11737: Multilingual NLP Fall 2023

**PI:** Lei Li (CMU)

## Primary Field of Science:

Artificial Intelligence and Intelligent Systems

## Introduction

This proposal requests the computing resources of the CMU 2023 Fall course, 11737: Multilingual NLP. CMU 11-737 is an advanced graduate-level course on natural language processing techniques applicable to many languages. Students who take this course should be able to develop linguistically motivated solutions to core and applied NLP tasks for any language. This includes understanding and mitigating the difficulties posed by lack of data in low-resourced languages or language varieties, and the necessity to model particular properties of the language of interest such as complex morphology or syntax. The course will introduce modeling solutions to these issues such as multilingual or cross-lingual methods, linguistically informed NLP models, and methods for effectively bootstrapping systems with limited data or human intervention. The project work will involve building an end-to-end NLP pipeline in a language you don’t know.

## Syllabus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Date | Topic | Material | Homework |
| 1 | 8/29 | Introduction | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-01-intro.pdf) | Reading List out |
| 2 | 8/31 | Typology: The Space of Languages | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-02-typology.pdf) |  |
| 3 | 9/5 | Words and Morphology | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-03-word.pdf) |  |
| 4 | 9/7 | Sequence Labeling | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-04-seqlabeling.pdf) | [HW1 out](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/assignment/assignment1_mt/index.html) |
| 5 | 9/12 | Machine Translation Overview and Evaluation | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-05-mt.pdf) |  |
| 6 | 9/14 | Neural Machine Translation Models | [Slides](https://www.cs.cmu.edu/~leili/course/11737mnlp23fa/slides/multiling-06-mtmodel.pdf) |  |
| 7 | 9/19 | Semi-supervised and Unsupervised MT | Slides |  |
| 8 | 9/21 | Multilingual NMT | Slides |  |
| 9 | 9/26 | Pre-training for NMT |  | HW1 Due, HW2 out |
| 10 | 9/28 | Language Contact and Change, Code Switching, Pidgins, Creoles |  |  |
| 11 | 10/3 | Automatic Speech Recognition |  |  |
| 12 | 10/5 | Sequence-to-sequence Speech Recognition |  | Project Proposal Due |
| 13 | 10/10 | Text-to-speech |  |  |
| 14 | 10/12 | Multilingual ASR and TTS |  |  |
|  | 10/17 | Fall Break |  |  |
|  | 10/19 | Fall Break |  |  |
| 15 | 10/24 | Speech Translation |  | HW 2 Due |
| 16 | 10/26 | Streaming Speech Translation |  |  |
| 17 | 10/31 | Guest Lecture: Juan Pino (Meta) |  |  |
| 18 | 11/2 | Morphological Analysis and Inflection |  | Mid-term Report Due |
|  | 11/7 | Democracy Day Holiday |  |  |
| 19 | 11/9 | Syntax and Parsing |  |  |
| 20 | 11/14 | Data Annotation and Data mining |  |  |
| 21 | 11/16 | Multilingual Question Answering |  | HW3 Due |
| 22 | 11/21 | Vocabulary Learning |  |  |
|  | 11/23 | Thanksgiving, no classes |  |  |
| 23 | 11/28 | Parallel Decoding |  |  |
| 24 | 11/30 | Guest Lecture and The LORELEI Project |  |  |
|  | 12/5 | Poster Presentations |  |  |
|  | 12/7 | Poster Presentations |  | Final Report Due |

## Computing Resource for the Course Assignments

This course will include two assignments and one course project. The assignments are training two systems for machine translation and speech recognition. Nowadays, all of these components are based on deep learning, and massive GPU resources are necessary.

## Request details

There are 31 active students in our class.

We plan also to provide the AWS credits for the student to work on their own development.

We will use the PSC resource for the course-term project to scale up their experiments for both training and inference. The GPU usages are critical for the success of the course.

We are expecting that additional GPU and CPU resources will enable the student to finish their course-term project with a sufficient level and can finalize their projects with scientific publications.