

# Language Technology from a European Perspective

<b>Hans Uszkoreit, Valia Kordoni</b>	<b>Vladislav Kubon</b>	<b>Michael Rosner</b>	<b>Sabine Kirchmeyer-Andersen</b>
Dept. of Computational Linguistics Saarland University	UFAL MFF UK Charles University	Dept. of Computer Science and A.I. University of Malta	Dept. of Computational Linguistics Copenhagen Business School
D-66041, Saarbruecken, Germany {uszkoreit, kordoni}@coli.uni-sb.de	Prague, Czech Republic vk@ufal.mff.cuni.cz	Msida, Malta mike.rosner@um.edu.mt	Copenhagen, Denmark ska.id@cbs.dk

## Abstract

This paper describes the cooperation of four European Universities aiming at attracting more students to European master studies in Language and Communication Technologies. The cooperation has been formally approved within the framework of the new European program “Erasmus Mundus” as a Specific Support Action in 2004. The consortium also aims at creating a sound basis for a joint master program in the field of language technology and computer science.

## 1 European higher education: Erasmus Mundus

The Erasmus Mundus programme [1] is a co-operation and mobility program in the field of higher education. It aims to enhance quality in European higher education and to promote intercultural understanding through co-operation with non-EU countries.

The program is intended to strengthen European co-operation and international links in higher education by supporting high-quality European Masters Courses, by enabling students and visiting scholars from around the world to engage in post-graduate study at European universities, as well as

by encouraging the outgoing mobility of European students and scholars towards non-EU countries.

The Erasmus Mundus program comprises four concrete actions:

**ACTION 1 - Erasmus Mundus Masters Courses:** high-quality integrated courses at masters level offered by a consortium of at least three universities in at least three different European countries.

**ACTION 2 - Erasmus Mundus scholarships:** a scholarship scheme for non-EU-country graduate students and scholars from the whole world.

**ACTION 3 - Partnerships:** Erasmus Mundus Masters Courses selected under Action 1 also have the possibility of establishing partnerships with non-EU-country higher education institutions.

**ACTION 4 - Enhancing attractiveness:** projects aimed at enhancing the attractiveness of the European higher education.

## 2 LATER

One of the projects approved for funding (and the only one in the field of language technology) in the 2004 call is called LATER – **L**anguage **T**echnology **E**rasmus Mundus [2].

LATER falls under action 4 of the program and hence addresses the need to enhance the attractiveness of European higher education in Language

Technology and Communication (LCT). This need will be met through dissemination of the combined LCT-related expertise in of a consortium of Universities whose members are as follows

#### **Saarland University in Saarbruecken (CoLi)**

The Department of Computational Linguistics and Phonetics (CoLi) of Saarland University (co-ordinator) has an excellent international reputation for graduate training in Language Technologies, and for leading-edge basic research in this area. CoLi offers a new M.Sc. program in Language Science and Technology [3]. This is an active program of basic, applied and cognitive research, which combines with state-of-the-art facilities to provide students with a rich and stimulating environment for their research. Moreover, CoLi offers a European Ph.D. program in Language Technology and Cognitive Systems. In the past 15 years, CoLi has provided postgraduate research training to 100 early-stage researchers [4].

#### **Charles University, Prague (ÚFAL)**

The Institute of Formal and Applied Linguistics (ÚFAL) at the Faculty of Mathematics and Physics of the Charles University in Prague offers a five-year master program in Computer Science with several specialized branches. One of the branches of this program is the masters in Computational and Formal Linguistics [7]. It focuses mainly on the following four topics: formal description of natural language, grammars and automata in linguistics, methods of artificial intelligence in linguistics, as well as methods of automatic natural language processing.

#### **University of Malta (UoM)**

The Department of Computer Science and Artificial Intelligence at the University of Malta, established in 1993, teaches both Bachelors and Masters degree programs. The 4-year BSc. (Hons) scheme include several streams relevant to Language Technology including NLP and Computational Linguistics itself, Information Retrieval, Semantic Web, Internet and Agent technologies. The Department also runs a, one-year research oriented M.Sc. program [10]. The areas of specialization include the development of computational tools, techniques and resources for Maltese, the only semitic language to enjoy official EU status.

#### **Copenhagen Business School (CBS)**

The Department of Computational Linguistics is part of the Faculty of Modern Languages at the Copenhagen Business School. The Department is actively involved in research in the following four core fields: formal descriptions of the Danish language, modeling of knowledge relevant for LSP, LSP databases, and Machine Translation. Embedded in this context is the Master of Language Administration (MLA) [9] that the Department of Computational Linguistics of the Copenhagen Business School offers in co-operation with the University of Southern Denmark in Roskilde

### **3 Overall aims of the project**

The overall aim of the project is to export the common educational experience currently embodied within existing Masters programs of the consortium to scholars and students of non-EU countries.

