

TECHNOLANGUE: A PERMANENT EVALUATION AND INFORMATION INFRASTRUCTURE

Valérie Mapelli, Maria Nava, Sylvain Surcin, Djamel Mostefa, Khalid Choukri

ELRA/ELDA

55-57, rue Brillat-Savarin, 75013 Paris, France
{mapelli, nava, surcin, mostefa, choukri}@elda.fr

Abstract

In France, Human Language Technologies (HLT) projects, though regularly supported by the French authorities, are usually limited in time and therefore do not offer long-term, continuous exploitation of their results. As an answer to these limitations and needs, the Technolangue programme, launched in 2002 by the three French funding bodies, namely the ministries of Research, Culture and Industry, aims at offering permanent infrastructures, allowing to capitalize on the results of industrial and academic R&D projects, both at national and international levels. The first section of this paper presents an overview of the Technolangue programme and some selected projects. In addition to its involvement in a number of LR projects, ELDA is coordinating two crucial projects in the Evaluation and Technology watch action lines of Technolangue, namely the organisation of an evaluation platform called EVALDA and the implementation of an HLT-oriented information portal known as Technolangue.net. As developed in this paper, the setting up of an evaluation infrastructure and the implementation of a HLT web portal both aim to fit one of the main objectives of the programme, i.e. the creation of permanent actions in the HLT field.

INTRODUCTION

In France, the language engineering field, now better known as Human Language Technologies (HLT), has been supported by the French authorities, in particular through the ministries of Research, Culture and Industry, for decades. Several actions are regularly supported by these three government bodies, in complement to the European programmes and other international funding bodies actions.

The HLT field, both in the written and spoken language areas, is considered as crucial for industrial and economical growth. Market globalisation, product and service localisation are worth-to-be-quoted frameworks particularly interested in HLT applications (e.g. with the use of internet applications and mobile phones systems). Beyond its technological, scientific and economical contribution, the HLT field is also considered as a crucial issue from a cultural and social point of view.

Though research and technology actions at the French ministries regularly support common industrial and research projects in the HLT field, such projects are usually limited in time and therefore do not offer long-term, continuous exploitation of their results.

As an answer to these limitations and needs, the Technolangue programme, launched in 2002 by the three French funding bodies (Research, Culture and Industry), aims at offering permanent infrastructures allowing to capitalize on the results of industrial and academic R&D projects, both at national and international levels.

This paper presents an overview of the Technolangue programme, focusing on the implementation of permanent evaluation and information dissemination infrastructures.

THE TECHNOLANGUE PROGRAMME

Objective

Technolangue is a French national programme on language technologies supported by a common action from the French ministries in charge of Research, Industry and Culture. The main objective of this action is to implement a permanent infrastructure for the

production and distribution of language resources, for the evaluation of text and speech technologies, for the contribution to national and international normalisation and standardisation activities, and for technology watch in the HLT field.

A call for proposals was launched in April 2002. It consisted of 4 separate action lines: Language resources and basic tools, Evaluation, Norms and standards, Technology watch.

1) Language resources and basic tools

This action line aims at reinforcing the infrastructure for the production, validation and distribution of language resources. These resources may consist of written or speech data, with possible added value (e.g. annotation at different levels), such as text corpora, electronic dictionaries or terminology databases. Such data are essential to develop and test systems. This action line is also dedicated to language processing tools or evaluation tools. This action takes into consideration technology and content issues, and it addresses economical, legal and distribution aspects.

2) Evaluation

This action line aims at developing an evaluation infrastructure which should take into account technologies for spoken (speech recognition, speech synthesis, speaker recognition, audio/video information retrieval, etc.) and written language processing areas (information retrieval, automatic summarisation, machine translation, etc.). It also focuses on the use of technologies in the framework of integrated applications or services, and research on evaluation methodologies (e.g. metrics) regarding the assessment of human language technologies performances.

3) Norms and standards

This action aims at facilitating the accession to knowledge on norms and standards developed in this field, by supporting the French participation in international forums and major standardisation bodies.

4) Technology watch

This action aims at setting up an intelligence watch portal offering information on the state of the art in HLT and on the use of human language technologies in several applications and worldwide.

A steering committee, consisting of representatives of the three national research and technological innovation networks (RNRT, RNTL and RIAM), and of representatives from the three ministries, was assigned the task to steer the programme and supervise the review of the proposals. The experts reviewed the 52 project proposals. The results were announced by the steering committee on 26 July 2002.

