Linguistic Essentials

Berlin Chen
Graduate Institute of Computer Science & Information Engineering
National Taiwan Normal University

References:
1. Speech and Language Processing, Chapter 3
2. Foundations of Statistical Natural Language Processing, Chapter 3
3. Natural Language Understanding, Chapter 2
Introduction

• Basic linguistic concepts
  – Word
    • Morphology
    • Part-of-speech (word categories)
  – Phrase and Syntax
    • Rewrite rules, parsing
  – Sentences and Discourse
  – Semantics and Pragmatics
Word Classes or Categories (1/3)

• Words are fundamental building block of languages

• Classify words into different classes (categories) based on their uses

• Two related areas of evidence
  – Semantic behavior
    • The word’s contribution to the phrase that contains it
  – Syntactic or grammatical behavior
    • The actual syntactic structures in which the word may play a role
    • Traditionally named part-of-speech (POS)
    • Four important parts-of-speech are nouns, verbs, adjectives, and adverbs
Word Classes or Categories (2/3)

• The syntactic classes of words (parts-of-speech) are traditionally divided into about 8 classes
  – E.g., noun, verb, adjective, adverbs, prepositions, conjunctions, determiners, pronouns,…

  – There are well-established sets of abbreviations for naming these classes, referred to as POS tags
    • E.g., noun (N), verb (V), adjective (A) …
    • Brown tag set (87 tags)
    • PenTreebank tag set (45 tags)
Word Classes or Categories (3/3)

- PenTreebank tag set (45 tags)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Coordin. Conjunction</td>
<td>and, but, or</td>
<td>SYM</td>
<td>Symbol</td>
<td>+, %, &amp;</td>
</tr>
<tr>
<td>CD</td>
<td>Cardinal number</td>
<td>one, two, three</td>
<td>TO</td>
<td>“to”</td>
<td>to</td>
</tr>
<tr>
<td>DT</td>
<td>Determiner</td>
<td>a, the</td>
<td>UH</td>
<td>Interjection</td>
<td>ah, oops</td>
</tr>
<tr>
<td>EX</td>
<td>Existential “there”</td>
<td>there</td>
<td>VB</td>
<td>Verb, base form</td>
<td>eat</td>
</tr>
<tr>
<td>FW</td>
<td>Foreign word</td>
<td>mea culpa</td>
<td>VBD</td>
<td>Verb, past tense</td>
<td>ate</td>
</tr>
<tr>
<td>IN</td>
<td>Preposition/sub-conj</td>
<td>of, in, by</td>
<td>VBG</td>
<td>Verb, gerund</td>
<td>eating</td>
</tr>
<tr>
<td>JJ</td>
<td>Adjective</td>
<td>yellow</td>
<td>VBN</td>
<td>Verb, past participle</td>
<td>eaten</td>
</tr>
<tr>
<td>JJR</td>
<td>Adj., comparative</td>
<td>bigger</td>
<td>VBP</td>
<td>Verb, non-3sg pres</td>
<td>eat</td>
</tr>
<tr>
<td>JJS</td>
<td>Adj., superlative</td>
<td>wildest</td>
<td>VBZ</td>
<td>Verb, 3sg pres</td>
<td>eats</td>
</tr>
<tr>
<td>LS</td>
<td>List item marker</td>
<td>1, 2, One</td>
<td>WDT</td>
<td>Wh-determiner</td>
<td>which, that</td>
</tr>
<tr>
<td>MD</td>
<td>Modal</td>
<td>can, should</td>
<td>WP</td>
<td>Wh-pronoun</td>
<td>what, who</td>
</tr>
<tr>
<td>NN</td>
<td>Noun, sing. or mass</td>
<td>llama</td>
<td>WP$</td>
<td>Possessive wh-</td>
<td>whose</td>
</tr>
<tr>
<td>NNS</td>
<td>Noun, plural</td>
<td>llamas</td>
<td>WRB</td>
<td>Wh-adverb</td>
<td>how, where</td>
</tr>
<tr>
<td>NNP</td>
<td>Proper noun, singular</td>
<td>IBM</td>
<td>$</td>
<td>Dollar sign</td>
<td>$</td>
</tr>
<tr>
<td>NNPS</td>
<td>Proper noun, plural</td>
<td>Carolinas</td>
<td>#</td>
<td>Pound sign</td>
<td>#</td>
</tr>
<tr>
<td>PDT</td>
<td>Predeterminer</td>
<td>all, both</td>
<td>“</td>
<td>Left quote</td>
<td>(‘ or “)</td>
</tr>
<tr>
<td>POS</td>
<td>Possessive ending</td>
<td>’s</td>
<td>”</td>
<td>Right quote</td>
<td>(‘ or ”)</td>
</tr>
<tr>
<td>PP</td>
<td>Personal pronoun</td>
<td>I, you, he</td>
<td>(</td>
<td>Left parenthesis</td>
<td>( [, {, &lt;)</td>
</tr>
<tr>
<td>PPS</td>
<td>Possessive pronoun</td>
<td>your; one’s</td>
<td>)</td>
<td>Right parenthesis</td>
<td>( ], } , &gt;)</td>
</tr>
<tr>
<td>RB</td>
<td>Adverb</td>
<td>quickly, never</td>
<td>,</td>
<td>Comma</td>
<td></td>
</tr>
<tr>
<td>RBR</td>
<td>Adverb, comparative</td>
<td>faster</td>
<td>;</td>
<td>Sentence-final punc</td>
<td>( . ! ?)</td>
</tr>
<tr>
<td>RBS</td>
<td>Adverb, superlative</td>
<td>fastest</td>
<td>;</td>
<td>Mid-sentence punc</td>
<td>( ; ... – )</td>
</tr>
</tbody>
</table>
Important Syntactic Classes of Words (1/5)

