LING 6390 - Linguistics Seminar:
Corpus Annotation and Dataset Coding

Fall 2011

Instructor:	Dr. Laurel Smith Stvan
Classroom:	014 Trimble Hall, Wednesday, 3:00 - 5:50 p.m.
Office:	410 Hammond Hall
Office Hours:	Mon. 3-4:30; Thurs. 12:30-2:00; and by appointment
Phone:	(817) 272-3133
Email:	**stvan@uta.edu** (preferred method of contact)

Course Description

In this class we'll analyze the management of data both for compiling or making use of an existing corpus and for coding aspects of the dataset that you cull from a corpus. Tools that we'll explore include choosing a Unicode character encoding, converting text files to spreadsheets, running basic UNIX commands for scripting, and annotating with xml tags, including gaining familiarity with guidelines from the Text Encoding Initiative. Students are encouraged to incorporate their own corpus data or data-coding questions from existing projects into the class discussions and presentations. Annotation issues will be investigated through readings, in-class discussion of these readings, and hands-on practice in using these tools to obtain sortable and reusable language data.

This course fulfills the seminar requirement for the 2009-and-later Linguistics PhD program, and can also serve as an elective for MA degrees. Depending on the topic of your final paper, this seminar can be applied to the requirements of a specialization in any of the five doctoral specialization areas. Prerequisite: permission of instructor.

Course Objectives

This graduate-level course is intended to help you achieve three objectives:

• Acquiring the knowledge and vocabulary to discuss (both orally and in writing) current approaches to gathering texts into electronic corpora.

• Practice in reading corpus linguistics and corpus annotation literature in order to gain insight into both the kinds of questions asked in this field and typical ways of researching the answers.

• Learning to construct and evaluate investigations that are intended to discover characteristics of a language based on frequency and distribution factors found in naturally occurring text selections.

Student Learning Outcomes

Upon successfully completing this course, students should be able to:

- identify different sections of a TEI coded document
- identify and describe key concepts in metadata selection
- demonstrate how to save text data in multiple formats
- demonstrate how to specify character encodings for text files
- demonstrate how to tag and sort data in a spreadsheet according to project-specific variables

Required Materials

There are two forms of required readings. The first is this slim anthology on annotation:

-- Wynne, Martin (ed.). 2005. Developing Linguistic Corpora: A Guide to Good Practice. Oxford: Oxbow Books.

A hard copy is available at the UT Arlington Bookstore (at the corner of Pecan and UTA Blvd.), but the chapters are all also available as a free PDF online: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm

--Additional required readings (consisting of websites and journal articles) will be provided on the course Blackboard site (https://elearn.uta.edu/webapps/login/). You are responsible for checking this Blackboard site regularly to access (among other things) class notes, updates, readings, and assignments.

Supplementary books for additional background reading will be on reserve at the UT Arlington Central Library as well as topics develop. You can search for them under my last name or the course number.

Course Policies

Regular attendance is expected in class, and your presence and thoughtful participation is included as part of the course grade. For each class, students will be expected to have completed prior to class the readings for each day's topic and any assigned problems to be discussed. Students assume responsibility for acquiring all notes, handouts, and announcements.

Because this is a seminar, students will all have a role to play in leading the discussion of readings and/or demonstrations. A schedule of presentations will be established once the students' current areas of research are established during the first class meeting. Regardless of whether a student leads the discussion or not, all students are expected to attend each class having completed the reading and being ready for graduate-level discussions.

Assignments are due at the beginning of class, on the day listed in the schedule, and no later. No late assignments will be accepted without PRIOR approval by the instructor.

Course Requirements

Grades will be determined as follows:

1. Active class participation	**10%**
2. Leading a discussion on one set of readings from the course subtopics	**15%**
2. Presenting an oral overview of a research problem that you are working on and ways that annotating the dataset may help you move forward with the project	**15%**
3. Writing one 2-page summary and response paper of one piece of primary literature on your topic	**20%**
4. Writing one 15-to-20-page term paper (or dissertation chapter) due in exam week	**40%**

Important Academic and Administrative Policies

Drop Policy: Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. **Students will not be automatically dropped for non-attendance**. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. Contact the Office of Financial Aid and Scholarships for more information.

