

FOR APPROVAL

PUBLIC

OPEN SESSION

TO: UTM Academic Affairs Committee

SPONSOR: Tracey Bowen, Vice-Dean, Teaching & Learning
CONTACT INFO: vdteachlearn.utm@utoronto.ca

PRESENTER: As above
CONTACT INFO:

DATE: October 11, 2023 for October 18, 2023

AGENDA ITEM: 4

ITEM IDENTIFICATION:

New Type 2 Certificate Program: Certificate in Computational Linguistics, UTM.

JURISDICTIONAL INFORMATION:

Under section 5.6 of its terms of reference, the Academic Affairs Committee is responsible for major and minor modifications to existing degree programs.

GOVERNANCE PATH:

1. **UTM Academic Affairs Committee [For Approval] (October 18, 2023)**

PREVIOUS ACTION TAKEN:

None.

HIGHLIGHTS:

The Department of Language Studies at the University of Toronto Mississauga (UTM) is proposing to create a Category 2, For-Credit Certificate in Computational Linguistics to consolidate existing interest in computational methodologies that address linguistic questions and process large quantities of unstructured language data. While the proposed certificate program will be open to all UTM undergraduate students, it is anticipated that the majority of students will be drawn from those already studying in a Linguistics or Computer Science program.

The intersection of linguistics and computer science is relevant to practices and research in both fields of study, both inside and outside of the university. Linguists increasingly make use of computational methods to study the nature of human language, and computer scientists are developing ever more intelligent computational systems that are able to extract patterns of practical relevance from large

amounts of unstructured text data. The proposed certificate program provides a consolidated structure for students to engage with this intersection of specializations, either as linguists or as computer scientists.

The certificate is composed of a total of 1.5 credits – 1.0 credit of foundation/ core courses (LIN340H5 Computing with Natural Language and LIN341H5 Linguistics and Computation) and 0.5 credit from a list of elective courses at the 300/ 400-level.

FINANCIAL IMPLICATIONS:

There are no net implications for the campus' operating budget.

RECOMMENDATION:

Be It Resolved,

THAT the New Type 2 Certificate Program: Certificate in Computational Linguistics at UTM, as detailed in the proposal date October 5, 2023, be approved, effective September 1, 2024.

DOCUMENTATION PROVIDED:

Proposal – New Type 2 Certificate Program: Certificate in Computational Linguistics, UTM.

University of Toronto Proposal to Create a Certificate in Conjunction With an Undergraduate Program

Certificates offered in conjunction with an undergraduate program are for-credit undergraduate certificates governed by the [Policy for Certificates \(For-Credit and Not-For-Credit\)](#).

Creation and closure of these certificates follow the protocols for minor modifications; are reviewed with the relevant undergraduate program; and are reported to the Provost through the Office of the Vice-Dean, Academic Programs. Successful completion of the certificate is recorded on the academic transcript. Students must be enrolled in a specific undergraduate program.

This template should be used to bring forward all proposals for new undergraduate, for-credit, certificates that will be offered in conjunction with an existing undergraduate degree program. The creation of the certificate follows a minor modification process and is reported to the VPAP office after approval.

Proposed certificate name: E.g., Certificate in Human Resources Management (Faculty of Arts & Science)	Certificate in Computational Linguistics
Undergraduate degree(s) the certificate will be offered in conjunction with:	Honours Bachelor of Arts – HBA Honours Bachelor of Science – HBSc
Faculty/academic division:	Department of Language Studies University of Toronto Mississauga
Dean’s office contact:	Marc Dryer Associate Dean, Academic Programs Marc.dryer@utoronto.ca Yen Du Manager, Academic Programs, Reviews & Quality Assurance yen.du@utoronto.ca
Version date:	October 5, 2023

1 Summary

- Please provide a brief summary of the certificate, including:
 - academic rationale for certificate
 - impetus for its development (including interest and demand)
 - how the certificate fits with unit/division's academic plans
 - any important or distinctive elements

From trend discovery through the large-scale analysis of social media data to the early detection of neurological disorders, language technology is all around us, and increasingly so. The completion of the Certificate in Computational Linguistics (CCL) ideally situates students to work or pursue further education in this field by providing a solid foundation at the interdisciplinary intersection of linguistics and the computing sciences.

