Term Information

Effective Term	Spring 2024
Previous Value	*Autumn 2022*

Course Change Information

What change is being proposed? (If more than one, what changes are being proposed?)

We are proposing to move this course from the GEN Mathematical and Quantitative Reasoning or Data Analysis (MQRDA) Foundation category to the Number, Nature, Mind Theme. We are also slightly updating the course description, and we have removed an exclusion that made reference to the quarter-based version of this course.

What is the rationale for the proposed change(s)?

The course fits the Goals and ELOs of the Number, Nature, Mind theme. Given the large number of other courses in the MQRDA Foundation category, including some courses from our own department, we expect GE enrollment to better support this offering as a Theme course.

Recent instructors requested to update the list of topics included in the course description, based on content that has typically been included over the last several years.

What are the programmatic implications of the proposed change(s)?

(e.g. program requirements to be added or removed, changes to be made in available resources, effect on other programs that use the course)?

none

Is approval of the requrest contingent upon the approval of other course or curricular program request? No

Is this a request to withdraw the course? No

General Information

Course Bulletin Listing/Subject Area	Linguistics
Fiscal Unit/Academic Org	Linguistics - D0566
College/Academic Group	Arts and Sciences
Level/Career	Undergraduate
Course Number/Catalog	3802
Course Title	Language and Computers
Transcript Abbreviation	Lang and Computers
Course Description	Introduction to human language technology, explaining the computational and linguistic principles behind such familiar technologies as web search, spam filtering, and text generation.
Previous Value	*Introduction to human language technology, explaining the computational and linguistic principles behind such familiar technologies as web search, machine translation, and spelling correction.*
Semester Credit Hours/Units	Fixed: 3

Offering Information

Length Of Course	14 Week, 12 Week, 8 Week, 7 Week, 6 Week, 4 Week
Flexibly Scheduled Course	Never
Does any section of this course have a distance education component?	No
Grading Basis	Letter Grade
Repeatable	No

Course Components	Lecture
Grade Roster Component	Lecture
Credit Available by Exam	No
Admission Condition Course	No
Off Campus	Never
Campus of Offering	Columbus, Lima, Mansfield, Marion, Newark, Wooster

Prerequisites and Exclusions

Prerequisites/Corequisites	Prereq: Soph standing or above.
Exclusions	
Previous Value	Not open to students with credit for Linguist 384.
Electronically Enforced	No

Cross-Listings

Cross-Listings

Subject/CIP Code

Subject/CIP Code	16.0102
Subsidy Level	Baccalaureate Course
Intended Rank	Sophomore, Junior, Senior

Requirement/Elective Designation

General Education course:

Mathematical or Logical Analysis; Number, Nature, Mind

The course is an elective (for this or other units) or is a service course for other units

Previous Value

General Education course:

Mathematical or Logical Analysis; Mathematical and Quantitative Reasoning (or Data Analysis)

The course is an elective (for this or other units) or is a service course for other units

Course Details

Course goals or learning objectives/outcomes

- Students will learn to apply quantitative techniques from fields such as statistics and formal language theory to describe and analyze linguistic data.
- Students will learn the basic principles that underlie machine learning systems.
- Students will draw upon their experiences as technology users and language speakers to reason about the strengths and weaknesses of language technologies.
- Students will think critically about the social and ethical implications of language technologies.

Previous Value

Content Topic List

- Search engines
- Spam filtering
- Text generation
- Spell checking
- Machine translation
- Dialog systems
- Text classification

Previous Value

- *Search engines*
- *Spell checking*
- *Machine translation*
- *Dialog systems*
- *Text classification*

Sought Concurrence
No
Previous Value

Attachments

- LING3802-Syllabus_2-20-23.pdf: Syllabus

(Syllabus. Owner: McCullough,Elizabeth Ann)

- LING3802-NNM-submission-form_2-20-23.pdf: NNM Theme worksheet

(Other Supporting Documentation. Owner: McCullough,Elizabeth Ann)

