European Commission
Directorate General XIII

TELEMATICS APPLICATIONS PROGRAMME 1994-1998

Language Engineering

Call for Proposals

Excerpts from the Work-programme

5

Language Engineering

The aim is to facilitate the use of telematics applications and to increase the possibilities for communication in and between European languages by integrating new spoken and written language-processing methods. Work will focus on pilot projects that integrate language technologies into information and communications systems and services. A key objective is to improve their ease of use and functionality and broaden their scope across different languages. Support will be provided for developing reusable language resources. Pre-normative guidelines, assessment methods and performance analysis techniques will be developed and their use promoted.

Users

The first group of users consists of:

- Administrations, utilities and services of public interest, including the fiscal, social, health care and transport sectors and the associated professionals and businesses. Other public utilities include cultural institutions and operators, tourist development agencies, libraries, and museums.

- SMEs and large enterprises, in particular in the areas of manufacturing (eg the aerospace, automotive and pharmaceutical sectors) and services (eg the software, financial, publishing, news and information, education and training, and travel and tourism sectors).

The second group consists of:

- Business and professional users.

- The general public.

Nature of the work

The push for economic growth and social integration, the wish to eliminate factors tending to exclude people from the information society, and the growing interest in new ways of decentralising work and other activities of daily life are all leading to demands for language-aware products and services to be made available to business and professional users and to the general public.

The focus of the Language Engineering sector of the TELEMATICS APPLICATIONS Programme is pilot projects that integrate language technology and components into information and communications systems and services, and aim to improve their ease of use and functionality and broaden their scope across different languages.

Appropriate support will be provided for the development of reusable language resources and for focused language engineering research efforts leading to medium to long-term applications. In addition, pre-normative guidelines, assessment methods and performance analysis techniques will be developed, and their use promoted within language engineering projects.

Support for multilinguality and interactivity concerns all the RTD work in this sector, which will be closely coordinated with work in other sectors of this Programme, other ICT programmes and relevant EUREKA actions.

Research tasks

1. Pilot Applications

Rationale and Scope

Pilot projects are the focus of this action and will be a major means for stimulating collaboration between users, providers and researchers of language technology. The domains covered are **document creation and management, information and communication services, and translation and foreign language acquisition**. In each of these domains there exist many opportunities for innovative applications based on well-mastered language and ICT components, as well for leading-edge applications bringing together existing and novel technologies expected to reach a sufficient degree of maturity over the next few years. The usage situations addressed are the workplace, the home environment and the mobile user.

Where appropriate, project clusters will be formed. These clusters will bring together projects that either address a coherent user community or share a common technology base, with the aim of improving the cohesiveness of individual actions and fostering cooperation and cross-fertilisation across projects and disciplines. These clusters will help to achieve critical mass, maximise the impact of individual projects, and ensure the wider dissemination, acceptance and exploitation of the results.

• Document Creation and Management

OBJECTIVES: Document creation and management, and in particular document authoring, are expensive and time-consuming processes. The ability to produce high-quality documents faster and more economically is a competitiveness factor for private enterprises and a step towards greater efficiency for administrations and services of public interest. Likewise, there is a growing need in organisations to shorten time-to-market for products and services, which demands more effective handling and localisation of product documentation. Opportunities for significant improvements exist in several areas, including more efficient text input techniques; faster and more accurate drafting in a non-native language; improved checking of style, grammar and comprehensibility; more consistent use of technical language and terminology; better re-use of existing texts; and transparent access to in-house and remote document translation services.

On the other hand, the developing group- and teleworking models require appropriate support for the management of collectively produced documents and documentation. Similar requirements arise from the fast-developing information exchange between EU and Member State administrations and from the needs for personal information management systems, which are growing with the widespread availability of powerful and affordable laptops and home computers. Opportunities for more cost-effective document management based on content analysis exist in several areas, including automatic indexing and categorisation of documents; ability to locate and browse documents in large distributed repositories; accurate classification and routing of incoming messages; content extraction from text and summarising; improved technical term detection and extraction; and the automated generation of documents from database records.

