

## LING 5550 Corpus Linguistics

Fall 2018

Instructor: Xian Zhang (Sean)

Time: Wed 6:30pm - 9:20pm

Location: Discovery Park D212

Email: [xian.zhang@unt.edu](mailto:xian.zhang@unt.edu),

Office location: Discovery Park G178C

Office Hours: Wed 5:00 pm – 6:00 pm, or by appointment

Objectives: This course offers a hand-on introduction to corpus linguistics. Students would learn to use the necessary tools and techniques to conduct corpus-based research. Specific objectives of the course include:

1. Understand corpus-based linguistic research.
2. Learn necessary statistical technique for analyzing corpus data.
3. Learn to use corpora to assist language teaching and assessment.
4. Implement corpus method to conduct a corpus-based research project.

### Required Textbook:

1. O'Keeffe A. & McCarthy, M. (2010). *The Routledge handbook of corpus linguistic* (**Handbook**). Available online via UNT library.

### Recommended Textbook:

I would recommend the first book given below if you are interested in text annotation and parsing. We would also read some chapters in the second book available online. Other reading materials would be available on blackboard.

2. Lu, X. (2014). *Computational Methods for Corpus Annotation and Analysis*. Springer.
3. Wynne, M. (Ed.) (2005). [\*Developing Linguistic Corpora: A Guide to Good Practice\*](#). Oxbow Books.

### Course Requirements:

1. (30%) **Assignments**. There would be three sets of assignments that would help students practice what has been learned in the class.
2. (10%) **Presentation**. On week 12 and Week 14, each student would need to find a corpus related article and present the paper to the class. Sign-up [here](#).
3. (50%) **Final project**. Select a project from the list below. You may need to propose the topic and get the instructor's approval. The final outcome of the project is a paper that can be developed into a publication (15-20 pages double space **Due 11:59 pm, Dec 8<sup>th</sup>**).
  - 1) Develop a word list for a specific field/area/genre (e.g. forensic, management, statistics, computer science). For this project, you will need to collect data to build a corpus so as to extract words/phrases that are unusually frequent in the corpus as compared to a general corpus. This project will have a strong implication for language teaching and learning. (Readings: wordlist.zip)

- 2) Compare two types of texts (e.g. novel vs. news). For this project, you will need to collect data to build two corpora. You will need to use corpus tools to analyze and compare the two types of texts.
  - 3) Test a linguistic theory. In this project, you would to test a linguistic theory/hypothesis using corpus data. (Reading: handbook chapter 15-21)
  - 4) Develop a program that can automatically retrieve data from a large corpus to produce exercise for English learners to practice grammars or vocabulary.
4. (5%) **Final project presentation.** Present your final project at the end of the semester (40 minutes).
  5. (5%) **Symposium.** You will need to write a summary of your project and post it on blackboard. You will also give comments to the projects of your fellow classmates.

### **Course Grade based on 100 points**

100-95	A
94-90	A-
89-85	B+
84-80	B
79-76	B-
75-70	C
69-66	C-
65-60	D
59-0	F

### **Academic Dishonesty and Integrity**

Students are expected to be familiar with UNT student standards of academic integrity that can be found at [https://policy.unt.edu/sites/default/files/untpolicy/pdf/7-Student\\_Affairs-Academic\\_Integrity.pdf](https://policy.unt.edu/sites/default/files/untpolicy/pdf/7-Student_Affairs-Academic_Integrity.pdf).

### **Disability Accommodation Statement**

“The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with an accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the Office of Disability Accommodation website at <http://www.unt.edu/oda>. You may also contact them by phone at 940.565.4323.”

## Tentative schedule

Weeks	Dates	Topics	Readings	Dues
Week 1	Aug. 29.	Overview		HW1
Week 2	Sep. 05.	Introduction	Chapter 1 & 2 (Handbook)	
Week 3	Sep. 12.	Building a corpus	Wynne (Chapter 1) Chapter 9 (Handbook) Chapter 3 (Handbook)	
Week 4	Sep. 19.	Corpus query I		HW1 due
Week 5	Sep. 26.	Corpus query II		
Week 6	Oct. 03.	Corpus tools I	ANTCONC help Chapter 11 (Handbook)	HW2(i)
Week 7	Oct. 10.	Corpus tools II	Chapter 2 (Wynne) Crossley et al (2011)	HW2(ii)
Week 8	Oct. 17.	Database I		
Week 9	Oct. 24.	Database II	<a href="#">Help file</a>	HW2(i) due
Week 10	Oct. 31.	Final project development	Wordlist.zip Collocation.zip Assessment.zip Formulaic language.zip	HW2(ii) due
Week 11	Nov. 07.	Corpus for assessment	Lu (2011) Lu(2013)	HW 3
Week 12	Nov. 14.	Corpus Statistics		
Week 13	Nov. 21.	A symposium	Chapter 15-21 (Handbook)	HW 3 due
Week 14	Nov. 28.	Corpus for teaching	Romer (2011) Liu & Jiang(2009)	
Week 15	Dec. 05.	Project presentation	Class evaluation	Final project due on Dec. 8