

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

**Classroom:** 014 Trimble Hall, T/Th 3:30-5:00 p.m.  
**Professor:** Laurel Smith Stvan  
**Faculty Profile:** <https://www.uta.edu/profiles/laurel-stvan>  
**Office:** 132 Hammond Hall  
**Office Hours:** 12:30 - 1:30 Tuesday and Thursday, and by appointment  
**Phone:** (817) 272-3133  
**Email:** [stvan@uta.edu](mailto:stvan@uta.edu) (preferred method of contact!)

**Course Description**

This course will explore some of the ways that computer science and linguistics can inform each other. We will be concerned in particular with the means by which computers can be used to both obtain the data we examine (a corpus of texts) and to provide the tools we use for analysis (concordance tools). A range of linguistic issues and problems that can benefit from computational approaches will be surveyed. These issues will be illustrated through readings and practical experience with several different software programs as well as sources of corpora. No programming experience is required.

This course fulfills a requirement for the PhD. in Linguistics, but is open as an elective to any graduate student. There is no prerequisite.

This course is intended to help you in achieving the following four objectives:

- Acquiring the knowledge and vocabulary to discuss (both orally and in writing) current and past approaches to corpus linguistics in particular, and computational linguistics in general.
- Practice in reading corpus linguistics literature in order to gain insight into both the kinds of questions asked in this field and typical ways of researching the answers.
- Practice in encountering and evaluating different software to find out how computers can automate many of the tasks you do that use language as data.
- Learning to construct and evaluate investigations whose goal is to discover fuller ways to describe and manipulate a body of naturally occurring language data.

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

**Student Learning Outcomes:**

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

- illustrate that you can open a text file in a corpus program and produce concordance lines
- illustrate that you can scan an image of text, run OCR on it and save it as editable text.
- create a frequency list and describe some of the distinctive aspects the list reveals about a language's vocabulary
- identify linguistic benefits of working with a corpus that is annotated with POS tags.
- describe and illustrate some of the factors that are useful to consider in compiling text samples for a corpus
- describe and illustrate how querying a corpus can offer information to linguistic description beyond what is available via intuition about a language
- describe a corpus search that would be useful in classroom lesson for second language learners.

**Required Course Materials**

There are two required sources of readings for this class:

--The book *Working with Specialized Language: A Practical Guide to Using Corpora* by Bowker and Pearson (Routledge, 2002) is available at the campus bookstore, or through any other bookseller of your choice (ISBN: 0-415-23699-1). (There is also a copy of the textbook on 2-hour reserve in the UTA Library.)

--We will also use a set of required articles that will be available online in the class Blackboard folder (<https://elearn.uta.edu>).

--You will also find it useful to have a USB flash drive and/or familiarity with using the school J-drive to save the work you do in the lab during the semester.

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

**Course Requirements**

Your course grade will be determined according to the following grading key:

|                                            |            |
|--------------------------------------------|------------|
| Attendance, preparation, and participation | 5%         |
| Applied exercises (6 X 8%)                 | 48%        |
| Reading and analysis assignment            | 10%        |
| Vocabulary quiz                            | 5%         |
| Final Paper                                | <u>32%</u> |
|                                            | 100%       |

The grades for each component will be determined as follows:

|             |            |          |           |               |
|-------------|------------|----------|-----------|---------------|
| A- 90-92 %  | B- 80-82 % | C- 70-72 | D- 60-62% | F 59 or lower |
| A 93-96 %   | B 83-86 %  | C 73-76  | D 63-66   |               |
| A+ 97-100 % | B+ 87-89   | C+ 77-79 | D+ 67-69  |               |

**Graded Assignments**

In addition to written exercises that will be required throughout the term, a final paper is required of graduate students, as an opportunity for you to produce a carefully crafted, extended piece of writing showing an application of computer analysis to data of your choice. Here you will demonstrate how the techniques we have discussed in class might assist you in analyzing your own material through gathering and analyzing data for a linguistic project, and briefly presenting this to others at the end of the term. The final paper should be 12-20 typewritten pages. No final exam will be given.

