

## Statistical NLP: Lecture 3

### Linguistic Essentials

(Ch 3)

## Parts of Speech and Morphology

- **Parts of Speech** correspond to **syntactic** or **grammatical categories** such as **noun, verb, adjective, adverb, pronoun, determiner, conjunction, and preposition**.
- Word categories are systematically related by **morphological processes** such as the formation of plural form from the singular form.
- The major types of morphological processes are **inflection, derivation** and **compounding**.

## Words' Syntactic Functions

- Typically, **nouns** refer to entities in the world like people, animals and things.
- **Determiners** describe the particular reference of a noun and **adjectives** describe the properties of nouns.
- **Verbs** are used to describe actions, activities and states.
- **Adverbs** modify a verb in the same way as adjectives modify nouns. **Prepositions** are typically small words that express spatial or time relationships. Prepositions can also be used as **particles** to create phrasal verbs. **Conjunctions** and **complementizers** link two words, phrases or clauses.

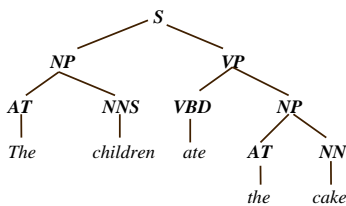
## Syntax or Phrase Structure: A simple context-free grammar

- |                 |                    |
|-----------------|--------------------|
| • S --> NP VP   | • AT --> the       |
| • NP --> AT NNS | • NNS --> children |
| AT NN           | students           |
| NP PP           | mountains          |
| • VP --> VP PP  | • VBD --> slept    |
| VBD             | ate                |
| VBD NP          | saw                |
| • P --> IN NP   | • IN --> in   of   |
|                 | • NN --> cake      |

The Grammar

The Lexicon

## Syntax or Phrase Structure: A Parse Tree



## Local and Non-Local Dependencies

- A **local dependency** is a dependency between two words expressed within the same syntactic rule.
- A **non-local dependency** is an instance in which two words can be syntactically dependent even though they occur far apart in a sentence (e.g., **subject-verb agreement; long-distance dependencies** such as **wh-extraction**).
- Non-local phenomena are a challenge for certain statistical NLP approaches (e.g., n-grams) that model local dependencies.

## Semantic Roles

---

- Most commonly, noun phrases are arguments of verbs. These arguments have **semantic roles**: the **agent** of an action, the **patient** and other roles such as the **instrument** or the **goal**.
- In English, these semantic roles correspond to the notions of **subject** and **object**.
- But things are complicated by the notions of **direct** and **indirect object**, **active** and **passive voice**.

## Subcategorization

---

- Different verbs can relate different numbers of entities: **transitive** versus **intransitive verbs**.
- Tightly related verb arguments are called **complements** but less tightly related ones are called **adjuncts**. Prototypical examples of adjuncts tell us time, place, or manner of the action or state described by the verb.
- Verbs are classified according to the type of complements they permit. This called **subcategorization**. Subcategorizations allow to capture syntactic as well as semantic regularities.

## Attachment Ambiguity and Garden-Path Sentences

---

- **Attachment ambiguities** occur with phrases that could have been generated by two different nodes in the parse tree.  
*The child ate the cake with a spoon.*
- Genuinely ambiguous: *Fruit flies like a banana.*
- **Garden-Path sentences** are sentences that lead along a path that suddenly turns out not to work.  
*The horse raced past the barn fell.*

## Semantics

---

- Semantics is the study of the meaning of words, constructions, and utterances.
- Semantics can be divided into two parts: lexical semantics and combination semantics.
- **Lexical semantics**: hypernymy, hyponymy, antonymy, meronymy, holonymy, synonymy, homonymy, polysemy, and homophony.
- **Compositionality**: the meaning of the whole often differs from the meaning of the parts.
- **Idioms** correspond to cases where the compound phrase means something completely different from its parts.

## Pragmatics

---

- Pragmatics is the area of studies that goes beyond the study of the meaning of a sentence and tries to explain what the speaker really is expressing.
- Understand the scope of quantifiers, speech acts, discourse analysis, anaphoric relations.
- The resolution of anaphoric relations is crucial to the task of information extraction.