Research seminar week 1

Tamás Biró
Humanities Computing
University of Groningen
t.s.biro@rug.nl

Aim of the research seminar

Skills to carry out a research (MA thesis). Emphasis on research using computer simulations.

- Introduction, background by B. T.
- Presentations by students on project
- Presentations by members of department

Website:

http://www.let.rug.nl/birot/courses/2008-research/

t . s . biro [at] rug . nl (always welcome!)

The project: computational models for language phenomena

- A model for *linguistic competence*.
- A model for *linguistic performance*.
- A model for *language learning*.
- A model for *language change*.

Prerequisites

- Programing language:
  $your$ choice and $your$ responsibility.
- Algorithmics.
- Openness to mathematical concepts.
- Openness to linguistic phenomena.

Requirements

- Presence and active participation.
- Minor tasks related to literature.
- (At least) two presentations *with slides*.
- Final paper.

Presentations

- 1st: Theoretical background, goals of the experiments, hypotheses to be tested + Details of implementation (lang., arch.).
- 2nd: Experiment results and discussion.

Each member of the group has to present something (3×15 min).

Final paper

Final paper covers:

- Everything presented + more experiments.
- Reaction to presentation feedback.
- Related work: other group, Niyogi, else.

Each student writes it up individually!

Structure of scientific paper

- Intro: general background, importance of research, motivations, past related work.
- Definition of problem.
  Formulating hypothesis.
- How to tackle the problem.
- Methodology, details of the experiment(s).

Structure of scientific paper (2)

- Results of the experiment(s).
- Discussion of results: why surprising? error analysis.
- Conclusion(s) and future directions.
- References: as much as you can, but only what you refer to.

Structure of scientific paper (3)

Abstract, captions of figures, intro and summary:

a concise summary of everything, with much redundancy (most readers will not read the paper).

The project: computational models for language phenomena

- A model for *linguistic competence*: best element of some candidate set, according to some function = *grammar*.
- A model for *linguistic performance*: an algorithm finding it.

The project: computational models for language phenomena

- A model for *language learning*: an algorithm finding the grammar accounting for observed forms.
- A model for *language change*: iterative learning.

Research methodologies (1)

- Collect literature in library.
- Collect data on the field: observe whatever you can (how?).
- Run *controlled* experiments on animals, subjects or atoms.

Research methodologies (2)

- Collect your data: 
- Systematize data, create typologies:

- Create a model describing your typology: a deeper "understanding" of phenomenon

Research methodologies (3)

- Data have been collected.
- Formulate concepts, theories.
- Formulate mathematical models.
- Verify models: falsification or corroboration.

Computer simulations in science

- Between mathematical-theoretical approaches and experimental ones.
- Test theories:
- More complex than analyzable in an analytical way (maths).

- Simplification from real life (real life too complex).
- Not enough time in real life.

Why run simulations for language evolution?

From hypothesis to hypothesis

1. Choose a research question.

2. Choose a theory to tackle it.

3. State exactly what your hypothesis is.

4. What observation would falsify it? What observation would corroborate it?

5. Run pilot experiment:
   debugging + get a general picture.

6. Run experiment.

7. Run improved experiment.

8. Result will always be surprising:
   discussion (the "why"s), error analysis.

9. Reformulate hypothesis, back to 3.

10. Alternatively: reformulate theory and go back to 2.

11. (Alternatively: give up on the topic, and go back to 1.)

Programming tricks

- Choice of prg language(s).
- Scripting languages: collect data, repeat many times, change parameters.
- Modularity: make changes in future.
- Parameters, std. input and output.

Playing with the simulation

- Behaviour of the model: e.g. run time, precision.
- Role of the parameters: change always one at a time.
- Plot the behaviour of the model as a function of parameters.

- Smooth vs. abrupt change in behaviour, as parameters change.
- Don't have too many parameters.

Save every information!

- Recall information months latter!
- Save results: file names, directories
- ...with LOTS of comments!
- Create tables, graphs immediately. Write up your motivations and observations immediately.

Moreover:

- Don't be shy to ask questions.
- Search for people working on related topic.
- Take every opportunity to present your topic: collect feedback, both experts and grandma. Understand their perspectives.
- Refresh your mind regularly.

By next week:

- Make groups.
- I finalize schedule.
- Read chapter 1 of Niyogi.
- Send to me in email 5 observations you have on Niyogi chapter 1.