Americans with Disabilities Act: The University of Texas at Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including the *Americans with Disabilities Act (ADA)*. All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of that disability. Any student requiring an accommodation for this course must provide the instructor with official documentation in the form of a letter certified by the staff in the Office for Students with Disabilities, University Hall 102. Only those students who have officially documented a need for an accommodation will have their request honored. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability or by calling the Office for Students with Disabilities at (817) 272-3364.

As a student, your responsibility rests with informing the instructor in writing or email and providing the official documentation from the Office for Students with Disabilities no later than the end of the second week of classes.

Academic Integrity: At The University of Texas at Arlington, academic dishonesty is a completely unacceptable mode of conduct and will not be tolerated in any form, including but not limited to) "cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an examination for another person, any act designed to give unfair advantage to a student or the attempt to commit such acts" (UT System Regents' Rule 50101, §2.2). Suspected violations of academic integrity

standards will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University.

While the Department of Linguistics and TESOL hopes to foster a sense of community in which students can enhance their educational experience by conferring with each other about the lectures, readings, and assignments, all work submitted must be the product of each student's own effort. Students are expected to know and honor the standards of academic integrity followed by American universities; ignorance of these standards is not an excuse for committing an act of academic dishonesty (including plagiarism). If you have questions, please speak with your instructor, your academic advisor, or the department chair.

Please be advised that departmental policy requires instructors to formally file charges with the Office of Student Conduct, following procedures as laid out for faculty (http://www.uta.edu/studentaffairs/conduct/faculty.html), as well as notify the department chair of the filing of the charges.

Student Support Services Available: The University of Texas at Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. These resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals to resources for any reason, students may contact the Maverick Resource Hotline at 817-272-6107 or visit www.uta.edu/resources for more information.

Final Review Week: A period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week unless specified in the course syllabus. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week.

Electronic Communication: UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. Grades or other confidential information can **only** be sent from and to UTA email accounts. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at http://www.uta.edu/oit/cs/email/mavmail.php.

Student Feedback Survey: At the end of each term, students enrolled in classes categorized as lecture, seminar, or laboratory will be asked to complete an online Student Feedback Survey (SFS) about the course and how it was taught. Instructions on how to access the SFS system will be sent directly to students through MavMail approximately 10 days before the end of the term. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback data is required by state law; student participation in the SFS program is voluntary.

Librarian to Contact: Jody Bailey, Central Library, Room 312: jbailey@uta.edu

Grade Grievance Policy:
http://www.uta.edu/gradcatalog/2012/general/regulations/#gradegrievances

Auditors: The Department of Linguistics and TESOL has a "no audit" policy. However, with instructor permission, Department of Linguistics and TESOL faculty, staff, and students enrolled in a linguistics/TESOL program may be able to sit in on a course. These courses cannot be used to satisfy any degree or program requirements/electives, nor will any credit (including retroactive credit) be granted.

LING 6390 Linguistics Seminar

Fall 2011

(Reading assignments and presenters will be updated on the Blackboard site after the first class.)

Week	Date	Topic	Reading	Assignments
1.	Wed. Aug. 31	Introduction		
2.	Wed. Sept. 7	Compiling a Corpus	Wynne: Ch. 1, 5, 6, & Appendix	
3.	Wed. Sept. 14			**Disc:**
4.	Wed. Sept. 21			**Disc:**
5.	Wed. Sept. 28			**Disc:**
6.	Wed. Oct. 5			**Disc:**
7.	Wed. Oct. 12			**S&R Due**
8.	Wed. Oct. 19			**Disc:**
9.	Wed. Oct. 26			**Disc:**
10.	Wed. Nov. 2			**Pres:**
11.	Wed. Nov. 9			**Pres:**
12.	Wed. Nov. 16			**Pres:**
13.	Wed. Nov. 23			**Pres:**
14.	Wed. Nov. 30			**Pres:**
15.	Wed. Dec. 7			**Pres:**

FINAL EXAM WEEK
| | Wed. Dec. 14 | **Papers due by 4:30 p.m.** | | **Final Paper Due** |

Special Dates to Note

Mon. Sept. 12	**Census Date** (last day to add a class)
Fri. Nov. 4	**Last Date to Drop a Course**
Wed. Dec. 21	**Grades Available via MyMav**