This aim will be realized by several different classes of activity under the rubrics of (i) workshops (ii) distance learning tools and (iii) coordination of a common Master program. We discuss these in the following sections.

#### **3.1 Workshops**

One of the most important types of activities of the project is organizing workshops and courses both for students from non-EU countries and for their teachers. The effect of these events is at least twofold – the students from countries or regions which do not have an access to any higher degree education in LCT get a chance to broaden their perspective by listening to lectures of prominent scientists and lecturers. The courses will also help the consortium to establish better contacts with non-EU Universities, teachers, and students which will turn out to be invaluable when disseminating the common European Master program in Language Technology discussed further below.

Both ÚFAL and CoLi have a long tradition in respect of offering such courses to students from the broadest possible range of countries.

ÚFAL has devoted a huge effort in the past to raise funding for the organization, once or twice a year, of a series of lectures by prominent scientists and lecturers from all over the world. This series of lectures, the Vilem Mathesius courses [6], have become well-known, especially among the Central

and East European students of computational and general linguistics.

This year's course, held in March under the auspices of LATER, was able to support the attendance of 50 students from Russia, Ukraine, Albania, Bosnia, Serbia, Croatia and Georgia to lectures by prominent individuals including two ACL award winners.

At CoLi, the Computational Linguistics Colloquium is also a traditional event attracting the attention of both well-known lecturers and a number of master and postgraduate students from various countries. A second series of lectures in the frame of our project was held at the University of Saarlandes in Saarbruecken in January.

A third event, organized by the CBS, will take place in June. The first day consists of information seminar on content management and language technology to promote CBS' newly-launched International Master of Language Administration, whilst the second will be devoted to diffusion of a various issues connected to the Erasmus Mundus course.

Finally, a fourth event, in the form of a workshop with invited guest lecturers, is being organized at the University of Malta that will take place in September 2005. The theme of the workshop will be Machine Translation which is currently very topical given the newly-achieved official European status that the local language now enjoys.

### 3.2 Coordination of Masters Programs

A second important aim of the LATER project is the definition, coordination and implementation of an integrated European Masters Programme in LCT by creating a common basis that will appeal to both European and non-EU students.

The rationale behind the creation of such a programme is the assumption that LCT now occupies a central position in research and education in Europe, being a key enabling technology for numerous applications related to the information society, although the shortage of qualified researchers and developers is slowing down the speed of innovation in Europe.

The proposed programme addresses this shortage by creating a directed education and training opportunity for the next generation of LCT innovators in that will in turn bring educational, social and economic benefits. Some specific aims of

Erasmus Mundus are also addressed: European education in LCT will be promoted worldwide and its competitiveness increased, increasing at the same time the competitiveness of European IT industries, creating a multilingual information society that is accessible for all, and turning the "information overload" into a wealth of accessible and useful knowledge.

### 3.3 Distance learning tools

A third aim of LATER is the development of effective methods of hosting and integrating non-EU students, for example by developing distance learning tools and joint distance education modules, in order to facilitate outreach by online dissemination of courses. An example of such modules, as well as for computer-based tools, is being developed on the basis of the virtual courses CoLi has developed in the last 3 years in the framework of the MiLCA project (Medienintensive Lehrmodule in der Computerlinguistik-Ausbildung<sup>1</sup>).

We also plan to explore the use of collaboration technologies based on SITESCAPE [16], that have been developed at CBS for academic collaboration, for the management of certain aspects of the proposed Masters programme.

The fruits of various initiatives already under way at UoM will be exploited and extended during the life of the proposed course. These include interactive web based course delivery [13], just-in-time support based on P2P architectures [14], XML-based frameworks for online courses [15], the latter being developed within as a part of the Mediterranean Virtual University (MVU) EUMEDIS project [17].

## 4 Integrated European LCT Masters Programme

Whilst many agree with the above assessment of the importance of LCT, they disagree on the definition of "integrated course". Fortunately, we can turn to the comprehensive definition supplied by the EU call, the central element of which is "a jointly developed curriculum or full recognition by the consortium of modules which are developed

---

<sup>1</sup> for more see <http://milca.sfs.uni-tuebingen.de/index.html>.

and delivered separately, but make up a common standard Masters course.”

Again, some turn away in horror at the notion of a standard curriculum in this area, the claim being that there is already enough standardization in the world, so why add to it? The point is, *any* programme dealing with LCT has to address the fact that it is highly interdisciplinary, including, at the core, computer science, computational and theoretical linguistics, and mathematics, and at the periphery, a wide variety of other subjects including electrical engineering, psychology, cognitive science artificial intelligence etc.

