1 Course Details

1.1 Calendar description

This course introduces the tools and methods of computational and corpus linguistics and discusses their application in the social sciences and the humanities. Specifically, the course examines how the tools and methods of computational and corpus linguistics could be used to answer research questions about society and culture in disciplines such as sociology, history, literary studies, area studies, journalism studies, and communication. Students also learn how to use computational and corpus linguistics tools through workshops and practical activities.

1.2 Course Description

This course provides a survey of the history and development of computational and corpus-based approaches for studying language and literature and introduces a range of topics of practical relevance to the compilation, management, and analysis of linguistic corpora. In doing so, the course establishes foundational knowledge and skills in the application of computational and corpus-based approaches for research on language in the social sciences and humanities. An emphasis is placed on using linguistic corpora for studying questions about the intersection of language, identity, society, culture, power, and politics. Pre-recorded (asynchronous) lectures and live (synchronous) seminars will be supplemented by hands-on workshops offered by Jennifer Marvin, Research & Scholarship Librarian at the McLaughlin Library.

1.3 Methods of Delivery

The course uses a blended model of delivery, combining asynchronous (pre-recorded) video lectures and synchronous (live on MS Teams) seminars. Video lectures will be posted on Course Link on Tuesday each week. In addition, we will meet each week for seminars on Thursdays between 1:00-2:00 pm. All course materials will be shared through Course Link and MS Teams.
1.4 Timetable

- **Video lectures** will be posted on MS Teams each Tuesday and linked to Course Link.
- **Live seminars** will be held on MS Teams each Thursday between 1:00-2:00 pm.

Course Link site: [https://courselink.uoguelph.ca/d2l/home/648388](https://courselink.uoguelph.ca/d2l/home/648388)

MS Teams channel: [https://tinyurl.com/yymrdczj](https://tinyurl.com/yymrdczj)

1.5 Final Exam

11:30AM - 01:30PM (2020/12/12) on Course Link

The exam will include (1) multiple-choice questions, (2) short-answer questions, and (3) one short essay question (300-500 words). The exam will take place on **December 12, 2020 between 11:30AM – 01:30PM** on Course Link. It will be an open-book exam, meaning that you can consult the course readings and other resources while answering questions. Three days (72 hours) before the exam, on **December 9, 2020 by 11:30AM**, I will share three essay questions, only one of which will come up on the exam. On the day of the exam, you will have a total of two (2) hours on Course Link to answer the multiple-choice and short-answer questions (approx. 30-40 min) and write the short essay (approx. 80-90 min).

________________________________________________________________________

2 Instructional support

**Instructor:** Asen O. Ivanov, PhD  
**Email:** ivanova@uoguelph.ca  
**Office Hours (on MS Teams):** Thursdays 2:00-3:00 pm or by appointment

________________________________________________________________________

3 Learning Resources

3.1 Recommended Textbooks


4 Learning Outcomes

4.1 Course Learning Outcomes

By successfully completing the course, you will be able to:

- Understand the difference between computational and corpus linguistics
- Analyze the development of ideas and technologies in computational and corpus linguistics
- Understand the role of computational and corpus linguistics in applied linguistics and the social science and humanities more broadly
- Analyze and evaluate the design and quality of linguistic corpora
- Evaluate the suitability of linguistic corpora for answering research questions of intellectual and social significance
- Create corpus, mark-up, and annotation and understand the principles of corpus compilation, management, and preservation
- Evaluate the theoretical and methodological soundness of research using computational and corpus linguistic tools and methods
- Create research proposals that use the methods and tools of computational and corpus linguistics.

5 Assessment

<table>
<thead>
<tr>
<th>#</th>
<th>Assignments</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Exam</td>
<td>20%</td>
</tr>
<tr>
<td>A1</td>
<td>Written Response &amp; Class Lead</td>
<td>15%</td>
</tr>
<tr>
<td>A2</td>
<td>Corpus Review and Design (group)</td>
<td>20%</td>
</tr>
<tr>
<td>A3</td>
<td>Research Project Design</td>
<td>30%</td>
</tr>
<tr>
<td>A4</td>
<td>Digital Tools Workshops</td>
<td>15%</td>
</tr>
</tbody>
</table>

5.1 Assessment details

Exam (20%)

The exam will include (1) multiple-choice questions, (2) short-answer questions, and (3) one short essay question (300-500 words). The exam will take place on December 12, 2020 between 11:30AM – 01:30PM on Course Link. It will be an open-book exam, meaning that you can consult the course readings and other resources while answering questions. Three days (72 hours) before the exam, on December 9, 2020 by 11:30AM, I will share three essay questions, only one of which will come up on the exam. On the day of the exam, you will have a total of two (2) hours on Course Link to answer the multiple-choice and short-answer questions (approx. 30-40 min) and write the short essay (approx. 80-90 min).
Assignment 1 Written response and class lead (15%)
In selected classes, a group of students will post on the course discussion board a personal 250-500-word analysis and reflection of the weekly readings. In addition, in that week, they are responsible for attending and leading class discussion during the Thursday seminar — i.e., engaging in conversation, asking questions, offering reflections on the course material, etc.

