public utilities, service and manufacturing industries, and publishing houses.

The proposals finally retained for funding address a number of highly relevant application domains, including voice response servers handling telephone inquiries, authoring workbenches for technical writers in the aerospace and automotive industries, document management and report generation for health-care operators, computer-aided foreign language learning, access to and navigation in large textual repositories, etc., as well as inter-related preparatory actions in the area of re-usable language resources, in particular speech and text corpora, lexica and terminology databanks, and the associated guidelines for encoding and interchange of language knowledge and data.

A public workshop will be held in October 1995, within the framework of the 1995 LE Convention and Exhibition, to present and discuss the results obtained by the consortia.

2. PROJECT SYNOPSES

2.1 TELE-SERVICES

RAILTEL

RAILTEL addresses the preliminary stages of a project which aims to provide interactive, speech-based passenger information services via telephone. These services will rely on a multilingual human-machine interface, designed to meet the inherently transnational requirements of the transportation field. An early field test will be based on existing speech technology, where this is sufficiently robust and mature. More advanced speech recognition and understanding techniques will be investigated if required. Two national railway operators are involved in the consortium, which also includes providers of speech technology and developers of information services, as well as research centres working in the field of man-machine communications. The consortium will eventually be enlarged to include as many European railway companies as possible, in collaboration with the MAIS project (see below).The service envisaged should be capable of handling telephone traffic of millions of calls each year in each country, thus creating a promising market for both service providers and equipment manufacturers.

Coordinator: CSELT- Mr Roberto BILLI
                I - 10148 Torino - Via G. Reiss Romoli 274
                Tel +39-11-2286209
                Fax +39-11-2286207

Partners: SARITEL - I
                EDINBURGH UNIVERSITY (CCIR) - UK
                LIMSI-CNRS - F
                FERROVIE DELLO STATO - I
                BRITISH RAIL, BUSINESS SYSTEMS - UK

MAIS

The aim of the MAIS project is to foster the deployment of multilingual inquiry systems, providing voice-controlled access to information and transaction services. The project will design a telephone-based information system for public transport, with a fully interactive, user-friendly interface. A mock-up demonstrator covering French, Dutch and English will be developed and field-tested. The demonstrator will offer substantial gains in usability and functionality as compared to current menu-based inquiry systems, and will stimulate a demand for other related applications. The long term goal of the consortium is to build multilingual information and transaction systems for transport and other services, which can then be used in their own language by European citizens in all countries. The consortium comprises national railway companies, technology providers, and public research centres.

Coordinator: PHILIPS GMBH FORSCHUNGSLABORATORIEN - Dr Martin OERDER
                D - 52066 Aachen - Weisshausstr. 2
                Tel +49-241-6003-571
                Fax +49-241-6003-518

Partners: SNCF, DIRECTION RECHERCHE - F
                NEDERLANDSE SPOORWEGEN - NL
                INST. DE RECHERCHE EN INFORMATIQUE, UNIV. PAUL SABATIER - F
                STICHT. KATHOLIEKE UNIV. NIJMEGEN, DEPT. LANGUAGE & SPEECH - NL
                RWTH AACHEN, UNIV. OF TECHNOLOGY - D

PP-TELELANG

PP-TELELANG encompasses the initial stages of a project which aims to create an integrated network of translation facilities, both human- and machine-based. The project will combine existing systems (machine translation systems, pre-translation and dictionary management tools, etc.) and services provided by e.g. translation bureaux, into a pan-European resource accessible via telematics networks, e-mail systems, etc. The main users of the envisaged service will be companies and organisations dealing with multilingual documents. These users will profit from the availability of a flexible and powerful translation service with a very short response time. An interest group of such users will be created at the outset of the project and will play a key role in the elaboration and validation of the functional and operational specifications of the demonstrator aimed at, as well as in the dissemination of project results. The consortium comprises system engineering, translation, and network services companies, The consortium is intended to be open and will be progressively enlarged to include suppliers of language systems and resources, network/service operators, and corporate customers.

Coordinator: SONOVISION ITEP TECHNOLOGIES - Mr Joël OLIVIER
                 F - 94700 Maisons-Alfort - 2, rue Louis Pergaud
                 Tel +33-1-45130500
                 Fax +33-1-45130559

Partners: SIETEC SYSTEMTECHNIK GMBH - D
                 INCYTA - E
                 THAMUS - I
                 ITALDATA - I
                 SARITEL - I

2.2 AUTHORING TOOLS

INDOCREN

The goal of INDOCREN is to make document production easier and more effective by providing assistance with tasks which are regarded as most time-consuming and error-prone. In particular, INDOCREN aims to integrate into a modular authoring workbench language tools and resources covering the entire life-cycle of technical documentation, from creation to production, through translation and revision. Such tools and resources are currently only available in isolation and in mutually incompatible formats. The INDOCREN workbench will provide easy access to information and knowledge bases, help with authoring and translation, provide mechanisms for quality and coherence checking, and a versioning mechanism for the maintenance and updating of language resources. The INDOCREN consortium comprises end-users from the aerospace, motor and electrical engineering industries, system integrators and technology providers. The consortium will organise at the very beginning of the project a special interest group, to be composed of organisations mainly from the commercial sector, which will permit validation and dissemination of results.

