Computational Linguistics Curriculum at CMU

Noah Smith
nasmith@cs.cmu.edu

TeachCL-08 June 19, 2008 Columbus, OH, USA

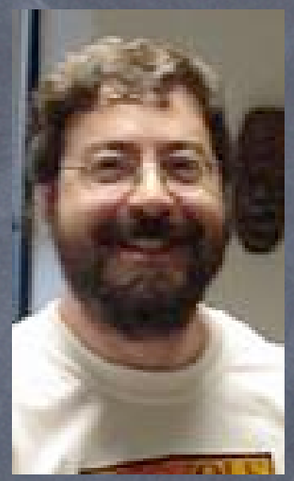
The Language Technologies Institute

- One of ~6 CS departments at CMU
- ~25 faculty (many cross-appointed)
- Graduate programs only: M.S. (~45) and Ph.D. (~84)
- All students: research from day one, two courses per semester (+ research), total ~8 courses, ~6 from LTI

NLP/CL Faculty at CMU


Black


Carbonell


Cohen


Fahlman

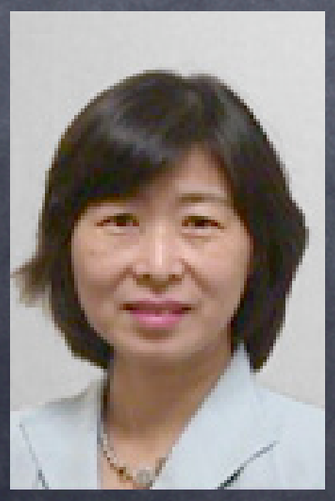

Frederking


Lafferty


Lavie


Levin


MacWhinney


Mitamura


Mitchell


Nyberg


Rosé


Rosenfeld


Smith

Offerings in 2008 (Spring + Fall)

- 4 + 6 graduate courses
  - Algorithms for NLP (Lavie & Frederking)
  - Grammars and Lexicons (Levin & Mitamura)
  - Lang. & Stats. I (Rosenfeld) & II (Smith)
  - Also: IR, Software Engineering I & II, ASR, Research Design & Writing, Services
- 3 + 4 seminar or project courses: Grammar Formalisms, MT, Summarization, Dialog (2), Speech Translation, IR
- Many non-LTI electives in CS, ML, Robotics …

Undergraduate Offerings

- New in 2007-2008: senior undergraduate courses in IR (Callan), NLP (Smith), and Speech (Black).
- Previous course: "Human Language Technologies," a senior/masters course covering IR, NLP, Speech, MT. Not consistently successful.

On Programming

- We're part of the School of Computer Science.
  - But everyone at CMU seems to know how to write code. (Business school professors, philosophy students in my course, ...)
- Students who don't know programming have to work much harder.
  - But to succeed, they do need to learn to program.
  - (And they usually do succeed.)

My Controversial View

- Teaching CL without any programming is as bad as teaching CL without any linguistics or corpora or statistical methods.
- CL is not a theoretical field.
- Students learn best by doing.
- Programming proficiency, especially in some modern high-level languages, may be attainable fairly quickly.

Parting Shots

- CMU has a widespread "computing culture."
- There's room in the LTI curriculum for 4 full CL/NLP courses and many topics courses.
  - Good news: can be leisurely, and some overlap is okay, even beneficial
  - Bad news: less pressure to evolve (just add more courses to the sequence?)
- We have it easy!

Extra Slides

Grammars & Lexicons
(Lori Levin & Teruko Mitamura)

- Breadth of cross-linguistic phenomena (morphology, syntax)
- Analyzing linguistic data
- Generalization and prediction in computational systems
- Lexical-Functional Grammar

Algorithms for NLP
(Alon Lavie & Bob Frederking)

- Algorithm analysis, complexity theory; languages and automata
- Fundamental symbolic algorithms and representations for morphological, syntactic, and semantic processing
- Emphasis on recognition and parsing algorithms

Language and Stats. I
(Roni Rosenfeld)

- Basics of statistical estimation (bias, variance, consistency, efficiency)
- Basics of information theory and the source-channel paradigm
- Language modeling with Markov and hidden Markov models
- Classification and regression, applications in NLP

Language and Stats. II
(Noah Smith)

- Statistical models for shallow and deep NLP (sequence labeling, segmentation, parsing)
- Generative and discriminative methods for prediction in structured spaces
- Disambiguation and inference algorithms
- Unsupervised and semisupervised statistical learning in NLP