

The 2nd GP meeting

11/30/2018

To start with

- All the important urls are at the CLMS policy page:
 - <http://depts.washington.edu/uwcl/clms/>
- GP meeting canvas site:
 - <https://canvas.uw.edu/courses/1263622/>
 - “Files”: slides, recordings, etc.
 - “Assignments”: submit your report or thesis proposal
 - “Discussions”: indicate your choice, and post your questions.
- Indicate your capstone choice by **Dec 20, 2018** by replying to the threads under “discussions”.

Outline

- Three options (Recap):
 - Internship option
 - Thesis option
 - Project option
 - Choosing between the three options
- CLMS-to-PhD transition
- Possible thesis topics

Internship option

Upcoming deadlines for internship in summer 2019

- Polishing your resumes: ASAP and no later than 1/15
- Sending out resumes: follow companies' deadlines
- Pre-internship report (Part 1): 1/15
- Career Services visit complete (recommended): 1/31
- Pre-internship proposals (Part 2): 6/15
- Internship: normally in summer of the year that you are graduating. Can be early for part-time students

Pre-internship report: part 1

- A few areas of interests: e.g., IE, MT
- A list of companies and why the companies targeted are good places to do this work
- A list of relevant references
- The typical length is 1-2 pages

Finding an advisor

- Normally, your advisor will be Gina or Fei
- If you have any preference, let us know by 12/20.
- You will be assigned an advisor in Winter 2019.

Internship position

- Must be relevant to CL
- At least 200 hours over 4 months
- Part-time positions are ok.
- Full-time jobs are ok, too.

Finding an internship position (cont)

- Check the CL job database and daily job post
- Attend CLMS Career talks, UW/MS symposia, etc.
- Write pre-internship report (e.g., areas of interest, list of companies, related work)
- Prepare a resume
- Visit UW Career Office: info about job fair, mock interview
- Keep your advisor informed
- Submit your resume before the companies' deadlines

Companies

- Microsoft
- Amazon
- Cambia Health
- Nuance
- ...
- An internship company list under
“Files” => “Graduation Planning General”

Prepare for interviews

- Know yourself:
 - Should be able to explain everything on your resume
 - What are your strengths/weaknesses?
 - What kind of job do you like?
- Know the companies and the position
- Have mock interviews
 - Common interview questions
 - Tips on interviews

Thesis option

Thesis Option: Deadlines for summer graduation

- late fall: 2nd GP meeting: thesis discussion
- 1/3: Thesis proposal (first draft)
- 2/1: Thesis proposal (second draft)
- 3/1: Thesis proposal (final draft), agreed w/advisor
- 4/1: Literature survey
- 5/1: Methodology chapter
- 6/1: Interim results presentation to other students
- 7/1: Results chapter
- 7/7: First full draft
- 8/5: Second full draft
- 8/10: Final draft for signature and submission

Deadlines for fall/winter graduation

- Discuss with your advisor
- A possible schedule for fall graduation:
 - 4/1: 1st thesis proposal
 - 5/1: 2nd thesis proposal
 - 6/1: Final thesis proposal
 - Summer and fall: Work full-time on the thesis

Finding an advisor

- Your committee consists of an advisor and a reader.
- Either your advisor or your reader must be CLMS faculty.
- If your advisor is not a CLMS faculty, you need to get approval from a CLMS faculty first.

The structure of a thesis

- Introduction: Problem description etc.
- Literature survey
- Methodology
- Algorithms, implementation, etc.
- Experiments
- Discussion
- Conclusion and future work

The 1st thesis proposal

- Problem description
- Related papers
- Methodology
- Evaluation
- Plan for the next step
- Advisor

Thesis topics

- Choose a thesis of the right size
 - You can discuss your idea with us.
 - You will be able to revise the idea in the 2nd and 3rd thesis proposal.
 - Some old CLMS theses are under dropbox/CLMA Thesis/thesis/ on patas.
 - Double dipping
 - More to come (past thesis titles and info for UW library access are at the end)
-
- ➔ Thesis option requires more work
 - ➔ Choose this when you truly want to do it
 - ➔ Need to find a faculty who is willing to be your thesis advisor

Project option

Project option

- Good option for those with full-time job
- Can also be backup for the internship or thesis option
- The option must be approved by the CLMS faculty
- Like the thesis option, you need to find a faculty member to work with, who can be from the Ling or other departments at UW.

