

# Joint Report from the International and Swedish Advisory Groups

21 October 2004

## **1 Executive summary**

GSLT seems to be functioning very well with respect to its mandate, currently having an enrolment of 30 PhD students (and an additional 3 associated PhD students). The graduate school seems to address a real national need as evidenced by the ability to attract excellent students and the active participation of researchers of international stature. The advisory group members appreciated the open and constructive atmosphere at the meeting.

The advisory groups are a little concerned about the apparent lack of sufficient physical and computational infrastructure and resources available at some of the participating institutions, and recommends that care be taken to ensure that a minimal reasonable research environment (material and computation means) are available. We highly approve the laudable practice of giving updated notebook computers to all new graduate students.

The group is pleased that the merits of the school have been recognized with the assured funding through 2012 and that new students have been able to join the program in 2004. The shared opinion of the groups is that the graduate school should continue to focus on what it does well, that is ensuring a top-level graduate education in Language Technologies, and not divert too much effort on other (possibly quite interesting) activities such as industrial courses and post-doctoral programs at the current time.

## **2 Introduction**

A joint meeting of the GSLT international advisory and the Swedish advisory groups met on October 21st 2004 at Goteborg University. The participants were the members of international advisory group: Lori Lamel (LIMSI-CNRS, Orsay), John Nerbonne (University of Groningen) and Stephen Pulman

(Oxford University); the members of the Swedish advisory group: Johan Boye (Telia Sonera), Eva Ejerhed (Uppsala), and Jesper Högberg (Hewlett Packard). GSLT was represented by Lars Ahrenberg (Linköping), Robin Cooper (Göteborg), Joakim Nivre (Vaxjö) and as well as two students: Svetoslav Marinov (Skövde) and Gustav Öquist (Uppsala). In 2003, separate meetings had been held with the two advisory boards. Both boards found the joint meeting to be productive allowing an open exchange of information and ideas. We recommend that future meetings also be held jointly.

This report summarizes the items addressed at the meeting, and provides some recommendations to GSLT.

### **3 Curriculum**

Concerning the GSLT curriculum, two main items were discussed: potential changes in Sweden resulting from the Bologna reform, and some of the difficulties confronted with assuring a critical mass for the third level courses. Some ideas for improving the cross-fertilization between the linguistics and speech areas was also discussed.

Most of the discussion focused on the possible education program reforms as a result of the Bologna accord. The current situation in Sweden is a 4-year undergrad program (with an optional bachelors after 3 years) followed by a 4-year PhD program. The reform proposes to replace this with a 3-year undergrad + a 2-year masters + and 3-year PhD program. It appears that the most logical mapping would be for the first year of the new master's to essentially correspond to what is now the 4th undergraduate year (the current Swedish masters) and the second year to what is now the first year of the PhD program (for GSLT this really means the core curriculum). Concerning teaching and funding resources, the new Masters would imply participation from both the undergraduate and graduate schools. Note that the program should also cater for longer undergraduate education programs, e.g. engineering.

There appear to be several advantages of the revised program: 1) it would offer the possibility that GSLT can accept applicants for the Masters program with no further commitment to the student of funding for a Doctorate; 2) Admissions in the PhD program can be based on the students' performance in the Masters; 3) PhD students can be recruited from external Masters programs (there still will be a need to ensure the equivalence of external MS programs); 4) this revised program would facilitate the equivalence of the GSLT degrees with those offered in other European countries; and 5) it would allow students to make more informed decisions before committing to the entire four-year period. A similar "decision point" is available in other programs in other countries.

In order to permit students from a variety of backgrounds to acquire a common skill base (an advantage of this is students with a wide knowledge base are often very attractive for industry) it was suggested that GSLT offer conversion courses, perhaps by re-packaging some Masters components, such as those offered at Cambridge and Edinburgh.

There was a strong consensus from both advisory groups that under no conditions should the 8 year total program be reduced in time as a result of the Bologna accord reducing the PhD portion to 3

years.

Another general recommendation was that the 8 year program be roughly equally focused on language (both linguistics and speech) and technology (computer science).

The perception was that most of the courses are either mainly speech or linguistics and it would be nice to see more cross-fertilization whenever possible (such as for course on Linguistic Resources, Machine Translation). One possible means is to have guest lecturers present the material from the other domain.

