

National Masters School in Language Technology*

GSLT

May 19, 2009

Introduction

Sweden is a relatively small country and competence in language technology is spread over a number of academic institutions which have been gathered together in GSLT, the national graduate school of language technology. GSLT would now like to propose that a nationally coordinated “masters school” in language technology would offer the possibility of combining the advanced training competencies of various institutions, thereby creating a critical mass of students taken from a broad range of subjects. This will increase the possibility of attracting both national and international students to a greater degree than is possible with our current national graduate programme where commitment to a four year research degree makes it difficult to admit students who are unknown to us or who have not yet be able to show their research potential in this field through an undergraduate degree. Our proposal is that each collaborating site would have its own masters programme in language technology which would then draw on the resources, mainly courses, of the national school. In this respect the organization would be similar to that of PhD studies in GSLT where each student is registered and examined at her home organization but takes part in courses and other activities organized by the national school. It does not seem feasible that several universities in Sweden could individually attract the numbers of students needed to make a viable masters programme in this specialist area or provide the breadth that is necessary for an internationally competitive programme. We believe that a national masters school would solve this problem and allow individual universities to exploit their specialist

*Contact people: Lars Ahrenberg (Linköping), Robin Cooper (Göteborg), Olov Engwall (KTH), Torbjörn Lager (Göteborg), Beáta Megyesi (Uppsala), Arne Ranta (CTH)

profiles within the area while sharing national resources in a way that would make smaller numbers of students registered in the local masters programmes an economically feasible proposition.

Opening the Masters courses to students from other Bachelor's programmes (and countries) is expected to attract more students from a wider range of backgrounds than is currently possible with the current degree structure.

Overview

Name of the school The name of the school shall be Masters School in Natural Language Technology (Mastersskola i Språkteknologi).

Content and goals Natural Language is becoming more and more important in IT applications. Some key areas are:

- extracting information from documents on the web
- producing documentation to products and systems in a multilingual environment
- localization of software to different languages and cultures
- speech-based communication with computers and also small devices (e.g. embedded in the car or home)

The goal is to provide a combination of courses that make it possible for students with an appropriate background to specialize in natural language technology. The specialization can be profiled towards computational linguistics or computer science, and towards a career in industry or academic research. There will, however, be a significant number of common, obligatory courses.

Number of credits The school will support two year programmes which carry 120 ECTS credits.

Targeted students Courses will be taught in English and aimed at the international student community. The programme will be open to students with

a background in language technology/computational linguistics, computer science, cognitive science and, with appropriate prerequisites, general linguistics, modern (or ancient) languages, philosophy and mathematics.

Detailed description

Degree structure and course content

A typical curriculum that the school would support has four semesters of study each of which corresponds to 30 ECTS credits. The first three semesters will normally have four courses each carrying 7.5 ECTS credits. The final semester will be devoted to the writing of a thesis (30 ECTS credits). Students are required to take 90 ECTS course credits (45 ECTS credits at level 2, 15 at level 3) and 30 ECTS credits of thesis work. A typical recommended study programme would be:

Semester 1:

Natural Language Processing (national, level 1)

A natural language oriented programming course (local/national, level 1)

Advanced methods course (local, level 1)

Linguistic resources (national, level 2)

Semester 2:

Speech Technology (national, level 1)

Statistical Methods (national, level 2)

two more level 2 courses

Semester 3:

Two level 2 courses à 7.5 ECTS credits

One level 3 course à 15 ECTS credits

Semester 4:

Thesis (30 ECTS credits)

Prerequisites for the courses

Students with an undergraduate degree in language technology, computational linguistics, cognitive science or computing science would be eligible to apply for level 1 courses without further prerequisites.

Students from general linguistics, languages, philosophy, mathematics and natural sciences are eligible, with the additional prerequisite of 15 ECTS credits in programming and 7.5 ECTS credits logic and/or discrete mathematics.

Examination and forms of study

Teaching will be largely web-based using technology such as Marratech, Google applications and Second Life, and will, in addition to traditional lectures delivered on the web, consist of “hands-on” lab-based exercises and projects with a problem-based orientation. The examination of courses will be a mixture of traditional examinations, course papers, practical lab exercises and programming projects. There will also be opportunities for students to meet physically, normally at least for initial and final sessions of courses.

Connection to research

All of the departments involved in this school are actively involved in externally funded research at both the national and European level and regularly collaborate with each other on such projects. Good masters students will be encouraged to relate their work to these projects.

Market demand for graduates

Language Technology is a rapidly growing area which creates a demand for experts internationally in both industry and research.

Recruitment base

GSLT has an international recruitment basis, and aspirants for GSLT might well be interested in our Masters programme. There are several GSLT applicants each year who are not admitted into the PhD programme because there is only a limited number of places available even though they are well qualified for advanced work. There is a further number of applicants each year who show promise but who have too little training or for whom we have too little information (in particular for foreign students) for us to be able to commit to a PhD. Such students could well be admitted to the Masters programme with a view to continuing to the PhD.

Another potential source of masters students comes to us through our work on NGSLT. For the introductory GSLT courses the number of students coming from NGSLT is often far greater than those from GSLT. Often 10–20 students from around the Nordic area (including the Baltic States and NW Russia) attend GSLT courses. Some of these could be interested in a masters degree.

