NLP and Language Learning

Xiaobin Chen

Tübingen University

17 May, 2017

Natural Language Processing

Natural Language Understanding (NLU)

- Mapping the given input in natural language into useful representations.
- Analyzing different aspects of the language: lexical, syntactic, discoursal, pragmatic...

Natural Language Generation (NLG)

- Producing meaningful phrases and sentences in the form of natural language from some internal representation.

NLP and Language Learning

- Analyzing learner language
  - Providing feedback on learner errors
  - Measuring language development
  - Automatic scoring of learner productions
  - ...
- Analyzing native language for language learning purposes
  - Providing on-demand, individualized learning materials
  - Providing automatic input enhancement
  - Generating exercises and tests
  - ...

Overview Articles

Meurers, D. (2012). Natural language processing and language learning. In C. Chapelle (Ed), *Encyclopedia of Applied Linguistics*. New York: Wiley.
Nerbonne, J. (2003). Natural language processing in computer computer-assisted language learning. In R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*, Oxford University Press.

Computational Analysis of Natural Language—A Premier

- Tools:
  - OpenNLP Premier Package (Donwload here)
  - CoreNLP Demo (`http://nlp.stanford.edu:8080/corenlp/`)
- Detecting/Segmenting sentences:
  ```
  $ bin/opennlp SentenceDetector models/en-sent.bin <
  input/text1.txt
  ```
- Tokenization:
  ```
  $ bin/opennlp TokenizerME models/en-token.bin <
  input/text1.txt
  ```
  or
  ```
  $ bin/opennlp SentenceDetector models/en-sent.bin <
  input/text1.txt | bin/opennlp TokenizerME
  models/en-token.bin
  ```

- POS tagging:
  ```
  $ bin/opennlp SentenceDetector models/en-sent.bin <
  input/text1.txt | bin/opennlp TokenizerME
  models/en-token.bin | bin/opennlp POSTagger
  models/en-pos-maxent.bin
  ```
- The Penn Treebank tag set

- Parsing:
  `$ bin/opennlp Parser models/en-parser-chunking.bin < input/text2.txt`
- The Penn Treebank tag set