Coordinator: GSI-ERLI - Mr Antoine OGONOWSKI
                 F - 94227 Charenton Cedex - 1, place des Marseillais
                 Tel +33-1-48 93 81 21
                 Fax +33-1-43 75 79 79
Partners: RENAULT - F
                 AEROSPATIALE - F
                 ALENIA - I
                 DATABASE INFORMATICA - I
                 RANK XEROX LTD. - UK

DOCSTEP

DOCSTEP concentrates on the documentation process in the aerospace industry, with the aim of defining an advanced authoring system to help overcome the major bottlenecks in current methods of producing and maintaining technical documents. Relevant methods and techniques will be investigated and a strategy drawn for their introduction into the end-user environment. A mock-up version of the system will integrate amongst others product information and resources, a terminology management system and a grammar checker for controlled languages. The system will be specified in direct co-operation with a group of corporate users, which will be progressively enlarged during the life-time of the project. Major European aerospace industries and satellite operators are part of the consortium, which also includes software and lingware developers, system integrators and engineering research centres.

Coordinator: CAP GEMINI INNOVATION - Mr Philippe LEVISSE
               F - 92513 Boulogne Billancourt - 86-90, rue Thiers
               Tel +33-1-49 10 53 96
               Fax +33-1-49 10 06 15

Partners: AIRBUS INDUSTRIE - F
               MATRA MARCONI SPACE - F
               INST. FÜR RECHNERANW. IN PLANUNG & KONSTRUK., U. KARLSRUHE - D

DOME

The project concerns the generation and processing of medical reports. Two different aspects of the report production cycle are addressed. The first concerns the standardised production and verification of reports by medical staff. The second concerns the extraction of information from reports, for storage and subsequent access by medical staff, hospital administrators, researchers, etc. The consortium comprises hospital departments, software houses active in the field, and research teams specialised in medical informatics and language technology. Work carried out during the project will be monitored by a medical advisory board. The principal goal of DOME is to contribute to an improvement in health care, and consequently the needs of the health care community will be taken as the starting point. A group of potential end-users will be set up and progressively enlarged in order to ensure critical evaluation of the system aimed at.

Coordinator: HÔPITAUX DE PARIS - CHU PITIE-SALPETRERIE - Mme Brigitte SEROUSSI
               F - 75634 Paris - 91, Boulevard de l'Hôpital
               Tel +33-1-45 86 56 84
               Fax +33-1-45 86 56 85

Partners: IRISH MEDICAL SYSTEMS - IRL
               HCRC, LANGUAGE TECHNOLOGY GROUP, UNIV. EDINBURGH - UK
               RAMIT VZW, UNIVERSITY HOSPITAL GENT - B
               CENTRE D'INFORMATIQUE HOSPITALIERE, HÔP. UNIV. GENEVE - CH

2.3 INFORMATION ACCESS AND MANAGEMENT

MEMORIA

The aim of the project is to provide a new generation of "reader's assistants", which will guide readers working in "virtual libraries" (i.e. large distributed repositories of electronic documents, including images as well as text). The project aims to promote the use of networked, multi-user models and tools that help the user find his way through large corpora, relying on knowledge acquired from textual information as well as human

expertise. Specifications will be drawn up on the basis of user requirements, taking into account a variety of user profiles, cultural differences, and the need for a European dimension. A special interest group will be set up. The group will be open to public and private libraries, museums, publishers, etc., who all share the need for efficient management and exploitation of document holdings in electronic form. Libraries and text archives, research centres and technology providers are part of the consortium, which will be enlarged to comprise a wider range of actors.

