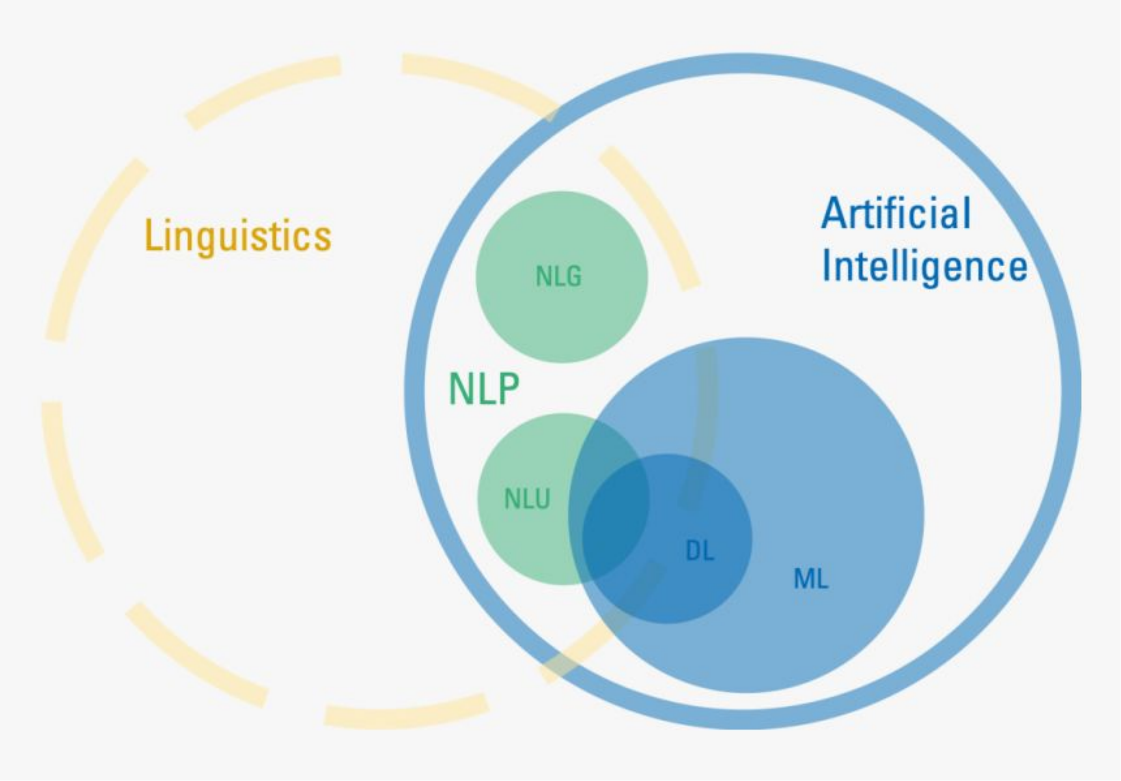
Week 11: Natural Language Processing

1. **Resources**
2. **Natural Language Processing**
3. **NLTK**
4. **SpaCy**
5. **Outlook**

