

Linguistics 684.02: Computer Grammar 2

Class location

JR 291: 9:30-11:18 Tuesday Thursday

Instructor Details

Chris Brew

583 Drees Labs

Columbus Ohio

Email: cbrew-that-at-thing-acm.org

Office hours: 4pm-5pm TWR

Aims

This course has two main aims: teach the basics of probabilistic modelling for language-related applications, and to expose you to ways in which these ideas have been used in Computational Linguistics. In the process we'll encounter some corpora and tools which you need to know about.

Statistical NLP

Course Book

The textbook is Roark and Sproat “Computational Approaches to Morphology and Syntax” ISBN 978-0-19-927478-9. We will use chapters 1 (introduction), 6-9 (computational syntax) and 5 (morphology learning), in that order. The rest of the material on Computational Morphology belongs in a different course.

The final assignment is a replication of a paper in morphology learning by Schone and Jurafsky. There are

Course Objectives:

Students in Linguistics 684.02 will have an opportunity to:

- Acquire a thorough knowledge of the fundamental terminology, concepts and techniques of language modeling for statistical NLP
- Gain command of algorithms that use dynamic programming for efficiency when dealing with complex search spaces. These will include the Viterbi algorithm for HMMs, the Forward-Backward algorithm for training HMMs and Part-of-speech taggers and the Inside-Outside algorithm for training probabilistic grammars.
- Develop an understanding of what an NLP person looks for in choosing research problems to work on.
- Gain experience in problem solving, in synthesizing ideas, and in writing reports.

Topics

1. **Probability**
2. **Statistical tests**
3. **Hidden Markov Models and Part-of-speech taggers**
4. **Context-free grammars**
5. **Probabilistic context-free grammars**
6. **Lexicalized CFGs**
7. **Learning of morphology**

Statistical NLP

Assessment

There will be 5 major assignments. These are intended to build up to a final project conducted by the whole class. The topic of this project has been chosen to exercise a wide range of skills and techniques that you will have picked up.

The Final Project

The final project is a group-based activity for the whole class and requires you to do the following

1. Program a replication of Schone and Jurafsky's "Knowledge-Free Induction of Inflectional Morphologies" Proceedings of the North American Chapter of the Association of Computational Linguistics (NAACL-2001), Pittsburgh, PA June 2001.
2. Evaluate the results against CELEX as they did.
3. If time permits, work out a way of following up on the suggestion to combine this work with Yarowsky and Wicentowski (2002)
4. Write a clear and organized report using the form of a submission for EMNLP.

Points table

Component	Score
Major assignments	60 points (5 at 12 points each)
Quizzes	5 points (5 at 1 point each)
Small assignments	5 points (5 at 1 point each)
Final project	24 points
Class participation	6 points
Total	100 points
Available extra credit	5 points

Statistical NLP

The table below shows the connection between grades and point scores. If you make the point score, you can count on getting **at least** the grade listed. I reserve the right to give a higher grade if this is warranted by something you do in the course, but I will never give a lower grade.

As a rough guide, you can get an A- by simply learning and understanding the material I teach in the course. To get an A you need to do that, but also show something extra.

Grade	Point Range
A	94-100
A-	90-93
B+	86-89
B	83-85
B-	80-82
C+	77-79
C	74-76
C -	70 – 73
D+	64-69
D	60-63
E	0-59

Your responsibilities

All class members are responsible for

- Keeping up with the assignments and reading
- Monitoring your own progress and understanding of the material. If there is something you don't understand, please do ask, preferably in class.
- Contributing to class discussion.
- Helping to form a "course community". This includes responding appropriately and helpfully to other class members.

Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "Academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but

Statistical NLP

not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct

(http://studentaffairs.osu.edu/resource_csc.asp).

Cheating is wrong, wastes your time and ours, and will not be tolerated.

Working together to find the answer is fine, but talking to someone who has already figured out the answer is cheating. You must also do your homework by yourself unless it is specifically designated as group work. We will assume that you are honest, but if we are confronted with clear evidence of cheating, it is our duty to take action.

Students with Disabilities

Ohio State is committed to extending access and opportunity to those who are disabled. Any student who feels s/he may need an accommodation based on the impact of a disability should contact me privately to discuss your specific needs. You may also contact the Office for Disability Services at 614-292-3307 in room 150 Pomerene Hall.

Spring 2008

The table below indicates when which piece of the course will be covered. Things may change as the course develops.

Week	Dates	Topic	Notes
1	Mar Tu 25 Th 27	Basics	Roark and Sproat ch 1
2	Apr Tu 1 Th 3	HMMs	Roark and Sproat ch 6
3	Apr Tu 8 Th 10	Viterbi and Forward Backward	
4	Apr Tu 15 Th 18	CFGs	Roark and Sproat ch 7
5	Apr Tu 22 Th 24	Enigma: the intelligence	
6	Apr Tu 27 Th 29	Enriched CFGs	Roark and Sproat ch 8
7	May Tu 6 Thu 8	Lexicalized CFGs	Roark and Sproat ch 9
8	May Tu 13 Th 15	Computational Morphology	Roark and Sproat ch 5
9	May Tu 20 Th 22	Morphology Learning	Schone Jurafsky 2001
10	May Tu 27 Th 29		
Exam	Jun Tu 3	Final project write up due	