A Proposal for a Computational Lingistics Training Programme at the Faculty of Arts of the ELTE University, Budapest

Balázs Kis
MorphoLogic
kis@morphologic.hu

Ádám Kis
SZAK Publishers Ltd.
adam.kis@szak.hu

This paper presents a proposed MA programme to be established at the future Department of Information Sciences and Library Sciences at the Faculty of Arts of the ELTE University, Budapest. Establishing such a programme is made quite necessary by the fact that, while there are several training programmes in language technology at science or technology departments, in Hungary it is very difficult to find a 'classical' linguist or philologist who is properly prepared to use, or even create computational means of linguistic research or modelling.

The proposed computational linguistics programme is planned to be an MA programme, taking place in the fourth and fifth year of university education. This programme is preceded by a three-year prerequisite BA programme that aims at training so-called humanities computing research assistants. This programme is presented in detail in the other brief presentation by Ádám Kis.

In planning the curriculum for the computational linguistics programme, it is crucial to ensure that this will be compatible both with similar national and well-known European computational linguistics programmes. Bearing this in mind, we are relying on two existing training programmes when designing the curriculum: one is the Theoretical Linguistics programme operated by the Linguistics Institute of the Hungarian Academy of Sciences, and the other is the Humanities Computing masters' and PhD programme run by the Rijksuniversiteit Groningen, The Netherlands.

There are some fundamental subjects that are not taught within the programme itself, but are considered as pre-requisites to successfully complete the computational linguistics programme. These basic subjects include those of fundamental linguistics (such as phonology or syntax), basic mathematics (set theory, statistics, formal languages), and basic programming. They are either taught as part of the preceding BA programme, or are expected to be delivered by other humanities programmes, such as those for foreign languages. The Groningen programme is actually based on this two-level scheme.

Main topics of the MA programme include: computational language models; computational morphology and syntax (algorithms and data structures); corpus linguistics; creation and management of linguistic resources such as corpora and lexicons; semantical databases and ontologies; knowledge representation; searching

and information extraction; development of Web applications; machine translation and computer-aided translation.

The presentation includes an extract of a possible curriculum for the programme described above.

The authors would like to thank those who play crucial roles in the preparation work, namely, László Lóth, acting head of the Institute for Computing and Library Sciences at the Faculty of Arts of the ELTE University, Budapest; and Zoltán Bánréti, assistant director of the Linguistics Institute of the Hungarian Academy of Sciences, for their ideas, thoughts and support.