Course Syllabus

11-721: Grammars and Lexicons
Fall Term 2013
Language Technologies Institute
Carnegie Mellon University

Time: Monday, Wednesday 1:30-2:50 pm
Place: GHC 4102

Instructors:

Lori Levin	Research Professor
	Language Technologies Institute
	GHC 5717
	`lsl@cs.cmu.edu`
	Office hours by appointment and by chance
Teruko Mitamura	Research Professor
	Language Technologies Institute
	GHC 6711
	`teruko@cs.cmu.edu`
	Office hours by appointment and by chance

Teaching Assistant

Matthew Gardner	Ph.D. Student
	Language Technologies Institute
	`mg1@cs.cmu.edu`
	Office hours to be announced and by appointment

Course Description:

Grammars and Lexicons is an introductory graduate course on linguistic data analysis and theory, focusing on methodologies that are suitable for computational implementations. The course covers major syntactic and morphological phenomena in a variety of languages. The emphasis will be on examining both the diversity of linguistic structures and the constraints on variation across languages. Students will be expected to develop and defend analyses of data in a variety of languages. The goal is for students to become familiar with the range of phenomena that occur in human languages so that they can generalize the insights into the design of computational systems. The theoretical framework for syntactic and lexical anaylsis will be Lexical Functional Grammar. Grades will be based on quizzes, problem sets, an exam, and a laboratory exercise.

Immediate Goals

- Students will be familiar with a basic set of morphological and syntactic phenomena and will be able to identify them in any language.

- Students will be able to read descriptive grammars – non-theoretical books that are written by linguists for linguists.

- Students will understand phrase structure and dependency treebanks.

- Students will be able to construct grammar rules and other models of syntactic knowlege for any language.

Latent goals for Language Technologies Students

- Students will be able to accurately label corpora with syntactic information in any language, and will be able to evaluate labeled corpora.

- Students will be able to perform qualitative error analysis on the output of language technologies such as machine translation.

- Students will understand how to design language technologies that work for all of the world's languages.

The readings will be excerpts from several textbooks. The excerpts will be available on the course web page. The pdf of the book by Emily Bender is available through CMU libraries or from the publisher's web site if you log in from a CMU machine. There are no books in the bookstore for this course. If you like the books you can buy them from your favorite book seller.

- *Linguistic Fundamentals for Natural Language Processing*, Emily Bender, Morgan and Claypool. (In the series *Synthesis Lectures on Human Language Technologies*, Graeme Hirst, ed.)

- *Describing Morphosyntax*, Thomas E. Payne, Cambridge University Press.

- *Analyzing Syntax*, Paul Kroeger, Cambridge University Press.

- *Natural Language Syntax*, Peter Cullicover, Oxford University Press.

- *Language Typology and Syntactic Description*, Timothy Shopen (ed.), Cambridge University Press, 1985.

- *Lexical-Functional Syntax*, Joan Bresnan.

- *An Introduction to Syntax*, Robert Van Valin, Cambridge University Press.

Prerequisites:

Introductory linguistics course
Familiarity with Unix/Linux and Emacs
Or permission of instructors

Requirements:

- Grades will be based on short quizzes, homework exercises (25%), a take-home midterm exam (40%), and a series of laboratory exercises (35%).

- Attendance, class participation, and other perceived effort will be an important factor in borderline cases. Coming to class is a really good idea. The course material is synthesized from several sources. What is presented in class is a synthesis and enhacement of the assigned reading.

- If you miss class, it is your responsibility to obtain handouts and assignments, and to find out any schedule changes announced in class.

- Reading assignments must be completed before coming to class, except for the first week of class. There will be quizzes.

- Please note that in order to control late submissions, late assignments may be assigned the next lower grade than they would otherwise have earned. In no case will late papers be accepted after graded papers are returned and discussed in class.

- If problems arise, talk to us.

Rules

- **Homework Assignments:** You are encouraged to discuss your homework with other students, but you must write your solution to the assignment independently in your own words. You may paraphrase sources with appropriate citations, but must not use material verbatim from your sources without citation.