Deadlines

- Pre-project proposals: 5/25
- Post-project report (initial version): 7/20
- Self-evaluation with supervisor's approval: 8/05
- Supervisor's evaluation: 8/10
- Post-project report (final version): 8/10
- Requires approval from CL faculty.

Choosing one of the three options

	Thesis option	Internship option	Project option
Assigned an advisor?	No	Yes	Yes
Independent research?	Yes	No	No
Position in	university	industry	university
Normal length	6-9 months	3 months	3 months
Need to find	thesis topic advisor reader	Internship (supervisor)	Research Project (supervisor)
Main documents	thesis proposal thesis	pre-internship report post-internship report evaluation letters	pre-project report post-project report evaluation letters
Followup	publications	full-time job	(publications)

Outline

- Three options (Recap):
 - Internship option
 - Thesis option
 - Project option
 - Choosing between the three options
- CLMS-to-PhD conversion
- Possible thesis topics

CLMS-to-PhD transition

Also see

<https://linguistics.washington.edu/clms-phd-transition>

CLMS=>PhD transition

- Application deadline is March 1.
- The decision is made by the whole ling faculty by April 1.
- In order to apply:
 - Choose the thesis option
 - In the year that the PhD program starts:
 - Complete your master thesis by March 1
 - Complete your CLMS program by the winter quarter
 - Take three of the required courses in the Ling Ph.D. program (CL track):
 - They cannot be ling570-573, but can be ling550 and ling566.
 - At least two out of the three courses should be taught by two distinct non-CL faculty.

Six required courses for PhD degree

- 2 courses in syntax: 566, 507, 508
- 2 courses in phonetics/phonology: 550, 551, 552, 553
- 1 course in semantics: 542, 579
- 1 course in sociolinguistics: 532, 533

Funding once enrolled in the PhD program

- Multiple CLMS=>PhD students may be accepted each year, but only one will receive a 4-year funding guarantee (TA/RA-ships) contingent on the student making satisfactory progress.
- Students without funding guarantee might still be eligible for TAsip, RAsip, or support from other sources; it is just that the funding is not guaranteed.
- With March 1 deadline, all the applicants will be evaluated together.
- If you are interested in the transition, talk to your advisor ASAP.

Application material

- UW transcript
- CV
- A one-page statement of research plans
- A brief letter of support from the UW CL faculty member who has agreed to be your PhD advisor
- A brief letter of support of another UW linguistics faculty member (CL or otherwise)

The typical timeline for full-time CLMS students

- Fall/winter/spring Y1: take 9 required courses
- Spring Y1: finalize thesis proposal
- Summer Y1 – Winter Y2: complete the thesis
- 3/1 in Y2: application for PhD conversion
- 4/1 in Y2: acceptance notice
- Winter/spring Y2: take additional courses (optional)
- Fall Y3: PhD program starts (w/ or w/o 4-year funding guarantee)

Timing

- CLMS alumni are eligible for the transition for **three** years after CLMS; that is, no more than three years may elapse between CLMS graduation date and the fall quarter following admission to the PhD.
- The 4-year funding package will start at the beginning of the PhD program.

Tips for the transition

- Course selection:
 - When choosing your ling elective, choose one of the courses required for PhD (e.g., do not use ling567 as your ling elective if you don't want to take another ling course for PhD application).
 - If you waive 550 or 566, replace it with one of the courses required for transition.
- Choose the thesis option for CLMS degree, and perform well in your coursework and thesis.
- Let your advisor know your plan soon.
- Apply for external funds (e.g., NSF Graduate Research Fellowship)
- Do not miss the March 1 application deadline and the CLMS graduation deadline in the winter quarter.
- This path is available only for 3 years between CLMS graduation and the start of PhD program.