Assignment 2 Corpus Review and Design (20%)
Working in groups, you will closely research at least two online corpora. Specifically, you will be asked to study and report on available documentation about the two corpora and describe their mode, origin, constitution, medium, style, topic, date, type of annotation and compiler(s) and/or current custodians. In addition, you will reflect on what research questions the corpora were created to answer and highlight previous studies with the corpora. You are encouraged to contact the compiler(s) of the two corpora to interview them over email or video call. Deliverables are (1) a written report (approx. 1000 words) and a short presentation (10 min) that profiles the two corpora and highlights key insights and learning points from your research.

Assignment 3 Research Project Design (30%)
Working individually, you will develop a research design and proposal (2000-2500 words) for a project at the intersection of your primary discipline(s) of specialization, and/or of research interest to you, and corpus linguistics. In other words, the central condition is that, irrespective of its disciplinary focus, your project must be based on the empirical study of linguistic corpus. Your proposal has to describe the research questions and anticipated contribution of the project; the mode, origin, constitution, medium, style, topic, date, and annotations of the corpus to be used; and the analytical approach and tools you plan to use. You will also be asked to evaluate and reflect on the theoretical and methodological soundness of your project and its potential contribution to knowledge.

This assignment requires independent research and in-depth study. Higher grades will be awarded to projects demonstrating scientific and/or societal contribution and interdisciplinary innovativeness (or the spirit thereof). Deliverables are (1) Topic, research questions, lit review, and research plan (1000 words; due after reading week); (2) final proposal (1500-2500 words, due at the end of the course).

Assignment 4: Digital Tools Workshops Participation (15%; 3.75% per workshop)
Assessment for the workshops will be based on the submission of a proof of workshop participation, including analysis and reflections.
6 Learning Activities

6.1 Video Lectures and Seminars

**Week 1**

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Sept</td>
<td>Introduction to the Course</td>
<td>The course goals, outline, learning outcomes, assessment, design &amp; philosophy; Computational and corpus linguistics: What is the difference?; The strengths and limitations of using computational approaches to the study of language and literature; Language, identity, society, and culture</td>
</tr>
</tbody>
</table>

**Readings:**

N/A

**Additional Readings:**


**Week 2**

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Sept</td>
<td>Lecture</td>
<td>Approaches and methods for language processing in computational linguistics and the adjacent fields of natural language processing in computer science and speech recognition in electrical engineering.</td>
</tr>
<tr>
<td>17 Sept</td>
<td>Seminar</td>
<td>The development of corpus linguistics and its relationship to computational linguistics, applied linguistics, social sciences, and the (digital) humanities.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A survey of major corpora available online</td>
</tr>
</tbody>
</table>
Readings:

Weisser (2016), Chapter 2

Additional Readings:
McEnery & Hardie (2011), Chapter 1
McEnery & Hardie (2011), Chapter 9 (pp. 227-233 only)
Weisser (2016), Chapter 1

Week 3
22 Sept
Lecture
24 Sept
Seminar

Corpus Design, Mark-up, and Annotation

Topics:
- Principles of corpus design: mode, sampling, size, balance, representativeness, extent, reach, depth, and strength.
- Best-practices and standards for corpora mark-up and annotation
- Typology of corpora annotations

Readings:
Weisser (2016), Chapters 3 & 11

Additional Readings:
McEnery & Hardie (2011), Chapter 2 (pp. 29-35 only)


Week 4  
29 Sept  
Lecture  
1 Oct  
Seminar  

Corpus Analysis Tools and Application  

Topics:  
- Computational tools for corpus analysis: Concordances, Word Frequencies, Keywords, Ngrams and more  
- Quantitative and qualitative application and use  
- Examples of application and use  

Readings:  

Weisser (2016), Chapters 5, 9, & 10  

Additional Readings:  

McEnery & Hardie (2011), Chapter 2 (pp. 37-48 only)  


Week 5  
6 Oct  
Lecture  
8 Oct  
Seminar  

Finding and Working with Plain Text Data + Intro to Advanced Query Languages  

Topics:  
- Working with plain text data  
- Finding and getting corpora online  
- Compiling corpora from library and internet resources  
- Using online interfaces to query mega corpora  
- Intro to RegEx and CQL  

Readings:  

Weisser (2016) chapters 4 & 6
Additional Readings:

Weisser (2016), chapter 8


<<<<<<<< FALL BREAK>>>>>>

Week 6
20 Oct
Lecture
22 Oct
Seminar

Week 7
27 Oct
Lecture
29 Oct
Seminar

Workshop 2
Counting Tools (Google Ngram, HathiTrust Bookworm, Voyant, AntConc)

Workshop 3
Context Tools (Voyant, Antconc)

Instructor: Jennifer Marvin, Research & Scholarship Librarian at the McLaughlin Library.

