Status: PENDING

PROGRAM REQUEST
Computational Analysis of language

Last Updated: Vankeerbergen,Bernadette
Chantal
02/09/2021

Fiscal Unit/Academic Org	Linguistics - D0566
Administering College/Academic Group	Arts and Sciences
Co-adminstering College/Academic Group	Arts and Sciences
Semester Conversion Designation	New Program/Plan
Proposed Program/Plan Name	Computational Analysis of language
Type of Program/Plan	Undergraduate certificate program
Program/Plan Code Abbreviation	
Proposed Degree Title	Computational Analysis of Language

Credit Hour Explanation

Program credit hour requirements		A) Number of credit hours in current program (Quarter credit hours)	B) Calculated result for 2/3rds of current (Semester credit hours)	C) Number of credit hours required for proposed program (Semester credit hours)	D) Change in credit hours
Total minimum credit hours required for completion of program				12	
Required credit hours offered by the unit	Minimum			6	
	Maximum			12	
Required credit hours offered outside of the unit	Minimum			0	
	Maximum			6	
Required prerequisite credit hours not included above	Minimum			3	
	Maximum			3	

Program Learning Goals

Note: these are required for all undergraduate degree programs and majors now, and will be required for all graduate and professional degree programs in 2012. Nonetheless, all programs are encouraged to complete these now.

Program Learning Goals
- Students will identify the tasks involved in the computational analysis of human language.
- Students will apply computational methods, statistical methods, and/or formal logic to the analysis of language data.
- Students will apply core grammatical concepts and principles to the analysis of language data.

Assessment

Assessment plan includes student learning goals, how those goals are evaluated, and how the information collected is used to improve student learning. An assessment plan is required for undergraduate majors and degrees. Graduate and professional degree programs are encouraged to complete this now, but will not be required to do so until 2012.

Is this a degree program (undergraduate, graduate, or professional) or major proposal? No

Program Specializations/Sub-Plans

If you do not specify a program specialization/sub-plan it will be assumed you are submitting this program for all program specializations/sub-plans.

Pre-Major

Status: PENDING

PROGRAM REQUEST
Computational Analysis of language

Last Updated: Vankeerbergen,Bernadette
Chantal
02/09/2021

Does this Program have a Pre-Major? No

Attachments

- Linguistics Chair certificate support ling.pdf: Chair Support Letter

  (Letter from Program-offering Unit. Owner: McGory,Julia Tevis)

- College Letter of Support D. Haddad.docx: College Support Letter

  (Letter from the College to OAA. Owner: McGory,Julia Tevis)

- CSE Letter of Support.png: CSE Support Letter

  (Support/Concurrence Letters. Owner: McGory,Julia Tevis)

- Stats_Concrrence.png: STATS Support Letter

  (Support/Concurrence Letters. Owner: McGory,Julia Tevis)

- computational-analysis-of-language-certificate.pdf: Updated Certificate Program Proposal

  (Program Proposal. Owner: McGory,Julia Tevis)

Comments

- See Panel feedback email sent 01-28-21 *(by Hilty,Michael on 01/28/2021 01:10 PM)*

- Please provide any comments to Julia McGory.1@osu.edu. 614-725-9217

  Thank you. *(by McGory,Julia Tevis on 12/11/2020 12:38 PM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	McGory,Julia Tevis	12/11/2020 12:38 PM	Submitted for Approval
Approved	McGory,Julia Tevis	12/11/2020 12:39 PM	Unit Approval
Approved	Haddad,Deborah Moore	12/11/2020 02:38 PM	College Approval
Revision Requested	Hilty,Michael	01/28/2021 01:10 PM	ASCCAO Approval
Submitted	McGory,Julia Tevis	02/09/2021 11:40 AM	Submitted for Approval
Approved	McGory,Julia Tevis	02/09/2021 11:40 AM	Unit Approval
Approved	Vankeerbergen,Bernadette Chantal	02/09/2021 12:29 PM	College Approval
Pending Approval	Jenkins,Mary Ellen Bigler Hanlin,Deborah Kay Oldroyd,Shelby Quinn Hilty,Michael Vankeerbergen,Bernadette Chantal	02/09/2021 12:29 PM	ASCCAO Approval

Certificate in Computational Analysis of Language

Proposal

1. Required Information

Types of Certificate:
- 1A (stand-alone post-high school diploma undergraduate certificate)
- 1B (embedded post-high school diploma undergraduate certificate)
- 2 (stand-alone post-bachelor degree undergraduate certificate)

Mode of Delivery: The certificate can be completed wholly in-person or via a combination of in-person and on-line courses.

Proposed implementation date: Autumn 2021

Academic unit responsible for administering the certificate program: Department of Linguistics

2. Rationale

In modern society we engage with human language technology on a daily basis in the form of search engines, predictive text messaging, virtual assistants, speech-to-text software (e.g. for automatic closed captioning of video), and automatic language translation services (e.g. Google Translate), among many other applications. Moreover, as the role of technology grows ever larger in people's daily lives, the need for good language technologies continues to grow, leading to strong demand for workers with language technology skills at tech companies such as Google, Facebook, Apple, Amazon, and Microsoft, and also in the area of security and intelligence technology (e.g. General Dynamics, Palantir Technologies), legal technology (e.g. Lex Machina), language learning technology (e.g. DuoLingo, Grammarly), and in a wide range of businesses. Computational linguistics is the academic discipline that most directly feeds this demand, combining elements of linguistics, computer programming and software engineering, data science, and artificial intelligence.