**Course Policies**

At The University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator in student success. Each faculty member is free to develop his or her own methods of evaluating students' academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I have made class attendance **required**. You are responsible for the material presented in class lectures and for any handouts passed out in class as well as for any group work done in class; for your own benefit, come to class. But if you must miss a lecture, do the reading and homework, get notes and information from another student, and then make an appointment to talk to me as soon as possible. However, while UT Arlington does not require instructors to take attendance in their courses, the U.S. Department of Education requires that the University have a mechanism in place to

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

mark when Federal Student Aid recipients "begin attendance in a course." UT Arlington instructors will report when students begin attendance in a course as part of the final grading process. Specifically, when assigning a student a grade of F, faculty report the last date a student attended their class based on evidence such as a test, participation in a class project or presentation, or an engagement online via Blackboard. This date is reported to the Department of Education for federal financial aid recipients.

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. Even approved late submissions will receive a reduction in points.

**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. For more information, contact the Office of Financial Aid and Scholarships (<http://web.uta.edu/aao/fao/>). (Note: Students enrolled in graduate courses may not repeat a class to "replace" a grade).

**Disability Accommodations:** UT 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)*, *The Americans with Disabilities Amendments Act (ADAAA)*, and *Section 504 of the Rehabilitation Act*. 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 disability. Students are responsible for providing the instructor with official notification in the form of a **letter certified** by the Office for Students with Disabilities (OSD). Only those students who have officially documented a need for an accommodation will have their request honored. Students experiencing a range of conditions (Physical, Learning, Chronic Health, Mental Health, and Sensory) that may cause diminished academic performance or other barriers to learning may seek services and/or accommodations by contacting:

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

**The Office for Students with Disabilities, (OSD)** [www.uta.edu/disability](http://www.uta.edu/disability) or calling 817-272-3364. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at [www.uta.edu/disability](http://www.uta.edu/disability).

**Counseling and Psychological Services, (CAPS)** [www.uta.edu/caps/](http://www.uta.edu/caps/) or calling 817-272-3671 is also available to all students to help increase their understanding of personal issues, address mental and behavioral health problems and make positive changes in their lives.

**Non-Discrimination Policy:** *The University of Texas at Arlington does not discriminate on the basis of race, color, national origin, religion, age, gender, sexual orientation, disabilities, genetic information, and/or veteran status in its educational programs or activities it operates. For more information, visit [uta.edu/eos](http://uta.edu/eos).*

**Title IX Policy:** The University of Texas at Arlington ("University") is committed to maintaining a learning and working environment that is free from discrimination based on sex in accordance with Title IX of the Higher Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex in educational programs or activities; Title VII of the Civil Rights Act of 1964 (Title VII), which prohibits sex discrimination in employment; and the Campus Sexual Violence Elimination Act (SaVE Act). Sexual misconduct is a form of sex discrimination and will not be tolerated. *For information regarding Title IX, visit [www.uta.edu/titleIX](http://www.uta.edu/titleIX) or contact Ms. Jean Hood, Vice President and Title IX Coordinator at (817) 272-7091 or [jmhood@uta.edu](mailto:jmhood@uta.edu).*

**Academic Integrity:** Students enrolled all UT Arlington courses are expected to adhere to the UT Arlington Honor Code:

*I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.*

*I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code.*

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

UT Arlington faculty members may employ the Honor Code in their courses by having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System *Regents' Rule* 50101, §2.2, suspected violations of university's standards for academic integrity (including the Honor Code) 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. Additional information is available at <https://www.uta.edu/conduct/>.

**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. 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>.

**Campus Carry:** Effective August 1, 2016, the Campus Carry law (Senate Bill 11) allows those licensed individuals to carry a concealed handgun in buildings on public university campuses, except in locations the University establishes as prohibited. Under the new law, openly carrying handguns is not allowed on college campuses. For more information, visit <http://www.uta.edu/news/info/campus-carry/>

**Student Feedback Survey:** At the end of each term, students enrolled in face-to-face and online classes categorized as "lecture," "seminar," or "laboratory" are directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student's feedback via the SFS database is aggregated with that of other students enrolled in the course. Students' anonymity will be protected to the extent that the law allows. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback is required by state law and aggregate results are posted online. Data from SFS is also used for faculty and program evaluations. For more information, visit <http://www.uta.edu/sfs> **Final**

**Final Review Week:** for semester-long courses, 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

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

this week *unless specified in the class 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. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

**Emergency Exit Procedures:** Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit: turn left as you exit the lab, and at the intersection go to the right and up the stairs. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities.