With such a large number of disciplines involved, it is practically impossible for a single University to excel in *all* of them. However if more than one University is involved, various kinds of curriculum sharing can be envisaged and so a much higher level of coverage becomes entirely achievable.

Put another way, curriculum sharing, together with common admission and assessment procedures envisaged, allows delivery of a complex course to be handled by what is effectively a “superuniversity”.

#### 4.1 Integration in practice

To put this idea into practice we are proposing that students will get the chance to attend a two years’ master program at two universities chosen from a larger consortium, which is currently being put together. It includes the four original partners of the LATER project and the following new partners: University of Amsterdam (UvA) in the Netherlands, Free University of Bolzano-Bozen (FUB) in Italy, the Universities of Nancy 1 and Nancy 2 in France, Roskilde University in Denmark and Utrecht University in the Netherlands.

Studying in multi-national groups at two universities in Europe, with English as instruction language, accompanied by language classes in another European language, will contribute to the students’ preparation for the increasing globalization of science, commerce and industry. The course also will also prepare students for follow-up Ph.D. studies provided by the participating partners and others.

The proposed programme follows the Bologna model for higher education in Europe and com-

prises 120 ECTS<sup>2</sup> credits, 30 of which make up the Masters dissertation, and 90 of which are coursework credits structured as follows:

- Compulsory modules in Computer Science (28 ECTS)
- Compulsory modules in Language Technology (28 ECTS)
- Advanced modules in Language Technology, Computational Linguistics and Computer Science (34 ECTS)

Coursework is distributed over three semesters, while the dissertation is supposed to be completed in the fourth semester

It is important to underline that this structure permits a considerable degree of variation. First, a module might be “implemented” by different set of courses at different Universities. Secondly, the advanced modules are electives, based on the specific strengths in research and teaching of individual partner institutions. There is no requirement that the advanced modules offered by different Universities should coincide.

Let us now introduce individual modules in more detail. Parentheses indicate ECTS credits.

#### Computer Science Modules

The Computer Science Modules are as follows:

- **Logic, Computability and Complexity (≥ 9)**  
Topics: Logic & inference; Computability theory; Complexity theory; Discrete mathematics
- **Formal Languages and Algorithms (≥ 9)**  
Topics: Formal grammars and languages hierarchy; Parsing and compiler design; Search techniques and constraint resolution; Automated Learning
- **Data Structures, Data Organization and Processing (≥ 6)**  
Topics: Algebraic data types; Relational databases; Semi-structured data and XML; Information retrieval; Digital libraries
- **Advanced Modules and Applications (≥ 6)**  
Topics: Artificial Intelligence, Knowledge Representation, Automated Reasoning, Semantic Web, Neural Networks, Machine Learning etc. Students are expected to obtain at least 9 ECTS credits from each of the first two

---

<sup>2</sup> European Credit Transfer System: a standard measure that is used in Europe for comparing the size of courses.

modules and 6 ECTS credits from each of the remaining two modules.

### Language Technology Modules

The Language Technology Modules are these:

- **Foundations of Language Technology ( $\geq 6$ )**  
Topics: Statistical methods; Symbolic methods; Cognition; Corpus Linguistics; Text and speech; Foundations of Linguistics
- **Computational Syntax and Morphology ( $\geq 9$ )**  
Topics: Finite state methods; Probabilistic approaches; Formal grammars; Tagging; Chunking; Parsing
- **Computational Semantics, Pragmatics and Discourse ( $\geq 6$ )**  
Topics: Syntax-semantics interface; Semantic construction; Dialogue; Formal semantics
- **Advanced Modules and Applications ( $\geq 6$ )** Topics: Machine Translation, Information Retrieval, Speech Recognition, Question Answering, Psycholinguistics etc..

### 4.2 Main issues to be addressed

Although it was not explicitly mentioned in the previous text, the integration of existing master programmes is done exclusively pair-wise. The students can't study at three universities (although the rules of the Erasmus Mundus programme allow such triangular cooperation). The restrictions within our consortia go even further – the students do not have a free choice of a combination of any two universities from within the consortium, they must choose one of the pairs offered by the consortium.

The reason for such a restriction is pretty simple - it turned out that although all members of the consortia in principle provide education both in Computer Science and in Computational Linguistics, they differ in the balance between these two fields. Within the consortium, there are universities with a strong stress on a Computer Science courses, aiming at a complex education including the sound theoretical background in the field, while other universities offer a more practically oriented educational scheme, stressing the concepts attracting a wider audience, e.g. various types of web technologies, databases, data mining etc.

