How to get a PhD (in Computational Linguistics)

Sebastian Pado
IMS, Stuttgart University

By way of introduction: My interests

- Lexical semantics
  - Selectional preferences
  - Semantic roles
  - Textual entailment
- Distributional models
  - Compositionality
  - Cognitive Data
- Cross-lingual aspects of semantics

By way of introduction: My interests

- Lexical semantics
  - **Selectional preferences**
  - Semantic roles
  - Textual entailment
- **Distributional models**
  - Compositionality
  - Cognitive Data
- **Cross-lingual** aspects of semantics
- Paper at NAACL 2010: **Cross-lingual selectional preferences with a distributional model**

Structure

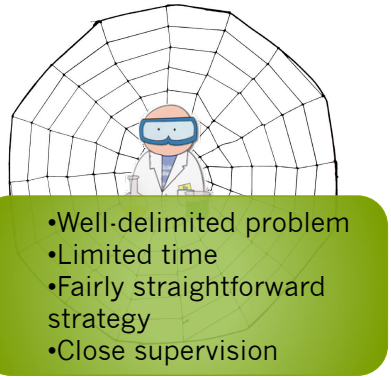
1. The situation and its challenges
2. Attitudes
3. Actions
4. Aims

Disclaimer

- There are no simple recipes for writing a PhD thesis
- I'm trying to point out parameters and decision points
  - Balance between high-level and low-level aspects
- All universal statements are wrong.
  - Many of the dichotomies in this talk aren't real dichotomies

1. The Situation

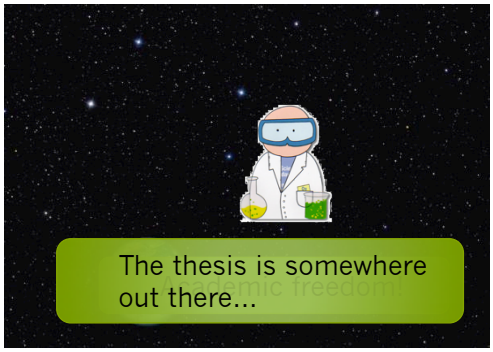
- Undergraduate student



- Well-delimited problem
- Limited time
- Fairly straightforward strategy
- Close supervision

The Situation

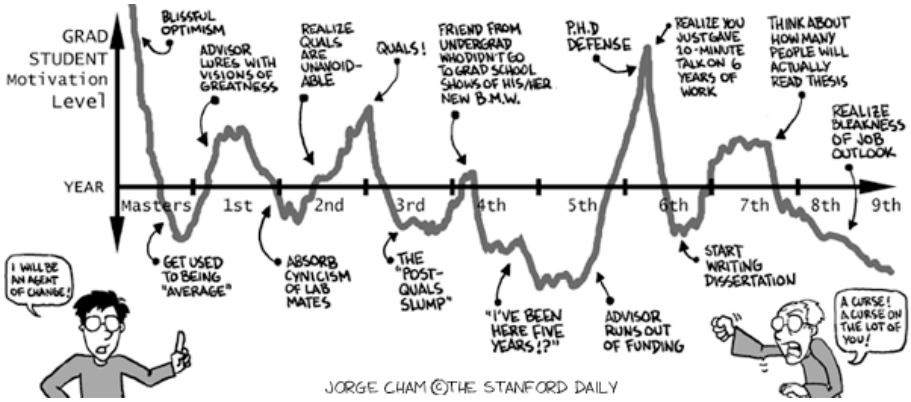
- PhD student



Its challenges

- **The attitude**
  - Can I do this?
  - Am I doing a good job?
- **The methods**
  - How to find a good topic
  - How to identify relevant prior work
  - How to present results
  - How solve problems X, Y, Z

2. Helpful attitudes



Helpful attitudes 1

- **Believe in yourself**

- Doing a PhD is, in fact, difficult
  - If the problems you work on were easy, they would be solved already
- You wouldn't be here if someone wouldn't think that you could do it

