

# LT5427: PROGRAMMING FOR LINGUISTS IN THE ERA OF GENERATIVE AI

---

## Effective Term

Semester A 2026/27

## Part I Course Overview

### Course Title

Programming for Linguists in the Era of Generative AI

### Subject Code

LT - Linguistics and Translation

### Course Number

5427

### Academic Unit

Linguistics and Translation (LT)

### College/School

College of Liberal Arts and Social Sciences (CH)

### Course Duration

One Semester

### Credit Units

3

### Level

P5, P6 - Postgraduate Degree

### Medium of Instruction

English

### Medium of Assessment

English

### Prerequisites

Nil

### Precursors

Nil

### Equivalent Courses

Nil

### Exclusive Courses

Nil

## Part II Course Details

### Abstract

This course is designed to provide linguistics students with the foundational programming skills essential for modern computational linguistics and Natural Language Processing (NLP). Recognizing the transformative impact of large language models, this course bridges fundamental programming concepts with their application in the era of generative AI. Students will begin with the Unix command line to efficiently manipulate text data, and then learn the fundamentals of Python programming, including core data structures. The curriculum will cover essential data science libraries like NumPy and Pandas, and NLP-specific libraries such as NLTK. A significant portion of the course is dedicated to introducing the Transformers library, enabling students to leverage pre-trained models for various NLP tasks. This hands-on, project-based course aims to equip students with the practical skills needed for advanced studies and research in computational linguistics.

### Course Intended Learning Outcomes (CILOs)

| CILOs |                                                                                                                                                                                              | Weighting (if app.) | DEC-A1 | DEC-A2 | DEC-A3 |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|--------|--------|--------|
| 1     | Operate the Unix command line to connect to remote servers and effectively inspect, process, and manipulate text corpora.                                                                    | 20%                 |        | x      | x      |
| 2     | Write well-structured Python scripts that utilize fundamental data structures (strings, lists, dictionaries) and control flow to solve linguistic problems.                                  | 30%                 | x      | x      |        |
| 3     | Apply key Python libraries, including NumPy, Pandas, and NLTK, for the quantitative analysis and processing of linguistic data.                                                              | 30%                 |        | x      | x      |
| 4     | Utilize the Transformers library to perform foundational NLP tasks (e.g., text classification, generation) with pre-trained language models and articulate their significance in modern NLP. | 20%                 | x      | x      | x      |

#### A1: Attitude

Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

#### A2: Ability

Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

#### A3: Accomplishments

Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

### Learning and Teaching Activities (LTAs)

| LTAs |          | Brief Description                                                                                                                                                      | CILO No.   | Hours/week (if applicable) |
|------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|----------------------------|
| 1    | Lectures | Concepts of Unix, Python programming, and NLP libraries will be introduced. The lectures will include live coding demonstrations to illustrate practical applications. | 1, 2, 3, 4 |                            |

|   |                                      |                                                                                                                                                                                              |            |  |
|---|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|--|
| 2 | Lab Sessions & Programming Exercises | Students will participate in hands-on lab sessions to practice writing Unix commands and Python code, reinforcing the concepts learned in lectures and applying them to linguistic datasets. | 1, 2, 3, 4 |  |
| 3 | Readings                             | Students will read assigned chapters from textbooks and articles from online documentation to deepen their understanding of the tools and theoretical concepts.                              | 1, 2, 3, 4 |  |

**Assessment Tasks / Activities (ATs)**

| ATs | CILO No.                | Weighting (%) | Remarks ("- for nil entry) | Allow Use of GenAI?                                                                                                                                                                                                                                                                                                                                                                                   |     |
|-----|-------------------------|---------------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1   | Programming Assignments | 1, 2, 3, 4    | 40                         | Students can use GenAI tools to brainstorm, gather information, solve problems, format references, and edit language. Students should give proper acknowledgment to all the content from GenAI tools that they incorporated into the assignment and keep a journal of their GenAI conversations. Students should not present texts generated or translated by GenAI tools as their own original work. | Yes |
| 2   | Quiz                    | 1, 2          | 20                         | A practical quiz to assess foundational Unix and Python skills.                                                                                                                                                                                                                                                                                                                                       | No  |

**Continuous Assessment (%)**

60

**Examination (%)**

40

**Examination Duration (Hours)**

2

**Assessment Rubrics (AR)**

**Assessment Task**

Programming Assignments

**Criterion**

Ability to apply Unix and Python programming concepts to solve problems in linguistic analysis.