Comments

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	McCullough,Elizabeth Ann	03/06/2023 12:01 PM	Submitted for Approval
Approved	McCullough,Elizabeth Ann	03/06/2023 12:01 PM	Unit Approval
Approved	Vankeerbergen,Bernadette Chantal	03/21/2023 10:46 PM	College Approval
Pending Approval	Jenkins,Mary Ellen Bigler Hanlin,Deborah Kay Hilty,Michael Vankeerbergen,Bernadette Chantal Steele,Rachel Lea	03/21/2023 10:46 PM	ASCCAO Approval

LING 3802: Language and Computers

Class Location | Class Meeting Time | Autumn 2023

Instructor

Instructor:
Email:
Office:
Mailbox:
Office Hours:

Teaching Coordinator

Dr. Hope Dawson
Email: dawson.165@osu.edu
Office: Oxley Hall 114
Phone: 614-292-5420

Prerequisite: Sophomore standing or above. No specific previous coursework in mathematics, statistics, computer science or linguistics is required for this course, although some experience with mathematical or formal reasoning (e.g. from Foundations GE coursework in Mathematical and Quantitative Reasoning) will be helpful. No knowledge of computer programming is required.

Course Description

Can a computer learn to comprehend and produce language like a human does? Current technologies like Alexa and Google Translate show impressive progress toward this goal, but with clear limitations. The potential of language technology is actively being researched and debated and is a matter of both practical and theoretical interest.

This course offers an undergraduate-level introduction to language technology, drawing upon the fields of computational linguistics and natural language processing. We will explore how computers can be used to represent human language and speech, to classify and search through documents, to uncover linguistic structure, and to translate between languages, among other applications. We will also spend some time discussing the social implications of these technologies, and what they show about the nature of intelligence.

General Education Fulfillment

Ohio State is launching a new General Education (GE) curriculum beginning in autumn 2022. Students entering the university in or after that semester are expected to follow the new curriculum, while students who entered the university earlier will continue with the legacy curriculum. LING 3802 meets GE requirements in both the old and new systems, which are outlined below.

Legacy GE Curriculum

In the legacy GE curriculum, this course satisfies the category of Quantitative Reasoning: Math and Logical Analysis. The goals of this category are for students to develop skills in quantitative literacy and logical reasoning, including the ability to identify valid arguments, use mathematical models, and draw conclusions and critically evaluate results based on data. The expected learning outcomes are the following:

- Students comprehend mathematical concepts and methods adequate to construct valid arguments.
- Students comprehend mathematical concepts and methods adequate to understand inductive and deductive reasoning.
- Students comprehend mathematical concepts and methods adequate to increase their general problem-solving skills.

This course will meet these objectives in the following ways:

- Students will apply foundational concepts from fields such as statistics and formal language theory to describe and analyze linguistic data.
- Students will learn quantitative methods to evaluate language technology systems.
- Students will solve linguistic problems using algorithms and statistical models.

New GE Curriculum

In the new GE curriculum, this course satisfies the theme of Number, Nature, and Mind. There are three goals associated with this theme:

- **Goal 1:** Successful students will analyze an important topic or idea at a more advanced and in-depth level than the foundations. In this context, "advanced" refers to courses that are e.g., synthetic, rely on research or cutting-edge findings, or deeply engage with the subject matter, among other possibilities.
    - **Expected Learning Outcomes:** Successful students are able to …
        - 1.1 Engage in critical and logical thinking.
        - 1.2 Engage in an advanced, in-depth, scholarly exploration of the topic or ideas within this theme.
- **Goal 2:** Successful students will integrate approaches to the theme by making connections to out-of-classroom experiences with academic knowledge or across disciplines and/or to work they have done in previous classes and that they anticipate doing in future.
    - **Expected Learning Outcomes:** Successful students are able to …
        - 2.1 Identify, describe, and synthesize approaches or experiences.

        - 2.2 Demonstrate a developing sense of self as a learner through reflection, self-assessment, and creative work, building on prior experiences to respond to new and challenging contexts.
- **Goal 3:** Successful students will experience and examine mathematics as an abstract formal system accessible to mental manipulation and/or mathematics as a tool for describing and understanding the natural world.
    - **Expected Learning Outcomes:** Successful students are able to …
        - 3.1 Analyze and describe how mathematics functions as an idealized system that enables logical proof and/or as a tool for describing and understanding the natural world or human cognition.