The CCL was conceived by the Department of Language Studies (DLS), with support from the Department of Mathematical & Computational Sciences (MCS). The proposed category 2 certificate program intends to consolidate the existing interest among students in both departments in learning about computational methodologies to address linguistic questions and process large quantities of unstructured language data (see details below).

The CCL provides students with learning that is truly complementary to their primary program of study in Linguistics, Computer Science, or adjacent programs of study. The interdisciplinary nature sets this certificate program apart from Computational Linguistics (CL) courses as they are typically taught in Computer Science departments (including the CL stream of the Computer Science Specialist at UTSG), in two ways. First, the certificate will be accessible to students in Linguistics programs who will develop elementary, but targeted (i.e., oriented to the analysis of language data) knowledge and skills in the domain of the computing sciences without committing to a CS program. Such knowledge and skills are increasingly and frequently used for linguistic analysis, both during a student's course of study at the university and in their practice once they have graduated. Second, the courses in the program place greater emphasis on concepts and questions from Linguistics regarding the nature, structure, and use of language. For computer scientists who intend to work with language data as the object of their computational work, such students will come away with a stronger disciplinary knowledge of linguistics, as well as a grounding in the core values of the discipline of Linguistics.

The certificate consists of two required courses, combined with one elective from among a select set of 300/400-level courses. With the certificate incorporating courses that are accessible from both the Computer Science Minor and the Linguistics Studies Minor, students may be encouraged to consider completing the

complementary minor program in conjunction with their major/ specialist and this certificate to take advantage of the course overlap (i.e. Linguistics students may choose to enrol in the Computer Science Minor; Computer Science students may choose to enrol in the Linguistics Minor).

2 Effective Date

September 1, 2024

3 Academic Rationale

- What are the academic reasons for the certificate, and how does it fit with the unit/division's academic plans?

The intersection of linguistics and computer science is relevant to practices and research in both fields of study, both inside and outside of the university. Linguists increasingly make use of computational methods to study the nature of human language, and computer scientists are developing ever more intelligent computational systems that are able to extract patterns of practical relevance from large amounts of unstructured text data. The proposed certificate program provides a consolidated structure for students to engage with this intersection of specializations, either as linguists or as computer scientists. The 'bridge' of the two required courses allows students specializing in Computer Science to develop an elementary understanding of the structure of human language, enabling them to address software engineering and machine learning problems in a more informed way and with a critical perspective reflective of the values of the field of Linguistics. Similarly, the elementary understanding of algorithmic thought and ability to apply computational techniques to questions of real language use provide students in Linguistics programs with the initial components of a highly transferable set of research skills that are applicable both inside and outside of the university. The structure of the proposed Certificate in Computational Linguistics thus supports the positioning of our graduates to make unique intellectual contributions in both fields. The critical lens on language technology provided in the program supports our core goal of bringing students to think and act critically.

4 Need and Demand

- Provide a brief description of the projected interest in and demand for the proposed certificate.
- Provide details regarding the anticipated yearly in-take.

We recently surveyed CSC and LIN students about their interest in the CCL and what they would like to see in such a program. The CCL matches a learning demand in both groups, with approximately 50 students (out of 65 respondents) indicating being interested in such a certificate (see Appendix C for more details). Students listed a great variety of reasons for their interest, as well as types of topics within the language-computing intersection that they are particularly interested in. Some of these are more applied, and others more theoretical. Responses indicate that both groups of students feel that this certificate could benefit them in their post-graduation career.