This 12 credit-hour undergraduate certificate will train students in basic concepts and methods of computational linguistics. It will introduce students to various tasks involved when computers process human speech and text, including speech recognition, text-to-speech conversion, machine translation (automatic translation of text from one human language to another), automated text analysis (e.g. question detection), and natural language generation (e.g. converting data tables into human language). Since computational linguistics is at the intersection of computer science and linguistics, the certificate will give students a basic understanding of both domains. The certificate is designed for current undergraduate students from any major who want to pursue a career related to the creation of language technologies, as well as for professionals in a related area who want to enhance their marketability.

The certificate is divided into two tracks, reflecting different kinds of preparation for a career in language technologies. Both tracks consist of courses that are already offered in Linguistics (and CSE), with the exception of Linguistics 3803 (Ethics of Language Technology), which is in development and has already been submitted to the ASCC. Linguistics expects to be able to meet demand for the certificate with existing teaching resources and could add sections of many of these courses as needed.

Track A introduces students to issues and methods in computational linguistics at a conceptual but mostly not a technical level and does not require computer programming, although students can *choose* to do coursework that involves computer programming. This track will prepare students for industry work as Language Specialists, Data Specialists, Localization Specialists, Speech Data Evaluators, Voice User Interface Designers, Language Annotators, and for similar entry-level positions.[1] These jobs generally require a Bachelor's degree in linguistics, a world language, English, or other relevant field. In hiring ads for jobs of this sort, basic knowledge of computational linguistics/natural language processing (NLP) is often an asset and preferred qualification, since workers in these positions will need to work in teams with language engineers and data scientists. This track is designed to provide students from any background a basic knowledge of the computational analysis of language data, which students can pair with a BA or BS degree in various fields.

Track B introduces students to issues and methods in computational linguistics at both a conceptual and a technical level and requires basic computer programming, which can be developed through certificate coursework. This track is designed primarily to prepare students for an MS or PhD program in computational linguistics. MS programs, in particular, have sprung up at many universities to feed industry demand. However, admission to these programs requires at least a basic background in linguistics; programming and computer science; and probability, statistics, and formal logic.[2] Track B is designed primarily for students who are pursuing a degree or otherwise have a background in one of these areas but not all three. It will help them bridge the gap between their background and these programs' admissions requirements, while at the same time allowing students to tailor their coursework to their particular needs. By preparing students for Master's or higher study in computational linguistics, this track is designed to lead ultimately to industry positions, for example as a Computational Linguist, Language Engineer, NLP Data Scientist, Analytical Linguist (a kind of data scientist), Human Language Technologist, or Research Scientist.[3] These jobs generally require an MS or PhD in computational linguistics, computer science, or statistics.

This certificate will complement and add value to a wide variety of majors/degrees: Linguistics, Computer Science & Engineering (CSE), Computer & Information Science (CIS), Data Analytics, Statistics, English, world languages (Russian, Chinese, Arabic, Spanish, German, French, etc.), and many others not listed here. Computational linguistics touches numerous fields in some way, and so knowledge of computational linguistics adds value to many areas of study. As noted in the program description for the University of Washington's MS in Computational Linguistics, "... a pre-med undergraduate degree plus a master's in computational linguistics will position [a student] well for a

---

[1] Sample job ads: Associate Linguist at Lionbridge: https://linguistlist.org/issues/31/31-3342/; Speech Data Evaluator at Google: https://linguistlist.org/issues/29/29-763/; Linguist Annotator at Appen: https://www.toplanguagejobs.com/Top-Language-Jobs/Bilingual-Jobs/Linguist-Annotator/Details/10559799?utm_campaign=google_jobs_apply&utm_source=google_jobs_apply&utm_medium=organic; Language Specialist at Nuance: https://linguistlist.org/issues/30/30-1273/; Junior Knowledge Engineer at Expert System USA (intelligence and security government contractor): https://linguistlist.org/issues/30/30-3732/; Data Specialist at Amazon: https://linguistlist.org/issues/29/29-3651/; Voice User Interface Designer at Voxify: https://linguistlist.org/issues/22/22-2096/; Technical Linguist at Artificial Solutions: https://linguistlist.org/issues/26/26-2809/.

[2] See, e.g., the advice from the University of Washington on preparing for their MS in Computational Linguistics program: https://www.compling.uw.edu/admissions/preparing-for-the-program/

[3] Sample job ads: Language Engineer at Facebook: https://linguistlist.org/issues/25/25-2511/; Language Data Researcher at Amazon: https://linguistlist.org/issues/30/30-2401/; Data Scientist at Bank of England: https://linguistlist.org/issues/27/27-3258/; Assistant Research Scientist at University of Maryland Applied Research Laboratory for Intelligence and Security: https://linguistlist.org/issues/31.2618/; NLP Scientist at AppTek: https://linguistlist.org/issues/31/31-1008/; Computational Linguist at Grammarly: https://linguistlist.org/issues/28/28-1628/; Linguist for Business Application at Gap International: https://linguistlist.org/issues/30/30-3841/

career in biomedical informatics. Similarly, legal studies are good background for NLP applications in the legal domain, and a degree in economics, business or marketing is good training for sentiment analysis, text analytics and other business-to-business NLP applications."[4]

Upon completion of the academic Certificate in Computational Analysis of Language, learners will be better prepared to:

1. Identify the tasks involved in the computational analysis of human language;
2. Apply computational methods, statistical methods, and/or formal logic to the analysis of language data;
3. Apply core grammatical concepts and principles to the analysis of language data.

3. Relationship to Other Programs / Benchmarking

The two CSE courses (6 credit hours) in Track B of this certificate (CSE 3521, 5525) can also be applied to the Artificial Intelligence Specialization within the BS in Computer and Information Science (CIS)/Computer Science & Engineering (CSE). However, other courses (totaling 9 hours, including the prerequisite) are unique to the certificate. Within the BA in CIS, students are required to take 12 hours of Related Field Core courses. This can presumably include courses on linguistics as relevant but is not required to. The proposed certificate is thus substantially distinct from the BA and BS degrees administered by CSE in its central focus on language analysis and linguistics.