This course will meet these objectives in the following ways:

- In this course you will learn to analyze human language using multiple types of mathematical reasoning drawn from fields such as statistics and formal language theory. In-class exercises, class discussions, and homework assignments will give you the opportunity to apply mathematical reasoning to specific sets of language data and to reason about what analyses of the data follow as logical consequences. You will think critically about what these types of reasoning can show about language as a human cognitive system, including its limitations. At the end of the course, you write a final essay on the social or ethical implications of language technology (specific topic of your choice). This will allow you to engage more deeply with the course material in an area of personal interest, to synthesize different perspectives on language technologies, and to draw on mathematical reasoning skills developed during the semester, in order to build a convincing argument about the role of language technology in modern society. Most topics involve issues that can be argued either way.
- You are encouraged to draw connections between the concepts introduced in class and your prior knowledge. This knowledge may come from previous coursework in related areas (e.g., Foundations GE coursework in Mathematical and Quantitative Reasoning), everyday experience interacting with language technologies, and/or introspection about how language works and how you as a speaker internally process language as a cognitive system. In-class exercises, class discussions, and homework assignments in this class will analyze language technologies that you probably use every day (e.g., predictive texting, voice assistants), so this class will give you a chance to connect course concepts to your experience using these technologies.
- In this course you will extensively practice applying mathematical tools to understand the natural phenomenon of human language. You will identify the advantages and disadvantages of studying language using these idealized, formal systems.

Course Materials

The main text for the first half of this course will be *Language and Computers* by Dickinson, Brew, and Meurers (2012), which is available for purchase online or at the Ohio State University Bookstore, 1598 N. High St. (ebook: $34, paperback: $39.95. Used copies should be available at less cost.) Other readings will be made available on Carmen.

Since this is a course about how computers work with language, and since course materials will be posted on Carmen, you will need reliable access to a computer and working internet. It will be most convenient to work on your own computer, but computing resources are also available through the university libraries, computer labs, etc. If you anticipate problems with access to computing resources, please talk to me ASAP so we can get it sorted out. Some assignments will involve working with software tools such as Excel and running basic commands on the command line.

Course Components

For simplicity, grading in this course will follow a point-based system, with a total of 1000 points possible to earn. Each point is worth 0.1% of your grade. The table below provides a breakdown of the point distribution for the various course components, and their conversion to letter grades.

Component	Value			
Assignments	300 pts	30%	A	930 - 1000 pts
Quizzes	100 pts	10%	A-	900 - 929
Midterm	150 pts	15%	B+	870 - 899
Final	150 pts	15%	B	830 - 869
Participation	150 pts	15%	B-	800 - 829
Paper	150 pts	15%	C+	770 - 799
Total	**1000 pts**	**100%**	C	730 - 769
			C-	700 - 720
			D+	670 – 699
			D	600 – 669
			E	0 - 599

Attendance and Participation

Because the assignments, quizzes, and exams will be based on material covered in lecture, attendance and participation are essential for doing well in this course. During lectures, I

will use a randomized class roster to call on students throughout the class period. When I ask you a question, it's fine if you don't always know the answer! All that is required for full participation credit is that (1) you are present when I call on you, and (2) you make a reasonable effort to respond, which could include telling me if you're confused. Along with encouraging everyone in the class to stay attentive, this method will help me know if I'm moving through material at an appropriate pace.

Homework Assignments

This course will include a total of 5 homework assignments. Each is worth 60 points, for a total of 300 points. Assignments will be submitted on Carmen, and will be due at 2 pm on the day indicated in the course schedule (which you can find on Carmen or at the end of this syllabus). Submissions should be in PDF format. I encourage you to get an early start on homework so that you will have time to ask for help in class and/or office hours if needed.

You are allowed to discuss homework questions with other students in the class. However, your assignment must be written up individually, in your own words. In addition, homework assignments **must** include the names of any students you collaborated with.

Quizzes

There will be 10 quizzes based on material from lectures and supplementary readings. Each quiz is worth 10 points, for a total of 100 points. Quizzes will be administered online through Carmen and are due at 2 pm on the day indicated on the course schedule. They will shut off automatically when this time is reached, so don't put them off until the last minute!