Transition vs. Genling admission path

	Transition path	Genling path
Application deadline	March 1	Mid-Dec
Application requirements	See slide #24-30	See ling website
Application fee	No	Yes
Decision process	Recommended by the CL faculty, final vote by the whole faculty	Decided by the ling admission committee
Acceptance rate	Much higher than the Genling path. Applicants are compared with CLMS students	About 5-10%. Applicants are compared with ones from all ling subareas
Funding	One with 4-year guarantee. Others w/o guarantee.	Accepted students all have 5-year guarantee.

Between CLMS graduation and the beginning of PhD program

- You can apply to the **General Linguistics Graduate** program:
 - The decision is made by the CL faculty
 - Once admitted, you can take additional courses to meet the requirements for PhD applications or PhD graduation.
 - Once admitted, you will pay the regular tuition rate, not the CLMS rate.
- Taking this step is useful if
 - you want to pay regular tuition rate
 - you need to maintain student status (e.g., for visa purposes)
- Application material: same as the one for CLMS-to-PhD transition application, except that the 2nd recommendation letter is not needed.

Possible thesis topics

Three dimensions of a Thesis

- Key components of a thesis:
 - Dimension 1: Task
 - Dimension 2: Algorithms
 - Dimension 3: Data/Domain/Language

Dimension 1: Task

- **What problem** are you trying to solve?
- Possible tasks: (some topics from Ling 570, 571)
 - Word segmentation
 - Morphological analysis
 - POS tagging
 - Named entity recognition
 - Grammar induction, etc....
- Application: (some covered in Ling 573)
 - Q&A, MT, ASR, sentiment detection, topic detection, etc

Dimension 2: Algorithms

- **How** are you going to solve the problem?
- Classes of approach:
 - Symbolic approaches (aka rule-based approaches)
 - Corpus-based approaches (machine learning, Ling 572)
 - Supervised learning:
 - Decision trees, transformation-based learning, MaxEnt, SVM, NN
 - Semi- or Un-supervised approaches:
 - Co-training, EM, Clustering, etc
 - Hybrid approaches:
 - Mix symbolic and corpus-based methods

Dimension 3: Data/Domain

- **What** materials will you use to evaluate your method?
 - **What data and evaluation techniques** will you use?
- Many types of data:
 - Raw data, treebanks, FrameNet, tagged corpora
 - Monolingual, bilingual, multi-lingual
 - Newswire, conversational speech
 - Limited domain: travel, bio-medical

Projects

- Draw from each of the dimensions
- What task?
- What algorithm?
- What data/domain/language?
- Novelty:
 - Novel aspects can come in any dimension or by novel combination, e.g. of data and algorithm

Faculty Research Interests

- See Appendix C in handout for the 1st GP meeting

Previous CLMS theses (1): 570

- Clustering:
 - Sergei Lushtak: “Unsupervised Morphological Word Clustering”
- POS tagging:
 - Alex Cheng: “Survey of unsupervised methods for Chinese POS tagging with Hidden Markov Models”
- WSD:
 - Eric Mead, “Associative Anaphora Resolution Using Word Sense Disambiguation and Spreading Activation in an Augmented Ontology”
 - Albert A. Bertram “An Application of a Connectionist Cognitive Model to Word Sense Disambiguation.”

Previous CLMS theses (2): 571

- Parsing:
 - Claire Jaja: “Leveraging Training Data from High-Resource Languages to Improve Dependency Parsing for Low-Resource Languages”
 - Zina Pozen: “Using Lexical and Compositional Semantics to Improve HPSG Parse Selection”
 - Glenn Slayden: “Array TFS Storage for Unification Grammars”
 - Matt Hohensee: “It's Only Morpho-Logical: Modeling Agreement in Cross-linguistic Dependency Parsing”
 - Ryan Georgi: “PCFG Induction using Interlinear-derived Prototypes”
 - Yi Zhu: “Dependency Parsing for Tweets”