**Student Support Services:** UT 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. Resources include [tutoring](#), [major-based learning centers](#), developmental education, [advising and mentoring](#), personal counseling, and [federally funded programs](#). For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to [resources@uta.edu](mailto:resources@uta.edu), or view the information at <http://www.uta.edu/universitycollege/resources/index.php>.

**The IDEAS Center** (2<sup>nd</sup> Floor of Central Library) offers **free** tutoring to all students with a focus on transfer students, sophomores, veterans and others undergoing a transition to UT Arlington. To schedule an appointment with a peer tutor or mentor email [IDEAS@uta.edu](mailto:IDEAS@uta.edu) or call (817) 272-6593.

**The English Writing Center (411 LIBR):** The Writing Center Offers free tutoring in 20-, 40-, or 60-minute face-to-face and online sessions to all UTA students on any phase of their UTA coursework. Our hours are 9 am to 8 pm Mon.-Thurs., 9 am-3 pm Fri. and Noon-6 pm Sat. and Sun. Register and make appointments online at <http://uta.mywconline.com>. Classroom Visits, workshops, and specialized services for graduate students are also available. Please see [www.uta.edu/owl](http://www.uta.edu/owl) for detailed information on all our programs and services.

**LING 5381 - 001 (Fall 2016)**  
**Corpus Linguistics**

The Library's 2<sup>nd</sup> floor Academic Plaza offers students a central hub of support services, including IDEAS Center, University Advising Services, Transfer UTA and various college/school advising hours. Services are available during the library's hours of operation. <http://library.uta.edu/academic-plaza>

**Librarian to Contact:** Ms. Jody Bailey [jbailey@uta.edu](mailto:jbailey@uta.edu) 817-272-7516  
A to Z List of Library Databases [libguides.uta.edu/az.php](http://libguides.uta.edu/az.php)  
Course Reserves [pulse.uta.edu/vwebv/enterCourseReserve.do](http://pulse.uta.edu/vwebv/enterCourseReserve.do)

**Auditors:** Department of Linguistics and TESOL faculty, staff, and students currently enrolled in a linguistics/TESOL programs may be able to audit a course (with the permission of the professor). Audited courses cannot be used to satisfy any degree or program requirements/electives, nor will any credit (including retroactive) be granted for audited courses.

**Emergency Phone Numbers:** In case of an on-campus emergency, call the UT Arlington Police Department at **817-272-3003** (non-campus phone), **2-3003** (campus phone). You may also dial 911. The non-emergency number is 817-272-3381.

## Corpus Linguistics—Fall 2016

Proposed Schedule: (Last Updated: **Aug. 25, 2016**)

### Background: Corpus Uses and Famous Corpora

Thur. Aug. 25      Introduction to the class; Introduction to the lab  
What is computational linguistics? What is corpus linguistics?

Tue. Aug. 30      The web as corpus; B&P Ch. 1 "Introducing Corpora and Corpus Analysis Tools" (9-24).

Thur. Sept 1      Fillmore, Charles J. (1992). "Corpus linguistics" or  
"Computer-aided armchair linguistics"; "Henry Kucera (1991)  
**Exercise 1 Due (Web Queries)**

Tue. Sept. 6      Kennedy (1998). "Design and Development of Corpora."  
Boyd& Crawford (2014) Critical Questions of Big Data

Thur. Sept. 8      B&P (2002). Ch. 2 "Introducing LSP." (25-41).  
Clear (1992). "Corpus Sampling." (21-31).  
Demo on using the scanner

### Compiling a Corpus; Inputting and Annotating Texts

Tue. Sept. 13      Crystal (2010). "Languages on the Web." (216-223).  
Bowker (2002). "Capturing Data in Electronic Form."

Thur. Sept. 15      Demo on doing OCR  
Douglas Hofstadter (1985) "Ch. 13: Metafont,  
Metamathematics, and Metaphysics."

Tue. Sept. 20      B&P (2002). Ch. 5 "Markup and Annotation." (75-91).  
Leech (1997). "Grammatical Tagging." (19-33)  
Leech (1997) Appendix III (The C7 and C5 Tagsets) (256-260).