**Excellent (A+, A, A-)**

Demonstrate excellent ability to write Unix and Python code for data analysis and NLP tasks.

**Good (B+, B, B-)**

Demonstrate significant ability to write Unix and Python code for data analysis and NLP tasks.

**Fair (C+, C, C-)**

Demonstrate moderate ability to write Unix and Python code for data analysis and NLP tasks.

**Marginal (D)**

Demonstrate basic ability to write Unix and Python code for data analysis and NLP tasks.

**Failure (F)**

Not even reaching marginal levels

---

**Assessment Task**

Quiz

**Criterion**

Ability to demonstrate competence in fundamental Unix commands and Python syntax.

**Excellent (A+, A, A-)**

Demonstrate excellent ability and understanding in writing Unix and Python code for data analysis and NLP tasks.

**Good (B+, B, B-)**

Demonstrate significant ability and understanding in writing Unix and Python code for data analysis and NLP tasks.

**Fair (C+, C, C-)**

Demonstrate moderate ability and understanding in writing Unix and Python code for data analysis and NLP tasks.

**Marginal (D)**

Demonstrate basic ability and understanding in writing Unix and Python code for data analysis and NLP tasks.

**Failure (F)**

Not even reaching marginal levels

---

**Assessment Task**

Examination

**Criterion**

Ability to demonstrate competence in applying Python libraries for NLP tasks and understanding of Transformer models.

**Excellent (A+, A, A-)**

Demonstrate excellent ability to write code for data analysis and NLP tasks.

**Good (B+, B, B-)**

Demonstrate significant ability to write code for data analysis and NLP tasks.

**Fair (C+, C, C-)**

Demonstrate fair ability to write code for data analysis and NLP tasks.

**Marginal (D)**

Demonstrate basic ability to write code for data analysis and NLP tasks.

**Failure (F)**

Not even demonstrating marginal ability to write code for data analysis and NLP tasks.

---

## Part III Other Information

**Keyword Syllabus**

- **Unix Environment:** Secure Shell (ssh), command-line navigation, file and directory manipulation, text processing tools (cat, head, grep, wc, sort, uniq), pipes (|), basic shell scripting.
- **Python Fundamentals:** Variables, data types (strings, lists, tuples, sets, dictionaries), control flow (loops, conditionals), functions, file I/O, regular expressions.
- **Python for Data Science:** NumPy for numerical operations, Pandas for data manipulation with DataFrames, Matplotlib for data visualization.
- **Natural Language Processing:** NLTK for basic NLP tasks (tokenization, part-of-speech tagging, frequency distributions). [1] (<https://tjzhifei.github.io/resources/NLTK.pdf>)[2] (<https://www.tableau.com/learn/articles/natural-language-processing-books/>)
- **Generative AI and Transformers:** Introduction to the Hugging Face ecosystem, using pipelines for zero-shot tasks, working with pre-trained models (e.g., BERT, GPT) for basic NLP tasks such as text classification, question answering, and text generation. [3] (<https://transformersbook.com/>)[4] (<https://www.oreilly.com/library/view/natural-language-processing/9781098136789/>)

**Reading List**

**Compulsory Readings**

| Title |                                                                                                                                                     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Tunstall, L., von Werra, L., & Wolf, T. (2022). Natural Language Processing with Transformers, Revised Edition. O'Reilly Media.                     |
| 2     | Janssens, J. (2021). Data Science at the Command Line: Obtain, Scrub, Explore, and Model Data with Unix Power Tools (2nd Edition). O'Reilly Media.  |
| 3     | Matthes, E. (2023). Python Crash Course (3rd Edition). No Starch Press.                                                                             |
| 4     | Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media. |

**Additional Readings**

| Title |                                                                              |
|-------|------------------------------------------------------------------------------|
| 1     | Hammond, M. (2020). Python for Linguists. Cambridge University Press.        |
| 2     | McKinney, W. (2022). Python for Data Analysis (3rd Edition). O'Reilly Media. |

|   |                                                                                                                                     |
|---|-------------------------------------------------------------------------------------------------------------------------------------|
| 3 | Jain, S. M. (2022). Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems. Apress.       |
| 4 | Hugging Face. (n.d.). Hugging Face Documentation and Courses. <a href="https://huggingface.co/docs">https://huggingface.co/docs</a> |