You will have 15 minutes to finish each quiz. Be aware that once you start the quiz, the timer will start, and it won't stop even if you exit Carmen. Quizzes must be done independently; **no collaboration is allowed**.

Late Work

As a general rule, late homework assignments and quizzes will not be accepted. I am willing to consider exceptions to this in extenuating circumstances, but only if the exceptions are fair to other students in the course. Please make arrangements with me in advance if you anticipate not being able to meet a deadline.

Midterm and Final Exams

This course will include midterm and final exams, each worth 150 points. The midterm will be given in class on date xxx, and the final exam will be given on date xxx, following the university-wide finals schedule. Both exams will be cumulative, although the final will focus more on content from the second half of the semester. Exams will include a mixture of problem-solving questions and short-answer questions testing conceptual knowledge.

Final Paper

The final paper will be an essay of between 1000 and 1500 words discussing a topic related to either (1) the social and ethical implications of language technology, or (2) the depiction of language technology in a recent movie or TV show. Details about these two options, plus a grading rubric, will be posted on Carmen.

Academic Integrity and Misconduct

Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, the Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the University's Code of Student Conduct, and that all students will complete all academic and scholarly assignments with fairness and honesty. Students must recognize that failure to follow the rules and guidelines established in the University's Code of Student Conduct and this syllabus may constitute Academic Misconduct. See the sections about assignments and quizzes, above, for my specific guidelines about academic integrity in the context of this course.

If I suspect that a student has committed academic misconduct in this course, I am obligated by University Rules to report my suspicions to the Committee on Academic Misconduct. If COAM determines that you have violated the Code of Student Conduct (i.e., committed academic misconduct), the sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the University. If you have any questions about the above policy or what constitutes academic misconduct in this course, please ask!

Official university statement on academic misconduct:
It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct.

Other Policies and Resources

- **Disability Accommodation:** The university strives to make all learning experiences as accessible as possible. In light of the current pandemic, students seeking to request COVID-related

accommodations may do so through the university's request process, managed by Student Life Disability Services. If you anticipate or experience academic barriers based on your disability (including mental health, chronic, or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

- The **Student Advocacy Center** is available to help with many problems you might have navigating OSU, including but not limited to dealing with bureaucratic issues, academic issues, health issues (including mental health and hospitalization), and financial issues. advocacy.osu.edu, 614-292-1111, advocacy@osu.edu, 001 Drackett Tower.
- **Student Mental Health**: As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614-292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available 24/7 by dialing 988 to reach the Suicide and Crisis Lifeline.
- **Title IX** prohibits discrimination on the basis of sex under any OSU program or activity. That includes assault, harassment, and limiting your enjoyment of opportunities or rights. You can find help, file grievances, and learn to help others via OSU's Title IX Coordinator**.** The Title IX office also has information about many other forms of discrimination. titleix.osu.edu, 614-247-5838, titleix@osu.edu, 21 East 11th Ave.

Tentative Course Schedule

This schedule may be subject to change throughout the semester. I will announce any updates in class and on Carmen.

The following abbreviations are used for readings:

- **DBM:** *Language and Computers* by Dickinson, Brew, and Meurers (2012)
- **NLTK:** *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit* by Bird, Klein, and Loper (available online at www.nltk.org/book)
- **JM:** *Speech and Language Processing* (3rd ed. draft) by Jurafsky and Martin (available online at web.stanford.edu/~jurafsky/slp3)