Technolanguage Projects

The Technolanguage steering committee ended up by selecting 27 projects out of 52 proposals. The projects' duration is between 24 and 36 months. The accepted projects include the participation of 35 industrials, 44 public research laboratories and 9 other organisations (associations, other institution categories).

For the "Language resources and basic tools" action 14 proposals were accepted. Among these proposals, 9 are dedicated to language resources, such as the production of a BLARK (Basic LAnguage Resource Kit) for the French language, including reference monolingual and multilingual corpora, sets of bilingual and multilingual dictionaries, a collection of children and adult recordings over the fixed telephone network for speech recognition. 5 projects are dedicated to tools, such as a free software platform for automatic speaker verification, and the development of text analysis tools (lemmatisers, guessers, taggers, parsers, etc.).

For the "Evaluation" action line, the proposal for an evaluation infrastructure was accepted and is being implemented by the organisation of 8 distinct evaluation campaigns. An evaluation campaign for speech verification technologies and a study on information retrieval systems were also selected.

As for the "Norms and Standards" action line, two projects were adopted, proposing a contribution to the definition of international norms and to the process of normalisation in the field of HLT.

The proposal for the setting up of an HLT information portal in the "Technology watch" action line was accepted.

In addition to its involvement in a number of language resources projects, ELDA is coordinating the evaluation platform and the Technolanguage portal. For the evaluation campaigns, a scientific expert is nominated to help streamline the technical work.

A complete presentation of the Technolanguage programme and related projects can be found at: <http://www.technolanguage.net>.

EVALDA: THE CREATION OF A PERMANENT INFRASTRUCTURE FOR HLT EVALUATION

The initiative

The aim of the EVALDA project is to establish a permanent evaluation infrastructure for both written and speech processing systems for the French language with a possible extension to the European scene.

A first step of the EVALDA programme is to set up evaluation campaigns for 8 language technologies for the French language and some associated languages when the scope is multilingual like in MT. The projects and related technologies selected for EVALDA are:

ARCADE II: Evaluation of multilingual corpora alignment systems (*Action de Recherche Concertée sur l'Alignement de Documents et son Evaluation*). New parallel corpora are currently being prepared.

CESART: Evaluation of terminology extraction systems (*Campagne d'Evaluation de Systèmes d'Acquisition de Ressources Terminologiques*). Specifications (tasks, protocols, corpora) have been produced.

CESTA: Evaluation of machine translation systems (*Campagne d'Evaluation de Systèmes de Traduction Automatique*). Corpora are now being collected and formatted for training and for two campaigns, while an integrated evaluation environment is being developed. The first evaluation campaign is expected to start by the end of Spring 2004.

EASy: Evaluation of parsers (*Evaluation d'Analyseurs Syntaxiques*). An annotation formalism has been agreed upon and a reference corpus is currently being tagged manually.

EQueR: Evaluation of question answering systems (*Evaluation en Questions-Réponses*). A general and a domain-dependent corpus are under construction and specifications are being prepared.

ESTER: Evaluation of broadcast news automatic transcribing systems (*Evaluation des Systèmes de Transcription enrichie d'Emissions Radiophoniques*). The first phase of the project is now ending, with the distribution of a training corpus and results from a dry run being collected.

EvaSy: Evaluation of speech synthesis systems (*Evaluation des Synthétiseurs de parole en français*). Protocols for this four-part campaign are currently under study. The preparation of a corpus for the first part (grapheme/phoneme translation) has started.

MEDIA: Automatic evaluation of man-machine dialogue systems (*Méthodologie d'Evaluation automatique de la compréhension hors et en contexte du Dialogue*). A protocol for the construction of a reference dialogue corpus has been agreed upon and corpus recording is well under way. Recordings are being transcribed and tagged.

In order to capitalize on the outcome of the project, the resources and tools produced within the campaigns are to be made available to external players, as evaluation packages available via the ELRA/ELDA catalogue, enabling them to reproduce the evaluation campaigns for benchmarking purposes.

This evaluation infrastructure is designed to assemble several components with an impact on evaluation in language engineering i.e. an organisational model, logistics, language resources, metrics, methodologies, evaluation protocols, scoring software, as well as major players within the field (scientific advisory committees, experts, project partners etc). On the one hand, it aims at capitalizing on the results of planned experiences and, on the other hand, at facilitating collaborative work and new evaluation actions. Capitalized results must allow to repeat experiments at any time.