Helpful attitudes 2

- **Get a hobby**
- Your job has an extremely delayed feedback loop.
  - Appreciation is very slow
- There can be a very strong connection between the success of your work and your self-confidence
- It helps to be able to concentrate on something else once in a while

Helpful attitudes 3

- **Accept ignorance as part of the job**
- You are trying to find out new things
  - Of course you don't know the answer in advance
- Trust that this is going to change

A paper you should read

Martin A. Schwartz. The importance of stupidity in scientific research. Journal of Cell Science 121, 1771 (2008).

http://jcs.biologists.org/cgi/content/full/121/11/1771

Helpful attitudes 4

- **Science is about questions, not about answers**
- Relevance for your own research:
  - Asking the right questions is often the first step towards solving the problem
- Relevance for others' research:
  - Questions are more general than answers

Helpful attitudes 5

- **Be proactive**

- You're on your own
  - If you want things to happen, make them happen yourself

3. Helpful actions

"Success is 10% inspiration and 90% perspiration." (T.A. Edison)

- You're the manager of a complex multi-year project
  - Success requires a lot of **methodical** and **social** competence

Overview

- Methodical competence
  - Identifying a thesis topic
  - Reading literature
  - Organizing your research
- Social competence
  - Learning
  - Getting feedback
  - Collaborations
- Thinking about the future

Overview

- **Methodical competence**
  - Identifying a thesis topic
  - Reading literature
  - Organizing your research
- Social competence
  - Learning
  - Getting feedback
  - Collaborations
- Thinking about the future

Identifying a thesis topic

- A PhD thesis is a piece of work that makes an **original contribution** and that solves a **coherent (set of) question(s)**

- An ideal thesis
  - Asks a new question (Relevance)
  - Can be solved within a finite time (Feasibility)

Types of Theses

- **Goal-driven**
  - Identify new problem; find a theory to solve it

- **Theory-driven**
  - Develop new theory; find problem that can profit from it

Relevance: Approaches

- **Bottom-up**: Start with a "local" that interests you (or that you have to work on anyway) and develop it into a "global" topic
  - Challenge: Relevance
- **Top-down**: Start with a global vision, work your way down to the local details
  - Challenge: Feasibility
- Iterating is okay. But try to get to a state where you can phrase the goal of your thesis in two sentences.

Side Node: Project PhDs vs. Scholarship PhDs

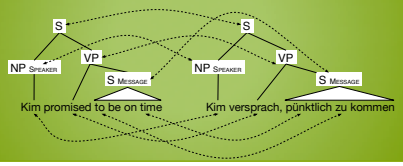
- Project PhDs: identify a topic that is close enough to the project goals
  - Ideally, work on your PhD counts as your contribution to the project
  - However, make sure that you are identifiable as the primary researcher
- Scholarship PhDs: choose from the broad range of possible topics
  - Attaching yourself to a project for some time may be a good idea

Example: me

SALSA project: German semantic role annotation

Goal: develop a semantic role labelling system for German (without German data)

Method: Annotation projection in parallel corpora

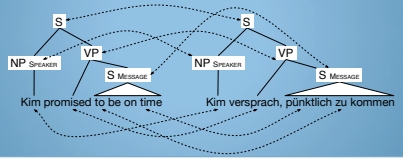


Example: me

SALSA project: German semantic role annotation

Goal: develop a semantic role labelling system for German (without German data)

Revised Method: Annotation projection in parallel corpora *for parallel sentences*



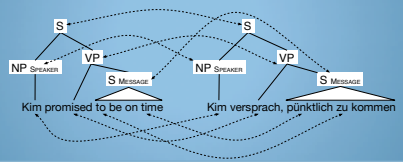
Problem 1: Translational divergence

Example: me

SALSA project: German semantic role annotation

Revised Goal: *create the data necessary to* develop a SRL system for German

Revised Method: Annotation projection in parallel corpora *for parallel sentences*