• Nouns (名詞)
  – Used to identify the basic types of objects (people and animal, etc.), things, concepts, or places being discussed
  – Mass nouns (不可數名詞) or count nouns (可數名詞)
    • mass nouns: west, water
    • Count nouns: book, fruit

• Verbs (動詞)
  – Used to express the action in a sentence

• Adjectives (形容詞)
  – Used to describe the properties of nouns
    • Qualify the object, thing, concept, or place
      The anger man waves his hands.
    – Noun modifiers: nouns used to modify another noun
      The cook book is just over there.
Important Syntactic Classes of Words (2/5)

- **Adverbs** (副詞)
  - Modify a verb in the same way that adjectives modify nouns
  - Specify **place** (here, everywhere), **time** (then, yesterday), **manner** (never, rarely), or **degree** (very, rather, too)

- **Pronouns** (代名詞)
  - A small class of words (it, he, she, they, ...) that act like variables in that they refer to a person or thing that is somewhat salient in the discourse context
  - They are the only words in English which appear different forms (cases) being used as the subject (nominative) and object (accusative) of a sentence
Important Syntactic Classes of Words (3/5)

- Pronoun Forms in English

<table>
<thead>
<tr>
<th>Tag(s)</th>
<th>Nominative</th>
<th>Accusative</th>
<th>Possessive</th>
<th>2nd Possessive</th>
<th>Reflexive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>I</td>
<td>me</td>
<td>my</td>
<td>mine</td>
<td>myself</td>
</tr>
<tr>
<td>2SG</td>
<td>you</td>
<td>you</td>
<td>your</td>
<td>yours</td>
<td>yourself</td>
</tr>
<tr>
<td>3SG MASC</td>
<td>he</td>
<td>him</td>
<td>his</td>
<td>himself</td>
<td>himself</td>
</tr>
<tr>
<td>3SG FEM</td>
<td>she</td>
<td>her</td>
<td>her</td>
<td>hers</td>
<td>herself</td>
</tr>
<tr>
<td>3SG NEUT</td>
<td>it</td>
<td>its</td>
<td>its</td>
<td>itself</td>
<td>itself</td>
</tr>
<tr>
<td>1PL</td>
<td>we</td>
<td>us</td>
<td>our</td>
<td>ours</td>
<td>ourselves</td>
</tr>
<tr>
<td>2PL</td>
<td>you</td>
<td>you</td>
<td>your</td>
<td>yours</td>
<td>yourselves</td>
</tr>
<tr>
<td>3PL</td>
<td>they</td>
<td>them</td>
<td>their</td>
<td>theirs</td>
<td>themselves</td>
</tr>
</tbody>
</table>

