How to write a master's thesis: a computational linguist's view

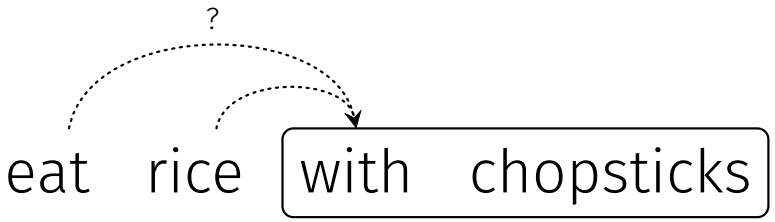
Simon Šuster

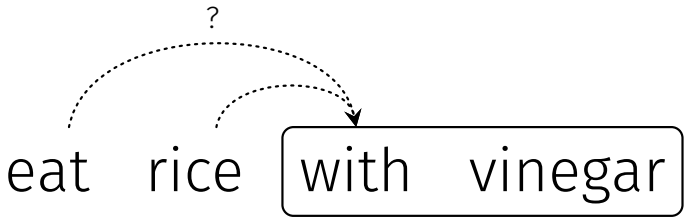
31 March 2015

http://simonsuster.github.io/

Background

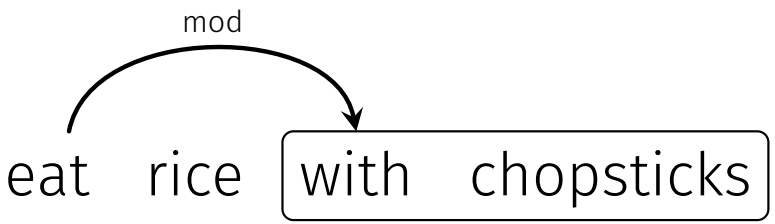
- master's thesis in computational linguistics
- resolving structural ambiguity automatically

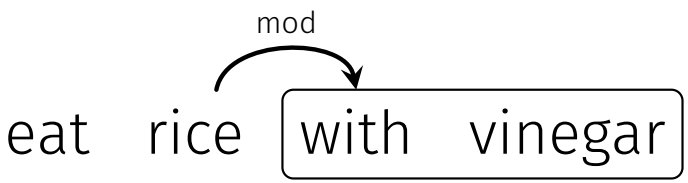




Background

- master's thesis in computational linguistics
- resolving structural ambiguity automatically





Background

Research & writing process could have been much better...

This talk:

- advice from my and others' experience
- valid for computational and experimental work

How to write a master's thesis?

7 suggestions

Suggestion #1: Write the paper first

Paper is paramount, thesis is secondary

- narrower, less intimidating than thesis
- chances of ever publishing your work are higher
  - and paper submission deadlines are a great motivation

From paper to thesis

Paper is thesis' skeleton around which you build:

- further experimental information
- related work
- general introduction to the field
- description of less successful experimental attempts

Separate core from details

Throughout, keep at least two documents:

- one with technical details of experiments you run
  - together with ideas you'd like to put into practice
- the paper (∼8–10 pages), to stay focused on main ideas

Thesis is a carefully thought-out compilation of both documents.

Suggestion #2: Writing before experimenting

Start writing *soon*, before doing experiments (assuming a good research problem)

- better planning
- more focus
- crystallization of the problem
- easier to weed out unnecessary experiments

Suggestion #2: Writing before experimenting

Start writing *soon*, before doing experiments (assuming a good research problem)

- better planning
- more focus
- crystallization of the problem
- easier to weed out unnecessary experiments

Start writing on:

- what is the problem and how you address it
- how you intend to run experiments
- why you run them, and what they will show
- (some) related work

Suggestion #3: Read a lot

Try hard to find papers close to your research problem

- to prevent duplication of effort
- and because replication is only a very small contribution

Related work sometimes hard to find due to varying terminology

- Help from your supervisor, conference, mailing lists (e.g. corpora-list)

Compounding

(adapted from R. Hamming):

The more you read, the more you know
The more you know, the easier to read
The more you know, the bigger the productivity

Note

Spending much time early on on writing literature review can block new views on the topic

---

"The reading is necessary to know what is going on and what is possible. But reading to get the solutions does not seem to be the way to do great research." (R. Hamming)

Suggestion #4: Choose a topic with care

Aim for a small contribution to a big, unanswered problem

- *small* contribution, concrete plan

It's OK to work on a topic predetermined by your supervisor

- if you're attracted by the topic
- if he/she knows the topic background very well (normally so)

It's great to come up with a topic yourself, but:

- talk to people knowing the field better than you do (includes your supervisor)

Suggestion #5: Tell a story

(Cf. talk by Simon Peyton Jones `http://youtu.be/g3dkRsTqdDA`)

Thesis/paper is a narrative:

- here is a problem
- it's an interesting problem
- it's an unsolved problem
- here is my idea 
- my idea works: data, experiments
- relation to other work

Suggestion #6: Painstaking documentation

Be able to re-run the experiments several years on from now

- everything needs to be in a single place, with a single entry point such as a readme with instructions to proceed
- you don't want to be figuring out which data set produced a specific plot, or what parameters were used

Back-up frequently or use a versioning system for all your documents and code (e.g. free private Bitbucket accounts)

Suggestion #7: Finalizing

- spell-check
- track down bad style:
  - frequent repetitions
  - verbose phrases & vagueness
  - useless adverbs, like “very”
  - avoid excessive hedging:
    “**seems** to **suggest** that the problem **might** be…”
- someone other than you or your supervisor should read the thesis
- `http://lemire.me/blog/rules-to-write-a-good-research-paper`

Further reading

- on writing good papers, in general:

  http://www.slideshare.net/lemire/write-good-papers

  http://www.slideshare.net/shawn_nordell/scholarly-writing-workshop-by-shawn-nordell?next_slideshow=1

  http://www.slideshare.net/ingermewburn/write-that-journal-article-in-7-days-12742195?next_slideshow=1

- advice on writing clear and concise sentences:

  http://homepages.inf.ed.ac.uk/sgwater/writing_advice.html

- finding research problems, how to read papers, and much more: http://www.cs.jhu.edu/~jason/advice/

- "How to be a successful PhD student" (also applicable to MA students):

  http://people.cs.umass.edu/~wallach/how_to_be_a_successful_phd_student.pdf

- "You and your research", by R. Hamming

N.B. Good thesis is a finished thesis.