

Ling 571 Computational Corpus Linguistics

San Diego State University

Schedule # 22196

Fall 2021

TuTh 9:30am–10:45am

NE-085

Advances in technology have revolutionized the way linguists approach their data. Using computers, extremely large bodies of text (“corpora”) can be collected and analyzed at a level of detail that only a decade ago would have been unthinkable. Sources like the British National Corpus, the Corpus of Historical American English, and Google Books collection allow us access to language use across an unprecedented range of time and space. For anyone studying human communication or culture, the accelerating growth of the World Wide Web and other natural language resources have made techniques for dealing with very large texts more important than ever.

Through a combination of lectures, demonstrations, and hands-on exercises, this course will give students an introduction to the skills necessary for computer-aided text manipulation. Students will learn to search text databases using on-line tools, to write Python programs to manipulate large natural language corpora, to apply quantitative linguistic measures to existing texts, and to formulate, carry out, and describe their own corpus-based linguistic research projects.

This class has no pre-requisites.

Prof. Rob Malouf

Website: malouf.sdsu.edu

Email: rmalouf@sdsu.edu

Office hours: Tue & Thu 1:00pm–2:00pm, or by appointment

Real office: SHW 244

Zoom office: SDSU.zoom.us/j/88612436627

Students are provided with an SDSU Gmail account, and this SDSU email address will be used for all communications. University Senate policy notes that students are responsible for checking their official university email once per day during the academic term. For more information, please see Student Official Email Address Use Policy here: senate.sdsu.edu/policy-file/policies/facilities#collapsed20e126_12

Student Learning Outcomes

By the end of the course, students will be able to:

- Articulate the principles of good corpus design
- Analyze patterns of word use in a corpus using concordancing tools
- Implement quantitative corpus analyses in Python
- Evaluate linguistic hypotheses on the basis of corpus data
- Apply corpus-based techniques to social media texts

Course Materials

You need to acquire two required textbooks for this course:

- Vaclav Brezina. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
- Martin Weisser. 2016. *Practical Corpus Linguistics: An Introduction To Corpus-Based Language Analysis*. Wiley.

These should be available from the campus bookstore (or the usual other places) and you should do your best to get them. Weisser is also available to check out from the SDSU library in digital form.

Other readings will be distributed via Canvas:

- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Press. www.nltk.org/book
- Allen B. Downey. 2016. *Think Python: How to Think Like a Computer Scientist (2nd edition)*. Green Tea Press. greenteapress.com/wp/think-python-2e
- Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.
- Anatol Stefanowitsch. 2020. *Corpus Linguistics: A Guide to the Methodology*. Language Science Press. langsci-press.org/catalog/book/148
- Martin Wynne (ed.). 2005. *Developing Linguistic Corpora: A Guide to Good Practice*. Oxbow Books. users.ox.ac.uk/~martinw/dlc/index.htm

Course Design

Course grades will be based on regular homeworks, a take-home mid-term project, and a final project:

- Homeworks (30%)
- Midterm project (30%)
- Final project (40%)

Tentative due date for the midterm is October 20 and for the final is December 17.

Grading Policies

The final grade will be based on homework assignments (30%), a midterm project (30%), and a final project (40%). Through the term, there will be several hands-on homework assignments in which students apply the techniques learned in class to actual corpus materials. Since it's important to not get behind on assignments, late assignments will be accepted for partial credit **for one week only** after the due date unless prior arrangements are made. Working in groups is encouraged, but please include the names of all coworkers on the assignment.

The midterm will be a take-home programming assignment, for which students will be required to replicate a published corpus analysis using Python. The final project should be a program (with documentation) to perform some substantial corpus processing task chosen by the student. Alternatively, the final project can be the collection and annotation of a new corpus, or a research project that makes crucial use of novel corpus data. More details about both projects will be given later in the term.

Schedule

Module	Topic	Readings
1	Intro	Stefanowitsch, ch. 2&3; Wynne, ch. 1
2	Word meanings	Hanks, ch. 1; Weisser, ch. 5&8
3	Frequency	Brezina, ch. 1&2; Weisser, ch. 9
4	Keywords	Brezina, ch. 3
5	Collocations	Stefanowitsch, ch. 7; Weisser, ch. 10
6	Annotation	Weisser, ch. 7; Wynne, ch. 2
7	Python	Bird, ch. 1&2; Downey
8	Corpus representation	Bird, ch. 3; Weisser, ch. 4&11; Wynne, ch. 3&4
9	Lexical dispersion	
10	Lexical diversity	
11	Lexicogrammar	Brezina, ch. 4; Stefanowitsch, ch. 8
12	Social media	Bird, ch. 5&11

University Policies

Effective Fall 2021, students who register for face-to-face classes are expected to attend as indicated in the course schedule. Faculty teaching face-to-face courses will not be required to create a new, alternative on-line class as an accommodation for any student.