In terms of what the certificate would offer students, we can consider both groups separately: for students in Linguistics programs, the experience with computational methodologies through the core courses LIN340H5 and LIN341H5 would help prepare them for graduate school in Linguistics, where such methods are increasingly necessary. Similarly, in other employment pathways, knowledge of language technology allows Linguistics graduates to engage with the technological industry and applications of language technology in government. Graduates from Computer Science programs are similarly supported by the experience and training offered through the certificate to pursue further study in Computational Linguistics (typically housed in Computer Science departments) and enter the job market (e.g., as a data scientist or data engineer) with disciplinary knowledge about language that allows them to approach natural language processing problems with disciplinary knowledge about language. With the challenges of interdisciplinary teaching and learning, the courses have been calibrated over the past years to be maximally attractive for both sets of students (CSC and LIN): LIN340H5 and LIN341H5 both build on prior experience (i.e., there are no prerequisite gaps) and offer mostly new course learning outcomes and concepts for both groups (i.e., there is no substantial catching up to do for one group, thereby taking away from the learning opportunity of the other group).

Based on the interest displayed in the survey, we anticipate a minimal annual in-take of around 10 students (50 interested students in the survey means 12.5 on average per year; if the program-structural entry requirements are simple and students are encouraged early on to consider the program through the LIN and CSC programs, we can anticipate at least 10 new students per year.) This would result in a steady state of approximately 30 students registered in this program.

5 Admission Requirements

- Provide the admission requirements for the certificate.

Admission is open to students in good standing in the HBA and HBA degree programs. Additionally, students will need to have the stated pre-requisites to enrol in the two required courses, LIN340H5 Computing with Natural Language and LIN341H5 Linguistics and Computation. Given the specific content and required courses, we expect that the majority of students in the Certificate will come from either enrolled in the Linguistics Major program and or the CSC Major and or Specialist Programs will be attracted to this Certificate. There are, however, other possible pathways for students to qualify for the required courses in the Certificate outside of the Linguistics and Computer Science programs. To ensure these students have access to this Certificate, we are leaving enrolment open to all HBA and HBA students in good standing.

6 Program Requirements

- This certificate will consist of a coherent sequence of for-credit undergraduate courses related to an identified topic or theme that may complement the degree program.
- Describe the academic requirements of the certificate and mechanism for the assessment of student performance.
- Clarify the certificate program length.
- Is this certificate linked to a particular undergraduate program or degree? Please explain the relationship.

The certificate will consist of two required 300-level H courses on the topic of language and computation and a 0.5 FCE elective:

- LIN340H5 -Computing with Natural Language: A more 'applied' course where students learn to apply existing techniques to natural language data to accomplish tasks that are linguistically or otherwise practically interesting.
- LIN341H5 -Linguistics and Computing: A more 'theoretical' course, where students learn about the basic concepts and research questions of computational linguistics (from a LIN-heavy angle compared to e.g. UTSG Computational Linguistics courses; this course has a limited programming component).

These two courses and their course learning outcomes are designed to directly support the program learning outcomes of the proposed certificate (see Appendix A). Student performance in them will be assessed through course assignments, including programming exercises, oral presentation in the form of a 'vlog' of a critical reflection

on a language technology, and writing assignments linking computational approaches to a theoretical problem in Linguistics.

The 0.5 FCE elective should be satisfied with a course from a list of 300-level and 400-level CSC and LIN courses, which encourages students to bring the interdisciplinary knowledge and skills to a topically adjacent course within one of their programs. (While the Courses courses listed below have prerequisites that must be met/ completed)., students in the Certificate program will have access to these courses as well as the stated pre-requisites, if they so choose, to enable them to complete the Certificate requirements.

- JLP384H5 - Speech Communication,
- LIN318H5 - Talking Numbers,
- LIN441H5 - Computing Meaning,
- CSC311H5 - Introduction to Machine Learning,
- CSC384H5 - Introduction to Artificial Intelligence,
- CSC363H5 - Computational Complexity and Computability
- CSC428H5 - Human-Computer Interaction

7 Consultation

- Outline any consultation undertaken with the Dean and chair/director of the relevant academic units and relevant programs.

