Language Technology: Research and Development

Introduction

Sara Stymne

Uppsala University
Department of Linguistics and Philology
sara.stymne@lingfil.uu.se

Teaching Team

- Course coordinator, examiner and lectures:
  - Sara Stymne
- Seminars
  - Meriem Beloucif
  - Beáta Megyesi
  - Johan Sjons

Course Content

Theory	**Practice**
Philosophy of science	Survey a research field
Research methods in LT	Plan and implement a project
Scientific writing	Write and review scientific papers

- Lectures covering theory (large group)
- Seminars devoted to practice (small group)
- Individual projects on a common theme (small group)

Research Themes

- Digging the Past: Digital Philology and the Analysis of Historical Sources [Bea]
- Language and Neural Networks [Johan]
- Low-resource languages [Meriem]

Research Themes

- Digging the Past: Digital Philology and the Analysis of Historical Sources [Bea]
- Language and Neural Networks [Johan]
- Low-resource languages [Meriem]
- More about the themes later!

Course Structure

1. Background part:
   - Philosophy of science and research methods [lectures]
   - Survey of the state of the art in research theme [seminars]
   - Planning an R&D project [lecture, seminar]
2. Project part:
   - Implementing an R&D project [seminars]
   - Writing a scientific paper [lecture, seminars]
   - Reviewing scientific papers [lecture]

Reading List

- Textbooks:
  - Okasha, S. (2002) *Philosophy of Science: A Very Short Introduction*. Oxford University Press.
    - **Obligatory**: buy it as soon as you can (either 1st or 2nd edition)
  - Zobel , J. (2004) *Writing for Computer Science*. Second Edition. Springer.
- Papers:
  - Available online from the course home page https://cl.lingfil.uu.se/kurs/rd22/

Assignments and Examination

1. Take home exam on philosophy of science (15%) [written]
2. Research paper presentation and discussion (15%) [oral]
3. Project proposal (15%) [written, oral]
4. Term paper (40%) [written, oral]
5. Review of term papers (15%) [written]

Assignments and Examination

1. Take home exam on philosophy of science (15%) [written]
2. Research paper presentation and discussion (15%) [oral]
3. Project proposal (15%) [written, oral]
4. Term paper (40%) [written, oral]
5. Review of term papers (15%) [written]

- Pass (G) = all assignments passed
- Distinction (VG) = at least 50% of 1, 3–5 with distinction

Deadlines

Choose your preferred topic	September 1, 13.00
Hand in take home exam	September 14
Project proposal	October 6
Present project proposal	October 11
First version of term paper	December 13
Peer review of (other) term papers	December 22
Final seminar	January 12
Final term paper	January 13

Deadlines

Choose your preferred topic	September 1, 13.00
Hand in take home exam	September 14
Project proposal	October 6
Present project proposal	October 11
First version of term paper	December 13
Peer review of (other) term papers	December 22
Final seminar	January 12
Final term paper	January 13

Backup deadlines available on course web page, but important to try to respect original deadlines! (This course is a prerequisite for the master thesis course.)

Seminars

- All seminars are obligatory!
- Group seminars:
  - Research papers
  - Project proposal (presentations with slides)
  - Progress reports (including ethics)
- Final seminar in full group
  - Full day "mini workshop" (on Campus)
  - Social event (if possible)

Seminars

- All seminars are obligatory!
- Group seminars:
  - Research papers
  - Project proposal (presentations with slides)
  - Progress reports (including ethics)
- Final seminar in full group
  - Full day "mini workshop" (on Campus)
  - Social event (if possible)
- If you miss a seminar, there will be a compensation task

Going for the Real Thing

- The goal is to do real research resulting in real publications
- Guidelines for submission and reviews:
  - Transactions of the Association for Computational Linguistics
    http://www.transacl.org/submission/
- Term papers may be revised and submitted for publication
- Actual submission is not a course requirement

Going for the Real Thing

- The goal is to do real research resulting in real publications
- Guidelines for submission and reviews:
  - Transactions of the Association for Computational Linguistics
    http://www.transacl.org/submission/
- Term papers may be revised and submitted for publication
- Actual submission is not a course requirement

- You are meant to function as a real research group
- Projects are individual, but you should support each other

Some Publications from Recent Years

Rafal Černiavski and Sara Stymne. *Uppsala University at SemEval-2022 Task 1: Can Foreign Entries Enhance an English Reverse Dictionary?*. SemEval-2022.