Week	Date	Topics	Reading	Deadlines
Week 1	(W)	Introduction		
Week 2	(M)	**Unit 1: Representing speech and text** Text encoding: Converting human writing systems into binary code	DBM 1.1–1.3	Background survey
	(W)	Speech encoding: Representing speech sounds on computers using acoustic properties such as frequency and amplitude	DBM 1.4–1.4.4, Under the Hood 1	
Week 3	(M)	NO CLASS—LABOR DAY		
	(W)	Automatic speech recognition: How probabilistic models can map speech to text despite ambiguity, speaker variation, and background noise Text-to-speech: How to synthesize speech by splicing together shorter audio segments—and why it's hard to make synthesized audio sound like a real human	DBM 1.4.5, Under the Hood 2	
Week 4	(M)	**Unit 2: Search** Search techniques: Efficiently locating relevant data for a	DBM 4.1–4.3	Quiz 1 (Representing Speech and Text). Students will answer

		user's information need using logical operations or keyword searches		basic questions about binary notation and acoustics.
	(W)	PageRank: A method to quantify the intuitive idea of a web page's popularity Evaluation: How to use quantitative evaluation metrics such as precision (a measure of quality of results) and recall (a measure of quantity of results) to compare results from search engines		Assignment 1 (Representing Speech and Text). Students will… • Practice converting between text and binary representations • Record their own speech and explore how it is represented by a computer • Evaluate strengths and weaknesses of current computational models for speech recognition and text-to-speech conversion
Week 5	(M)	Regular expressions (a concise means of identifying matching patterns of characters) and finite-state machines (models used to simulate sequential logic): How mathematically defined formal languages can serve as a practical tool for searching through text (and a shortcut for solving Wordle puzzles)	DBM 4.4, Under the Hood 7	
	(W)	More regular expressions and finite-state machines		
Week 6	(M)	More regular expressions and finite-state machines		Quiz 2 (Search). Students will practice calculating the quantitative evaluation metrics of precision and recall and writing short regular expressions (character-matching patterns).
	(W)	**Unit 3: Classification**	DBM 5.1–5.4;	

		Text and document classification: How tasks varying from grammar error checking to language identification to spam filtering can be approached as variants of the same underlying machine learning problem	NLTK Book Ch. 6, Sec. 1	
Week 7	(M)	Naive Bayes: A statistical classification technique that exploits the properties of conditional probabilities	DBM 5.5.1, Under the Hood 9; NLTK Ch. 6, Sec. 5	Assignment 2 (Search). Students will… • Analyze how the notions of precision and recall can apply to situations beyond search • Convert between regular expressions (character-matching patterns) and finite-state machines, two formal methods for describing text patterns • Examine limitations of search techniques by finding examples of information that is difficult to search for online
	(W)	Decision trees: A probabilistic classification technique that imitates how humans often make decisions	NLTK Ch. 6, Sec. 4	
Week 8	(M)	Perceptrons: A probabilistic classification technique inspired by the interactions between neurons in the brain, and which uses error-driven learning	DBM 5.5.2–5.6	Quiz 3 (Classification). Students will practice the steps involved in the probabilistic methods of Naive Bayes and decision tree classification.
	(W)	More perceptrons		
Week 9	(M)	Review		
	(W)	**MIDTERM**		

Week 10	(M)	**Unit 4: Language modeling** *n*-gram models: How computers predict upcoming words (e.g., for predictive text) based on mathematical modeling of the preceding context	JM Ch. 3 up through Sec. 3.2.1	
	(W)	Smoothing and backoff: How computers' predictions about text can be made robust to rare word sequences and unfamiliar vocabulary	JM Ch. 3, 3.3–3.5	Assignment 3 (Classification). Using a small collection of restaurant reviews, students will… • Choose a technique for classifying the reviews as positive or negative out of the options introduced in class • Determine whether the selected classification model can learned generalizable patterns • Compare the strengths and weaknesses of their classification model compared to alternatives • Compare the classification steps taken by the model to those a human would take to classify a restaurant review as positive or negative
Week 11	(M)	More smoothing and backoff		
	(W)	Text generation: Using computational language models to produce novel text and imitate particular styles (e.g., plays by Shakespeare or fantasy novels)		Quiz 4 (Language modeling). Students will practice estimating the parameters of a probabilistic *n*-gram model and using this model to calculate sentence probabilities.
Week 12	(M)	**Unit 5: Analyzing linguistic structure**	NLTK Ch. 8, Secs. 1 and 2	