Table 3.2 Pronoun forms in English. Second person forms do not distinguish number, except in the reflexive, while third person singular forms distinguish gender.

Mary saw her in the mirror

- Relative pronouns (關係代名詞)
  - who, which, that
Important Syntactic Classes of Words (4/5)

• Prepositions (介系詞) (in, on over, about, …)
  – Small words that express spatial or temporal relationships
  – Prepositions can also used as particles (質詞) in the formation of phrasal verbs (片語動詞)

  He looks over the paper.
  I ran up a hill.
  I ran up a bill.

• Conjunctions (連接詞) (or, and, but, if, because, …)
  – Conjoin or coordinate (or subordinate) two words or phrases of (usually) the same category
  – Coordinating conjunctions (對等連接詞) (or, and, but, …), subordinating conjunctions (從屬連接詞) (if, because, …)
Important Syntactic Classes of Words (5/5)

• **Determiners** (定詞/限定詞)
  - Determiners describe the particular reference of a noun

  - **Articles** (冠詞) (the, a/an) are a subtype of determiners
    - The article *the* indicates that we are talking about something or someone that we already known about or can uniquely determined
    - The article *a* (or *an*) indicates that thing or person we are talking about was not previously mentioned

  - **Demonstratives** (指示代名詞) (*this*, *that*) are another kind of determiners
Other Syntactic Classes of Words (1/2)

- Interrogative pronouns/determiners (疑問代名詞/疑問定詞)
  - Used for questions and relative clauses
  - Interrogative pronouns:
    - Subject cases: who, which, what
    - Object cases: whom, which, what
  - Interrogative determiners:
    - E.g.: what, which

What questions will you be asked?
Which book will you read?
Other Syntactic Classes of Words (2/2)

• Proper Nouns (Proper Names 專有名詞)
  – Names referring to particular persons, things, or places, which are usually **capitalized**
  – E.g.: George W. Bush, 911 Attack on America, Unite States

• Compound Words: merge two or more words into a new word
  • In English: noun-noun compound words, or other combinations
    – E.g.: college degree (N), disk driver (N), downmarket (A), overtake (V), mad cow disease (PN)
An Example Sentence

Children ate sweet candy.

- The noun children refer to a group of people
- The noun candy refer to a particular type of food
- The verb ate describes what children do with candy
- The adjective sweet tells us about the property of candy
Substitution Test and Multiple POS

- **Substitution test**
  - The most basic test for word belong to the same class
    
    \[
    \text{sad} \quad \text{intelligent} \\
    \text{green} \quad \text{fat} \\
    \text{.....} \\
    \text{The} \quad \text{one is in the corner.}
    \]

- **Multiple part-of-speech of words**
  - E.g.: a noun can be a verb or a modifier, a adjective can be a noun
    
    Too much boiling will **candy** the molasses. (**candy**: verb)
    There is a **book** worm. (**book**: noun modifier)
    That **green** is lighter than the color. (**green**: noun)
Substitution Test and Multiple POS (1/2)

• Multiple part-of-speech of words (cont.)
  – Adjectives: can be further divided into
    • Those that can also used to describe a concept or quality directly
      – E.g.: The hot are on the table. (the hot plates are on the table)
    • Those that can’t
      – E.g.: green
Open and Close Word Classes (2/2)