Professor Andrew Peterson was consulted in his capacity as Acting Vice-Dean, Teaching & Learning, Acting Associate Dean, Pedagogical Development and Scholarship and as a faculty member in Mathematical and Computational Sciences (MCS) involved with the program development from the outset. Previous versions of the program design and learning outcomes were discussed with him over the period 2018-2021. A previous complete draft of this document was circulated among the relevant departments: From MCS, Professors Jessica Burgner-Kahrs and Daniel Zingaro were consulted in their capacity as members of the MCS curriculum committee, and Professor Konstantin Khanin in his capacity as (former) chair of the department. From the Department of Language Studies (DLS), the Linguistics faculty were consulted, as was the program coordinator of the Linguistics program. Furthermore, Professor Emmanuel Nikiema, the former chair of DLS has been consulted on the program as well.

Additional consultation occurred through the UTM Office of the Dean with the UTM Department of Mathematical & Computational Sciences (MCS), the UTM Department of Psychology, and through the Dean's Offices within the Faculty of Arts & Science (St.

George campus) and UTSC. Feedback received was incorporated into the proposal, as appropriate.

8 Resources

- Describe any resource requirements including, but not limited to, faculty complement, space, libraries and enrolment/admissions.
- Indicate if the certificate will affect any existing agreements with other institutions, or will require the creation of a new agreement to facilitate the certificate (e.g., Memorandum of Understanding, Memorandum of Agreement, etc). Please consult with the Provost's office (vp.academicprograms@utoronto.ca) regarding any implications to existing or new agreements.

The certificate program will not have any additional resource requirements in terms of faculty complement or facilities. The Certificate includes course offerings taught by current faculty in our existing facilities, such that it does not entail the creation of any new courses or additional teaching resources, including teaching assistants. The administration of the certificate (eligibility checks for students, maintenance and updating of eligible courses and program requirements as part of the normal curriculum review process) will be taken on by the DLS Academic Advisor & Undergraduate Program Administrator. These duties will fall within their current workload. The certificate program does not affect existing agreements with other institutions.

9 Oversight and Accountability: Review

- Category 2 certificates are subject to periodic reviews with the relevant undergraduate program. Please provide details. This will be tracked by the VPAP office.

As the Certificate in Computational Linguistics will be housed in the DLS, this unit will have the administrative and academic responsibilities to 1) inform and connect with partners departments for regular updates to the list of eligible courses and 2) review the academic quality of the Certificate along with the mandated review of its academic programs. The certificate will first be reviewed as part of the UTQAP review of the Department and its programs in 2030-2031. The DLS will also conduct annual consultations with MCS, in order to update the list of eligible courses as per the current UTM Calendar.

10 Process Steps and Approvals

The pathway is summarized in the table below.

	Approving Body	Approval Date
Development & Consultation within Unit	Salvatore Bancheri Chair, Department of Language Studies	October 10, 2023
Consultation with Dean's Office (and VPAP)	Marc Dryer Associate-Dean, Academic Programs	October 6, 2023
	VPAP sign-off	October 6, 2023
Divisional Governance Approval	UTM Academic Affairs Committee	October 18, 2023
Submission to Provost's Office		May 2024
AP&P (report)		July 2024

Appendix A: Proposed Learning Outcomes

Certificates offered in conjunction with an undergraduate program will have a sub-set of complementary learning outcomes in relation to the program. Divisions are responsible for developing the outcomes and expectations for certificates in the context of divisional norms. Please outline in the table below how the design, structure, requirements and delivery of the certificate support the certificate learning outcomes and expectations.

Certificate Expectations	Certificate Learning Outcomes	How the Design/Structure Supports the Certificate Expectations
<p>1. Depth and Breadth of Knowledge</p> <p><i>Breadth of Knowledge:</i> In the course of their studies, students will gain an awareness and appreciation of the variety of modes of thinking, methods of inquiry and analysis, and ways of understanding the world that underpin different intellectual fields. Students will engage in critical thinking and analytical skills – including with respect to equity, diversity, and inclusion – through courses within and beyond their core field(s) of study, across the humanities, the social and behavioural sciences, and the natural sciences.</p> <p><i>Depth of Knowledge:</i> Students will attain depth of knowledge in their core field(s) of study through a progression of introductory, core, and specialized courses.</p>	<p>Identify and explain the central insights of the interdisciplinary field of computational linguistics, including the equity, diversity and inclusion implications of language technology (Breadth of knowledge)</p>	<p>The two core courses (LIN340H5 and LIN341H5) are designed to support and assess both outcomes: lectures (LIN341H5) and demonstrations through live-coding (LIN340H5) introduce the students to the elements of knowledge of the interdiscipline. Students practice with their understanding of the basic concepts through in-class exercises and homework exercises. Student performance on these outcomes will be assessed through course assignments, including programming tests (in LIN340H5) and quizzes (in LIN341H5), and writing assignments linking computational approaches to a theoretical problem in Linguistics (in LIN341H5). Developing an understanding of the EDI dimension of language technology is supported by lecture components in LIN341H5 and assessed through a class presentation. Students furthermore practice with such aspects in the practical sessions of LIN340H5.</p>

