**601.467/667 Introduction to Human Language Technology**

**Fall 2022**

**Coordinator:** [Philipp Koehn](http://www.cs.jhu.edu/~phi/) (phi@jhu.edu)  
**TA:** Yunmo Chen (yunmo@jhu.edu)  
**CAs:** Rosalyn Shin (hshin36@jhu.edu), Venkat Mukthineni (vmukthi1@jhu.edu)  
**Class:** Tuesday and Thursday 9:00-10:15am, Gilman 50  
**Office hours:** Coordinator: on request  
**Office hours:**  
Yunmo: Friday, 11:00 am - 12:00 pm, Malone 216 / [Zoom](https://jhubluejays.zoom.us/j/91893338600)  
Venkat: Monday, 3:00 pm - 4:00 pm, Malone 216 / [Zoom](hhttps://wse.zoom.us/j/7760470402?pwd=T2J4YUF5QkZZcmFUV2d6dVR6Z1ZEUT09)  
Rosalyn: Tuesday, 3:30 pm - 4:30 pm, Malone u216 / [Zoom](https://jhubluejays.zoom.us/j/3893623446)  
(Zoom-only session would be announced in advance on Piazza)  
[**Gradescope (entry code: DJKDYD)**](https://gradescope.com/) ☆ [**Piazza**](https://piazza.com/jhu/fall2022/601467667) ☆ [**Lecture recordings**](https://docs.google.com/document/d/1pNyOjVTUeagBYFAgyrGAjqUVv4OW1NptgwxY8PIUc80/edit?usp=sharing)

**Assignments**

**Note:**For Fall 2022, please \*do not\* proceed before assignments being announced. They are subject to change.  
You can confirm whether the homework is released by checking the date listed on each homework.  
  
**Late submissions:** For each student, we allow a total of 10 days of late submission for all homeworks.  
It is counted on a daily basis, for example, if you submit a homework even a few minutes late, you will lose 1 day of your quota.  
After you use up all 10 days of late submission, each late day would cost 5% points penalty.  
Late submission for teamwork would use 1-day for each teammate.  
For each homework, you \*are not allowed to submit\* after 20 days.

1. [N-gram language modeling, CYK parsing](https://jhu-intro-hlt.github.io/assignments/intro-hlt-fall22-hw1.html): Due on September 28 (Wednesday)
2. [RNNLMs, word2vec](https://jhu-intro-hlt.github.io/assignments/intro-hlt-fall22-hw2.html): Due on October 22 (Saturday)
3. [Seq2seq for pronunciation prediction](https://jhu-intro-hlt.github.io/assignments/intro-hlt-fall22-hw3.html): Due on November 20 (Sunday)
4. [Speech recognition with CTC](https://jhu-intro-hlt.github.io/assignments/intro-hlt-fall22-hw4.html): Due on December 14 (Wednesday)

**Exam**

There will be two mid-terms and final exam. You are allowed to bring 1 sheet of paper with notes to the exam.

* First midterm
* Second midterm
* Final

The final exam time is December 14, 6-7:15pm

* [First midterm exam from 2021.](https://jhu-intro-hlt.github.io/exam/midterm-2021.pdf)
* [Second midterm exam from 2021.](https://jhu-intro-hlt.github.io/exam/second-midterm-2021.pdf)
* [Final exam 2021.](https://jhu-intro-hlt.github.io/exam/final-2021.pdf)
* [First midterm exam from 2020.](https://jhu-intro-hlt.github.io/exam/first-midterm-2020.pdf)
* [Second midterm exam from 2020.](https://jhu-intro-hlt.github.io/exam/second-midterm-2020.pdf)
* [Final exam 2020.](https://jhu-intro-hlt.github.io/exam/final-2020.pdf)
* [Midterm exam from 2019.](https://jhu-intro-hlt.github.io/exam/midterm-2019.pdf)
* [Example questions for the final exam (2019).](https://jhu-intro-hlt.github.io/exam/final-example.pdf)

