Introduction and warm-up unit:
- General introduction to the course (today)
- Lab: Python and relevant libraries (Friday)

Today:
- Computational linguistics in the context of related fields
- Syllabus and course details
- Basic vocabulary of computational linguistics
I. Traditional methods
   A. Introduction (2 days)
      ▶ Python
      ▶ NLTK and other libraries
      ▶ NLP terminology
   B. Regular expressions (2 days)
      ▶ Definition, rules, and limitations
      ▶ Regex-based chatbots
   C. Edit distance (1 day)
   D. Information theory (2 days)
      ▶ Concepts and terminology
      ▶ Noisy channel model
      ▶ Compression
      ▶ Character-based language models
   E. Statistical language models (4 days)
      ▶ n-grams and other language statistics
      ▶ Lexical language models
      ▶ Smoothing and interpolation
      ▶ Applications and evaluation
   F. Part-of-speech tagging and hidden Markov models (3 days)
      ▶ Lexical categories
      ▶ Hidden Markov model definition
      ▶ Viterbi algorithm
   G. Parsing (3 days)
      ▶ Context-free grammars
   H. Lexical semantics (2 days)
      ▶ Word senses and relationships
      ▶ WordNet

II. Machine-learning methods
   A. Basic training in machine learning (2 days)
      ▶ Concepts and tasks
      ▶ Bag-of-words model
   B. Naive Bayes classification and sentiment analysis (3 days)
      ▶ Bayes’s theorem
      ▶ Sentiment analysis
   C. Stylometry and authorship attribution (3 days)
      ▶ Concepts and problem statement
      ▶ Techniques and practice
   D. Vector semantics and word embeddings (3 days)
      ▶ The idea of word vectors
      ▶ Word2vec and other approaches
   E. Neural nets and related models (3 days)
      ▶ Neural net basics
      ▶ Recurrent neural nets
      ▶ Long short-term memory
   F. Machine translation (2 days)
   G. Large language models (3 days)
Tokens, types, and lemmas

\[ I \text{ rose to saw off the still rose that I saw still grew near the still.} \]

16 tokens.

Types occurring once: to, off, that, grew, near.

Types occurring twice: I, rose, saw, the.

Type occurring three times: still

Distinct lexemes/lemmas: I, rise, to, saw (verb-cut), off, the, still (adjective), rose (noun), that, see, still (adverb), grow, near, still (noun).
Coming up:

- Learn Python
- Take Python quiz (Thurs, Aug 24)
- Do Python warm-up assignment (Mon, Aug 28)
- Read J&M, Sections 2.(0–4) (Mon, Aug 28)

Next time: Introduction to Python and NLTK *in the lab.*