• Open or lexical classes (categories)
  – Words like nouns, verbs, and adjectives (adverbs), which have a large number of members, and to which new words are commonly added as language evolves
  – Used to form the basis of a phrase
    • The head of the phrase

• Closed or functional classes (categories)
  – Words such as prepositions (e.g. in, on, over, …) and determiners (e.g. a, an, the, …), which have only a few members, and members of which normally have a clear grammatical use
  – New words in these classes are rarely introduced
Morphology (1/2)

• What is morphology (構詞學)？
  – Study the way words are built up from smaller meaning-bearing units, morphemes (詞素)
    • A morpheme is the minimal meaning-bearing unit in a language, e.g.,
      - *fox* consists of a single morpheme *fox*
      - *cats* consists two: *cat* and *-s* (singular → plural)
        
          stem    affix
          
    – Many new words are morphologically related to known words
      • We can infer a lot about the syntactic and semantic properties of new words if we understand the morphological process

          Japan    →    Japanese
          Taiwan   →    Taiwanese
          Hong Kong →    Hong Kongese?
Morphology (2/2)

• Two broad categories of morphemes
  – Stems (詞幹)
    • The main morpheme of the word, supplying the main meaning
  – Affixes (詞綴)
    • Add additional meanings of various kinds
    • Can be further divided into prefixes, suffixes, infixes, and circumfixes
      - prefixes, suffixes: concatenative morphology
• Concatenative morphology vs. non-concatenative morphology
  - Concatenative: English,
  - Non-concatenative: Arabic, Hebrew, Tagalog, ...
Three basic ways to form words from morphemes

• Inflection
  – The combination of a word stem with a grammatical morpheme, usually resulting in a word of the same syntactic class as the original stem (does not change word class or meaning significantly), e.g.:
    • The plural on English nouns, “dog” → “dogs”
    • The past tense on English verbs, “walk” → “walked”
  – Systematic, relatively simple in English

• Derivation
  – The combination of a word stem with a grammatical morpheme, usually results in a word of different syntactic classes, e.g.:
    • “computerize” → “computerization”
  – Less systematic, quite complex in English

• Compounding
  – Refer to merging of two or more words into a new word
    • White house, blueprint, etc.
Inflectional Morphology (1/5)

- Only **nouns**, **verbs**, and sometimes **adjectives** can be inflected in English

- Nominal inflection
  - Inflections for nouns: **number**, **case**, **gender**
  - Only two kinds of inflections first discussed here:
    - plural (number)
    - possessive/genitive (case)
  - The plural suffixes can be regular or irregular

<table>
<thead>
<tr>
<th></th>
<th>Regular Nouns</th>
<th>Irregular Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td>cat, thrush (-sh, -ch, -x)</td>
<td>butterfly, mouse, ox</td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td>cats, thrushes</td>
<td>butterflies, mice, oxen</td>
</tr>
</tbody>
</table>
Inflectional Morphology (2/5)

• Nominal inflection (cont.)
  – Possessive/genitive suffix:
    • Realized by apostrophe (’) plus -s for regular nouns and plural nouns not ending in -s
      – Singular noun: llama’s
      – Irregular plural noun: children’s
    • Realized by a lone apostrophe after regular plural nouns and some names ending in –s or –z
      – Regular plural noun: llamas’
      – Names ending in -s: Euripides’ comedies
Inflectional Morphology (3/5)

• Verbal inflection
  – More complicated than nominal inflection
    • Three kinds of verbs
      – Main verbs (eat, sleep, impeach, …)
      – Primary verbs (be, have, do)
      – Modal verbs (can, will, should, …)
  – Main verbs (can be regular or irregular)
    • Regular verbs: with three predictable endings
**Inflectional Morphology (4/5)**

- The regular class is **productive**: new words can be automatically included, e.g., *fax*