Antonia Karamolegkou and Sara Stymne. *Investigation of Transfer Languages for Parsing Latin: Italic Branch vs. Hellenic Branch*. NoDaLiDa 2021.

Harm Lameris and Sara Stymne. *Whit's the Richt Pairt o Speech: PoS tagging for Scots*. VarDial 2021.

Sebastian Reimann and Daniel Dakota. *Examining the Effects of Preprocessing on the Detection of Offensive Language in German Tweets*. KONVENS 2021.

Marsida Toska, Joakim Nivre and Daniel Zeman. *Universal Dependencies for Albanian*. UD workshop 2020.

Learning Outcomes 1

The student should at least be able to do the following, in relation to a scientifically organised language technology project:

- explain the basic principles of scientific work and research methodology in general and in relation to a current project
- make an overview of earlier research and the state of the art within the field that the project treats and identify its most urgent research issues,
- show an ability to identify and formulate research questions in a critical, independent, and creative way

Learning Outcomes 2

The student should at least be able to do the following, in relation to a scientifically organised language technology project:

- plan and carry out research tasks based on sound methodological principles and within given time limits,
- evaluate results and partial results with current validation methods,
- present the purpose of the project and its results in a professional manner, both for scientists and for the general public, orally and in writing, taking the target audience into consideration.

Student Feedback

- 2021 students were very happy with the course: (4.3/5)
- Some comments:
  - Seminars in small groups were good
  - Good to discuss each other's work at seminars and in reviews
  - Good to mix lectures and seminars
  - The literature seminars were helpful
  - Good with hybrid lectures
  - The communication and interaction between students could have been better
  - Too many students in each group
    Smaller groups in 2022
  - It would have been good with more concrete project ideas
- Two new teachers and themes this year!

Teaching mode

- We will mainly have Campus teaching
  - Seminars and lectures will be on campus
  - Some lectures may be hybrid (if you want that)
  - The occasional activity may be on Zoom due to teacher availability (like today)
- Campus activities may be cancelled on short notice
  Check your email+Studium before going to Campus!
- We will follow current regulations

Computing resources

To implement your project, you will need a varying amount of computing resources

- Your own computers
- Our Linux system
  - Smaller needs
  - kasus and tempus available
- UPPMAX
  - Larger needs, GPUs and CPUs
  - Instructions and project number on the course web page
  - Introductory lecture about UPPMAX and visit to computer halls:
    September 13, 13-15: Ångström, Polhemssalen 10134

Course information and assignments

- Main information on course web page:
  `https://cl.lingfil.uu.se/kurs/rd22/`
  - Annotated schedule
  - Course information and instructions
- Studium
  - Handing in most assignments
  - Grades
- External systems for reviewing (more info later)

Coming up

- Now: introduction to the topics
- Wish for your preferred topics
  - Rank your preference for the 3 topics, plus indicate if your first choice is a strong preference
  - By email to Sara: deadline Thursday September 1, 13.00
- Lecture on science, research and NLP: Friday
- Debates on philosophy of science and NLP: next Monday
- First research paper seminar: September 7
- Take home exam: September 8-14
- UPPMAX lecture: September 13

Research Paper Seminars

- Obligatory attendance
- All students are expected to have read all articles, to bring discussion points, and actively discuss the articles
- Each student is responsible for introducing one article each
  - briefly summarize the paper (MAX 2 min)
  - discuss the main points being made
  - bring up difficult to understand parts
  - initiate a discussion by proposing themes to discuss
- Bring the articles to the seminars (on paper or electronically)
- The list of articles and presenters will be available on the web page by Friday

Take home exam

- On philosophy of science, based on Okasha
- Distributed in Studium on September 8
- Due September 14, 23:59
- Sign up in Ladok, will be done automatically for everyone we register
- Anonymous, write your Ladok code in your document
- Individual, no collaborations

Questions?