A permanent evaluation service can be useful for several reasons:

- at a scientific level: evaluation works help promote technological advances and collaborative work within the scientific community;

- at an economical level: evaluation helps determining the pros and cons of individual systems in the framework of their main application domain; it helps promote technology transfer between research institutions and

industry and between industry and market;

- at an institutional level: HLT evaluation in France and Europe is not as regular as it can be in the US or Japan, where it enjoys institutional and financial supports, which highlight its importance.

Furthermore, the creation of a permanent infrastructure aims at filling the gaps. The EVALDA initiative inspires from previous efforts to offer a general evaluation framework and proposes:

- to promote data reuse;
- to work on the French language but also offer a multilingual approach, bearing in mind how to extend this approach to other languages;
- to start new evaluation campaigns, in order to develop alternatives to existing methods and metrics or approach unexplored evaluation domains.

Web pages dedicated to a general description of the campaigns and their participants are available from <http://www.elda.fr> and <http://www.technolangu.net>.

EVALDA structure

For the time being, EVALDA consists of a number of autonomous work packages with, on the one hand, a project dedicated to building the general infrastructure as illustrated in Figure 1, and on the other hand the setting up of distinct evaluation campaigns.

In order to make the project efficient, scale-economic and capitalizable, organisational and managing tasks are grouped as much as possible.

Each evaluation campaign has been designed independently from the others. However, the production of common resources is envisaged. Presently, several campaigns allow to set up such an evaluation infrastructure. These campaigns are dedicated to various topics, though corresponding to a coherent set, and which cover both the speech and written fields.

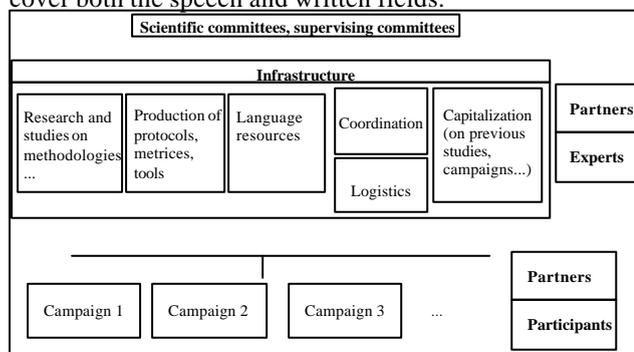


Figure 1: Organisational framework of the infrastructure for HLT evaluation

Most of the campaigns follow a « black box » protocol and quantitative evaluation methods, e.g. inspiring from what was done in previous evaluation campaigns.

Some of these campaigns propose original exploratory tasks, for instance:

- detection of parallelism breaks in corpora where language groups are quite different from French language (Arabic, Farsi, Chinese), in ARCADE II;
- user-oriented aspect in tasks of identification of named entities and theme indexing, in ESTER;
- use of a medical corpus, in EqueR;
- use of a new protocol for the recording of a reference dialogue corpus, through a « wizard of Oz » system, in MEDIA;
- investigation on three different metrics to be

applied during an automated evaluation process, including the famous NIST/BLEU metrics (as described by Papineni K., Roukos S., Ward T. and Wei-Jing Z., in *Bleu: a method for automatic evaluation of machine translation*. IBM Research Report RC22176. Yorktown Heights, NY: IBM, 2001), in CESTA.

Organisation of the campaigns

Setting up campaigns requires three steps. The first step consists of an exploration phase and the setting up of a general frame for the evaluation. This implies the following activities:

- state-of-the-art studies;
- definition of evaluation tasks;
- implementation of metrics and tools;
- specification or validation of adequate resources to be used within the envisaged experiments.

Step two corresponds to the definition and setting up of the campaigns. Each campaign consists of:

- a call for participation;
- a phase of training;
- a phase of testing, where results are made accessible only to the participants;
- a final workshop.

Step three consists of evaluating and capitalizing on the results of the experiment. This implies the following activities:

- publications;
- distribution of evaluation packages which include resources, protocols, scoring tools, results of the official campaign, etc., that were used or produced during the campaigns.

For ELDA, capitalizing on the knowledge acquired during the previous campaigns, means to be able to re-iterate an evaluation campaign on demand. Furthermore, as said above, ELDA envisages to distribute the resources produced or enriched during the campaigns, in evaluation packages, almost “at cost”, to allow the continuity of the resources and their access to the HLT community, so that each actor of the field can repeat the evaluation on a given topic.