This certificate also overlaps with the Computational Analytics specialization within the BS in Data Analytics. Courses for that specialization (specifically, the Linguistics and Text Analytics Focus) are drawn mostly from the Linguistics curriculum. Most of the courses in this focus can also be applied towards the proposed certificate. However, the Linguistics and Text Analytics Focus forms only a small part of the Data Analytics major (10 credit hours out of 61 in total), and the proposed certificate offers additional/unique training specifically in linguistics and the computational analysis of language data, separately from the Data Analytics BS specialization.

The proposed certificate does not overlap with any other certificates at OSU that we are aware of. It has not previously been submitted for approval.

Comparison to Programs at Other Universities: There are no similar certificates or comparable programs at other universities in Ohio. There are a few undergraduate computational linguistics certificates or concentrations at other universities in the U.S., including:

- San Diego State University (4-course undergraduate Basic Computational Linguistics Certificate): https://ces.sdsu.edu/open-university-certificate-in-computational-linguistics
  - Required courses: Fundamentals of Linguistics (Ling 501), Computational Corpus Linguistics (Ling 571), Computational Linguistics (Ling 581), and Python Scripting for Social Science (Ling 572)
- San Francisco State University (5-course undergraduate Certificate in Computational Linguistics): http://bulletin.sfsu.edu/colleges/liberal-creative-arts/english/certificiate-in-computational-linguistics/
  - Requires courses: Introduction to the Study of Language (English 420), Syntax (English 421), Introduction to Computational Linguistics (English 620), and Applied

---

[4] https://www.compling.uw.edu/admissions/preparing-for-the-program/

Computational Linguistics (English 680). A fifth course is a choice between Phonology and Morphology (English 424) and Natural Language Technologies (Comp Sci 620)

- Montclair State University (5-course optional undergraduate concentration in Language Engineering within the Linguistics BA): https://www.montclair.edu/linguistics/programs-of-study/language-engineering-and-computational-linguistics/linguistics-major-with-a-concentration-in-language-engineering/
- Rochester Institute of Technology (3-course undergraduate Human Language Technology and Computational Linguistics Immersion): https://www.rit.edu/study/human-language-technology-and-computational-linguistics-immersion
  - Housed in English
- San Jose State University (6-course undergraduate Certificate in Computational Linguistics): https://www.sjsu.edu/linguistics/academic_programs/linguistics/computational_linguistics/
  - Includes a separate programming requirement
- University of Utah (9-course undergraduate Computational Linguistics Certificate): https://linguistics.utah.edu/certificates-and-programs/comp-ling.php

There are also a number of graduate certificates at other universities, including: Montclair State University (6 courses), Texas A&M (5 courses), University of Colorado Boulder (5 courses), University of Illinois (6 courses), University of North Carolina (3 courses + speaker series), University of North Texas (4 courses), University of Washington (3 courses).

The program at San Diego State University is most similar to Track B in the proposed certificate but does not seem to include coursework that is as advanced as what is available (optionally) to students in our proposed certificate. The SFSU certificate, where the Linguistics Program is housed in the English Department, is closest to our proposed Track A but it places greater emphasis on (non-computational) linguistic theory courses. Our proposal also allows students more flexibility to tailor the certificate to their needs and background, compared to both SDSU's and SFSU's certificates.

Rob Malouf, who runs SDSU's certificate, reports that this semester in their corpus linguistics course, which can be applied to either the Computational Linguistics certificate or a separate Text Analytics certificate, there are "…2 big data students, 10 statistics majors, 8 linguistics majors, and 3 open university. I assume all of them are probably going to get at least one of the two certificates." The Computational Linguistics certificate is the greater draw for linguistics students and the Text Analytics certificate draws more statistics students. The former also draws a few members of the community looking to get jobs in the tech industry.

According to Anastasia Smirnova, SFSU's certificate has "… attracted a variety of majors from different disciplines and colleges, including Anthropology, Classics, Journalism, Philosophy, Psychology, Creative Writing, Math, Computer Science, and Business. We also have received inquiries from non-matriculated students, but the enrollment for this particular group has been low." Students who earn the certificate tend to go on to careers in the local Bay Area tech industry; some find "data science / technical linguist careers in industry. Others find non-technical linguistic jobs in tech companies. The comp ling classes are useful because even non-technical positions often have a technical component and require good quantitative skills." At Monclair State, program head Prof. Anna Feldman reports that about 20 of 120 Linguistics majors choose the Language Engineering concentration within the major.

4. Student Enrollment

We hope that about 15 students per year will earn the certificate. This would be consistent with demand at institutions with similar certificates, taking into account the overall larger number of students at OSU and larger number of Linguistics majors/minors.

5. Curricular Requirements

This certificate has Linguistics 2000(H) (Introduction to Linguistics) or English 3271 (Structure of the English Language) – as a prerequisite. The certificate is divided into two tracks – a less technical Track A and a more technical Track B. Twelve credit hours are required in each track, of which six can overlap a degree program, per university rules for certificates.

The certificate is expected to take 2-4 semesters to complete, depending on the particular pathway through the certificate that a student chooses. Since students have some freedom to choose among course options, course availability is not expected to be an issue. Courses at the 2000- to 4000-level are generally offered every year in both Autumn and Spring semesters. Most of the Linguistics courses at the 5000-level are offered once per year. Ling 5803 and English/Ling 5804 are offered less than once per year but these courses are one of multiple options for fulfilling the certificate requirements.