Students with medical conditions that would present a COVID-related risk in a face-to-face instructional setting should contact the Student Ability Success Center (sdsu.edu/sasc) to begin the process of getting support. Students who do not adhere to the Covid19 Student Policies (csrr.sdsu.edu) or the directives of their faculty will be directed to leave the classroom and will be referred to the Center for Student Rights and Responsibilities.

Do not come to campus if you do not feel well. Remain home and monitor your symptoms and seek medical attention as needed.

Campus vaccination policy

On July 27, 2021, the California State University (CSU) system announced that students, faculty, and staff, including auxiliary employees will need to be immunized against SARS-CoV-2, the virus that causes COVID-19, with a vaccine record on file in order to access campus this fall. As outlined by the CSU, this requirement is not contingent on the full U.S. Food and Drug Administration (FDA) approval, and therefore removes some of the earlier uncertainty regarding the policy's effective date.

SDSU will continue to operate in accordance with all federal, state, and county public health guidelines, and in compliance with CSU policies. The university will continue to prioritize the safety of students, faculty, staff, and community, while seeking to fulfill its educational mission. Visit the university's COVID-19 website (sacd.sdsu.edu/student-health-services/covid-19) frequently, as the site is updated with current information.

Campus facial covering policy

SDSU's full facial covering policy is available online (sacd.sdsu.edu/student-health-services/covid-19/general-prevention#facial-covering-policy) and may be updated again in the fall. Please refer to this URL for any fall 2021 updates, as the site will be kept current.

Medical-related absences

Students are instructed to contact their professor/instructor/coach in the event they need to miss class, etc. due to an illness, injury or emergency. All decisions about the impact of an absence, as well as any arrangements for making up work, rest with the instructors. Student Health Services (SHS) does not provide medical excuses for short-term absences due to illness or injury. When a medical-related absence persists beyond five days, SHS will work with students to provide appropriate documentation. When a student is hospitalized or has a serious, ongoing illness or injury, SHS will, at the student's request and with the student's consent, communicate with the student's instructors via the Vice President for Student Affairs and may communicate with the student's Assistant Dean and/or the Student Ability Success Center.

Resources for students

Need help finding an advisor, tutor, counselor, or require emergency economic assistance? The SDSU Student Success Help Desk (studentsuccess.sdsu.edu) is here for you. Student assistants are available during the academic term via Zoom Monday through Friday, 9:00 AM to 4:30 PM to help you find the office or service that can best assist with your particular questions or concerns.

A complete list of all academic support services—including the Writing Center and Math Learning Center—is available on the Student Affairs' Academic Success website (go.sdsu.edu/student_affairs/academic_success.aspx). Counseling and Psychological Services (619-594-5220, sdsu.edu/cps) offers a range of psychological services for students. Emergency support is available after hours at the same phone number. The San Diego Access and Crisis Line can also be accessed 24 hours/day (1-888-724-7240).

Academic honesty

The University adheres to a strict policy prohibiting cheating and plagiarism. Examples of academic dishonesty include but are not limited to:

- copying, in part or in whole, from another's test or other examination;
- obtaining copies of a test, an examination, or other course material without the permission of the instructor;
- collaborating with another or others in work to be presented without the permission of the instructor;
- falsifying records, laboratory work, or other course data;
- submitting work previously presented in another course, if contrary to the rules of the course;
- altering or interfering with grading procedures;
- assisting another student in any of the above;
- using sources verbatim or paraphrasing without giving proper attribution (this can include phrases, sentences, paragraphs and/or pages of work);
- copying and pasting work from an online or offline source directly and calling it your own;
- using information you find from an online or offline source without giving the author credit;
- replacing words or phrases from another source and inserting your own words or phrases.

Unauthorized recording or dissemination of virtual course instruction or materials by students, especially with the intent to disrupt normal university operations or facilitate academic dishonesty, is a violation of the Student Conduct Code. This includes posting of exam problems or questions to on-line platforms. Violators may be subject to discipline.