<table>
<thead>
<tr>
<th>Morphological Form Classes</th>
<th>Regularly Inflected Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>stem</td>
<td>walk</td>
</tr>
<tr>
<td>-s form</td>
<td>walks</td>
</tr>
<tr>
<td>-ing participle(分詞)</td>
<td>walking</td>
</tr>
<tr>
<td>past form or -ed participle</td>
<td>walked</td>
</tr>
</tbody>
</table>
Inflectional Morphology (5/5)

- Irregular verbs: have some more or less idiosyncratic forms of inflection (3~8 forms)
  » In general, the most frequent a word form, the most likely it’s to have idiosyncratic properties

<table>
<thead>
<tr>
<th>Morphological Form Classes</th>
<th>Regularly Inflected Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>stem</td>
<td>eat</td>
</tr>
<tr>
<td>-s form</td>
<td>eats</td>
</tr>
<tr>
<td>-ing participle (also for a gerund 動名詞)</td>
<td>eating</td>
</tr>
<tr>
<td>past form (preterite)</td>
<td>ate</td>
</tr>
<tr>
<td>-ed participle (prefect construction, passive construction)</td>
<td>eaten</td>
</tr>
</tbody>
</table>
Derivational Morphology

- A very common kind of derivation in English is the formation of new nouns, often from verbs or adjectives

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Base Verb/Adjective</th>
<th>Derived Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ation</td>
<td>computerize (V)</td>
<td>computerization</td>
</tr>
<tr>
<td>-ee</td>
<td>appoint (V)</td>
<td>appointee</td>
</tr>
<tr>
<td>-er</td>
<td>kill (V)</td>
<td>killer</td>
</tr>
<tr>
<td>-ness</td>
<td>fuzzy (A)</td>
<td>fuzziness</td>
</tr>
</tbody>
</table>

- Adjectives derived from nouns and verbs

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Base Noun/Verb</th>
<th>Derived Adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>-al</td>
<td>computation (N)</td>
<td>computational</td>
</tr>
<tr>
<td>-able</td>
<td>embrace (V)</td>
<td>embraceable</td>
</tr>
<tr>
<td>-less</td>
<td>clue (N)</td>
<td>clueless</td>
</tr>
</tbody>
</table>

- Generally less productive!
Morphological Comparatives and Superlatives

• In English, only some, mainly short, adjectives form morphological comparatives and superlatives by suffixing -er or -est, e.g.:
  - rich, richer, richest
  - trendy, trendier, trendiest

• For the rest adjectives, periphrastic form are used
  - Intelligent, more intelligent, most intelligent

• “Semantically” superlative adjectives
  - chief, main, top
Case Inflection

- **Case:**
  - Nouns or pronouns appear in different forms when they have different functions (subject, object, etc.) in a sentence, and these forms are called cases
    - Nominative (subject case) personal pronouns
      - E.g.: he, she
    - Accusative (object case) personal pronouns
      - E.g.: him, her
    - Genitive (possessive case)
      - Have a systematical indication
      - Explained previously
      - E.g.: his, hers, mine, ours,
Word Order and Phrase Structure

- Word do not occur in just any old order, but language have constraints on word order
  - Words in a sentence are organized into phrases
    - **Phrases**: groupings of words (called constituents) that clumped as a unit
  - E.g.:
    - I put the bagels in the freezer
    - The bagels, I put in the freezer.
    - I put in the freezer the bagels.
Syntax

• Meanings
  – From the Greek: “setting out together or arrangement”
  – The way words are arranged together
    • Study the regularities and constraints of word order and phrase structure

• Things to deal with
  – Constituency
    • Group of words may behave as a single unit or phrase
  – Grammatical Relations
    • E.g.: subjects and objects
  – Sub-categorization and dependencies (e.g. verbs)
    • Certain kinds of relations between words and phrases
Phrase Structure and Syntax