For most courses no particular facilities or equipment is required in order to complete the certificate. In the 5000-level classes (relevant mostly to Track B), access to the computing lab in the Linguistics Department (Oxley Hall) may be needed. We anticipate that existing resources will be sufficient to meet this need. We do not anticipate any impact on other existing programs.

Course Requirements

The course requirements for the two certificate tracks are listed below. All courses are currently offered except for Ling 3803, which is in development and has been submitted for College review and approval.

See Appendix B for sample pathways through the certificate curriculum, Appendix C for the certificate completion sheet, and Appendix D for course descriptions and prerequisite courses.

Track A (less technical)

Prerequisite: Ling 2000(H) or English 3271
Four courses (3 credit hours each), as follows:

1. One course on linguistic analysis (* = also offers an introduction to formal logic):
   a. Ling 2001: Language and Formal Reasoning*
   b. Ling 4100: Phonetics
   c. Ling 4200: Syntax
   d. Ling 4300: Phonology
   e. Ling 4350: Morphology
   f. Ling 4400: Linguistic Meaning*
2. Introduction to human language technology (core course):
   a. Ling 3802(H): Language and Computers

3. One language and technology elective:
    a. Ling 3801: Code Making and Code Breaking
    b. Ling 3803: Ethics of Language Technology
4. One course on methods and tools for computational analysis of language:
    a. Ling 2051(H): Analyzing the Sounds of Language
        i. Note: Although at the 2000 level, this course requires students to use R functions for statistical analysis of language data[5]
    b. Ling 5050: Technical Tools for Linguists
    c. English/Ling 5804: Analyzing Language in Social Media

No prior knowledge of computer programming is required for courses in Track A.

Track B (more technical)

Prerequisite: Ling 2000(H) or English 3271
Four courses (3 credit hours each), as follows:

1. One course on linguistic analysis or introduction to human language technology:
    a. Ling 3802(H): Language and Computers
    b. Ling 3803: Ethics of Language Technology
    c. Ling 4100: Phonetics
    d. Ling 4200: Syntax
    e. Ling 4300: Phonology
    f. Ling 4350: Morphology
    g. Ling 4400: Linguistic Meaning
2. Introduction to computational linguistics (core course):
    a. Ling 5801: Computational Linguistics 1
3. One upper-division course on computational linguistics methods and tools:
    a. Ling 5050: Technical Tools for Linguists*
    b. Ling 5802: Computational Linguistics 2
    c. Ling 5803: Computational Semantics
    d. English/Ling 5804: Analyzing Language in Social Media
    e. CSE 3521: Survey of Artificial Intelligence 1
    f. CSE 5525: Foundations of Speech and Language Processing
4. One additional course from either 1. or 3.

*For students without a background in computer programming, this course (or another introduction to computer programming) is strongly recommended prior to taking Ling 5801.

6. Assessment

We have identified three learning outcomes students are expected to attain after completing the Certificate in Computational Analysis of Language. All courses included in the program are mapped onto each of these learning outcomes and are provided below. Course descriptions are included in Appendix D.

---

[5] R is a programming language for statistical analysis and data visualization

Outcome 1: Students will identify the tasks involved in the computational analysis of human language.
- Courses: Ling 3802(H); Ling 5801; Ling 3801; Ling 3803; Ling 5802; Ling 5803; Ling 5804; CSE 3521; CSE 5525

Outcome 2: Students will apply computational methods, statistical methods, and/or formal logic to the analysis of language data.
- Courses: Ling 3802(H); Ling 5801; Ling 4100; Ling 2051(H); Ling 5050; Ling 5802; Ling 5803; Ling 5804; CSE 3521; CSE 5525

Outcome 3: Students will apply core grammatical concepts and principles to the analysis of language data.
- Courses: Ling 2001; Ling 4100; Ling 4200; Ling 4300; Ling 4350; Ling 4400

Evaluation

The Linguistics Undergraduate Curriculum Committee (LUCC) will conduct an assessment using several metrics, using both direct and indirect measures, to evaluate the viability of the certificate, attainment of learning outcomes, and student satisfaction. First, the LUCC will track enrollments of students in the certificate program and their completion rates over time. It is important to the department that our efforts be worth the time invested in overseeing the program, while it is important to students that they are able to complete the program within a reasonable timeframe. Second, student performance as determined by final grades in certificate courses will be used to compare those students completing the certificate to those who are not. Given that The Department of Linguistics has not offered a certificate up until now, we would like to be aware of potential differences between certificate and non-certificate students. Third, to ensure that each expected learning outcome is met, the LUCC will work with course instructors to develop a set of questions that align with each of the ELOs. Given that these are program learning outcomes, and not course outcomes, these questions will be included in an exit survey required of students upon completing the certificate. The Undergraduate Coordinator, also part of the LUCC, will oversee the entire assessment by tracking enrollments, completion rates, and student grades. The coordinator will also administer the exit survey as part of a requirement for completing the certificate program. The coordinator will share results from the assessment with the LUCC, and then later to the faculty within a department meeting at the onset of each academic year.

Appendix A

Advising Sheet
Certificate in Computational Analysis of Language
Offered by The Department of Linguistics

Coordinating Advisors
Dr. Andrea Sims sims.120@osu.edu   Dr. Julie McGory mcgory.1@osu.edu

Department of Linguistics website: http://www.ling.osu.edu

This 12 credit-hour undergraduate certificate will train students in basic concepts and methods of computational linguistics. It will introduce students to various tasks involved when computers process human speech and text, including speech recognition, text-to-speech conversion, machine translation (automatic translation of text from one human language to another), automated text analysis (e.g. question detection), and natural language generation (e.g. converting data tables into human language). Since computational linguistics is at the intersection of computer science and linguistics, the certificate will give students a basic understanding of both domains. The certificate is designed for current undergraduate students from any major who want to pursue a career related to the creation of language technologies, as well as for professionals in a related area who want to enhance their marketability.

