Introduction to NLP and NLG
Introduction to NLP

- Rules or Statistics??
- Lexical Analysis, Syntax Analysis, Semantic Analysis, Pragmatics
- Speech Processing (Phonetics, Punctuations, Prosody)
- Ambiguity
Concepts

- Tokenization
  
  "Dr. Watson, Mr. Sherlock Holmes", said Stamford, introducing us.
  "’Dr.’ ‘Watson’ ‘,’ ‘Mr.’ ‘Sherlock’ ‘Holmes’ ‘’ ‘,’ ‘said’ ‘Stamford’ ‘,’
  ‘introducing’ ‘us’ ‘.’

- n-grams, collocations
  
  *United States, Vice President*

- Stemming
  
  “lying” -> “lie”

- Lemmatization
  
  “women” -> “woman”
Concepts

- Segmentation of words and sentences
- Parts of Speech Tagging (POS)
  I/PRP picked/VBD up/RP a/DT ball/NN.
- Chunking : Partial Parsing
  [NP He ] [VP reckons ] [NP the current account deficit ] [VP will narrow ] [PP to ] [NP only $1.8 billion ] [PP in ] [NP September ].
- Named Entity Recognition
  ○ Organization, Location, Date, Time, Person, Money, Facility, …
- Quantifier : Quantity Detector and Standardizer
  http://cogcomp.cs.illinois.edu/page/software_view/Quantifier
Concepts

- Relation Extraction
  [ORG: ‘RPI’] in [LOC:’Troy’]
- Temporal information Extraction
- Event extraction
- Word Sense Disambiguation
  “I ducked as he hurled the stone at me”, “These ducks are beautiful”, “He ranged from ducks to centuries.”
- Coreference Resolution : Anaphora, Cataphora, Split Antecedent, Coreferring noun phrases, pleonasticity.
- Pronoun Resolution
Concepts

- **Dependency Parsing**: Linking of close semantic relationships together
  
  *She looks very beautiful*  
  
  `nsubj(looks, She)`  
  
  `acomp(looks, beautiful)`  
  
  `advmod(beautiful, very)`

- **Semantic Role Labeling**
  Parsing text into agents, actions etc.
Tools

- Illinois NLP Tools (http://cogcomp.cs.illinois.edu/page/software/)
- Stanford CoreNLP (http://nlp.stanford.edu:8080/corenlp/process)
- OpenNLP (https://opennlp.apache.org)
- NLTK (http://www.nltk.org)

Shop for NLP tools throughout the web.
Natural Language Generation

- Definition
- Architecture
- Example of NLG system
Definition

“Process of deliberately constructing a natural language text in order to meet specified communicative goals.”

~~[MacDonald, 1992]
The month was cooler and drier than average, with the average number of rain days,
Architecture

Communication Goals

Knowledge Source

Document Planning

Micro-planning

Surface Realisation

Content Determination, Document Structuring

Aggregation, Lexicalisation, Referring Expression Generation

Linguistic Realisation

NL Text
NLG Systems

- OpenCCG (http://openccg.sourceforge.net)
- NLGen (http://novamente.net/example/nlp.html)
- ASTROGEN (http://people.dsv.su.se/~hercules/ASTROGEN/ASTROGEN.html)
- CLINT (http://www.cs.bgu.ac.il/~elhadad/clint.html)
- NaturalOWL (http://www.cs.bgu.ac.il/~elhadad/clint.html)
- KPML (http://www.purl.org/net/kpml)
- Multimodal Unification Grammar (http://www.david-reitter.com/compling/mug/)
Combinatory Categorial Grammar

- Categorial Grammar: mapping from words to categories (atomic, complex)

S: Sentence/Clause
NP: Noun phrase
N: Noun
PP: Prepositional phrase

(S\NP): verb phrase, intransitive verb
((S\NP)/NP): transitive verb
(PP\NP): Preposition
(NP/N): Determiner