The California State University system requires instructors to report all instances of academic misconduct to the Center for Student Rights and Responsibilities. Academic dishonesty will result in disciplinary review by the University and may lead to probation, suspension, or expulsion. Instructors may also, at their discretion, penalize student grades on any assignment or assessment discovered to have been produced in an academically dishonest manner.

Classroom conduct standards

SDSU students are expected to abide by the terms of the Student Conduct Code in classrooms and other instructional settings. Prohibited conduct includes:

- Willful, material and substantial disruption or obstruction of a University-related activity, or any on-campus activity.
- Participating in an activity that substantially and materially disrupts the normal operations of the University or infringes on the rights of members of the University community.
- Unauthorized recording, dissemination, or publication (including on websites or social media) of lectures or other course materials. This includes posting of exam problems or questions to on-line platforms.
- Conduct that threatens or endangers the health or safety of any person within or related to the University community, including physical abuse, threats, intimidation, or harassment; or sexual misconduct.

Violation of these standards will result in referral to appropriate campus authorities.

Accommodations

SDSU via the Student Ability Success Center (SASC) provides accommodations for students with documented disabilities or medical conditions covered under the Americans with Disabilities Act (ADA). In keeping with current public health guidance, I cannot provide arrangements to students without an ADA-qualified disability or medical condition.

If you are a student with a disability and are in need of accommodations for this class, please contact the Student Ability Success Center at sascinfo@sdsu.edu (or go to sasc.sdsu.edu) as soon as possible. Please know accommodations are not retroactive, and I cannot provide accommodations based upon disability until I have received an accommodation letter from the Student Ability Success Center. SASC registration and accommodation approvals may take up to 10–14 business days, so please plan accordingly.

Student privacy and intellectual property

The Family Educational Rights and Privacy Act (FERPA) mandates the protection of student information, including contact information, grades, and graded assignments. I will use Canvas to communicate with you, and I will not post grades or leave graded assignments in public places. Students will be notified at the time of an assignment if copies of student work will be retained

beyond the end of the semester or used as examples for future students or the wider public. Students maintain intellectual property rights to work products they create as part of this course unless they are formally notified otherwise.

Religious observances

According to the University Policy File, students should notify the instructors of affected courses of planned absences for religious observances by the end of the second week of classes.

Sexual violence / Title IX mandated reporting

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I am a mandated reporter in my role as an SDSU employee. It is my goal that you feel able to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep the information you share private to the greatest extent possible. However, I am required to share information regarding sexual violence on SDSU's campus with the Title IX coordinator, Gail Mendez (619-594-6464). She (or her designee) will contact you to let you know about accommodations and support services at SDSU and possibilities for holding accountable the person who harmed you. Know that you will not be forced to share information you do not wish to disclose and your level of involvement will be your choice. If you do not want the Title IX Officer notified, instead of disclosing this information to your instructor, you can speak confidentially with the following people on campus and in the community. They can connect you with support services and discuss options for pursuing a University or criminal investigation. Sexual Violence Victim Advocate 619-594-0210 or Counseling and Psychological Services 619-594-5220, psycserv@sdsu.edu. For more information regarding your university rights and options as a survivor of sexual misconduct or sexual violence, please visit titleix.sdsu.edu or sdsutalks.sdsu.edu.

SDSU Economic Crisis Response Team

If you or a friend are experiencing food or housing insecurity, or any unforeseen financial crisis, visit sdsu.edu/ecrt for more information or to submit a request for assistance.

SDSU's Economic Crisis Response Team (ECRT) aims to bridge the gap in resources for students experiencing immediate food, housing, or unforeseen financial crises that impact student success. Using a holistic approach to well-being, ECRT supports students through crisis by leveraging a campus-wide collaboration that utilizes on- and off-campus partnerships and provides direct referrals based on each student's unique circumstances. ECRT empowers students to identify and access long-term, sustainable solutions in an effort to successfully graduate from

SDSU. Within 24 to 72 hours of submitting a referral, students are contacted by a member of ECRT and are quickly connected to the appropriate resources and services.

For students who need assistance accessing technology for their classes, visit our ECRT website (sdsu.edu/ecrt) to be connected with the SDSU library's technology checkout program. The technology checkout program is available to both SDSU and Imperial Valley students.