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
  - Mats Dahllöf
  - Joakim Nivre
  - Sara Stymne

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

- Universal Dependencies [Joakim]
  - Cross-linguistically consistent treebank annotation
  - Used for research on parsing, tagging, linguistic typology etc.
- Multilinguality [Sara]
  - Multilingual systems involve more than one language
  - Useful especially for low-resource languages
- Author identification and profiling [Mats]
  - Who has written a text?
  - What kind of person has written a text? Age, gender, personality, mental health, economy, etc.

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.
  - Zobel , J. (2004) *Writing for Computer Science*. Second Edition. Springer.
- Papers:
  - Available on line from the course home page https://cl.lingfil.uu.se/kurs/rd19/

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 6, 13.00
Hand in take home exam	September 20
Project proposal	October 4
Present project proposal	October 8–9
First version of term paper	December 13
Peer review of (other) term papers	December 20
Final seminar	January 15
Final term paper	January 17

Deadlines

Choose your preferred topic	September 6, 13.00
Hand in take home exam	September 20
Project proposal	October 4
Present project proposal	October 8–9
First version of term paper	December 13
Peer review of (other) term papers	December 20
Final seminar	January 15
Final term paper	January 17

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
- Final seminar in full group
  - Full day "mini workshop"
  - First-year master students also invited
  - Social event

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

Publications from Recent Years (1)

Artur Kulmizev, Miryam de Lhoneux, **Johannes Gontrum, Elena Fano** and Joakim Nivre. Deep Contextualized Word Embeddings in Transition-Based and Graph-Based Dependency Parsing – A Tale of Two Parsers Revisited. *The Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*.

Ailsa Meechan-Maddon and Joakim Nivre. How to Parse Low-Resource Languages: Cross-Lingual Parsing, Target Language Annotation, or Both? *International Conference on Dependency Linguistics (Depling 2019)*.

Gabi Rolih. K-means Clustering for POS Tagger Improvement. *Language Technologies and Digital Humanities 2018*.

Publications from Recent Years (2)

Gongbo Tang, Fabienne Cap, Eva Pettersson and Joakim Nivre. An Evaluation of Neural Machine Translation Models on Historical Spelling Normalization. *The 27th International Conference on Computational Linguistics*.

Allison Adams and Sara Stymne. Learning with Learner Corpora: Using the TLE for Native Language Identification. *The Joint Workshop on NLP for Computer-Assisted Language Learning and NLP for Language Acquisition*.

Joakim Nivre and **Chiao-Ting Fang**. Universal Dependency Evaluation. *NoDaLiDa Workshop on Universal Dependencies*.

Rebeca Padilla López and Fabienne Cap. Did You Ever Read About Frogs Drinking Coffee? Investigating the Compositionality of Multi-Emoji Expressions. *The 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

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

- 2018 students were very happy with the course: (4.6/5)
- Some comments:
  - The course was very useful in preparation for the master's thesis
  - The workload was heavy at the end of the course
  - Stressful to have work to do over Christmas
  - Do not split into topics so early
  - Seminars were useful
  - Doing a full project from proposal to reviewing was very useful
- Redistributed the workload 2019, to allow more time for projects
- Partly new topics and teachers 2019

Student Feedback

- 2018 students were very happy with the course: (4.6/5)
- Some comments:
  - The course was very useful in preparation for the master's thesis
  - The workload was heavy at the end of the course
  - Stressful to have work to do over Christmas
  - Do not split into topics so early
  - Seminars were useful
  - Doing a full project from proposal to reviewing was very useful
- Redistributed the workload 2019, to allow more time for projects
- Partly new topics and teachers 2019
- All students in 2018 passed the course on time!

Coming up

- Now: introduction to the topics
- Choose your preferred topics
  - By email to Sara: deadline Friday Spetember, 13.00
- Lecture on science and research: tomorrow
- Lecture on language technology R&D: next week
- Research paper seminars: next week

Research Paper Seminars

- Obligatory attendance
- Each student is responsible for introducing one article each
- All students are supposed to have read and actively discuss all articles
- Bring articles to the seminars (on paper or electronically)
- The list of articles and presenters will be available on the web page early next week

Questions?