International Co-operation in the field of Language Resources and Evaluation Panel Session at LREC2000

Antonio Zampolli
Università di Pisa – Department of Linguistics
Istituto di Linguistica Computazionale del CNR - Pisa
intpan@ilc.pi.cnr.it

Lynette Hirschman
The MITRE Corporation -- Information Technology Center
lynette@mitre.org

Abstract

The goal of this Panel session is to discuss international co-operation in the field of Language Resources and Evaluation with particular reference to past and current experiences. The panel will represent the different scientific, technical, and strategical priorities of the major funding agencies, as well as their different organisational requirements. This will be followed by short presentations from groups involved in international cooperation, reporting on their experiences.

1. Background

The issue of international co-operation was extensively discussed at the first LREC in Granada (1998), with emphasis on the following issues:

- Language resources (LR) are essential components of HLT activity, supporting research, system development and training, and evaluation in both the mono- and multilingual context.
- A key enabling condition of integration of different technologies and languages requires that LR are shared among the different sectors and applications.
- The richness of the multilingual capabilities associated with a language depends on the number of languages for which adequate LR exist.
- The high cost and effort of the production of LR should be shared, in order to make them more affordable. The creation of multilingual LR requires agreement on a co-ordination policy, to ensure the reuse of existing monolingual resources and to facilitate access to native speakers of the various languages.

The situation in the field of evaluation is rather different in Europe and in the United States, where American and European expertise seem to be complementary. The question of co-operation in the field of evaluation therefore arises very naturally, in particular because many experts believe that is often only through such evaluations as TREC and MUC that research finds a common focus and makes easily quantifiable progress.

Two events of the first LREC have particularly stimulated discussion on these topics:

(1) the Panel on "Co-operation between EU and other countries in the field of Language Resources and Evaluation" (see A. Zampolli, "Panel of the Funding Agencies", in ELRA Newsletter, vol. 3, n° 3 (August 1998, Special Issue on 1st LREC);

(2) the Closing Session of the post-Conference Workshop on "Cross-lingual Information Management (see E. Hovy, A. Zampolli, "Governments: Policy and Funding", ch. 10, in E. Hovy, N. Ide, R. Frederking, J. Mariani, A. Zampolli, [Eds]., "Multilingual Information Society: Current Levels and Future Abilities", to be found at http://www.cs.cmu.edu/~ref/mlim/index.html).

The following areas of Language Technology emerged in the Granada debates as being in urgent need of international co-operation:

- Standards: de facto, best practices
- Language Resources and Related Tools
- Core Technologies
- Evaluation
- Selected vertical sector domains

These were endorsed in the session dedicated to HLT at the International Conference on "New Vista in Transatlantic Scientific and Technical Cooperation," organised on the occasion of the signing of the transatlantic technical and scientific co-operation agreement (Washington DC, June 1998).

2. Objectives of the Panel

The panel will explore different ways in which the international community can cooperate in creating and sharing resources for language processing.

We will discuss the needs and the possible priorities of international co-operation in the short and medium term. We will focus on the various phases of LR design, production and distribution, and their relationship to evaluation. The panel will consider the current policy and programmes of international and national funding agencies, reviewing a range of strategies including jointly funded projects, such the EU Fifth Framework Projects, the joint EU-NSF projects, the recently established North

American – South American collaboration, and the possibility of collaborative efforts with Asian countries.

In addition, the panel will also explore less formal approaches to resource sharing, including sharing of tool sets, sharing of data resources, and sharing of evaluations.

The goal of this session is to make participants aware of the various formal and informal strategies for resource sharing, to learn from participants what the obstacles are, and what strategies have been successful, and to discuss priorities and collect suggestions for the future.

The outcome of the discussions will be published as a report in the ELRA Bulletin and on the web; the report is intended as a contribution to the formation of new co-operation initiatives and as feedback and input for the funding agencies.

3. Overall Structure of the Panel

PART 1. PANELISTS: "The big picture" (50')

A. Zampolli (UPI, ILC-CNR)
R. Cencioni (EC-DGXIII)
G. Strong (DARPA)
L. Hirschman (MITRE)
J. Tsujii (University of Tokyo)

PART 2: DISCUSSANTS (tentative list to be confirmed and completed) (50'):

J. Mariani (LIMSI-CNRS, Paris)
H. Hoege (SIEMENS, Munich)
N. Bel (gilcUB, Barcelona)
L. Levin (CMU, Pittsburg)

PART 3: GENERAL DISCUSSION (30')

4. Post-Panel Discussion

Those wishing to further contribute to the discussion, for example reporting on experience of international co-operation, highlighting general or specific needs, suggesting priorities, or commenting on policy and organisational problems, are invited to send messaged to the discussion list intpan@ilc.pi.cnr.it. If appropriate, we will channel comments and suggestions to the relevant funding agencies.