Coordinator: CAP GEMINI INNOVATION - Mr Eric BILANGE
                 F - 92513 Boulogne Billancourt - 86-90 rue Thiers
                 Tel +33-1-49 10 53 36
                 Fax +33-1-49 10 06 15

Partners: OXFORD UNIVERSITY, OXFORD TEXT ARCHIVE - UK
                 AIS S.A. - F
                 BIBLIOTHEQUE NATIONALE DE FRANCE - F
                 INST. DE RECHERCHE EN INFORMATIQUE DE TOULOUSE - F
                 CONSORZIO PISA RICERCHE - I

TAMIC

TAMIC investigates the feasibility and requirements of natural interfaces meant to greatly facilitate the access to administrative information, in particular when this information is heterogeneous and/or held at geographically dispersed sites. The project scenario envisages an information desk staffed by a clerk who is able to access multiple information sources, and with whom the end-user interacts. The clerk would then translate the user's inquiries into a form understood by the system. This approach would help overcome two major problems in current methods of accessing information handled by public services and utilities, namely fragmentation (different help-desks deal with different domains) and lack of transparency (the relevant information can be difficult to retrieve through menu-based inquiries). State-of-the-art language technology can be adapted for use in such a human- and technology- mediated context. Moreover, users who are trained in relevant domains will be able to interact directly with the system, either from home or from the workplace. Potential users, that is administrations and public utilities, will be involved in each stage of the project, and the interest group set up within the project will allow different bureaucratic cultures and languages across Europe to be taken into account. Software houses and research centres make up the consortium.

Coordinator: QUINARY SpA - Mr G. ROCCA
                 I - 20121 Milano - Via Crivelli 15/1
                 Tel +39-2-58 30 27 12
                 Fax +39-2-58 30 53 74

Partners: IRST - I
                 CAP GEMINI INNOVATION - F
                 DFKI - D

2.4 FOREIGN LANGUAGE ACQUISITION

LOTOS

LOTOS aims to devise a management and authoring system for second language learning. The system is intended to facilitate the re-packaging and delivery of comprehensive language learning materials. As well as a number of software modules which provide monitoring and control facilities to tutors, and feedback to learners, the system will offer interfaces for accessing external databases, for communicating with remote tutors and learners, and for customising LOTOS packages to suit the needs and purposes of individual learners. The design and specifications of a LOTOS demonstrator will be based upon user, market and technology studies, as well as on careful examination of relevant learning and tutoring scenarios. A consortium composed of software houses, software and telematics service providers, academic institutions, and end users will co-operate in the definition of the system. The results of the project will be validated by, and disseminated through, an interest group comprising user and industrial organisations, which will be set up in co-operation with the ILAM1999 project.

Coordinator:	CRYSTAL PRESENTATIONS Ltd. - Mr Joseph FARRELL 80 Sherlock Street - UK B5 6LT Birmingham Tel +44-21 666 6988 Fax +44-21-666 6994
Partners:	BERLITZ SCHOOLS OF LANGUAGES GMBH - D INSYS S.A. - F UNIV. TÜBINGEN - D CORMORANT TELEMATIC SYSTEMS - IRL PROLOGIA - F

ILAM1999

The project's objective is to carry out preparatory work for the development of tools to assist in spoken language acquisition. The main target market is that of second language learning for business purposes. ILAM1999 tools are intended to be used either in stand-alone, self-learning mode, or as part of a tutored environment, and will incorporate multimedia and telecommunications features. The student will be able to interact spontaneously with the system in realistic situations, via innovative components based on interactive video and speech recognition technologies. The consortium is composed of major providers of speech and educational technologies, language teaching and training centres, and cultural institutions. The consortium will work in consultation with a user group, which will ensure that tools match user needs, and an industrial group which will investigate market opportunities and commercial prospects.

Coordinator:	ELEMEDIA - Mr Marcello FRANZITTA I - 80078 Pozzuoli (Napoli) - Via Campi Flegrei 34 Tel +39-2-87 84 85 Fax +39-2-86 56 82
Partners:	DRAGON SYSTEMS UK LTD. - UK BERLITZ SCHOOLS OF LANGUAGES GMBH - D AURALOG S.A - F THE BRITISH COUNCIL - UK GOETHE INSTITUTE. - D

2.5 REUSABLE LANGUAGE RESOURCES

POINTER

The project aims to facilitate the development and widespread availability of terminology data across Europe, and to foster co-operation and exchange of results between developers, holders and users of terminology. In order to attain these objectives, POINTER will specify the key characteristics of a European-wide collaborative infrastructure for terminology, and carry out identification, prioritisation and cost/benefit analysis of the actions required to build the recommended infrastructure. The emphasis is on meeting the needs of both intermediate and end-users across Europe, particularly in industry. The POINTER consortium includes a wide range of actors, to achieve a balanced coverage among different needs and perspectives in the multilingual terminology area. Participants include SMEs active in the field of consultancy and documentary engineering, providers of language services, associations of professionals in the terminology area, international clearing houses for information on terminological data and activities, research centres and academic institutions across Europe. The result will be a set of concrete recommendations that specify a common terminology infrastructure to support users throughout Europe in the creation, dissemination and (re-)use of multilingual terminology.

