University of Guelph
School of Languages and Literatures

LING*3020
Language & Technology
Fall 2020

Instructor: Asen O. Ivanov, PhD
Email: ivanova@uoguelph.ca
Prerequisite: LING*1000

Please note: This is a preliminary web course description only. The department reserves the right to change without notice any information in this description. The final, binding course outline will be distributed in the first class of the semester.

Course description:

This course introduces the tools and methods of computational and corpus linguistics and discusses their application to the social sciences. 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. Parallel to that, students will learn how to use computational and corpus linguistics to answer research questions in applied linguistics. Pre-recorded (asynchronous) lectures and live (synchronous) seminars will be supplemented by hands-on workshops in corpora analysis, mark-up, and annotation offered in collaboration with the McLaughlin Library.

Methods of Delivery:

Blended model—synchronous (live) and asynchronous (pre-recorded). Pre-recorded (asynchronous) weekly lectures and online learning materials will be posted on Course Link for each module. In addition, we will meet twice each week live on Zoom: once a week on Thursdays for live (synchronous) seminars (50 min), and once a week on Tuesday for live (synchronous) open office hours (50 min). The seminars will be used to discuss and clarify the course readings, and the open office hours will provide additional opportunity to discuss any course-related matters beyond the course readings (including the exam, capstone assignment, group work, etc.).

Instructor bio:

Asen O. Ivanov is the Michael Ridley Postdoctoral Fellow in Digital Humanities at the College of Arts at the University of Guelph, where he is affiliated with the THINC Lab, the School of Languages and Literatures, and the Research & Scholarship team at the McLaughlin Library.

Modules & Classes:
1. **Module I — Linguistics and Technology: Foundations and Application of Computational and Corpus Linguists (length: four weeks)**

This module provides an overview of the history and development of probabilistic, logic-based, and machine-learning models and algorithms developed in computational linguistics and the adjacent fields of computer science and electrical engineering. The module further provides an overview of the history and application of computational approaches in the domain of corpus linguistics. The module, thus, clarifies the relationship between computational and corpus linguistics—two leading research orientations at the intersection of linguistics and technology. The module concludes with three classes examining the opportunities and challenges of using computational and corpus approaches in applied linguistics.

2. **Module II — Linguistic, Society, & Culture: Theoretical Frameworks, and Approaches for empirical research (length: four weeks)**

This module provides a selective overview of analytical perspectives that have had a lasting influence on linguistic research and the study of culture and society more broadly. Including, generative grammar and systemic-functional linguistics; theoretical perspectives in sociolinguistics; multimodal discourse analysis; sociology and history.

3. **Module III — Data and Materials/ Sources and Evidence: Creating and Using Linguistic Corpora (length: two weeks)**

This module provides students with readings and hands-on exercises in identifying, analyzing, and using linguistic corpora. Students will learn how to find and use already existing corpora, how to create their own corpus, and the basics of corpora mark-up and annotation. They will also learn about principles of, the need for, corpora curation and preservation.

4. **Module IV — Linguistics, Technology, and Empirical Research (length: two weeks)**

This module provides students with hands-on training on installing and using turnkey tools for linguistic analysis, including word frequencies, collocation, concordance, n-gram, sentiment analysis, and topic modelling. In addition to the hands-on training, for each workshop, the students are assigned case-study readings, which demonstrate how the data output of a tool (e.g., topic modelling) can be used to answer specific research questions about society and culture. Four (4) of the five (5) workshops will be delivered by Jennifer Marvin, Research & Scholarship Librarian at the McLaughlin Library.

**Technology:** Instructions on how to install and use the digital tools will be provide in advance of the workshops. Students are required to have access to working laptop or desktop computer on which they can install and run software.

**Learning outcomes:**

By successfully completing the course, students will:

- Understand the difference between computational and corpus linguistics (module I)
- Analyze the historical development of ideas and technologies in computational and corpus linguistics (module I)
- Understand the role of computational and corpus linguistics in the social science and humanities and applied linguistics more generally (module II)
- Analyze and evaluate the design and quality of linguistic corpora (module III; A2)
• Evaluate the suitability of linguistic corpora for answering specific research questions (module II & III; A3)
• Create corpus mark-up and annotation and understand the need for, and principles of, corpus curation and preservation (module III)
• Evaluate the theoretical and methodological soundness computational and corpus linguistics research proposals (A3)
• Create research proposals that leverage the tools of computational and corpus linguistics (A3, module II, module IV)

Methods of Evaluation:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exam</td>
<td>20%</td>
</tr>
<tr>
<td>2</td>
<td>A1- Written Response &amp; Class Lead</td>
<td>15%</td>
</tr>
<tr>
<td>3</td>
<td>A2 - Corpus Review and Design</td>
<td>20%</td>
</tr>
<tr>
<td>4</td>
<td>A3 - Research Project Design</td>
<td>30%</td>
</tr>
<tr>
<td>5</td>
<td>A4 - Digital Tools Workshops</td>
<td>15%</td>
</tr>
</tbody>
</table>

* Additional instructions and rubric for all exams and assignments will be provided in advance of the due dates.

(1) Exam: Multiple choice test
Virtual-proctored exam on Course Link. Based on the course readings.
**Weight:** 20%

(2) Assignment 1: Written response & class lead
Each class, randomly selected group of students will individually post on the course discussion board a 250-500-word summary and reflection on the weekly readings. In addition, on that week, they are responsible for attending and leading/engaging in class discussion—i.e., asking questions, offering reflections on the course material, etc.

Students will post to a maximum of two posts (depending on final course enrollment)
**Weight:** 15% — 10% written response; 5% class discussion

(3) Assignment 2: Corpus Review and Design (Report & Group Presentation)
Working in groups of 5-6, the students will research and develop a written report (1000-1500 words) on at least two online corpora (a list to choose from will be provided). They will research all available documentation online to report on the mode, origin, constitution, medium, style, topic, date, type of annotation and author(s) of the corpora. In addition, they will contact the authors of the two corpora to interview them (over email or on Zoom/Skype). Interview topics and questions will be provided, but students are also expected to develop their own questions. Students will record a ten-min presentation highlighting the key learning points in their report and post it on Course Link.
**Weight:** 20% — 15% written report (1000-1500 words) and 5% pre-recorded presentation (10 min) posted on Course Link.
(4) Assignment 3: Research Project Design

Working individually, students will develop a research design (2000-2500 words) for a project in their primary discipline of specialization that uses computational and cuprous linguistic tools. The proposal will describe the project’s research questions and anticipated contribution; the mode, origin, constitution, medium, style, topic, date, and type of annotation of corpus to be used; and will identify and discuss an analytical approach and tools. In addition, the students will be asked to evaluate and reflect on the theoretical and methodological soundness of their project within their primary discipline of specialization and/or topic of their project.

Please Note: This capstone assignment requires students to conduct extensive independent research and to substantively integrated into their project knowledge and perspectives drawn from their primary areas of specialization at the university. Higher grades will be awarded to projects demonstrating scientific and/or societal contribution and interdisciplinary innovativeness (or the spirit thereof).

Weight: 30%

(5) Assignment 4: Digital Tools Workshops Participation

Assessment will be based on the submission of a proof of workshop completion (3% per workshop).

Weight: 15%