As a rough estimate, we could expect to get 15–25 students every year.

Organization and budget plan

The school will be administered by a board (*ledningsgrupp*) which includes at least one representative from each participating university. One of the members of the programme's board will be director of the programme and 20% of the director's time will be devoted to running the programme and advising students on general matters relating to the programme, which courses to take etc. The director will also be provided with administrative and systems assistance to support course and admissions administration, maintenance of the programme's website and systems administration associated with course software and the like (perhaps 20% of full-time for each of these functions). Funds for teaching will be administrated centrally in the school and distributed to those sites conducting the teaching. A possible model for this is that each participating university contributes those funds which would be assigned to the masters students they have registered (possibly a fixed sum per student per year so that each participating university contributes the same per year per student). Students should have their travel and living expenses covered when taking part in the programme's activities and also be provided with a laptop computer. If national funding is not available for this, a fixed sum per student per year should be contributed by the participating universities to a central

fund which is disbursed to students who need to travel. This will ensure that none of the participating sites are financially disadvantaged by geographical location.

The following budget is a rough estimate of approximate costs that could be involved in such a school.

Annual cost when the Masters school is fully operative assuming an yearly intake of 25 students

<i>Cost of instruction and instructors' travel and accommodation expenses</i>				
per course	80,000 SEK	number of courses	10 Total	800,000 SEK
<i>Student travel and accommodation</i>				
per student	4,500 SEK	number of students	25 Total	112,500 SEK
<i>Supervision of masters thesis paid locally</i>				
<i>Administration</i>				
Director, administrator, computer support, board travel				500,000 SEK
				1,412,500 SEK

Concrete plans for immediate future

Uppsala is committed to starting a masters degree in language technology in the autumn of 2010. This degree relies on the existence of national courses. A proposal for what should be place in Uppsala by the autumn of 2010 is given in Figure 1 showing planned dependencies on offerings of national courses.

GSLT has so far planned national course offerings up to Spring 2011 and they have been designed to harmonize with the plans for the Uppsala masters programme. Fig. 2 shows the planned national courses for Autumn 2010 and Spring 2011.

Considerations on the form of courses and course materials

We aim to produce courses of excellence which can be used not only by Swedish masters programmes but also by programmes in other Nordic countries and elsewhere. (Considerable interest has already been expressed by our colleagues involved in the Nordic Graduate School of Language Technology on collaborating with this.) The greatest practical problem we face is the travel involved for

National Master in LT	Weeks	BA in language technology ----- BA in general linguistics/modern languages/ancient languages AND 15 ECTS in programming and 7.5 ECTS in logic/discrete mathematics	BA in computer science
Semester 1	W 36-40	Computer Science [B] (Local)	Linguistics I 7.5 [B] (Local)
	W 40-44		Phonetics 7.5 [B] (Local)
	W 44-48		Grammar 7.5 [B] (Local)
	W 36-03		Natural Language Processing 7.5 [A] (GU)
Semester 2	W 04-13	Speech Technology 7.5 [A] (KTH)	
		Statistical Methods 7.5 [A] (UU)	
	W 14-22	Advanced Method Course 7.5 [A] (Local)	
		Project Work 7.5 [A] (Local)	
Semester 3	W 36-44	Language Resources 7.5 [A] (GU)	
		Machine Learning 7.5 [A] (UU)	
	W 46-03	Student's choice 15 (Local)	
Semester 4	W 03-22	Thesis Work 30 [A] (Local)	

B – Basic course

A – Advanced course

Local – Given by the university where the student is registered

Courses in blue – Given by the responsible university (GU/KTH/UU) on the national level (existing GSLT course)

Figure 1: Proposal for Uppsala's programme starting autumn 2010

Semester	Level 1	Level 2
Autumn 2010	NLP	Linguistic Resources Machine Translation Natural Language Generation Java Development for HLT
Spring 2011	Speech Technology	Statistical Methods Speech Synthesis Treebanks Lexical Semantics

Figure 2: Planned national courses 2010–11

students if we follow the GSLT model of courses with intensive periods of teaching in Gothenburg or elsewhere. It is also not so clear that intensive teaching periods are the most suitable form of delivery for masters students. We therefore aim to produce course materials that are flexible with respect to the extent that traditional classroom teaching is employed. We aim at the following:

- Course materials should be in English
- They will be made available on a website
- They will include a downloadable video presenting the course (perhaps a filming of a classroom instance of the course). In addition to being available on the national website, the video will also be published on iTunes U to increase visibility for the course.
- They will include a complete set of exercises, projects etc. for the course
- Also a complete set of course notes, overhead presentations used and bibliographical references
- The course should be in a form so that in principle it could be delivered by a local teacher (e.g. an advanced graduate student) who is not necessarily an expert in the particular subject matter of the course.
- The course website should name an expert or several experts (e.g. the course coordinator) willing to consult with local teachers if questions about the material should arise.
- Course materials should be reviewed by the expert at three yearly intervals to determine whether changes should be made to keep the course up-to-date.

In this way we hope to maximize the use of the expert knowledge we have without creating an impossible teaching situation for the experts.