Thur. Sept. 22      B&P (2002). Ch. 3. "Designing a Special Purpose Corpus."  
B&P (2002). Ch. 4. "Compiling a Special Purpose Corpus."  
**Exercise 2 Due (Scanning and Using OCR)**

## Corpus Linguistics—Fall 2016

### Tools to Use on a Corpus: Concordance Software and Indexing Software

Tue. Sept. 27      B&P (2002) Ch. 7 "Introduction to Basic Corpus Processing Tools"  
AncConc demo

Thur. Sept. 29      **Exercise 3 Due  
(Tagging by Hand Using Different Tagsets)**

Tue. Oct. 4.      Working with Antconc  
Zhang (2013) Semantic Prosody Change

Thur. Oct. 6      Using regular expressions

### Applying Corpus Tools: Terminology, Translation, Language Teaching

Tue. Oct. 11      B&P (2002) Ch. 6 "Bilingual and Multilingual corpora:  
Preprocessing, Alignment and Exploitation"

Thur. Oct. 13      **Exercise 4 Due (Using AntConc with Untagged Texts)**

Tue. Oct. 18      B&P (2002) Ch. 8 "Building Useful Glossaries"

Thur. Oct. 20      B&P (2002) Ch. 9 "Term Extraction"  
**Exercise 5 Due Regular Expressions**

Tue. Oct. 25      Stvan (2014) "Truth Is,...

Thurs. Oct 27      B&P (2002) Ch. 10 "Using LSP Corpora as a Writing Guide"  
**Exercise 6 Due (Concordancing with Tagged Texts)**

### Corpus Linguistics—Fall 2016

|                |                                                                                                                                               |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Tue. Nov. 1    | B&P (2002) Ch. 11 “Using LSP Corpora as a Translation Resource”                                                                               |
| Thur. Nov. 3   | B&P (2002) Ch. 12 “Other Applications and Future Directions”                                                                                  |
| Tue. Nov 8     | Granger (1998). “The Computer Learner Corpus: A Testbed for Electronic EFL Tools.” Pp. 175-188.<br><b>Reading and Analysis Exercise 1 Due</b> |
| Thur. Nov.10   | Biber, Conrad, and Reppen (1998). “Language Acquisition and Development.” Pp. 172-201.                                                        |
| Tue. Nov. 15   | Biber, Conrad, and Reppen (1998). “Historical and Stylistic Investigations.” Pp. 203-229 + Pp. 252-253.                                       |
| Thur. Nov. 17  | Biber (1992). “Using Computer-based Text Corpora to Analyze the Referential Strategies of Spoken and Written Texts.” Pp. 213-255.             |
| Tue. Nov. 22   | Lindquist (2000) “Livelier or More Lively? Syntactic and Contextual Factors Influencing the Comparison of Disyllabic Adjectives.”             |
| Thur., Nov. 24 | --Thanksgiving Holiday—no classes--                                                                                                           |
| Thur. Nov. 29  | Stvan (2006). “Diachronic Change in The Discourse Markers Why and Say in American English.”                                                   |
| Tue. Dec. 1    | <b>Undergrad Reading and Analysis Exercise 2 Due</b><br>Wind-up and evaluations                                                               |

### Corpus Linguistics—Fall 2016

Tuesday, Dec. 6

**Vocabulary Quiz**

### EXAM WEEK

Thur. Dec. 15 2:00 -4:30 pm. (note longer meeting time)

In-class presentations on final projects

**Undergrad Reading and Analysis Exercise 3 Due**  
**Grad students’ final written paper Due**

### Additional Dates to Note

**Mon. Sept. 12** Census date (last day to add a class)  
**Wed. Nov. 2** Last day to drop a course  
**Wed. Dec. 21** Grades available: <http://www.uta.edu/mymav>

### Schedule updates

As the instructor for this course, I reserve the right to adjust this schedule in any way that serves the educational needs of the students enrolled in this course. If there are any changes from the paper copy given out on the first day of class, the most current course schedule of readings and assignments and any additional links to citations and readings that come up in class will be uploaded in Blackboard. You are responsible for checking the site regularly.