At least 12 credit hours, excluding prerequisites will lead to Certificate in Computational Analysis of Language.

The certificate program must be approved by the Undergraduate Programs Advisor in the Linguistics Department.

The certificate is divided into two tracks, reflecting different kinds of preparation for a career in language technologies.

TRACK A

Track A introduces students to issues and methods in computational linguistics at a conceptual but mostly not a technical level and does not require computer programming, although students can *choose* to do coursework that involves computer programming. This track will prepare students for industry work as Language Specialists, Data Specialists, Localization Specialists, Speech Data Evaluators, Voice User Interface Designers, Language Annotators, and for similar entry-level positions. These jobs generally require a Bachelor's degree in linguistics, a world language, English, or other relevant field. This track is designed to provide students from any background a basic knowledge of the computational analysis of language data, which students can pair with a BA or BS degree in various fields.

No prior knowledge of computer programming is required for courses in Track A.

Coursework (5 courses (1 prerequisite + 4 core courses)
A. Prerequisite: Ling 2000(H) or English 3271 (3 credits)
B. Four Core courses (3 credit hours each), as follows:
     5. One course on linguistic analysis
          a. Ling 2001: Language and Formal Reasoning
          b. Ling 4100: Phonetics
          c. Ling 4200: Syntax
          d. Ling 4300: Phonology
          e. Ling 4350: Morphology
          f. Ling 4400: Linguistic Meaning
     6. Introduction to human language technology (core course):
          a. Ling 3802(H): Language and Computers
     7. One language and technology elective:
          a. Ling 3801: Code Making and Code Breaking
          b. Ling 3803: Ethics of Language Technology
     8. One course on methods and tools for computational analysis of language:

a. Ling 2051(H): Analyzing the Sounds of Language
    i. Note: Despite the 2000-level number, this course requires students to use R functions for statistical analysis of language data
b. Ling 5050: Technical Tools for Linguists
c. English/Ling 5804: Analyzing Language in Social Media

TRACK B

Track B introduces students to issues and methods in computational linguistics at both a conceptual and a technical level and requires basic computer programming, which can be developed through certificate coursework. This track is designed primarily to prepare students for an MS or PhD program in computational linguistics. Admission to these graduate programs requires at least a basic background in (1) linguistics; (2) programming and computer science; and (3) probability, statistics, and formal logic. Track B is designed primarily for students who are pursuing a degree or otherwise have a background in one of these areas but not all three. The certificate will help them bridge the gap between their background and these programs' admissions requirements, while at the same time allowing students to tailor their coursework to their particular needs. This track is designed to lead ultimately to industry positions, for example as a Computational Linguist, Language Engineer, Data Scientist, Analytical Linguist (a kind of data scientist), NLP Scientist, Human Language Technologist, or Research Scientist.

Consult with the Undergraduate Advisor in Linguistics to determine and seek approval for coursework required to complete the certificate.

Coursework (5 courses (1 prerequisite + 4 core courses)
A. Prerequisite: Ling 2000(H) or English 3271 (3 credits)
B. Four Core courses (3 credit hours each), as follows:
    1. One course on linguistic analysis or introduction to human language technology:
        a. Ling 3802(H): Language and Computers
        b. Ling 3803: Ethics of Language Technology
        c. Ling 4100: Phonetics
        d. Ling 4200: Syntax
        e. Ling 4300: Phonology
        f. Ling 4350: Morphology
        g. Ling 4400: Linguistic Meaning
    2. Introduction to computational linguistics (core course):
        a. Ling 5801: Computational Linguistics 1
    3. One upper-division course on computational linguistics methods and tools:
        a. Ling 5050: Technical Tools for Linguists*
        b. Ling 5802: Computational Linguistics 2
        c. Ling 5803: Computational Semantics
        d. English/Ling 5804: Analyzing Language in Social Media
        e. CSE 3521: Survey of Artificial Intelligence 1
        f. CSE 5525: Foundations of Speech and Language Processing
    4. One additional course from either 1. or 3.

*For students without a background in computer programming, this course (or another introduction to computer programming) is strongly recommended prior to taking Ling 5801.

Certificate in Computational Linguistics program guidelines
The following guidelines govern the Certificate in Computational Linguistics.

Credit hour required: Minimum of 12 credits.

<u>Overlap with courses in degree</u>
- The certificate must be in a different subject than the major.
- Max 50% overlap with major, minor, other certificate, or GE.

<u>Grades required</u>
- Minimum C- for a course to be counted on the certificate.
- Minimum 2.00 cumulative point-hour ratio required for the certificate.

<u>X193 credits</u>: Not permitted.

<u>Approval Required</u>: The certificate program must be approved by the Undergraduate Programs Advisor in the Linguistics Department.

<u>Consult with the Linguistics Advisor</u> for filing deadlines, and for changes or exceptions to the certificate.

Appendix B: Semester-by-semester sample program

In the following, courses in green are generally offered twice per year, in fall and spring semesters. Courses in orange are offered at least once per year (or in the case of Ling 3803, we anticipate offering it at least once per year). Courses in blue are offered less than once per year.

Sample pathways through Track A

The following are illustrations of some of the ways students could move through the certificate curriculum for Track A. Arrows indicate courses that must be sequenced because of formal prerequisites. Commas/semicolons indicate courses that can be taken in any order or concurrently.