As a result of this, each university participates in an average of four bilateral partnerships. We think

that the fact that the consortium consists of universities which are not identical greatly increases the variety of options available. They have a chance to choose those universities which are best suited to their preferences whether these are in terms of subject area emphasis or geographical region.

The preparation of the integrated Master programme doesn't stop at matching the universities and lectures offered. Erasmus Mundus is not just a cooperation, it is really a completely new scheme which must also address practical issues as grades, examination procedures, admission procedure, tuition fees, defense of the thesis, local specialties existing at some partner universities etc.

The proposed Masters programme is something new. It is the first attempt to create a comprehensive Masters degree in this subject area that conforms to all the legalistic requirements of each participating University. Students completing the course will possess a Masters degree delivered by two of the participant Universities. This is in contrast to the existing European Master in Language and Speech [11], which is implemented through a certification procedure that does not replace any legal degree that a student may obtain from a University.

### 5 Conclusion

Although the process of establishing a new European Master programme in Language Technology was really very complicated, time consuming and painful, there are definitely already at this stage very positive results.

In order to submit a proposal, our consortium has managed to overcome all formal and structural differences among all partners, it has found a reasonable model of cooperation, it has developed a high-quality master programme open both to European and non-EU students.

The wide variety of modules and topics offered combined with a relatively high degree of freedom of choice for students allows for individual pairs of partner universities to promote those courses and fields in which they excel. The students are of course offered individual guidance from consortium members in order to allow them to identify that pair of universities which best suits their individual needs and preferences

The strategy we have chosen – the initial cooperation of a smaller consortium in the LATER project, promoting LTC education among the students from outside the EU and testing our ability both to offer a coordinated high-quality education and to attract a reasonable amount of interested students, has turned to be a sound one. It also helped to solve some issues in the larger consortium based on the experience from the smaller one.

## References

- [1] [http://europa.eu.int/comm/education/programmes/mundus/index\\_en.html](http://europa.eu.int/comm/education/programmes/mundus/index_en.html) (Erasmus Mundus web page)
- [2] <http://europa.eu.int/comm/education/programmes/mundus/projects/2004/47.pdf> (The description of the LATER project)
- [3] <http://www.coli.uni-saarland.de/msc/> (the MSc website at the University of Saarlandes in Saarbruecken)
- [4] <http://www.coli.uni-saarland.de/kvv/> (courses at the Dept. of Computational Linguistics at the University of Saarlandes in Saarbruecken)
- [5] <http://www.coli.uni-saarland.de/courses/late2/> (the web page of the Language Technology II course in Saarbruecken)
- [6] [http://ufal.mff.cuni.cz/vmc/vmc\\_ls20.html](http://ufal.mff.cuni.cz/vmc/vmc_ls20.html) (the web page of the Vilem Mathesius Lecture Series)
- [7] <http://www.mff.cuni.cz/toUTF8.en/studium/bcmgr/ok/ilb53.htm> (the master programme in Mathematical Linguistics at the Charles University in Prague)
- [8] [http://web.cbs.dk/stud\\_pro/clmdatauk.shtml](http://web.cbs.dk/stud_pro/clmdatauk.shtml) (the master program at the Copenhagen Business School)
- [9] <http://uk.cbs.dk/mla> (Master of Language Administration at the Copenhagen Business School)
- [10] <http://www.cs.um.edu.mt/research/pgEnquiries.html> (the master program at the University of Malta)
- [11] <http://www.cstr.ed.ac.uk/euromasters> (European Masters in Language and Speech)
- [12] A.Burchardt, S. Walter and M. Pinkal. 2004. "MiLCA -- Distance Education in Computational Linguistics". In Szucs, Andras and Bo, Ingeborg (eds.), *New Challenges and Partnerships in an Enlarged European Union – Proc. 2004 EDEN Conference, Budapest*, pp. 351-356.
- [13] Ellul, C., 2002, “Just-in-Time Lecture Delivery, Management and Student Support System”, BSc. Project report, Dept. CSAI, University of Malta.
- [14] Bezzina, R., 2002, “Peer-to-Peer Just-in-Time Support for Curriculum based Learning”, BSc. Project report, Dept. CSAI, University of Malta.
- [15] Cachia, E., and Micallef, M., forthcoming, “A Universal XML/XSLT Framework for Online Courses”, Proc. International Conference on IT-Based Higher Education And Training (ITHET), Dominican Republic.
- [16] [www.sitescape.com](http://www.sitescape.com) : SiteScape corporate website.
- [17] <http://www.eumedis.net/en/project/22>: Mediterranean Virtual University (MVU) description.