<p>Methodological awareness; application of knowledge; professional capacity.3. Application of Knowledge</p> <p>Students will be able to frame relevant questions for further inquiry within or beyond the core field(s) of study. They will be able to identify and apply the appropriate tools with which they can address such questions effectively. This includes a knowledge of how historical and present discrimination (including, but not limited to, discrimination on the basis of race, religion, sexuality, gender, and ability) affect these questions, problems, and solutions.</p>	<p>Assess the appropriateness of, and apply existing computational linguistic techniques to simplified theoretical and applied research problems. (Methodological awareness; application of knowledge; professional capacity)</p>	<p>The two core courses (LIN340H5 and LIN341H5) are designed to support and assess this outcome: the application of knowledge will take centre stage in LIN340H5, where students work on ongoing course projects as part of the practical sessions – these projects form the portfolio that the students submit as (part of) their assessment for the course. In LIN341H5, the application of knowledge is supported through pencil-and-paper problem sets, evaluated as part of the quizzes and through the written reflection assignments.</p> <p>Elective courses provide further opportunities in which students are able to apply the learned methodologies</p>
<p>5. Awareness of Limits of Knowledge</p> <p>Students will acknowledge and appreciate the limits of their own knowledge. They will also gain an awareness of the uncertainty, ambiguity, and limits of our collective knowledge and how these might influence analyses and interpretations.</p>	<p>Recognize and explain the limitations of computational techniques as applied to language, in particular concerning issues of social bias.</p>	<p>The two core courses (LIN340H5 and LIN341H5) are designed to support and assess this outcome by demonstrating the limitations of the techniques, having students explore them in practice using scaffolded exercises (LIN340H5) and on a conceptual level (LIN341H5) through reflection exercises. In LIN340H5, the PLO is assessed through a graded programming exercise, whereas in LIN341H5, a presentation (vlog, class presentation, public-facing write-up) constitutes the assessment for this PLO</p>

Appendix B: Proposed Calendar Copy

The Certificate in Computational Linguistics is open to students who seek a deeper understanding of the diverse ways in which Linguistics and the Computing Sciences interact. The requirements include two core advanced courses in which students will learn about theoretical and practical aspects of the intersection between Linguistics and the Computing Sciences, and a 0.5 FCE elective from a set of topically fitting LIN and CSC courses. While Enrolment enrolment in the Certificate in Computational Linguistics is open to all students completing any programs at UTM program, interested students should note that there are significant pre-requisites which must be met before enrolment in the required foundational courses. It is expected that this Certificate will be of most interest to students in Computer Science and Linguistics programs.

1.5 credits are required.

1. Foundation: LIN340H5 and LIN341H5

2. 0.5 credit to be chosen from the following list: JLP384H5 or LIN318H5 or LIN441H5 or CSC311H5 or CSC384H5 or CSC363H5 or CSC428H5

Appendix C: Summary of survey results

Survey CL programming

Basic demographics

In the survey, we heard from 65 students (9 first-year, 12 second-year, 44 upper-year). 57/65 students go to UTM, 7 to UTSG, and 1 to UTSC. In terms of POST enrollment, we see that:

- 1 9 of the students are enrolled in both a LIN and a CSC (or similar: MAT, STA, Bioinformatics, GIS) program;
- 2 17 are enrolled in a LIN program but not a CSC program or similar;
- 3 26 in a CSC or similar program;
- 4 13 are enrolled in neither (most have another POST, some are yet undeclared).