Ling 2000, Ling 2001, Ling 3802, Ling 3801, Ling 2051
Ling 2000 → Ling 4100; Ling 3802, Ling 3803, Ling 2051
- E.g. someone interested in phonetic aspects of speech processing
Ling 2000, Ling 2001, Ling 3802, Ling 3803, Ling 5050
Ling 2000H → Ling 4200; Ling 3802, Ling 3801, Ling 5050
Ling 2000 → Ling 4400; Ling 3802, Ling 3803, English/Ling 5804
- E.g. someone interested in text analysis

Sample pathways through Track B

The following are illustrations of some of the ways students could move through the certificate curriculum for Track B. Arrows indicate courses that must be sequenced because of formal prerequisites or which are recommended to be taken in sequence. Commas/semi-colons indicate courses that can be taken in any order.

The following list of majors is not meant to be exhaustive. However, Track B is designed primarily for students who independently have background in an area of study related to computational linguistics and who are seeking a course of study that can lead to MS or PhD study in computational linguistics.

For Linguistics majors:

Ling 2000; Ling 3802, Ling 5050 → Ling 5801; English/Ling 5804
- e.g. someone with no prior computer programming experience
Ling 2000; Ling 3802, Ling 5050 → Ling 5801; Ling 4350
Ling 2000 → Ling 4300; Ling 3802 → Ling 5801 → Ling 5802
- e.g. someone with prior computer programming experience, seeking more foundation in NLP
Ling 2000H → Ling 4200, Ling 4400 → Ling 5801 → Ling 5803
Ling 2000H → Ling 4400 → Ling 5801; English/Ling 5804, Ling 3803

For CIS/CSE majors:

Ling 2000, Ling 3803, Ling 5801, CSE 3521, English/Ling 5804
Ling 2000 → Ling 4200; Ling 3803; Ling 5801 → Ling 5802
Ling 2000 → Ling 4200, Ling 4400; Ling 5801 → Ling 5802
- e.g. someone seeking more foundation in linguistic analysis/formal grammar

Ling 2000H, Ling 3802 → Ling 5801 → Ling 5803; CSE 3521
Ling 2000, Ling 3803, Ling 5801, CSE 3521 → CSE 5525

For Data Analytics, Statistics, or similar majors:

Ling 2000 → Ling 4100; Ling 3802, Ling 5050 → Ling 5801
Ling 2000; Ling 3802 → Ling 5050 → Ling 5801; English/Ling 5804
Ling 2000 → Ling 4300, Ling 4400 → Ling 5801 → Ling 5802
Ling 2000 → Ling 4200 → Ling 5801 → CSE 3521; Ling 3803
Ling 2000H → Ling 4350 → Ling 5801 → Ling 5802; English/Ling 5804

Appendix C   **Certificate in Computational Analysis of Language**
Offered by The Department of Linguistics

COLLEGE OF ARTS AND SCIENCES

STUDENT NAME: _____

STUDENT OSU EMAIL: _____

CERTIFICATE ADVISOR NAME: _____

REQUIRED PREREQUISITE, ONE OF THE FOLLOWING (3 CREDITS)

Course (Hours)	Course Grade	Term Completed
Ling 2000 (H)		
English 3271		

TRACK A: FOUR CORE COURSES (12 CREDITS)

Students will select four or more courses from among those listed below.

Course (Hours)	Course Grade	Term Completed
A: ONE COURSE ON LINGUISTIC ANALYSIS		
Ling 2001: Language & Formal Reasoning		
Ling 4100: Phonetics		
Ling 4200: Syntax		
Ling 4300: Phonology		
Ling 4350: Morphology		
Ling 4400: Linguistic Meaning		
B: ONE COURSE ON INTRODUCTION TO HUMAN LANGUAGE TECHNOLOGY		
Ling 3802(H): Language and Computers		
C: ONE LANGUAGE AND TECHNOLOGY ELECTIVE		
Ling 3801: Code Making and Code Breaking		
Ling 3803: Ethics of Language Technology		
D: ONE COURSE ON METHODS AND TOOLS FOR COMPUTATIONAL ANALYSIS OF LANGUAGE		
Ling 2051(H): Analyzing the Sounds of Language		
Ling 5050: Technical Tools for Linguists		
English/Ling 5804: Analyzing Language in Social Media		

SUBSTITUTIONS APPROVED:

_____

TRACK B: FOUR CORE COURSES (12 CREDITS)

Students will select four or more courses from among those listed below.

Course (Hours)	Course Grade	Term Completed
A: ONE COURSE ON LINGUISTIC ANALYSIS OR INTRODUCTION TO HUMAN LANGUAGE TECHNOLOGY		
Ling 3802(H): Language & Computers		
Ling 3803: Ethics of Language Technology		
Ling 4100: Phonetics		
Ling 4200: Syntax		
Ling 4300: Phonology		
Ling 4350: Morphology		
Ling 4400: Linguistic Meaning		
B: ONE COURSE ON INTRODUCTION TO COMPUTATIONAL LINGUISTCS		
Ling 5801: Computational Linguistics 1		
C: ONE UPPER-DIVISION COURSE ON COMPUTATIONAL LINGUISTIC METHODS AND TOOLS		
Ling 5050: Technical Tools for Linguists		
Ling 5802: Computational Linguistics 2		
Ling 5803: Computational Semantics		
English/Ling 5804: Analyzing Language in Social Media		
CSE 3521: Survey of Artificial Intelligence 1		
CSE 5525: Foundations of Speech and Language Processing		
D: ONE ADDITIONAL COURSE FROM EITHER A. OR C.		