Case Studies (read selectively based on your interests):


Corpus-Based Studies of Discourse: Theoretical Foundations

Topics:

- The functionalist view of language as a medium of communication and the influential systemic-functional linguistic perspective (SFL)
- Discourse Analysis—studying the functions of language in context
- Corpus-based studies of discourse
- Heuristics for research
Readings:


Additional Readings:

McEnery & Hardie (2011), Chapter 6 & 7


---

Week 8
3 Nov
Lecture - 5 Nov
Seminar

Corpus-Based Studies of Discourse: Research Applications

Topics:

- We continue our discussion from the previous week by surveying corpus-based studies of discourse.

Readings:


Additional Readings:

McEnery & Hardie (2011), chapter 8


Workshop 4: Natural Language Processing Tools (NER & Sentiment Analysis)

Workshop 5: Clustering Tools (Topic Modelling, Gephi)

Instructor: Jennifer Marvin, Research & Scholarship Librarian at the McLaughlin Library.

Case Studies (read selectively based on your interests):


Multimodal Discourse Analysis

Topics:
- Introduction to multimodal discourse analysis
- Multimodal texts
- The corpus-based study of multimodal discourse

Readings:


Additional Readings:


**Corpus-based studies in Applied Linguistics and Sociolinguistics**

**Topics:**
- Overview of the opportunities and challenges of using corpora in applied linguistics and sociolinguistics

**Readings:**


**Additional Readings:**


**Computer-Assisted Language Learning & Course Wrap-up**

**Topics:**
- Overview of the history, application, and current developments of the use of corpus linguistics in language teaching and learning
- Synthesis of the knowledge and skills gained in the course
- Final exam prep and discussion

**Readings:**


Additional Readings:


7 Course Statements

7.1 Course Schedule

The course reflects a mixture of professor-supported independent research (office meetings) and in-class interactions (class meetings). You are strongly advised to make use of individual consultations with the professor to discuss your research.

To fully gain from the course, you are expected to engage with the course material for the duration of the course. This includes: (1) reading the required weekly readings, (2) reviewing the lecture notes and videos, (3) completing some or all of the weekly course activities, (4) and regularly participating in live seminar discussions.

7.2 Online Interactions

Inappropriate behaviour in class cannot be tolerated in keeping with the University of Guelph’s code of ethical conduct. In an online environment, inappropriate behaviour includes (but is not limited to) using obscene, threatening, inflammatory, or offensive language; plagiarism (in any form); selling course materials online; recording class discussions without the explicit permission of all parties.

7.3 Recording of Lecture Materials

Thursday's live seminars on MS Teams will not be recorded. However, depending on the circumstances and your comfort level on that day, feel free to turn off your camera, mute your microphone, and ask questions through the chat.

Presentations that are made in relation to course work - including lectures - cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.
8 University Statements

8.1 Email Communication

As per university regulations, all students are required to check their e-mail account regularly: e-mail is the official route of communication between the University and its students.

8.2 When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. The grounds for Academic Consideration are detailed in the Undergraduate and Graduate Calendars.

Undergraduate Calendar - Academic Consideration and Appeals
https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml

Graduate Calendar - Grounds for Academic Consideration
https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml

8.3 Drop Date

Students will have until the last day of classes to drop courses without academic penalty. The deadline to drop two-semester courses will be the last day of classes in the second semester. This applies to all students (undergraduate, graduate and diploma) except for Doctor of Veterinary Medicine and Associate Diploma in Veterinary Technology (conventional and alternative delivery) students. The regulations and procedures for course registration are available in their respective Academic Calendars.

Undergraduate Calendar - Dropping Courses
https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml

Graduate Calendar - Registration Changes
https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml

8.4 Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

8.5 Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.
When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required; however, interim accommodations may be possible while that process is underway. Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability. Use of the SAS Exam Centre requires students to book their exams at least 7 days in advance and not later than the 40th Class Day.

For Guelph students, information can be found on the SAS website https://www.uoguelph.ca/sas

8.6 Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community-faculty, staff, and students-to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that encourages academic integrity. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar - Academic Misconduct
https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml

Graduate Calendar - Academic Misconduct
https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml

8.7 Recording of Materials

Presentations that are made in relation to course work - including lectures - cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

8.8 Resources

The Academic Calendars are the source of information about the University of Guelph’s procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.

Academic Calendars https://www.uoguelph.ca/academics/calendars

8.9 Disclaimer
Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings and academic schedules. Any such changes will be announced via CourseLink and/or class email. All University-wide decisions will be posted on the COVID-19 website (https://news.uoguelph.ca/2019-novel-coronavirus-information/) and circulated by email.

8.10 Illness

The University will not normally require verification of illness (doctor's notes) for fall 2020 or winter 2021 semester courses. However, requests for Academic Consideration may still require medical documentation as appropriate.