Libraries and tools

- **NLTK**
  - Small but useful datasets with markup
  - Preprocessing tools: tokenization, normalization…
  - Pre-trained models for POS-tagging, parsing…
- **Stanford parser**
- **spaCy**: python and cython library for NLP
- **Gensim**: python library for text analysis, e.g. for word embeddings and topic modeling
- **MALLET**: Java-based library, e.g. for classification, sequence tagging, and topic modeling
- …