		Syntax and structural ambiguity: Using formal syntactic rules to describe ambiguities that arise in human language		
	(W)	Context-free grammars (a type of formal language model): How computers can represent—and produce—an infinite variety of sentences using a finite set of rules	NLTK Ch. 8, Sec. 3	Assignment 4 (Language modeling). Students will… • Apply smoothing and backoff techniques to design a robust computational language model • Computationally generate text in the style of an author of their choosing using a Colab notebook • Qualitatively compare text generated by humans and computers
Week 13	(M)	Parsing: How a computer can algorithmically figure out the underlying linguistic structure in a sentence, and how this relates to how humans process language	JM Ch. 13 up through Sec. 13.2	
	(W)	**Unit 6: Social context and contemporary trends** Ethics of artificial intelligence: Concerns about bias, fairness, privacy, and environment cost that arise in contemporary applications of computational linguistics	NYT: Racial divide in ASR; MIT Tech Review: What does GPT-3 know about me?	Quiz 5 (Analyzing linguistic structure). Students will identify cases of lexical (word-level) and structural (sentence-level) ambiguity and answer questions about a simple type of formal language model known as a context-free grammar.
Week 14	(M)	Distributional semantics: How the meaning of a word can be defined mathematically based on the contexts in which it appears	JM Ch. 6 up through Sec. 6.5	Assignment 5 (Analyzing linguistic structure). Students will… • Walk through how a computer parses an ambiguous sentence

				and identify how the ambiguity arises • Propose modifications to a set of syntactic rules to allow a computer to parse a wider variety of English sentences
	(W)	NO CLASS—THANKSGIVING		
Week 15	(M)	word2vec: A popular algorithm for mathematically deriving word representations based on the distributional semantics approach	JM 6.8	Final paper description. A short proposal and list of sources for the final paper about the social context of language technology.
	(W)	Neural networks: How popular "deep learning" computational systems achieve success at a variety of language-related tasks, and the challenges that arise in interpreting the behavior of these systems	JM Ch. 7 up through Sec. 7.5	Quiz 6 (Social context and contemporary trends). Students will practice deriving word representations and using these to estimate the similarities of words (e.g., synonyms and antonyms).
Week 16	(M)	Current research: How language models and statistical computational parsers can be applied to the study of human language acquisition and comprehension		
	(W)	Review		Final paper. An essay of 1000–1500 words critically examining and synthesizing either (a) a contemporary ethical issue in computational linguistics, such as fairness or privacy, or (b) the depiction of language technology in entertainment media (e.g., a recent movie or TV show)
FINAL EXAM: Date and time				

GE Theme course submission worksheet: Number, Nature, Mind

Course: Linguistics 3802

Overview

Courses in the GE Themes aim to provide students with opportunities to explore big picture ideas and problems within the specific practice and expertise of a discipline or department. Although many Theme courses serve within disciplinary majors or minors, by requesting inclusion in the General Education, programs are committing to the incorporation of the goals of the focal theme and the success and participation of students from outside of their program.

Each category of the GE has specific learning goals and Expected Learning Outcomes (ELOs) that connect to the big picture goals of the program. ELOs describe the knowledge or skills students should have by the end of the course. Courses in the GE Themes must meet the ELOs common for all GE Themes and those specific to the Theme, in addition to any ELOs the instructor has developed specific to that course. All courses in the GE must indicate that they are part of the GE and include the Goals and ELOs of their GE category on their syllabus.

The prompts in this form elicit information about how this course meets the expectations of the GE Themes. The form will be reviewed by a group of content experts (the Theme Advisory) and by a group of curriculum experts (the Theme Panel), with the latter having responsibility for the ELOs and Goals common to all themes (those things that make a course appropriate for the GE Themes) and the former having responsibility for the ELOs and Goals specific to the topic of this Theme.

Briefly describe how this course connects to or exemplifies the concept of this Theme (Number, Nature & Mind)

In a sentence or two, explain how this class "fits" within the focal Theme. This will help reviewers understand the intended frame of reference for the course-specific activities described below.

LING 3802 students will analyze human language using multiple types of mathematical reasoning drawn from fields such as statistics and formal language theory. They will think critically about what these types of reasoning can show about language, drawing from their own experience with language technology systems to evaluate how well such systems account for language data.