Tasks

LE 1.1 AUTHORING OF OFFICE DOCUMENTS

Develop flexible and modular authoring environments for individual users, SMEs and operating units of larger organisations, supporting direct input such as dictation and providing advanced writing and proof-reading aids for foreign language authoring, extended lexical and terminological guidance, and access to machine-aided translation and other language processing facilities, including the ability to maintain parallel versions of the same document in different languages. The focus will be on generic office documents and single authors or small work-groups. These applications will typically run on low-cost platforms and will be compatible with popular office software, while exhibiting innovative features, such as voice commands and controls. They should provide the openness required to operate as *"clients"* of a broader, organisation-wide document authoring system.

LE 1.2 COLLABORATIVE AUTHORING OF TECHNICAL DOCUMENTS

Facilitate distributed document authoring by highly structured work groups through integrated document authoring environments. In addition to the facilities provided by applications developed under the previous heading, these environments will provide support for the specific needs of work-groups and teleworking. Document editing, storage and re-use functions will be based upon a tight integration of linguistic contents and structural and work-flow related information, and will support, amongst other things, content-based version control and the selective handling of document elements. Support will be provided for controlled language, including style guidance and specialised technical vocabulary and terminology. The focus will be on technical documents, such as product documentation, characterised by periodic revisions and for which parallel versions in several languages must be maintained. Interactive assistance for the effective use of controlled language by technical writers, with special attention paid to features such as the unambiguous use of words and standard grammatical usage, is envisaged..

LE 1.3 DOCUMENT MANAGEMENT

Facilitate distributed document access and management by developing applications designed to handle textual and multimedia information repositories such as those maintained by corporate documentation centres, online news providers, libraries and cultural institutions. Provide applications with advanced capability to browse, locate, retrieve and filter textual information, according to users' preferences, by means of linguistic and stochastic analysis methods. Improve automated indexing of documents, linking of related documents, and the cross-referencing of text, images and sound (eg voice annotations). Support more effective and reliable document capture, such as robust OCR and refined bit-mapped-to-coded text conversion.

LE 1.4 DOCUMENT INTERCHANGE AND MESSAGE HANDLING

Ensure more effective and selective handling of incoming and outgoing messages, and document attribution and distribution within organisations, by means of automated content analysis. Provide techniques for the automatic categorisation, summarising and dispatching of messages and texts retrieved from in-house and external sources, such as online information services and technical and scientific libraries.

LE 1.5 REPORT GENERATION

Provide support for the interactive generation of reports and other documents relevant for the business and professional user, through the extraction and combination of information from

different textual sources (eg articles and technical or scientific reference materials) and formalised database records (eg medical, administrative, insurance and bank records). Support will be provided for easy input and control (eg by voice), with user-friendly access and interface to databases, selection and extraction of appropriate information items, and summary generation.

•   **Information and Communication Services**

OBJECTIVES: The increasing availability of electronic information in a variety of forms requires better and more selective data access. The approach taken in this work-area is to develop advanced applications characterised by more intuitive language interfaces and content-based information analysis and filtering. Typical examples of services delivered in digital form which can profit from easier access include online directory and news services, financial transactions, home shopping and interactive travel planning. Public authorities and utilities wish to offer a more transparent service through natural and flexible access to information concerning procedures, tariffs, office hours, etc. Teleworking and group-working call for improved technology-mediated interpersonal communications, including electronic mail and teleconferencing. Major benefits can be expected from the provision of information and communication services with advanced spoken and written language capabilities, with support for multilingual multimedia interfaces to telematics systems and services, and provision for content-based information browsing and navigation. Multilingual products and services presuppose flexible and effective methods and tools for human-computer interface and software internationalisation and localisation, leading to faster and more widespread user acceptance. There is consequently a need to develop guidelines and techniques that help to achieve the maximum localisation of human-computer interfaces.

Tasks

LE 1.6   **TEXT AND VOICE-BASED COMMUNICATION SERVICES**

Electronic mail: Improve the quality and effectiveness of person-to-person communication within professional and business work-groups by developing electronic mail services supporting the creation and handling of multimedia messages (eg messages with voice annotations or encapsulated speech), including conversion from text to speech and vice-versa, and message structuring tools enabling facilitated translation and summarising based on message understanding. *Teleconferencing:* Explore the effectiveness of in-session provision of language-based facilities such as computer-aided translation and sub-titling for studio and desktop-based conferencing, thus broadening the accessibility and diffusion of such services in a multilingual environment.