Concerning the Level 3 courses, there are apparently some difficulties on deciding what these should be and getting enough critical mass for them. Some of the possible solutions discussed at the meeting were:

- proposing courses on basic methodologies that can be applied to speech and linguistics
- proposing courses that are task/system oriented, such as shared application prototype development which requires the collaboration of participants in different research areas
- proposing courses that are problem-oriented, inviting students to address a given difficult problem using the method of their choice
- carrying out common shared task development and evaluations using common corpora and methodologies
- facilitating the circulation of students, allowing them to spend time at other institutions, particularly for the Level 3 courses
- proposing that students study some hot research topics, which may require possible contact with external researchers at other laboratories
- the possibility of an MIT-like Area Exam which requires a critical analysis of published work in an area with written report and oral presentation to a small committee. This could be done individually or in a small group (2-3 students).
- pursuing a closer cooperation with Nordic Graduate school to provide a larger student base

#### **4 Information Dissemination and Relation to Industry**

The advisory groups recommend that the GSLT maintain its focus on graduate training (potentially open to industrial participants) but not to disperse their resources (human, support and financial) to provide specific programs for industry.

Though industrial courses may not be the appropriate channel, contacts with industry is encouraged in order to market GSLT and create awareness. Opportunities to obtain coverage in the media should be grasped.

Concerning the employment of an information officer, the advisory groups were under the opinion that this is a positive step, and consider that the information officer should be of service to the internal functioning of GSLT and as well as being responsible for the external dissemination of GSLT activities for both the industrial and research communities. One concrete suggestion was to include some research highlights on the web site. This should require only a minimal effort as the web pages should be proposed by the students, with the agreement of their supervisors.

Another relatively low-cost but potentially quite effective dissemination means is to create a group of Industrial Supporters, modeled after that of the Cambridge University Masters in Computer Speech and Language Technology program. The basic idea is to hold a annual one-day meeting during which the GSLT presents an overview of the program and some research highlights, and in the afternoon the industrial participants present some of their ongoing LT activities. This is attractive for the students to learn about what is going on in industry, and to forge potential links for joint projects and future employment possibilities. It is also quite attractive for the industrial representatives that get the opportunity to hear what is going on at the GSLT and at other companies. A proposal was that the GSLT approach SICS, CTT and KTH CID to try to come up with an initial list of industrial contacts. The web site could also provide a means for interested industrials to sign up for an Industrial Supporters mailing list.

## **5 Recommendations for the Annual Report and General Reporting Practices**

We also discussed the form of the Annual report and the general reporting practices. Some of the comments on the form of the draft Annual report circulated prior to the meeting are of course due to the preliminary stage of the report. It was felt that an effort should be made to improve the homogeneity of the descriptions of the student thesis subjects. Additional information that would also help place the each student's situation, specifically to provide the year of entry in the doctoral program and the year of the previous degree in addition to the already listed program, as well as the Assistant Supervisor (which is missing for some of the students).

It would also help if there were a table providing summary information of the students' progress, such as percent completion as a function of the entry year. General tendencies should be noted and explanations should be provided for slow/fast situations such as the student's background or the research environment (infrastructure, support) or other factors. The criteria used to evaluate student progress should be listed. The advisory groups suggested having a mid-stream progress review possibly with external reviewers (other than the supervisors) when an elaborated thesis proposal is available. This ideally should occur after the 2nd annual report by the student/advisor to minimize any additional overhead.

The report would also gain in value if historical summaries were added. It is difficult to see the development of GSLT over the past several years because two reports cover two semesters each while the last one covers three. It would useful to see summaries such as the number of students admitted, number of course offered (including initial enrollment and completion, as is presently reported), totals for all courses, number of degrees awarded and number of withdrawals, and perhaps useful alterna-

tive perspectives (e.g., perhaps speech vs. language). Naturally, this would require standardizing or normalizing the reporting period.

Although it is apparent that many of the recommendations and concerns expressed in the previous advisory group reports have been taken seriously by the GSLT, it would help us if a response to these points was addressed in the following annual report.

## **6 Other topics**

Some other topics were briefly discussed.

### **6.1 Nordic Graduate School of Language Technology (NGSLT):**

Concerning NGSLT, it was suggested that GSLT try to clarify the relationship and collaboration with it, especially in relation to the possibility for joint activities such as level 3 courses. There was concern that the collaboration may introduce complications in addition to the expected added benefits.

### **6.2 The GSLT Repository:**

Our impression is that the original vision of the GSLT Repository to provide resources available on a central GSLT server no longer seems to be feasible (that it is too costly and difficult to maintain). We still strongly encourage the sharing of tools and linguistic resources when possible, particular to support students at institutions that may not have means for infrastructure development. One possible relatively low cost option is to start by collecting and maintaining links to relevant sites with open-source or publicly available software and linguistic resources. The advice of the groups is that GSLT further consider the costs and benefits of such a repository (what can be done at a reasonable cost).

### **6.3 Postdoctoral program:**

The general consensus was unanimous that GSLT should focus their effort on the graduate education and not on post-doctoral formation.