Connect this course to the Goals and ELOs shared by *all* Themes

Below are the Goals and ELOs common to all Themes. In the accompanying table, for each ELO, describe the activities (discussions, readings, lectures, assignments) that provide opportunities for students to achieve those outcomes. The answer should be concise and use language accessible to colleagues outside of the submitting department or discipline. The specifics of the activities matter – listing 'readings' without a reference to the topic of those readings will not allow the reviewers to understand how the ELO will be met. However, the panel evaluating the fit of the course to the Theme will review this form in conjunction with the syllabus, so if readings, lecture/discussion topics, or other specifics are provided on the syllabus, it is not necessary to reiterate them within this form. The ELOs are expected to vary in their "coverage" in terms of number of activities or emphasis within the course. Examples from successful courses are shared on the next page.

Goal 1: Successful students will analyze an important topic or idea at a more advanced and in-depth level than the foundations. In this context, "advanced" refers to courses that are e.g., synthetic, rely on research or cutting-edge findings, or deeply engage with the subject matter, among other possibilities.

Goal 2: Successful students will integrate approaches to the theme by making connections to out-of-classroom experiences with academic knowledge or across disciplines and/or to work they have done in previous classes and that they anticipate doing in future.

	Course activities and assignments to meet these ELOs
ELO 1.1 Engage in critical and logical thinking.	Throughout the course, students will develop skills needed to engage in critical and logical thinking about human language and language technology via the following: 1. Readings describing how different kinds of language technology work. For example, in Week 12 students learn how specific formal models of natural language syntax, such as context-free grammars, can represent syntactic ambiguities exploited by various jokes and found in amusingly misleading headlines, requiring them to think logically about when these grammars support multiple analyses. Then in Week 13 (Day 1), students learn how these analyses can be derived by parsing algorithms and the role these can play in applications such as grammar checkers and machine translation systems. 2. In-class activities providing hands-on exploration and discussion of language technology systems and language data. For example, in an activity building on the

	syntax and parsing reading, students use online demos of probabilistic parsers to analyze sentences that people find difficult to process and understand (so-called "garden path" sentences), then reflect on the extent to which the machine-derived analyses can help explain why people may be led astray in reading the sentences. Subsequent class discussion about the variety of factors that can come into play with difficult-to-process sentences reinforces their ability to think critically about language and language technology. 3. Assignments exploring these topics in greater detail, typically via exercises that require applying mathematical reasoning to specific sets of language data. For example, the syntax and parsing assignment (Assignment 5) requires students to syntactically analyze a variety of sentences according to specific context-free grammars, determining when the grammars suffice to identify salient ambiguities as well as when the grammars are insufficient for this purpose; in the latter case, students must revise the grammars to support analyses that make the ambiguity explicit. Again, doing so requires students to reason about what analyses follow as logical consequences of a specific grammar specification.
ELO 1.2 Engage in an advanced, in-depth, scholarly exploration of the topic or ideas within this theme.	Students will analyze human language using multiple types of mathematical reasoning drawn from fields such as statistics and formal language theory. They will think critically about what these types of reasoning can show about language. To do so, students must comprehend mathematical concepts and methods in order to construct valid arguments, understand inductive and deductive reasoning, and increase their general problem-solving skills. For example: • In order to automatically classify whether a message (e.g., a tweet) or a document (e.g., a movie review) primarily expresses positive or negative sentiment, students

	will learn (Week 7, Day 1) how word frequency and conditional probability can be used to define a simple Naïve Bayes classifier (a simple kind of probabilistic model) for making sentiment predictions. (A Naïve Bayes classifier is naïve in assuming that words are independent predictors and Bayesian in the way it updates prior probabilities considering current evidence; despite their simplicity, such models can be surprisingly strong predictors.) They will then explore how such models can fail to make accurate predictions by ignoring syntactic structure or irony. • Students will learn (Week 5, Day 1) how regular expressions (a concise means of specifying character-matching patterns) can be used to define precise searches for sets of strings in documents (e.g., all the words ending in *-ing*.) and subsequently how they provide strictly less descriptive power than the formal language models known as context-free grammars mentioned earlier. After becoming familiar with these methods from formal language theory through in-class activities, they will explore their limitations via sets of language data that violate the simplifying assumptions these methods require, such as cross-serial (non-nested) syntactic dependencies found in some languages (Assignments 2 and 5).
ELO 2.1 Identify, describe, and synthesize approaches or experiences.	Students are encouraged to draw connections between the concepts introduced in class and their prior knowledge (e.g., Foundations coursework). This knowledge may come from previous coursework in related areas, everyday experience interacting with language technologies, and/or introspection about how they (the students) internally process language. Drawing connections with prior knowledge is emphasized in both in-class activities and homework assignments. Examples: • When analyzing the limitations of speech recognition and synthesis systems,