Coordinator: BJL CONSULT - Mr Corentin ROULIN
                B - 1160 Bruxelles - boulevard du Souverain 207/12
                Tel +32-2-6756068
                Fax +32-2-6756351

Partners: CL SERVICIOS LINGÜISTICOS - E
                DEUTSCHES INST. FÜR TERMINOLOGIE - D
                INST. NATIONAL DES TELECOMMUNICATIONS - F
                UNIVERSITY OF SURREY - UK
                INTERNATIONAL INFORMATION CENTRE FOR TERMINOLOGY - A

SPEECHDAT

Spoken language resources (SLR), for example annotated speech databases, pronunciation lexica and language models, are necessary both for present day speech driven applications, and for the development of a technology capable of handling real, spontaneous speech. SLR are costly to produce because high levels of skill and manpower are needed for each language. The goal of the project is to minimise these costs by creating an efficient infrastructure capable of ensuring the production of SLR for both research and product development, carrying out the distribution of such SLR taking into account quality control and reusability issues, and defining SLR for future research purposes. The above infrastructure is to be tested by launching, in collaboration with other relevant projects, an initiative aimed to bring production centres together and handle the distribution of SLR. The consortium is strongly positioned to contribute to this infrastructure and exploit its results in that it comprises leading European players in the field: telecommunications companies, speech technology providers and research centres.

Coordinator: SIEMENS AG - Mr Harald HOEGE
                D-81739 München - Otto Hahn Ring 6
                Tel +49-896 36 33 74
                Fax +49-896 36 49 802

Partners: ALCATEL ITALIA DIV. TELETTRA - I
                GEC MARCONI LTD. - UK
                JYDSK TELEFON - DK
                PHILIPS GMBH FORSCHUNGSLABORATORIEN - D

CENTRO STUDI E LABORATORI TELECOMUNICAZIONI - I
PORTUGAL TELECOM - P
VOCALIS LTD. - UK
UNIVERSITY COLLEGE LONDON - UK
LIMSI-CNRS - F
UNIV. AMSTERDAM, INST. OF PHONETIC SCIENCES - NL
UNIV. AUTONOMA BARCELONA, DEPT. FILOLOGIA ESPANYOLA - E
UNIV. MÜNCHEN, INST. FÜR PHONETIK & SPRACHLICHE KOMMUNIK. - D
UNIV. AALBORG, CENTER FOR PERSONSKOMMUNIKATION - DK
INST. NAT. POLYTECHN. GRENOBLE, INST. DE LA COMMUNIC. PARLEE - F
DEFENSE RESEARCH AGENCY - UK
INST. DALLE MOLLE D'INTELLIGENCE ARTIFICIELLE PERCEPTIVE - CH

PAROLE

The project aims to define and contribute to the establishment of a collaborative infrastructure which will stimulate, throughout the Union, the creation and reuse of harmonised textual and lexical resources and tools. The infrastructure aimed at will ensure the reuse of existing generic resources for written language, and will contribute to their development where such resources do not exist or are not yet widely available. These activities will be undertaken by an initial group of national language centres working in association with relevant research and industrial actors. A collaborative network will be set in place at the very beginning of the project, and will be progressively enlarged during the life-time of the project. It will rely, in the first instance, on resources the partners already have available as a result of their long standing participation in lexicon and corpus work. Users, data producers and industrial partners across languages will be widely and actively consulted throughout the duration of the project. Synergies will be established with other relevant European projects and with activities in non-EU countries. The consortium comprises companies, national language centres and research laboratories.

Coordinator: SIETEC SYSTEMTECHNIK GMBH - Mr T. SCHNEIDER
                 D - 13629 Berlin - Nonnendammallee 101
                 Tel +49-89-63 64 18 54
                 Fax +49-89-63 64 96 46

Partners: IST. DI LINGUISTICA COMPUTAZIONALE, CONSORZIO PISA RICERCHE - I
                 UNIVERSITY OF BIRMINGHAM - UK
                 DET DANSKE SPROG-OG LITTERATURSELSKAB - DK
                 INST. FOR DUTCH LEXICOLOGY - NL
                 REAL ACADEMIA ESPANOLA, INST. DE LEXICOGRAFIA - E
                 INSTITÜT FÜR DEUTSCHE SPRACHE - D
                 INST. NATIONAL DE LA LANGUE FRANCAISE - F
                 INST. D'ESTUDIS CATALANS, E
                 SPRAKDATA, DEPT. OF SWEDISH, UNIV. GÖTEBORG - S
                 SOCIETA GENERALE DI INFORMATICA - I
                 GRUPO CL SERVICIOS LINGUISTICOS S.A. - E
                 ALCATEL CIT, F