Previous CLMS theses (3): 571

- Semantics:
 - Jason Shaw: “Learning for Resource-Poor Languages: Building a Language-Independent Model for Frame-Semantic Annotation”
 - Liam McGrath: “Semantic Features for Semantic Role Labeling”
- Sentiment analysis:
 - Anthony Gentile: “Investigation of sentence structure in domain adaptation for sentiment classification”
 - Maria Alexandropoulou: “Agreement/Disagreement detection and the role of linguistic coordination and conversational features”
 - Ethan Roday: “Three Cheers for Partisanship: Lexical Framing and Applause in U.S. Presidential Primary Debates”

Previous CLMS theses (4): 566/567

- Grammar engineering:
 - Michael Lockwood: “Automated Gloss Mapping for Inferring Grammatical Properties”
 - Woodley Packard: “Full Forest Treebanking”
 - T. J. Trimble: “Adjectives in the LinGO Grammar Matrix”
 - David Wax, “Automated Grammar Engineering for Verbal Morphology”
 - Safiyyah Saleem, “Augment Optionality: A New Library for the Grammar Matrix Customization System”
 - Kelly O’Hara: “A Morphotactic Infrastructure for a Grammar Customization System”

Previous CLMS theses (5): 575

- MT:
 - Ronald Lockwood: “A Linguist-Friendly Machine Translation System for Low-Resource Languages”
 - Francesca Gola: “An Analysis of Translation Divergence Patterns using PanLex Translation Pairs”
 - Spencer Rarrick: “Machine Translation Detection and Automatic Filtering of Web-Extracted Parallel Corpora”
 - Ankit Srivastava: “Learning a Translation Lexicon from Non-parallel Corpora”
 - Achim Ruopp: “Finding and Evaluating Structured Bilingual Corpora on the Web”

Previous CLMS theses (6): 573/575

- Information extraction:
 - Todd Lingren: “Detecting Adverse Events in Clinical Trial Free Text”
 - Nicholas Flacco: “Using Semantic Information to Improve Protein Interaction Detection”
 - Eric Bell: “Event Detection and Classification by Sense Disambiguation”
 - Scott Russell Halgrim: “A Pipeline Machine Learning Approach to Biomedical Information Extraction”
- Summarization:
 - Prescott Klassen: “Calculating LLR Topic Signatures with Dependency Relations for Automatic Text Summarization”

Previous CLMS theses (7): 575

- IR:
 - Adam Rhine: “Information Retrieval for Clinical Decision Support”
- Speech:
 - Ahmed Aly: “Submodular Data Selection in ASR language modeling”
 - Leanne Rolston: “An Independent Assessment of Phonetic Distinctive Feature Sets used to Model Pronunciation Variation”
 - Ping Yu: “GMM-based Automatic Identification between Mainland Mandarin and Taiwan Mandarin
 - Stella Podgornik: “Automatic Detection Of Language Levels in L2 English Learners”

Previous CLMS theses (8)

- Tool development:
 - Michael Goodman: “Egad: Efficiently Evaluating and Extracting Errors from Deep Grammars”
 - David Arthur Bullock: “TreeTran: A Tool for Visual Selection a Testing of Transfer Rules for Machine Translation”
- Others:
 - Andrea Kahn: “New Methods for Detecting Deceptive Product Reviews”
 - Maria Antoniak: “Extracting Topically Related Synonyms from Twitter using Syntactic and Paraphrase Data”
 - Megan Schneider: “Comparative Analysis of DeepBank and the Penn Treebank”
 - Chad Mills: “Learning Board Game Rules from an Instruction Manual”

Accessing recent CLMS theses

- Go to <http://www.lib.washington.edu/services/graduate>
- Click "ProQuest dissertations & theses", the third link under "Your research"
- Enter your userid and password
- Search by author name, advisor name, etc. Remember to enter "University of Washington" in the "University/institution" field

Final note

- Upcoming deadlines:
 - 12/20: indicate your decision w.r.t. which group and which option
 - 1/3: 1st thesis proposal (thesis option for summer graduation)
 - 1/15: resume and part 1 of pre-internship report (internship option)
- It is very important to follow the deadlines
- Keep your advisor informed