LIN7076 – Foundations of Computational Linguistics Syllabus (Autumn 2025)

Adèle Hénot-Mortier

Credits: 15.0

When: Tuesdays 12-2, from 23/09/2025 to 02/12/2025

Where: Bancroft 1.01.2

Instructor: Adèle Hénot-Mortier, ArtsOne 1.13, qp252142@qmul.ac.uk

Course Description: This module introduces students to the basic linguistic ideas needed for computational processing of text (morphology, syntax, semantics, discourse cohesion) and to how their properties can be understood computationally. With this in hand, it then introduces computational methods for text processing, for example, symbolic parsing models, basic text processing, probabilistic models, vector semantics, and deep learning architectures as well as applications in text classification, sentiment analysis, word prediction and text generation. It considers some of the ethical, environmental and research issues raised and includes discussion of the range of career routes that students' learning is relevant to.

Ressources:

- All course contents: https://github.com/AdeleMortier/LIN7076
- Labs: here
- Speech and Language Processing, Third Edition draft (January 2025).

Course schedule:

- Week 1 (23/09/2025): general introduction
  - overview of the concepts covered and their applications
  - overview of Python and Colab (Lab 1)
- Week 2 (30/09/2025): n-gram language models
  - the Markov assumption, probability estimation, perplexity.
  - Creation and evaluation of small scale n-gram models (Lab 2)
- Week 3 (07/10/2025): naive Bayes
  - the Naive Bayes classifier, precision, recall, F-score.
  - Creation and evaluation of small scale Naive Bayes classifier (Lab 3)
- Week 4 (14/10/2025): linear regression
  - linear and ridge regression, evaluating learning algorithms
  - Linear regression with Python and R (Lab 4)

- Week 5 (21/10/2025): logistic regression
  - the sigmoid function, the softmax function (for multiclass logistic regression), gradient descent
  - Logistic regression with Python and R (Lab 5)
- Week 6 (28/10/2025): vector semantics
  - word vectors, cosine similarity, TF-IDF, word2Vec
  - Visualizing word vectors (Lab 6)
- Week 7 (4/10/2025): reading week, no class
  - Problem Set 1 due
- Week 8 (11/11/2025): regular expressions
  - regular expressions, text processing and tokenization
  - Text processing in Python (Lab 7)
- Week 9 (18/11/2025): feedforward neural networks
  - basic neural net architecture, backpropagation
  - Building and training a small neural network for classification (Lab 8)
- Week 10 (25/11/2025): recurrent networks
  - RNNs, LSTMs, attention
  - Language modeling with LSTMs (Lab 9)
- Week 11 (02/12/2025): large language models
  - architectures, language generation, prompting
  - Language modeling with GPT-2 (Lab 10)
  - Project due

Assessment

- Midterm problem set: 50%
- Final project: 50%