SUBSTITUTIONS APPROVED:

_____

CERTIFICATE LINGUISTICS ADVISOR SIGNATURE

_____

DATE _____

Appendix D: Course Descriptions and Prerequisites

Ling 2000(H): Introduction to Linguistics

This course examines language as a system of human communication. It also provides students with the tools needed for the recording, investigation, and close analysis of language. The course consists of a general survey of language and linguistics. A number of topics relating to man's knowledge and use of language are systematically investigated. Examples are drawn primarily from the English language, although other languages are used to illustrate certain concepts. Nevertheless, the focus of the course is not on any specific language or languages; rather, it is on properties common to all languages and on ways in which languages may differ. 3 credit hours. Prerequisites: None

Ling 2001: Language and Formal Reasoning

The goal of this course is to lead students to think analytically about syntax, meaning, and reasoning in terms abstract enough to encompass both natural languages (like English) and artificial formal languages (in this case, first-order logic) to see underlying structural similarities and to understand some fundamental differences as well. This goal is accomplished by (1) introducing students to two kinds of formal systems, first-order logic and formal phrase-structure grammars, (2) using these systems to analyze syntax and reasoning, in symbolic form and in English, and (3) examining differences between artificial and natural language in principles of cooperative communication. 3 credit hours. Prerequisites: Math 1075 or equivalent, or Math placement level R

Ling 2051(H): Analyzing the Sounds of Language

In this course, we will introduce pertinent ideas and results from research in the various disciplines that have contributed to our understanding of the sounds of language. We will introduce some of the quantitative analytical tools that are used in the phonetic sciences, and do several experiments in class, to give a flavor of the diverse research methods that speech scientists have developed to try to determine how speech is produced and perceived by humans. 3 credit hours. Prerequisites: Math 1075 or equivalent, or Math placement level R

Ling 3801: Code Making and Code Breaking

This course has two main aims. It introduces old and new technologies for code making and code breaking, and it shows how good and bad choices in how codes are used can affect whether they succeed or fail. Students will learn what codes are, how they work and how they are used. The topics discussed will include code breaking, digital signatures, quantum cryptography and the decipherment of ancient languages. 3 credit hours. Prerequisites: None

Ling 3802(H): Language and Computers

This is an introduction to human language technology. In this subject area we study whether and how it is possible for humans and computers to communicate in ordinary language. The widening use of computers has had a profound influence on the way ordinary people communicate, search and store information. For the overwhelming majority of people and situations, the main vehicle for such information is human language. Text and speech are crucial encoding formats for the information revolution. This course will give students insight into the fundamentals of how computers are used to represent, process and organize textual and spoken information, as well as provide tips on how to

effectively integrate this knowledge into their work. The course will cover the theory and practice of human language technology. Topics include text encoding, search technology, tools for writing support, machine translation, dialog systems, computer aided language learning and the social context of language technology. 3 credit hours. Prerequisites: Not open to students with first-year standing

Ling 3803: Ethics of Language Technology

Rapid increases in the capabilities of Natural Language Processing (NLP) systems and other language technologies are leading us toward a world in which computers make many of the decisions which affect our everyday lives. NLP systems are already involved in hiring workers, filtering our words online and deciding how political campaigns choose to approach us. These systems have immense power--- but all too often, they make unfair decisions that reflect or even amplify the biases of the society that created them. In this course, we'll learn about how language processing systems are created, and at what steps in the process bias and unfairness might creep in. We'll learn about efforts to define, detect and quantify bias, and how different ethical principles can lead to different results. Finally, we will discuss different ways to remedy the ethical problems of language technology, to what extent they can be 'fixed', and whether there are problems for which it is too dangerous to use NLP at all. 3 credit hours. Prerequisites: None

Ling 4100: Phonetics

Cross-linguistic survey of the sounds of the world's languages. 3 credit hours. Prerequisites: Ling 2000, 2000H, or 5000

Ling 4200: Syntax

Basic elements of syntactic description and an overview of syntactic structure across languages. 3 credit hours. Prerequisites: Ling 2000, 2000H, or 5000

Ling 4300: Phonology

Introduction to phonological analysis and description, and an overview of phonological structure across languages. 3 credit hours. Prerequisites: Ling 2000, 2000H, or 5000

Ling 4350: Morphology

The grammatical analysis of words, and an overview of morphological structure across languages. 3 credit hours. Prerequisites: Ling 2000, 2000H, or 5000

Ling 4400: Linguistic Meaning

Introduction to linguistic meaning across languages, including word meaning, the contribution of syntactic structure, and the role of context in interpretation. 3 credit hours. Prerequisites: Ling 2000, 2000H, or 5000

Ling 5050: Technical Tools for Linguists

Practical training in standard computational tools for tackling different kinds of linguistic research. Students will learn computational techniques to access, search and format linguistic datasets, including text corpora, speech and audio, structured representations such as parse trees, and experimental

measurements. The course will also cover data exploration and basic modeling. 3 credit hours. Prerequisites: None

Ling 5801: Computational Linguistics 1

Symbolic computation applied to the structure of words and sentences, models of morphology and syntax, parsing algorithms. 3 credit hours. Prerequisites: Ling 3802, Ling 5000, CSE 3321, CSE 3522, or CSE 5052; or permission of instructor

Ling 5802: Computational Linguistics 2

Computational models of semantic interpretation, and the role of pragmatic knowledge in sentence processing; implementation of current grammatical theories. 3 credit hours. Prerequisites: Ling 5401 and Ling 5801

Ling 5803: Computational Semantics

Methods for construction semantic representations for fragments of natural language and performing inference with such representations. 3 credit hours. Prerequisites: Ling 5801

English/Ling 5804: Analyzing Language in Social Media

Course gives students experience analyzing language in social media. It covers theoretical issues arising in digital communication and provides hands-on practice at computational data analysis, applicable across fields. Students gain an understanding of the sociolinguistic dynamics of online communication and the technical skills to conduct research on them. No previous experience in linguistics or programming is required, though some background in the study of language will be helpful. Team taught. 3 credit hours. Prerequisites: None

English 3721: Structure of the English Language
Students learn basic characteristics of English linguistics focusing on the basic building blocks of language; the sounds of English and how they are put together, word formation processes, and rules for combining words into utterances/sentences. Students investigate and explore linguistic variation, accents of American English, and the implications of language evaluation in educational settings. 3 credit hours. Prerequisite: English 1110.01

CSE 3521: Survey of Artificial Intelligence 1

Survey of basic concepts and techniques in artificial intelligence, including problem solving, knowledge representation, and machine learning. 3 credit hours. Prerequisites: CSE 2331 or 5331, and enrollment in CSE, CIS, ECE, or Data Analytics major.