Problem 2: Developing an SRL system takes time

Relevance: Reading

- The primary way of assessing relevance of a work (plan) is by reading prior work
- Challenge: don't be overwhelmed
- Read breadth-first, not depth-first
  - Try to judge relevance
    - Rules of thumb: Citations, Reputation of venue, Originality
- Focus on finding the original idea that **you** can contribute

Keeping an Open Mind

- **CL/NLP is by nature an interdisciplinary field**
- Ideas from other fields make major impressions onto CL/NLP
  - Theoretical CS: Formal Languages
  - Linguistics: Semantic roles
  - Machine Learning: Structured Probabilistic Models
  - Psycholinguistics: WordNet

Organization is key

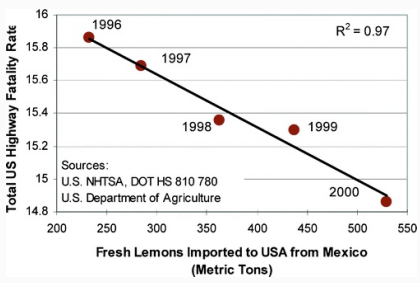
- A multi-year project requires careful planning
- Set medium-term goals for yourself (~several months)
  - Evaluate your plans regularly (~several weeks)
  - In terms of local success, in terms of effort, in terms of success towards the global goal
- Keep track of what you have done

The demands of an empirical discipline

- Central requirement: **Accountability**
  - Your results are as much a function of the data as they are of your model
  - You must be able to reproduce your results
- Keep a lab book
  - Document each command, your code, your representation formats, the version of external software, results...
  - Keep all stages of your data (as far as possible)
  - Avoid manual fiddling
    - Dan Ramage's Research Assistant (ask Google)

Data, data, data

- Analyse the data
  - when it doesn't work well
  - when it works well

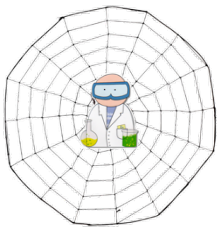


Overview

- Methodical competence
  - Identifying a thesis topic
  - Reading literature
  - Organizing your research
- **Social competence**
  - Learning
  - Getting feedback
  - Collaborations
- Thinking about the future

The social aspect of research

- The social aspect of research is often underestimated
- You need to build a new web for yourself



- Outbound: **Dissemination**
  - A result that's just as good in your head is almost as good as no result:
- Inbound: **Learning through Interaction**

Learning through Interaction

- As PhD student, you're in a weird situation
- You are among the top experts on your topic in the world
  - But you don't have a lot of experience (yet) in making productive use of this situation
  - Especially at the beginning, talking about your work can be daunting
- Different groups of people can help you in different ways

Interaction 1: Fellow PhDs

- Set up seminars
  - Learn to prepare presentations
    - Audience of semi-experts
  - Learn to give presentations
  - Learn to answer questions
  - Get feedback (and, if you are lucky, inspiration)
- Form a reading circle
  - Acquire background knowledge that you're missing
  - Learn to discuss merits and problems of scientific studies

Interaction 2: Mentors

- Find someone who you can come to regularly with all your questions
  - Ideally your advisor, but postdocs or other researchers work as well
- Tap into their longer experience
  - Discuss your global and local PhD plans
  - Ask them about of problems or phenomena that you don't understand
  - Have them comment on texts you wrote
- The more pre-processed the input, the more helpful the feedback

Interaction 3: Cooperations

- There may be no one at your place who is an expert on your particular problem
- Find out who is, and write them
- Ours is a very friendly community
  - For problem-specific questions, your chance for a reply is quite good
  - I also know a number of successful collaborations that began in this manner

Interaction 4: Participate

- Take advantage of relevant offers
  - "Soft skills" courses
  - Programming skills
  - Event organisation
  - Become a **conference reviewer**