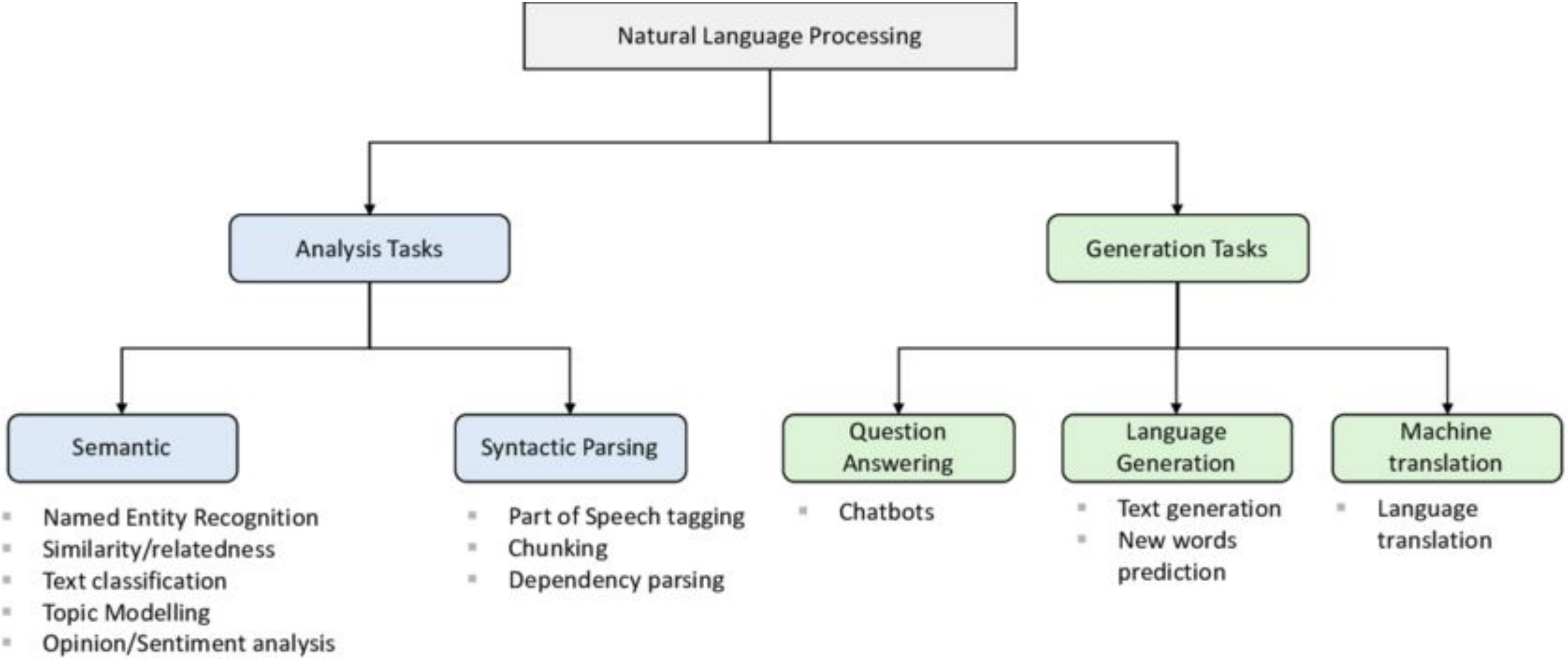
Resources

- NLTK API Documentation: https://www.nltk.org/api/nltk.html
- SpaCy API Documentation: https://spacy.io/api
- NLTK Book: https://www.nltk.org/book/
- SpaCy Usage: https://spacy.io/usage
- Python for Text Analysis: https://github.com/cltl/python-for-text-analysis

Natural Language Processing



Natural Language Processing



Natural Language Processing

Analysis Tasks

Generation Tasks

Semantic

- Named Entity Recognition
- Similarity/relatedness
- Text classification
- Topic Modelling
- Opinion/Sentiment analysis

Syntactic Parsing

- Part of Speech tagging
- Chunking
- Dependency parsing

Question Answering

- Chatbots

Language Generation

- Text generation
- New words prediction

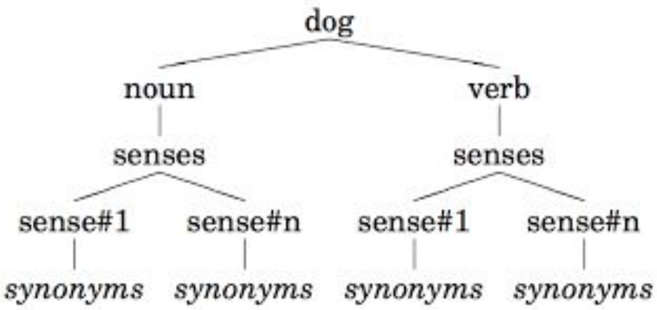
Machine translation

- Language translation

NLTK

Properties:

- Old but time-tested
- Many model choices
- Developed by academia

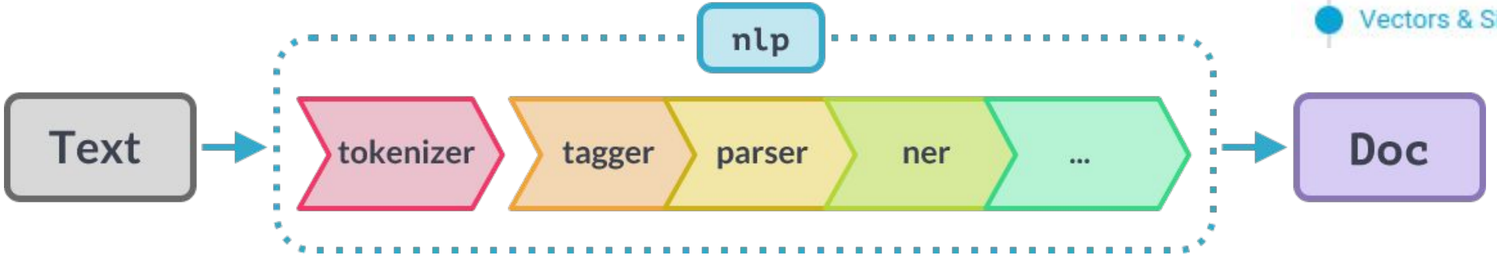


SpaCy

Properties:

- Fast because of Cython
- Few options (“just works”)
- Developed by industry

- POS Tagging
- Morphology
- Lemmatization
- Dependency Parse
- Named Entities
- Entity Linking
- Tokenization
- Merging & Splitting
- Sentence Segmentation
- Vectors & Similarity



Outlook

Application Lectures:

- ~~Week 11: Natural Language Processing~~
- Week 12: Experiment Design
- Week 13: Webscraping

Final Project Ideas:

- **Time & Performance Comparison of SpaCy & NLTK**
- **Authorship Attribution Service**
- **Re-implement Paper on NLP**

Regular Assignment:	2021-homework11	(link in StudIP announcement)
Bonus Assignment:	2021-homework10-bonus	(still ongoing until June 30th)
Final Project Registration Deadline:	July 15th	(mail)