THE TECHNOLOGUE.NET HLT PORTAL

Objective

The central objective of the Technolangu.net project is the development and maintenance of a portal in the French language, dedicated to HLT both in the written and spoken language domain. This portal aims at enhancing information dissemination in the HLT field, providing information on HLT community players, clarifying HLT issues and alerting the community to principal scientific, technological, industrial and normative developments.

The work, coordinated by ELDA, is carried out by creating a distributed network of web sites, composed of a central portal (<http://www.technolangu.net>) with its set of associated, specialised sites. This network will hopefully lead to the creation of a « community » of experts and users, French-speaking or not.

Technolangu.net is bound to become:

- a continuously informed and hierarchised state-of-the-art of the ever-growing HLT field, focusing, in

particular, on its improvements and limitations;

- a reference site on issues and specificities of the field, to help future producers make adequate choices corresponding to their needs;
- a resource centre which will present definitions, useful information, advice, technical expertise;
- a place of exchange where researchers, teachers, engineers, technicians, producers, journalists and users will converge on.

This portal will work on a mutualised effort from the project partners, which are well-known associations in their knowledge field (AFCP, AFNOR, APIL, ATALA) by linking not only the central site to their sites but also by cross-linking these sites on a “ring” principle, as illustrated in Figure 2.

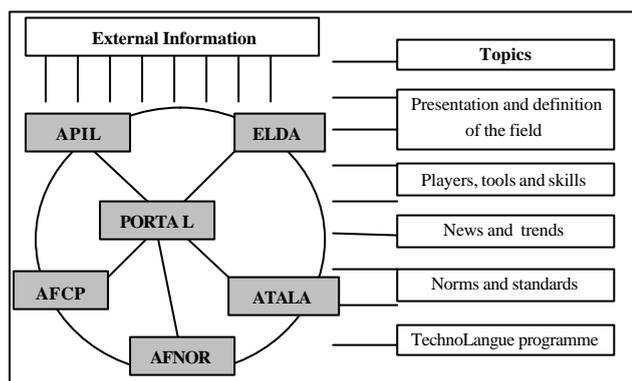


Figure 2: Technolange.net “ring” principle .

This portal presents validated and up-to-date sources of information. It addresses two types of users:

- current players of the HLT field who wish to obtain specific information;
- newcomers to the field, i.e. persons or organisations who wish to identify resources, products, technology providers and research centres or to acquire basic knowledge on the field.

Obviously, this portal will be open to everyone who wishes to know more about HLT.

The portal will cover several sectors and geographical areas from France perspectives (French-speaking countries, the European Union, the USA, Japan). Moreover, it will present the field of written and spoken language processing, current issues and applications, an overview of industrial and technological calls for tenders and proposals, etc.

Organisation

To implement the information portal, several actions must be undertaken:

- Analysis and collection of existing sources;
- Identification of shortcomings;
- Choice of original contents either on the portal site or on partners’ sites;

- Structuring of the portal and the “ring” of network sites;
- Design of the entire site and setting up of a minimum set of information;
- Test phase;
- Correction and content adjustment;
- Implementation of the on-line web site and launching of the network;
- Enhancing the value of the network.

On a technical point of view, the site is set up with a content management tool, namely SPIP, which allows an easy and user-friendly maintenance of the site, through an Internet interface.

On a content point of view, a first set of information has been completed. The web site was officially open on 24 November 2003, at the occasion of the LangTech 2003 Forum (see <http://www.lang-tech.org>). Visitors and participants at LangTech 2003 were given the opportunity to test the web portal and discuss with Technolange.net partners.

The current information sections give access to the following information in French:

- HLT: introduction of the field, methods & techniques, stocks and limits, key figures, case studies
- Overview: players of the field (companies, research organisations, associations), training, tools & resources, national & European initiatives
- News: Technolange news, calls for bids & proposals, events
- The Technolange programme: presentation of the Technolange programme and information on related projects.
- Norms & Standards: issues & institutions

CONCLUSION

As required within the Technolange programme, the setting up of an evaluation infrastructure and the implementation of an HLT web portal both fit one of the main objectives of the programme, i.e. the creation of permanent actions in the HLT field.

Indeed, creating a whole set of evaluation services as a permanent infrastructure should benefit from the organisation of specific campaigns in the first place.

The questions and problems encountered in the organisation of EVALDA campaigns, as well as the solutions found within this framework, will help consolidate the working model for the permanent evaluation service.

From the point of view of field information and monitoring, the Technolange.net portal will not only offer centralised information on the Technolange project on a short-term period but also a repository of information to all communities interested in HLT in the long term.