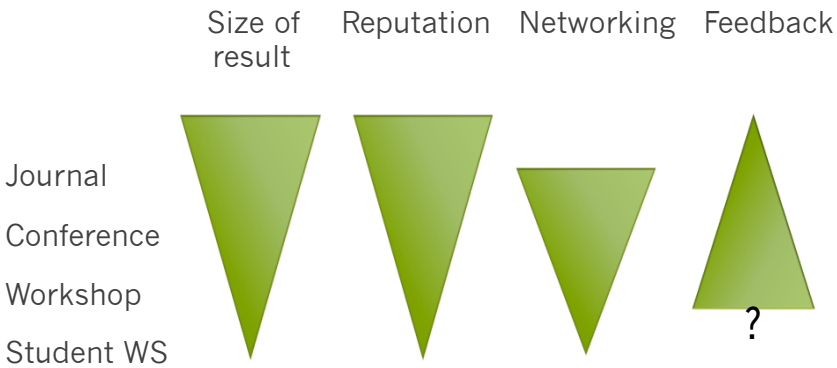
Interaction: the social aspect of keeping an open mind

- Do some activities that don't seem directly relevant to your work
  - Sit in on interesting lectures
  - Go and see invited talks
  - Join invited speakers for dinner
- Try to get out of your institution for some time, ideally abroad
  - See a different system, ask different questions
  - There are a number of opportunities for funding research visits

Dissemination

- Ideally, at the end of your PhD, people in your area have heard your name and **connect it with a topic**
- Collaborating and communicating is a good first step
- Primary strategy: having papers at events
  - Twin effect: Works gets disseminated, and you get to network

The hierarchy of venues



Find the best venue for your result

- Sometimes, a paper at the right workshop does more for your name than one at a big conference
- However, the "big" conferences carry more reputation, also due to their low acceptance rate
  - In Core CL: *ACL
- Aim at at least one (preferably more) conference papers on your topic
  - Papers in good journals need a good story and take long to get published

Writing tips for *ACL papers

- Start writing early
- Know what you say
  - State your claims precisely
  - Make your case comprehensible to "semi-experts"
- Don't let it get rejected on formal grounds
  - Write comprehensible English
  - Try to identify the "standard structure" of ACL papers
- Tell one story, but tell it convincingly
  - A good paper contains ~12 pages of material, summarized into 8 pages

More publication strategies

- Many studies in CL require the development of infrastructure
  - Identify an interesting technical point and make it into a paper at a workshop or smaller conference
- ACL is trying right now to draw in papers of different kinds
  - Short papers (work in progress)
  - Negative results
  - Opinion/Discussion pieces
- Aim for the earliest conference, and resubmit if necessary

Overview

- Methodical competence
  - Identifying a thesis topic
  - Reading literature
  - Organizing your research
- Social competence
  - Learning
  - Getting feedback
  - Collaborations
- **Thinking about the future**

The future

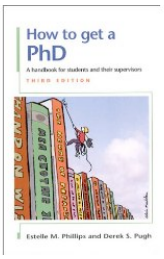
- Contrary to opinion among PhD students…
- …there is a life after the PhD!
- What do you want to do after you finish?
  - Large company: R & **D**
  - Small company: R & D & **EE**
  - University: **R** & T
- Other criteria:
  Work style, Work/Life balance

Think about the future

- Most of what I said was geared towards an academic career
- Find out what is best for you
  - **Do an internship**!
- Options require different skill sets
  - Start getting the right qualifications as soon as possible
  - Might even influence your topic…?

More literature

- *How to get a PhD*. E. Philips, D. Pugh. McGraw-Hill.



- *The craft of research*. W. Booth, G. Colomb, J. Williams. University of Chicago Press.

SEMEVAL 2010 task on classifying semantic relations

My<e1>hat</e1> is in the <e2>car</e2>.

- 10.000 sentences of 10 relations
  Product-Producer, Entity-Origin, Entity-Destination, Instrument-Agency, Content-Container, Component-Whole, Member-Collection, Communication-Topic
- Step 1: Get the data on Feb 15th/22nd
- Step 2: Submit results: April 2nd
- Step 3: Go to beautiful Uppsala