MESSAGE OF THE CO-CHAIRS BENTE MAEGAARD AND KHALID CHOUKRI

Khalid Choukri, ELRA/ELDA – Evaluations and Language Resources Distribution Agency, France	Bente Maegaard, CST – Center for Sprogteknologi, University of Copenhagen, Denmark

Dear Participant and Dear Reader,

Welcome to this Second International Conference on Arabic Language Resources and Tools, which is organized by the MEDAR project (Mediterranean Arabic Language and Speech Technology), a follow-up of the NEMLAR project (Network for Euro-Mediterranean LAnguage Resources) and which takes place again in this fabulous city of Cairo, Egypt.

The NEMLAR project was funded by the European Commission (2003-2005) with partners from the European Union and from Arabic-speaking countries in the Mediterranean region. The project aimed at surveying the state-of-the art of Language Resources and Tools for Arabic in the region, at developing a Basic Language Resource Kit (BLARK) definition and specification for Arabic, and at starting development of language resources or updating of existing language resources. The project achieved its major goals, in particular it identified the requirements of Human Language Technology (HLT) players regarding Language Resources and created a number of language resources that are made available via ELRA (the European Language Resources Association) and widely used within the R&D community. The project gave some visibility for Arabic language technology, through networking, through a web site, a newsletter and through the first edition of this International Conference.

The other achievements of the NEMLAR project are related to the strong networking activities it managed to trigger. The NEMLAR core network, consisting of the 14 project partners, has proven to be very well suited for the tasks defined within the project but also acted as an important vector to convey the HLT messages within the Arabic world. The NEMLAR project and consortium has extended its network in order to promote the NEMLAR ideas and to give more people access to information about Arabic language technology and shareable Arabic language technologies resources.

In order to understand better the situation within the Arabic region, NEMLAR conducted a couple of surveys on Arabic Language Resources and Tools in Mediterranean Countries that led to an accurate overview of existing Arabic LRs and tools in the region. As a derivative of this survey, a list of institutions and companies involved in Language Technology Development and the production and distribution of LRs and tools has been made. Another survey managed to gather information about the needs and requirements of Industrial players for Language Resources and Tools. Such requirements were used to prioritize the development of LRs within NEMLAR.
When the NEMLAR project ended in 2005, a very good development had begun: Experts were identified, working methodologies were exchanged and refined. A few language resources were produced, but in order to have a major impact on the availability of Arabic language resources and tools, the collaboration had to continue and even extended. The NEMLAR network had continued to exist beyond the project, and some of its members managed to continue the organization of regular events related to Arabic HLT, such as specialised workshops (i.e. organized as a satellite event to LREC 2006).

The goal of the new project **MEDAR**, also supported by the European Commission's ICT programme, is to go beyond the achievements of NEMLAR. MEDAR is structured around 3 pillars, 4 main objectives, and a number of instruments.

The 3 pillars are:

1) Producing a knowledge base on Human Language Technology (HLT) players, existing language resources (LRs) and processing tools, activities and products for Arabic,
2) Designing a strong cooperation roadmap between the EU and Arabic countries, within the Arabic countries, and between academia and industry,
3) Focusing on Machine Translation (MT) and Multilingual Information Retrieval (MLIR) for which required technology components, LRs, and benchmarking methodologies will be identified, customized and distributed as open source kits.

The 4 objectives are:
1) Consolidating the NEMLAR network of players in all areas of HLT,
2) Developing the Cooperation Roadmap based on a clear picture of the foreseeable technological trends, market potentials, and cooperation possibilities. Human resources will be at the heart of our recommendations. Information will be disseminated through e.g. website and e-newsletter, and this major conference on LRs, tools and Evaluation for Arabic.
3) Updating the Basic Language Resource Kit: the minimum set of resources and tools necessary for carrying out research and training on LRs and HLT, with a focus on MT and MLIR, and
4) Supporting the development of tools and resources for MT on the basis of partners' technologies and open source code (e.g. Statistical MT, MLIR, and speech recognition) and the set up of a framework for their benchmarking.

A variety of instruments are being used to achieve the objectives: Surveys have been conducted as well as market analysis. For this purpose, partners' background has been exploited to identify the best Machine Translation open-source tools to customize to Arabic. Expertise on evaluation within the consortium is used to create the right benchmarking framework.
The project key achievements will be highlighted for each strategic partnership in addition to the actions to be taken. Human resources will be at the heart of our recommendations. Information will be disseminated through e.g. website and e-newsletter, and this major conference on LRs, tools, and evaluation for Arabic.
The issue of HLT for processing Arabic language is now getting prominent; the lack, on the one hand, of resources, and, on the other hand, of real-world applications, highlights the need for improving R&D in this area and for promoting the use of HLT among the potential partners, in particular to safeguard some of the cultural heritages of this geographical area.
We hope that this Second edition of the International Conference on Arabic Language Resources and Tools will modestly contribute to this.

We hope that you will appreciate and get the many benefits of this event and that altogether, we will draw some useful conclusions.
On behalf of the MEDAR Consortium and on your behalf, we would like to warmly thank our colleagues from RDI, Egypt, the local team in Cairo, responsible for the practical aspects of this event. We also would like to thank all members of the Program and Scientific Committees for their valuable help to review the submitted papers and demos.

Let us take this opportunity to thank a number of institutions which have helped or contributed to the organization of this event, in particular the official sponsors, Microsoft Innovation Lab Cairo, ITIDA (Information Technology Industrial Development Agency), and our media sponsor Media International. We are also grateful to the institutions that are supporting the conference, in particular the Faculty of Computers and Information, Cairo University (FCI-CU), The European Language Resources Association (ELRA) and its operational body, the Evaluations and Language resources Distribution Agency (ELDA), and the University of Copenhagen.

Last but not least we are very grateful to the Egyptian minister of CIT, his Excellence, Dr. Tarek Kamel who kindly agreed to hold this conference under his auspices. This without any doubts recognizes the importance of Human language Technologies in the information society. We are also very grateful to the European Commission, which funds MEDAR, and that supports a large number of R&D projects devoted to Arabic ICT issues, paving the way to efficient and fruitful cooperation within the Mediterranean area.