LE 1.7   **ACCESS TO INFORMATION AND TRANSACTION SERVICES**

Information services: Provide easy access to specialised information (eg legal, financial, yellow pages) needed by large enterprises and SMEs, as well as general information (eg timetables, travel, tourist and cultural services, news) for the wider public, including powerful content-based browsing through large amounts of data. Voice-controlled access to, selection of and navigation through structured multilingual information from fixed and mobile terminals, outputting results through different means (eg voice, fax, screen display) will also be supported. *Transaction services:* Provide more advanced and natural interactivity for transaction services such as telebanking, teleshopping, tele-administration and telemarketing, through multimodal dialogue, including voice and text, for a broad range of users. Special attention will be paid to confidentiality issues, including automatic speaker identification.

LE 1.8   MOBILE INFORMATION AND COMMUNICATION SERVICES

Increase the usability and functionality of portable devices for business and professional users and the general public by extending current language capabilities of laptops, handheld computers and in-vehicle devices, in particular voice input-output, dictation and hand-writing, for eg electronic mail, audioconferencing and access to online information services. Support interaction with a variety of teleservices (eg voice dialling) and in-vehicle navigation aids, and provide built-in language facilities such as advanced dictionary and terminology look-up.

LE 1.9   INTERFACE AND SOFTWARE GLOBALISATION

Develop methods and tools ensuring the effective and timely design and development of multilingual human-computer interfaces to telematics applications through better internationalisation and localisation (ie globalisation) techniques. Due attention will be paid to technical requirements such as character sets (see also task TE 2.2 in the previous chapter covering the Telematics Engineering sector), complex layout and compound (ie multimedia) information, and ergonomic aspects such as online help and hypertext tutorials.

• Translation and Foreign Language Acquisition

OBJECTIVES: One of the major characteristics of the Union is undoubtedly the rich variety of culture and languages of its member states. A prerequisite for integration and more intense collaboration is the lowering of language barriers.

Translation is of critical importance in a multilingual environment. By its very nature, translation is of potential relevance in many phases of the document creation, management and distribution process, as well as for information access and communication in general. As such, it is an activity that cuts across most of the domains covered by the Language Engineering sector. Translation can be significantly improved through more effective support for professional and occasional translators, including transparent access to in-house and remote translation facilities, better support for the translation of technical documents, and speedier translation of typical business documents and correspondence.

In order to achieve the goals of more effective foreign language acquisition, there are opportunities for the more effective learning of lexical and grammatical knowledge; more efficient foreign language practice, including pronunciation; and more effective foreign language reading and writing aids.

Tasks

LE 1.10   TRANSLATOR'S TOOLBOX AND TRANSLATION AIDS

Integrate tools for professional and occasional translators into distributed document authoring environments with the aim of easing the translation process. Such environments will provide access to personal, organisation-wide and remote language databases, including dictionaries and domain-specific terminology; provide support for the translation of technical language and the re-use of translations, source- and target-language spelling, and grammar and consistency checking; and support direct input (eg dictation and intelligent OCR), as well as the ability to align and maintain parallel texts in different languages. The establishment of telematic links between authors, decentralised translation teams and translation bureaux will be addressed.

LE 1.11   MACHINE-AIDED TRANSLATION

Tune machine translation for effective use in enterprises and administrations, either for high-volume translation of stereotypical documents (eg contracts, tenders, patents) and technical

materials (eg maintenance manuals), or for high-speed information gathering from incoming messages or documents in foreign languages. The stress will be on process engineering, system integration and field validation, aiming at a seamless and cost-effective integration of in-house and networked machine translation facilities into the users' working environment and the organisations' information and communications infrastructure.