- **Take-Home Midterm Exam:** You must work alone. You must not have any communication with anyone except Lori Levin about the exam. You must refuse to communicate with other students who try to communicate with you about the exam. You may consult the course readings and other materials supplied by Lori Levin, but you must not use the internet or any other material.

- **Laboratory Exercises:** You must work alone on laboratory exercises and the work that you hand in must be your own work.

- **What to do if you are stuck?** Your other classes are demanding, there is a conference deadline, you are ill, you have family obligations, it is your most important religious or national holiday, you have job interviews, your roommate is noisy, and you haven't been keeping up with the course reading. In fact, you don't understand the course reading and the lecture notes are gibberish to you. The due date is approaching and you realize that you are not going to make it.

    - **WRONG:** You think that if you had some time you really could master this material, so you will just borrow some time by copying someone else's work; just for now. You really intend to learn the material later when you have time. Anyway you think your English isn't very good so copying good English is better than writing your own English with errors.

    - **RIGHT:** Send email to Lori Levin or Teruko Mitamura. Ask for a meeting, more time, or an "incomplete" in the course. You will probabaly get it or we will suggest other reasonable arrangements. We will not criticize you for your difficult circumstances. We will commend you for doing the right thing.

- **Consequences of breaking rules:** Breaking rules can result in failing an assignment or exam, failing the course, being expelled from the LTI, and being expelled from CMU and being banned from being on CMU's campus. Foreign students who are expelled generally lose their visas and return to their home countries. Privacy laws prevent us from making examples of specific students who have broken rules in the past, but you should know that students who break rules in LTI classes suffer consequences and all of the consequences listed above have been used.

  The LTI evaluated six cases of rule breaking (not all in this class) in the 2012-2013 academic year. Consequences were very severe, especially for second offenses.

  If the instructors of this course detect evidence of rule breaking, we will report the evidence to the LTI's academic integrity committee. It will be out of our hands at that point. The committee will decide on the consequences.

Schedule and Reading Assignments:

Holidays:

- There is no class on the following days:

    - Labor Day, Monday, September 2, 2013.
    - Day before Thanksgiving, Wednesday, November 27, 2013.

- Please tell the instructors if your national or religious holidays conflict with an assignment or an exam.

Part One: Language Typology and Morphosyntax

August 26:	**Syllabus, course requirements, what is linguistics**
	Reading
	Bender, Chapters 1 and 5
	"The Weirdest Languages" http://idibon.com/the-weirdest-languages/
August 28:	**Parts of speech**
	protoypes and argumentation
	Reading
	Bender, Chapter 6
	Schacter, "Parts of Speech Systems", Section 1, pages 3-23, in Shopen (ed.)
September 2:	**Labor Day, no class**
	Reading
	Bender, Chapters 2, 3, and 4, pages 11-51.
September 4:	**Phrasal categories**
	constituency, transformations
	Reading
	Bender, Chapter 7, Section 51, page 61
	Kroeger, Chapter 2, pages 22-52

September 9-11:	**Dependency trees**
	Reading
	heads, arguments, adjuncts, semantic roles, grammatical relati
	Bender, pages 63-91
September 16:	**X-bar theory**
	Reading
	Cullicover, Chapter 4
September 18:	**Word Order**
	Reading
	Payne, Chapter 4
September 23-25:	**Lexical Functional Grammar**
	Reading
	Bresnan, Chapters 4 and 5
September 30-October 2:	**Case marking and agreement**
	Reading
	Bender, pages 91-97
	Van Valin, pages to be announced
October 7-14:	**Voice and Valence**
	Reading
	Bender pages 97-106
	Keenan, Passive in the World's Languages
	Kroeger, pages 53-80 and pages 192-219
October 16-21:	**Clause Combinations**
	Reading
	Bender, pages 107-113
	Kroeger, pages 103-132
	Davies and Dubinsky, pages 3-16
October 23-28:	**Long Distance Dependencies**
	Bender, page 116
	Kroeger, Pages 165-189
	Keenan, Relative Clauses, available on Blackboard
October 28-30:	Take Home Midterm Exam

Part 2: Lab Exercise, Grammar Writing:
October 30-December 4