	students are encouraged to analyze how the writing systems for other languages they have learned or are familiar with affect the performance of these systems (Week 2 in-class activity; Assignment 1). For instance, students might investigate the extent to which the lack of vowels in the standard orthography for Arabic or Hebrew affects the ability of a speech synthesis system for these languages to say the intended word. • To explore how well computational neural network models of word meaning capture the effect of context on intended meaning, students are expected to draw on their own experience with language to come up with contexts where the intended meaning of a word would be clear to humans but likely to be misinterpreted by the machine (Week 15, in-class activity). Students then test empirically whether these models of word meaning make mistaken predictions as hypothesized.
ELO 2.2 Demonstrate a developing sense of self as a learner through reflection, self-assessment, and creative work, building on prior experiences to respond to new and challenging contexts.	Students are expected to become increasingly sophisticated in their ability to analyze the abilities and limitations of various language technologies through in-class activities and discussions and assignments. At the end of the course, students write a final essay on a topic of their choice, allowing them to take control of their learning through the choice of topic and the choice of outside readings for the topic. Essay topics address social or ethical implications of language technology, enabling students to reflect on material covered earlier in the course and how this material informs their views on topics of interest to them. Most topics involve issues that can be argued either way, requiring the student to offer careful argumentation to be convincing. As an example, students may argue that machine translation will revolutionize the sharing of information across cultures, or argue that inevitable errors in machine translation will ultimately lead to miscommunication in ways that inhibit cross-cultural understanding.

	The final essay also provides an opportunity for students to reflect on their learning and their ability to think critically about issues surrounding language technology. At the outset of the course, some of these issues are discussed in class, with the discussions generally rather unsophisticated as for most students language technologies are magic black boxes. Students are reassured that by the end of the course, they will be in a better position to participate in informed discussion of such issues, and indeed students do manage to make arguments in their final essay that are informed by knowledge of how the language technology in question works.

Goals and ELOs unique to Number, Nature, Mind

Below are the Goals and ELOs specific to this Theme. As above, in the accompanying Table, for each ELO, describe the activities (discussions, readings, lectures, assignments) that provide opportunities for students to achieve those outcomes. The answer should be concise and use language accessible to colleagues outside of the submitting department or discipline. The ELOs are expected to vary in their "coverage" in terms of number of activities or emphasis within the course. Examples from successful courses are shared on the next page.

GOAL 3: Successful students will experience and examine mathematics as an abstract formal system accessible to mental manipulation and/or mathematics as a tool for describing and understanding the natural world.

	Course activities and assignments to meet these ELOs
ELO 3.1 Analyze and describe how mathematics functions as an idealized system that enables logical proof and/or as a tool for describing and understanding the natural world.	Students will extensively practice applying mathematical tools to understand the natural phenomenon of human language. They will identify the advantages and disadvantages of studying language using these idealized, formal systems. The course goes beyond merely subjective evaluation of systems, emphasizing analysis and reasoning to draw and argue for valid conclusions about the design, capabilities and behavior of natural language systems. Several examples of such analyses have been given above. Beyond understanding how language technologies work, the mathematical tools

	students learn to employ enhance their appreciation of (and even awe for) their own cognitive abilities with language. For example, by examining the errors that automatic parsers can make in analyzing the structure of sentences, students come to appreciate that these sentences could in principle have these structures, but analyzing them in the way the machine did would yield readings so implausible that these structures are not even consciously entertained in normal reading or conversation. Such exercises illustrate how the human language faculty is able to effortlessly integrate many sources of information, including common sense reasoning, in ways that remain difficult for computers.