CSE 5525: Foundations of Speech and Language Processing

Fundamentals of natural language processing, automatic speech recognition and speech synthesis; lab projects concentrating on building systems to process written and/or spoken language. 3 credit hours. Prerequisites: CSE 3521 or 5521; and CSE 5522, Stat 3460, or Stat 3470.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Linguistics

100 Oxley Hall
1712 Neil Avenue
Columbus, OH 43210

614-292-4052 Phone
614-292-8833 Fax

linguistics.osu.edu

November 13, 2020

Dear Colleagues:

I am writing to express the support of the Department of Linguistics of the proposed Certificate in Computational Analysis of Language. This certificate will provide an excellent opportunity for our students to develop skills in computational linguistics and data science, as well as for students in other majors to engage with linguistic data from a computational perspective.

Sincerely,

Cynthia G. Clopper
Professor and Chair

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Offices of the Associate
and Assistant Deans

114 University Hall
230 North Oval Mall
Columbus, OH 43210

614-292-1667 Phone
asc.osu.edu

November 18, 2020

Vice-Provost Randy Smith
Office of Academic Affairs
Ohio State University

Dear Randy and Committee Members,

I am pleased to convey the support of the College of Arts and Sciences for the proposed certificate, *Computational Analysis of Linguistics*. The strength of this proposal is evident in its cohesive structure, the solid undergraduate student preparation in the subject matter, and the sound collaboration between the Linguistics Department and the Computer Science and Engineering Department.

The curriculum will be inviting to students throughout the university due to the choice between two tracks; one of which delivers a foundational introduction to the field, while the other provides students with sufficient background to pursue graduate work. Moreover, we anticipate an external audience among those in the workforce seeking to improve their current positions or to increase their marketability.

We feel this unique, well-conceived undergraduate certificate will be an important addition to our growing catalog of certificates. The College appreciates your consideration of the proposal which we urge you to approve.

Sincerely,

THE OHIO STATE UNIVERSITY

Deborah Haddad, PhD
Assistant Dean, Curriculum
College of Arts and Sciences

CSE feedback on Computational Analysis of Language Certificate proposal

Sivilotti, Paul
Tue 11/24/2020 12:38 PM

To: Sims, Andrea
Cc: Wenger, Rephael; Fosler-Lussier, Eric; Arora, Anish; McGory, Julia; Clopper, Cynthia

📄 computational-analysis-of-la...
313 KB

Hi Andrea--

I'm the chair of the Undergraduate Studies committee here in the Department of Computer Science. We received your proposal for a new certificate in "Computational Analysis of Language". I'm attaching here the version that we reviewed (named FINAL_11-16-20).

We do have one suggestion for modification: In Track B, CSE 5522 (Survey of Artificial Intelligence 2) should be replaced with CSE 3521 (Survey of Artificial Intelligence 1). This course sequence has recently undergone a curricular redesign which makes 3521 the more appropriate choice for the certificate. In addition, CSE 5522 is likely to be withdrawn in the near future.

I'll also note that we cannot guarantee that students in the certificate program will be able to enroll in the two CSE courses listed as technical electives for the certificate. Because of current enrollment pressures in CSE, students who need courses for their major and for graduation requirements are given higher enrollment priority than those completing certificates. Given the number of alternative courses from which students can choose technical electives, however, this limitation should not pose a significant challenge for students to complete the proposed certificate.

With this caveat, and the change noted above, the CSE department will concur with the proposal.

I'm happy to discuss further if you have any questions. I am also cc'ing Rafe Wenger, our associate chair, who--if you wish--can also provide a concurrence on the modified final-final version.

Best wishes,
--paul

_____
Paul A. G. Sivilotti, Associate Professor
College of Engineering, Dept. of Computer Science & Engineering
395 Dreese Labs, 2015 Neil Avenue, Columbus, OH 43210
614-292-5835 paolo@cse.ohio-state.edu
https://go.osu.edu/sivilotti

From: "Craigmile, Peter" <pfc@stat.osu.edu>
Date: Tuesday, November 10, 2020 at 10:29 PM
To: "Sims, Andrea" <sims.120@osu.edu>
Cc: "Clopper, Cynthia" <clopper.1@osu.edu>, "MacEachern, Steven" <snm@stat.osu.edu>, "Hans, Christopher" <hans@stat.osu.edu>
Subject: Re: concurrence request -- Computational Analysis of Language Certificate proposal

Good evening Andrea.

The Department of Statistics gives concurrence for this certificate, and wishes Linguistics well in offering it.

However we do have one comment:

We think that the following statement is misleading:

"By preparing students for Master's or higher study in computational linguistics, this track is designed to lead ultimately to industry positions, for example as a Computational Linguist, Language Engineer, Data Scientist, Analytical Linguist (a kind of data scientist), NLP Scientist, Human Language Technologist, or Research Scientist. These jobs generally require an MS or PhD in computational linguistics."

Data Scientists and Research Scientists do not generally require an MS or PhD in computational linguistics -- a MS/PhD degree in Statistics, Machine Learning, or Computing Science is normally more common in our experience. We think this should be rephrased in the proposal.

Best wishes,
Peter

Peter Craigmile, Ph.D.,
Professor, Department of Statistics, The Ohio State University.