LE 1.12   TUTORED AND SELF-TRAINED FOREIGN LANGUAGE ACQUISITION

Develop innovative applications for foreign language learning for large enterprises and SMEs, administrations and course suppliers such as universities and language training centres, with support for the tutor and the learner, facilitating the acquisition of basic and advanced foreign language skills. The main emphasis will be placed on business and professional users and languages. Furthermore, provide natural language interfaces and embedded applications and resources (eg spelling and grammar checkers, pronunciation aids, and special vocabularies) enabling effective on-the-job learning of written and spoken foreign languages, with appropriate support for self-evaluation. RTD actions established under this heading will be carried out in close collaboration with the Education and Training sector.

---

2.   Re-usable Language Resources

OBJECTIVES: The main objective of the actions established in this work area is to ensure harmonised and balanced development and widespread availability of general-purpose, monolingual and multilingual language resources and development tools, for all Union languages. Due attention will be paid to coordination with national efforts, and to less-widely used Union languages. These actions encompass activities which due to their nature are relevant not only to Language Engineering, but to the whole of language processing industry and R&D community. Actions in this work area complement the development of application specific resources and are to be coordinated with work carried out in pilot projects.

Tasks

LE 2.1   CREATION OF RE-USABLE RESOURCES

Develop harmonised and re-usable language data, including corpora, lexica and terminology data banks, designed to meet clearly identified needs arising from a range of applications, and which are neither covered by existing resources nor by resources developed within specific pilot projects. Actions will be carried out in close coordination with national and industrial programmes, and due attention will be paid to adding value and generality to existing resources, and to the owners' rights. Language data produced under this heading will comprise basic reference materials for Union languages, as well as data used for systems training, testing and benchmarking purposes, and should be suitable for wide dissemination. Portable software tools will be developed to ease the conversion of existing data collections, the creation of new normalised data collections, the exploitation and interchange of large-scale re-usable resources, and to facilitate the related quality-control processes.

LE 2.2   COLLECTION AND DISTRIBUTION OF RESOURCES OF GENERAL INTEREST

Establish a collaborative network of resources management centres entrusted with the gathering of existing and newly created resources and tools of general interest, their documentation, quality assurance, packaging and distribution to research and industrial users through appropriate licence agreements. The participating centres will maintain electronic catalogues and repositories of

available resources, and establish appropriate links with national research centres and with major users of language data.

---

3. Language Engineering Research

OBJECTIVES: Research actions established under this heading will address topics which are relevant to the whole of Language Engineering and focus on problems common to a number of application domains, thus avoiding duplication of effort within pilot projects. At the same time they will ensure the provision of adequate language technology for the next generation of applications. Special attention will be devoted to system and software engineering aspects and to the re-usability of results across applications and languages. These actions will aim to establish best practices for the design and implementation of complex language processing applications, addressing in particular the problem of modularity and scalability, with a view to ensuring a smoother integration of language technology and components into telematics applications.

<div style="text-align: center;">

Tasks

</div>

LE 3.1 **ADVANCED MODELS OF INTERPERSONAL AND HUMAN-COMPUTER DIALOGUE**

Enhance language processing models for improved human and human-computer dialogue, based on appropriate concept and meaning representations independent of specific languages, paying appropriate attention to multilingual and multimodal aspects of communication, in limited domains and for specific tasks.

LE 3.2 **ADVANCED INFORMATION TECHNOLOGY FOR LANGUAGE PROCESSING**

Integration of speech and natural language: Enhance current language processing capabilities, moving towards multifunctional systems by integrating signal processing, speech recognition and synthesis, language analysis and generation, understanding and dialogue into global processes responding to the users' information and communication needs. Special attention will be paid to modular system architectures in order to reduce complexity and facilitate development and maintenance. Likewise, scalability will be enhanced in order to ensure a longer life-cycle of language components and technologies. *Automated acquisition of language models:* Progress towards automating the development of language-aware systems, aiming in particular at gradually inferring specific aspects of language modelling from the analysis of concrete usage situations, thus ensuring more adequate systems and improved user acceptance.

LE 3.3 **ADVANCED METHODS FOR LANGUAGE PROCESSING**

Robust language processing: Improve language processing techniques providing robust, accurate and efficient treatment of everyday language in a range of user situations, thus broadening the range of potential applications and the target user population. Robust language processing systems will support more natural user interaction, improved accuracy of interpretation and better real-time behaviour. *Free-text and spoken language translation:* Advance towards automated free text and spoken dialogue understanding and translation, based amongst other things on more advanced speech recognition and understanding capability, for a number of specific usage situations (eg teleconferencing). Due to the complexity of the problems addressed and the likely scale of investment, synergy and complementarity with other initiatives in this area (eg other ICT programmes, Eureka and national projects) will be sought.