- **Paradigmatic Relationship (範例關係)**
  - All elements that can be replaced for each other in a certain syntactic position, e.g. the following noun phrase constituents

```
She
The woman
The woman with sad eyes
....
```

```
him
the man
the man with red hair
...
```

- **Syntagmatic Relationship (結構體關係)**
  - Two words in syntagmatic relationship can form a phrase like clothes in *sewed clothes* and a dress in *sewed a dress*
  - sewed wood?
Typical English Sentence Structure

- A sentence normally rewrites as a subject noun phrase and a verb phrase

```
S
  NP
    That man
  VP
    VBD caught
    NP the butterfly
    PP with
      NP a net
```
Noun Phrases (NP)

• Noun phrase is a syntactic unit of the sentence in which information about the noun is gather
  – A noun is usually embedded in a noun phrase (NP)
  – The noun is the head of the noun phrase, the central constituent that determines the syntactic character of the phrase

• Noun phrases usually exist along with verbs

• Determiners, adjectives, post-modifiers, prepositional phrases may occurred in noun phrases

The homeless old man in the park that I tried to help yesterday goes away.

determiner  adjectives  prepositional phrase  post-modifier
Verb Phrases (VP) (1/2)

• The verb phrase organize all elements of the sentence that depend syntactically on the verb
  – The verb is the head of the verb phrase

  *Getting to school on time* was a struggle.
  *He was trying to keep his temper.*
  *That woman quickly showed me the way to hide.*

• Subject-verb agreement
  – The subject and verb of a sentence agree in number and person
Verb Phrases (VP) (2/2)

- **Sub-categorization**
  - **Transitive and Intransitive**
    - Transitive: the verb with a following noun phrase (or a complement)
      
      John loves Mary.
    - Intransitive: the verb may stand alone
      
      The women walked.

- **Arguments and Complements**
  - Subject (NP) and (direct/indirect) objects (NP), PP, etc., are arguments of a verb
    
    – Centrally involved in the action of the verb
    
    We deprived him of food.
  - All non-subject arguments are complements

- **Adjuncts**
  - Phrases that have a less tight link to the verb
    
    She saw a Woody Allen movie in Paris.
  - Specify time, place, manner of the action
    
    She saw a Woody Allen movie with a couple of friends.
Prepositional Phrases (PP)

- Headed by a preposition and contain a noun phrase complement
  - Express spatial and temporal locations and other attributes
- Can appear within other major phrase types
  - Nominal modifier prepositional phrases
  - Verbal modifier prepositional phrases

Jack gave the book inside the box to me.

Jack put the book inside the box.
Adjective Phrases (AP)

- Adjectives can be grouped into a phrase
  - Can have an adverb before the adjective
- Complex adjective phrases are less common

It is the least expensive fare.
She is very sure of herself.
He seemed (a man who was) quite certain to succeed.

- But most commonly found as the complements of verbs such as be or seem
- May take a degree modifier preceding the head
Phrase Structure Grammars

- A syntactic analysis (parsing) of a sentence tells us how to determine the meaning of the sentence from meaning of words

  Mary gave Peter a book.
  Peter gave Mary a book.

- In English, the basic word orders are
  - Declaratives (直述句): Subject - Verb - Object
    The children (subject) should (auxiliary verb) eat spinach (object).
  - Interrogatives (詢問句): (question)
    Did he cry?
    Yes/No question
    Wh-question
  - Imperatives (祈使句): (requests/commands)
    Eat spinach!
Showing Syntactic Constituency

• Three ways to show the syntactic constituency
  – Rewrite rules
  – Parsing trees
  – (Labeled) Bracketing
Rewrite Rules (1/3)

• The regularities of word order are often captured by means of rewrite rules
  - Generate sentences
  - Parsing: the process of reconstructing the derivation(s) or phrase structure tree(s) for a particular sequence of words
    • A phrase structure tree is called a “parse”
    • Multiple parses → “syntactic ambiguity”