**Lectures**

|  |  |  |
| --- | --- | --- |
| **Date** | **Topic** | **Instructor** |
| Tu Aug 31 | [Introduction](https://jhu-intro-hlt.github.io/slides/lecture-introduction.pdf) | [Koehn](http://www.cs.jhu.edu/~phi/) |
| **Text** | | |
| Th Sep 1 | [Words and Language Models](https://jhu-intro-hlt.github.io/slides/language-modeling.pdf) | [Yarowsky](https://www.cs.jhu.edu/~yarowsky/) |
| Tu Sep 6 | [Morphology](https://jhu-intro-hlt.github.io/slides/morphology.pdf) | [Yarowsky](https://www.cs.jhu.edu/~yarowsky/) |
| Th Sep 8 | [Syntax](https://jhu-intro-hlt.github.io/slides/2022-09-08-IntroHLT-Syntax.pdf) | [Post](https://mjpost.github.io/) |
| Tu Sep 13 | [Semantics](https://jhu-intro-hlt.github.io/slides/semantics-2020.pdf) | [Lippincott](http://www.logical-space.org/) |
| Th Sep 15 | [Machine Translation](https://jhu-intro-hlt.github.io/slides/machine-translation-2021.pdf) | [Post](https://mjpost.github.io/) |
| Tu Sep 20 | [Deep learning I](https://jhu-intro-hlt.github.io/slides/deep-learning-1-2021.pdf) | [Murray](https://kentonmurray.com/) |
| Tu Sep 22 | [Distributional Semantics](https://jhu-intro-hlt.github.io/slides/distributional-semantics.pdf) | [Koehn](http://www.cs.jhu.edu/~phi/) |
| Th Sep 27 | [Deep learning II](https://jhu-intro-hlt.github.io/slides/deep-learning-2-2021.pdf) ([Python notebook](https://github.com/KentonMurray/IntroHLT/blob/main/DeepLearning/IntroHLT-checkpoint.ipynb)) | [Murray](https://kentonmurray.com/) |
| Th Sep 29 | [Information Retrieval](https://jhu-intro-hlt.github.io/slides/information-retrieval.pdf) | [Duh](http://cs.jhu.edu/~kevinduh/) |
| Tu Oct 4 | [Information Extraction](https://jhu-intro-hlt.github.io/slides/lecture-information-extraction.pdf) | [Koehn](http://www.cs.jhu.edu/~phi/) |
| Th Oct 6 | First Midterm Exam | - |
| **Speech** | | |
| Tu Oct 11 | [Auditory system](https://jhu-intro-hlt.github.io/slides/auditory-system-2022.pdf) | [Elhilali](https://engineering.jhu.edu/ece/faculty/elhilali-mounya/) |
| Th Oct 13 | [Speech basics](https://jhu-intro-hlt.github.io/slides/2022-10-13-SpeechBasics.pdf) | [Moro-Velazquez](https://pages.jh.edu/lmorove1/index.html) |
| Th Oct 18 | [Speaker recognition](https://jhu-intro-hlt.github.io/slides/speaker-id-2022.pdf) | [Villalba](https://www.clsp.jhu.edu/faculty/jesus-villalba/) |
| Tu Oct 25 | [Enhancement and Diarization](https://jhu-intro-hlt.github.io/slides/JSALTDiarization-Tutorial%201.pdf) | [Garcia](https://www.clsp.jhu.edu/faculty/paola-garcia/) |
| Tu Oct 27 | [Classic speech recognition](https://jhu-intro-hlt.github.io/slides/IMA-Math-Seminar.pdf)[1](https://jhu-intro-hlt.github.io/#footnote1) ([additional slides](https://jhu-intro-hlt.github.io/slides/2016-06-WS16-Tutorial.pdf)) | [Khudanpur](https://www.clsp.jhu.edu/faculty-pages/sanjeev/) |
| Th Nov 1 | [End-to-end neural speech recognition](https://jhu-intro-hlt.github.io/slides/NeuralASR4IntroHLT.pdf) | [Khudanpur](https://www.clsp.jhu.edu/faculty-pages/sanjeev/) |
| Tu Nov 3 | Hands on: Kaldi ([K2](https://colab.research.google.com/drive/15AZID1NItmKgISXVX5Dx39COOSk0sIlK?usp=sharing), [ESPnet](https://colab.research.google.com/drive/13IAEX--ceFxzu_Romeeuv30L93McZfjr?usp=sharing)) | [Khudanpur](https://www.clsp.jhu.edu/faculty-pages/sanjeev/) |
| Tu Nov 8 | Hands on: Kaldi (continued) |  |
| Th Nov 10 | Second Midterm Exam |  |
| **Applications** | | |
| Tu Nov 15 | [Question Answering](https://jhu-intro-hlt.github.io/slides/question-answering.pdf) | [Duh](http://cs.jhu.edu/~kevinduh/) |
| Th Nov 17 | [NLP for Digital Humanities](https://jhu-intro-hlt.github.io/slides/digital-humanities-hlt-2020.pdf) | [Lippincott](http://www.logical-space.org/) |
| Tu Nov 29 | (no class) |  |
| Th Dec 1 | [Ethical Problems](https://jhu-intro-hlt.github.io/slides/intro-hlt-ethics-lec_2022-LM.pdf) | [Moro-Velazquez](https://pages.jh.edu/~lmorove1/index.html) |
| Tu Dec 6 | [Is Scale all You Need?](https://jhu-intro-hlt.github.io/slides/lecture-large-language-models-khashabi.pdf) | [Daniel Khashabi](https://danielkhashabi.com/) |
| Th Dec 8 | [Clinical NLP](https://jhu-intro-hlt.github.io/slides/dredze-clinical-nlp-2020.pdf) | [Dredze](https://www.cs.jhu.edu/~mdredze/) |