Students who are not enrolled in LIN indicate not having considered it because the job prospects of other programs are better. However, some indicate a secondary interest:

- *"I might have given it more consideration had I known how similar the studies were to computer science theory."*
- *"I have considered it, especially to complement computer science, as computational linguistics and things like natural language processing interest me."*

Students who are not enrolled in CSC cite an insufficient CGPA, the perceived toughness of the program, the financial considerations of it being a deregulated program as reasons.

Certificate in CL

When asked about Certificate programs in general, 78% of the students indicated not knowing what Certificate programs are. Among those who do, however, the vast majority (93%) expect the annotation on their degree to benefit them after graduation.

When asked about a CL Certificate (*If there were a Certificate in CL that you would get by taking two level-300 CL courses plus a level-300 or level-400 elective, would you be interested to take it?*), 54 (83%) of the respondents indicated being interested in the program. Respondents volunteered the following remarks about such a program:

- *"I think they would be interesting and offer the best of both worlds."*
- *"Yes because that was what I originally wanted to do, CSC with LIN."*

- *"I do believe future RPG games will rely on NLP and computational linguistics to create more immersive worlds."*

Of note is that some CSC students indicate that this would be a great option, as they would not have to take Linguistics, and similarly LIN students who'd like to learn about this without taking computer science:

- *"Please don't gate CL courses behind linguistics. Because a lot of CS people are interested primarily because of advanced natural language processing. They are not interested because they want to learn about synchronous context grammars. At the end of the day, people want to build chatbots and do stuff like sentiment analysis. Beyond basic syntax, a lot of linguistics concepts are not very relevant for the newer neural network based technologies."*
- *"Yes. I can't dedicate myself to major in linguistics but I would be open to taking a couple courses in computational linguistics."*
- *"YES. It is just such a huge and, with growing technology, an important part of linguistic so I would love to have the chance to learn about it even if I don't necessary major or minor in CSC."*

Topics for future CL courses

Part of the CL Certificate program as it stands is to have two courses and electives. This allows us to explore particular course and program outcomes. Given the breadth of the field, many possible outcomes can be imagined.

The students were asked about their interest in three broad topics that can be said to form different 'flavours' of CL:

- mathematical linguistics (discrete mathematics, set theory, logic, some graph theory),
- quantitative linguistics (using 'big data' and statistics to study language; psycholinguistics, sociolinguistics), and
- practical CL, or 'NLP' (Natural Language Processing; machine learning, sentiment analysis, customer research).

The table below presents the results broken down over students enrolled in LIN (and possibly CSC or similar), CSC or similar (and possibly LIN), and neither. Students are overall interested in these topics, but the interest in the flavour geared more strongly towards traditional linguistics (mathematical linguistics) is unsurprisingly stronger among the LIN-enrolled students than among the CSC-enrolled ones. Vice versa, NLP is more popular among the CSC'ers than among the LIN'ers.

	mathematical linguistics	quantitative linguistics	practical CL (NLP)	total
LIN	22 (85%)	20 (77%)	20 (77%)	26
CSC/similar	23 (66%)	25 (71%)	30 (86%)	35
neither	8 (73%)	10 (91%)	9 (82%)	11

Table: Number and proportion of students indicating being 'interested' or 'very interested' (as opposed to 'very uninterested', 'uninterested', or 'indifferent') about a certain flavour of CL.

When asked which topics students would further care about in an open question, students mentioned the full breadth of possibilities: mathematical linguistics, learning to use practical programming libraries, state-of-the-art machine learning applied to language, speech synthesis, cognitive modeling, ethics of CL, and studying linguistic diversity and language history with computational techniques.

Concluding

The survey indicates that there is substantial interest in the CL Certificate Program. That interest is furthermore broad, meaning that there is interest in all the flavours of CL. As such, the survey supports our proposal in general, and supports in particular a program that is broadly set up (as it currently is). Importantly, certificate programs are not very well known; an important part of our mission should be to advertise this option early on in the various POSTs.