• A rewrite rule has the form:
  \[ \text{Category} \rightarrow \text{category}^* \]
  – The symbol on the left side can be rewritten as the sequence of symbols on the right side
Rewrite Rules (2/3)

• To produce a language
  – We can start with the start symbol ‘S’ (for a sentence)

• A property of the most formalizations of natural language in terms of rewrite rules is recursive
Rewrite Rules (3/3)

- A simple set of rewrite rules

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
\text{NP} & \rightarrow \{ \text{AT NN} \} \\
\text{AT} & \rightarrow \{ \text{the} \} \\
\text{NNS} & \rightarrow \{ \text{students} \} \\
\text{VP} & \rightarrow \{ \text{VBD} \} \\
\text{VBD} & \rightarrow \{ \text{ate} \} \\
\text{PP} & \rightarrow \{ \text{IN NP} \} \\
\text{IN} & \rightarrow \{ \text{of} \} \\
\text{NN} & \rightarrow \{ \text{cake} \}
\end{align*}
\]

- The rules on the right hand side rewrite one of the syntactic categories (part-of-speech symbols) into a word of the corresponding category
- The lexicon: words with pronunciations and POS tags
Rewrite Rules and Context-Free Grammar

• Examples:

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
& \rightarrow \text{AT NNS VBD} \\
& \rightarrow \text{The children slept.}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
& \rightarrow \text{AT NNS VBD NP} \\
& \rightarrow \text{AT NNS VBD AT NN} \\
& \rightarrow \text{The children ate the cake.}
\end{align*}
\]

• Context-free grammar
  – The possibilities for rewriting depend solely on the category, not the surrounding context
Representing Phrase Structures as a Tree (1/3)

- The tree has a single root node which is the start symbol of grammar

```
  S  
 / \ 
NP  VP
 / 
AT  NNS  VBD
    The  Children  slept
```

```
  S  
 / 
NP  VP
 / 
AT  NNS  VBD
    The  Children  ate
      AT  NN
           the  cake
```

- Nonterminal/Terminal nodes
  - Each nonterminal node and its immediate children (known as a local tree) corresponds to the application of a rewrite rule
Representing Phrase Structures as a Tree (2/3)

- Two words that were generated by common rewrite rules and syntactically linked can become separated by intervening words as the derivation of a sentence proceeds
  - Non-local dependencies

Figure 3.1 An example of recursive phrase structure expansion.
Representing Phrase Structures as a Tree (3/3)

- Non-local dependencies
  - Two words can syntactically dependent even though they occur far apart
  - Two examples
    - Subject-verb agreement
      
      The women who found the wallet were given a reward.

- Long-distance dependence

  Should Peter buy a book?
  Which book should Peter buy?

  An argument of “buy”
Representing Phrase Structures via a Bracketing

- Bracketting
  - Sets of brackets delimit constituents and may be labeled to show the categories of tree nonterminal nodes

```
[S[NP[AT The][NNS children]][VP[VBD ate][NP[AT the][NN cake]]]]
```
Dependency: Arguments and Adjuncts (1/2)

• Frequently, noun phrases are arguments of verbs, which can be described at various levels
  - **Semantic roles**
    • *Agent*: the person or thing that is doing something
    • *Patient*: the person or thing that is having something done to it
  
  - **Syntactic roles** (grammatical relations)
    • *Subject*: the noun phrase that appears before the verb
    • *Object*: the noun phrase that normally appears immediately after the verb
Dependency: Arguments and Adjuncts (2/2)

• Examples

- **Children eat sweet candy.**
  - Agent/subject: Children
  - Patient/object: sweet candy

- **She gave him the book.**
  - Recipient/indirect object: him
  - Patient/direct object: the book

- **She gave the book to him.**
  - Patient: the book (accusative case)
  - Recipient: him (dative case)

- **Bill received a package from the mailman.**
  - Patient/subject: Bill
  - Agent/indirect object: the mailman