4.   Support Issues Specific to Language Engineering

This section refers to activities specific to the Telematics for Language Engineering sector that are not covered by the actions described in the chapter on Support Actions (which deals with, for example, telematics watch and assessment, trans-sector consensus development and coordination, awareness, dissemination of results and promotion of telematics, international cooperation, and training). The activities described aim to provide focus and support for the sector actors.

Tasks

LE 4.1   PRE-NORMATIVE STANDARDS

Carry out pre-normative standards work aiming at the establishment of guidelines for the encoding and interchange of language knowledge and data, thus accelerating the adoption of widely recognised standards. Definition, experimentation and validation will be implemented through actions bringing together a broad cross-section of interested parties. Work will build on the results of previous and concomitant industrial and RTD projects, and appropriate links will be established with standards bodies, industrial and professional associations, and similar initiatives at international level. Exploratory actions on system-level interworking between modules and components of language processing applications will be carried out.

LE 4.2   ASSESSMENT AND EVALUATION

Develop and integrate assessment and evaluation methods into the RTD project life-cycle with a view to promoting innovation through qualitative and quantitative assessment. Aspects to be addressed include evaluation of functional and operational adequacy with respect to the user's requirements and working environment; progress assessment during the application development and maintenance cycle; and performance analysis and benchmarking of alternative approaches and technologies, in selected fields and for tasks of general interest.

LE 4.3   USER AND INDUSTRIAL INTEREST GROUPS

In order to further the deployment of language-aware telematics systems and services, special interest groups bringing together enterprises, industrial and professional associations, research centres and universities will be established, with the aim of contributing to individual projects and playing an active role in the validation of RTD results. These groups will play a fundamental role in ensuring the required level of concertation and cross-fertilisation across individual actions.

LE 4.4   APPLICATION COORDINATION STRUCTURE

A coordination and support structure will be set up to ensure that applications be developed based upon open industry standard platforms, and that all projects contribute to the definition of implementation guidelines aimed at ensuring portability, interoperability and cost-effective deployment and replication of RTD results, thus fostering the integration of national and EU language engineering actions into trans-European telematics applications and facilities. Appropriate links will be established with the Telematics Engineering sector.

LE 4.5   PROJECT EVALUATION CRITERIA AND METRICS

Define and integrate quality control and validation methods and procedures into the RTD cycle with a view to facilitating widespread acceptance and exploitation of results. Aspects to be addressed include approaches, methods and procedures to be used for a reasoned and consistent evaluation of Language Engineering actions along such dimensions as satisfaction of user

requirements, improvement of industrial competitiveness, development of language infrastructure, furthering of existing and emerging standards, contribution to innovation and technological development, and relative value with respect to comparable developments at international level.

LE 4.6 AWARENESS

Awareness activities will be undertaken to improve the user perception of the services which the deployment of language technologies can provide. They will include the organisation of workshops and conferences and the publication of newsletters and annual reports. The possibility of setting up national focal points in collaboration with relevant national bodies and other telematics sectors will also be examined.

LE 4.7 EXCHANGE OF RESULTS AND TRAINING

In order to ensure an adequate level of expertise and demonstrate the advancement of European projects, support will be granted for fellowships, specialist workshops and participation in relevant international events, with the aim of promoting contacts and exchanges between language engineers active in different environments, such as industry, research centres and universities. Where appropriate, support will be granted for the training of users involved in pilot projects. Adequate support will be provided for the inclusion of language engineering in BSc and MSc curricula.

LE 4.8 USER REQUIREMENTS AND MANAGEMENT STUDIES

Preparatory and prospective actions will be undertaken, including studies aiming at detailing user requirements in specific professional or business sectors, market analyses and surveys of new technology trends and business opportunities, analyses of the socio-economic and legal impact of new language technologies, products and services, and cost/benefit analyses and models.