- **Candy is eaten by children.**
  - Patient/subject: Candy
  - Agent/prepositional by-phrase: by children

Active voice

Passive voice
Phrase Structure Ambiguity (1/4)

• Example: “Our company is training workers” has 3 syntactic analyses (parses)

• “List the sales of the products produced in 1973 with the products produced in 1972” has 455 syntactic analyses (parses)

• Therefore, a practical NLP system must be good at making disambiguation decisions of word sense, word category, syntactic structure, and semantic scope
Phrase Structure Ambiguity (2/4)

a. S
   |   NP  VP
   |   Our company  Aux  VP
   |       is  V  NP
   |          training  workers

b. S
   |   NP  VP
   |   Our company  V  NP
   |       is  V  NP
   |          training  workers

(Cf. Our problem is training workers.)
Phrase Structure Ambiguity (3/4)

c.

\[
\begin{array}{c}
\text{S} \\
\text{NP} \\
\text{Our company} \\
\text{VP} \\
\text{is} \\
\text{AdjP} \\
\text{training} \\
\text{NP} \\
\text{workers} \\
\end{array}
\]

(Cf. Those are training wheels.)

* The last two parses (b. and c.) are semantic anomalous!
Phrase Structure Ambiguity (4/4)

Figure 3.2 An example of a prepositional phrase attachment ambiguity.
Semantics

• The meaning of words, constructions and utterances
  – The study of individual words (lexical semantics)
    • Lexical hierarchy
  – The study of how meanings of individual words are combined into the meaning of sentences (or even larger units)

• In most current systems semantic analysis is done only after syntactic analysis!
Lexical Semantics (1/2)

- **WordNet** defines the lexical hierarchy
  - **Hypernyms** (上義詞) and **hyponyms** (下義詞)
    - **Hypernym**: a word with a more general sense, e.g., animal is a hypernym of cat
    - **Hyponym**: a word with more specialized meaning
  - **Antonyms** (反義詞): words with opposite meanings
  - **Synonyms** (同義詞): Words have the same (very similar) meanings
  - **Homonyms** (同形異義詞): Words are written the same way but have different meanings which seems unrelated (e.g.: bank, suit, bass,...)
    - **Homophones** (同形同音異義詞): two word are not only written the same way but also same pronunciation (bank, suit, ...)

- **thesaurus**
  - cause
  - dress/suite
  - instrument
  - fish
  - tree
Lexical Semantics (2/2)

Figure 1 shows an example of a poset representing geographic locations and sub-locations using a tree structure to show the partial ordering relation.

Figure 1: Example of Geographic Semantic Poset
Discourse/Dialogue

• Elucidate the covert between sentences in a text
  – The anaphoric (前後照應) relations
  – How the immediately preceding sentences affect the interpretation of the next sentence

• Model the relationship between turns and the kinds of speech acts involved
  – Speech acts: questions, statements, requests, acknowledges etc.
  – Important for interpreting pronouns and for interpreting temporal aspects of information conveyed

*Hurricane Hugo* destroyed 20,000 Florida homes. At an estimated cost of one billion dollars, *the disaster* has been the most costly in the state’s history.

Which hurricanes caused more than a billion dollars worth of damage?
Pragmatics

- The study of how knowledge about the world and language conventions interact with literal meaning
  - How sentences are used in different situations
  - How use affects the interpretation of the sentence

Make hay while the sun shine.
Yes, I don’t.
.....
Other Areas

• **Phonetics** (語音學)
  – The study of speech sounds and their production, classification, and transcription
  – Include the phenomena like consonants, vowels and intonation

• **Phonology** (音韻學)
  – The structure of the sound systems
  – The tacit rules governing the speech pronunciation

• **Language acquisition**
  – Investigate how children learn (comprehend) language

• **Psycholinguistics**
  – Focus on issues of real-time production and perception of language and the way language is presented in the brain