

# **Linguistic Essentials**

Berlin Chen 2003

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
    - Part-of-speech (word categories)
    - Morphology
  - Phrase and Syntax
    - Rewrite rules, parsing
  - Sentences and Discourse
  - Semantics and Pragmatics

# Word Classes or Categories

- 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 part-of-speech are nouns, verbs, adjectives, and adverbs

# Word Classes or Categories

- The syntactic classes of words (part-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 as POS tags
    - E.g.: noun (N), verb (V), adjective (A) ...
    - Brown tags

# Important Syntactic Classes of Words

- **Nouns (名詞)**
  - Used to identify the basic types of objects (people and animal, etc.), things, concepts, or places being discussed
  - mass nouns or count nouns
- **Verb (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

- 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

- Pronoun Forms in English

	主格代名詞 Nominative	受格代名詞 Accusative	所有格代名詞 Possessive	所有格代名詞 2nd Possessive	反身代名詞 Reflexive
Tag(s)	PPS (3SG) PPSS (1SG,2SG,PL)	PPO	PP\$	PP\$\$	PPL (PPLS for PL)
1SG	I	me	my	mine	myself
2SG	you	you	your	yours	yourself
3SG MASC	he	him	his	his	himself
3SG FEM	she	her	her	hers	herself
3SG NEUT	it	it	its	its	itself
1PL	we	us	our	ours	ourselves
2PL	you	you	your	yours	yourselves
3PL	they	them	their	theirs	themselves

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.

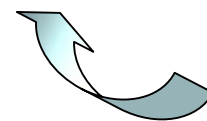
- Relative pronouns

- who, which, that

# Important Syntactic Classes of Words

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

He looks over the paper.



look up  
look over  
look out  
take off

....

- 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

- Determiners (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 know about or can uniquely determine
    - 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

- 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*

# Other Syntactic Classes of Words

- Proper Nouns (Proper Names 專有名詞)
  - Are names refer 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 combination
    - E.g.: college degree (N), disk driver (N), downmarket (A), overtake (V), mad cow disease (PN)

# An Example Sentence

Children ate sweet candy.

noun                  verb    adjectives          noun

- 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

The { sad  
intelligent  
green  
fat  
..... } 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

- 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

- 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

- 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



# Morphology

- 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, ...

# Two 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 resulting a word of **different syntactic classes**, e.g.:
  - “computerize” → “computerization”
- Less systematic, quite complex in English

# Inflectional Morphology

- Only nouns, verbs, and sometimes adjectives can be inflected in English
- Nominal inflection
  - Inflections for nouns: <sup>English</sup> number, case, gender
  - Only two kinds of inflections first discussed here:
    - plural (number)
    - possessive/genitive (case)
  - The plural suffixes can be regular or irregular

	Regular Nouns			Irregular Nouns	
Singular	cat	thrush (-sh, -ch, -x)	butterfly	<u>mouse</u>	ox
Plural	<u>cats</u>	<u>thrushes</u>	<u>butterflies</u>	<u>mice</u>	<u>oxen</u>

# Inflectional Morphology

- 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

- 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

Quirk et al., 1985

# Inflectional Morphology

Morphological Form Classes	Regularly Inflected Verbs			
stem	walk	merge	try	map
-s form	walk <u>s</u>	merg <u>e</u> s	tr <u>y</u> s	map <u>s</u>
-ing participle(分詞)	walk <u>ing</u>	merg <u>ing</u>	tr <u>y</u> ing	map <u>ing</u>
past form or -ed participle	walk <u>ed</u>	merg <u>ed</u>	tr <u>ied</u>	map <u>ped</u>

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

# Inflectional Morphology

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

Morphological Form Classes	Regularly Inflected Verbs		
stem	eat	catch	cut
-s form	eats <u> </u>	catches <u> </u>	cuts <u> </u>
-ing participle (also for a gerund 動名詞)	eating <u> </u>	catching <u> </u>	cutting <u> </u>
past form (preterite)	ate <u> </u>	caught <u> </u>	cut
-ed participle (perfect construction, passive construction)	eaten <u> </u>	caught <u> </u>	cut

# Derivational Morphology

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

Suffix	Base Verb/Adjective	Derived Noun
-ation	computerize (V)	computeriz <u>ation</u>
-ee	appoint (V)	appoint <u>ee</u>
-er	kill (V)	kill <u>er</u>
-ness	fuzzy (A)	fuzz <u>iness</u>

- Adjectives derived from nouns and verbs

Suffix	Base Noun/Verb	Derived Adjective
-al	computation (N)	computat <u>ional</u>
-able	embrace (V)	embrace <u>able</u>
-less	clue (N)	clue <u>less</u>

- 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

# Case Inflection

- **Case:**

- Nouns or pronouns appear in different forms when they different functions (subject, object, etc.) in a sentence, and these forms are called cases

Only for  
pronoun  
in English

- 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

# 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*.

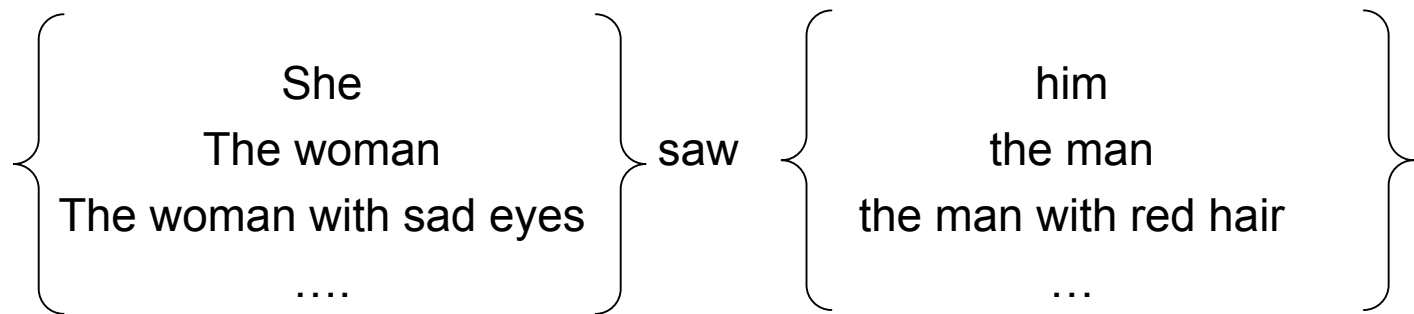
# Syntax

- 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



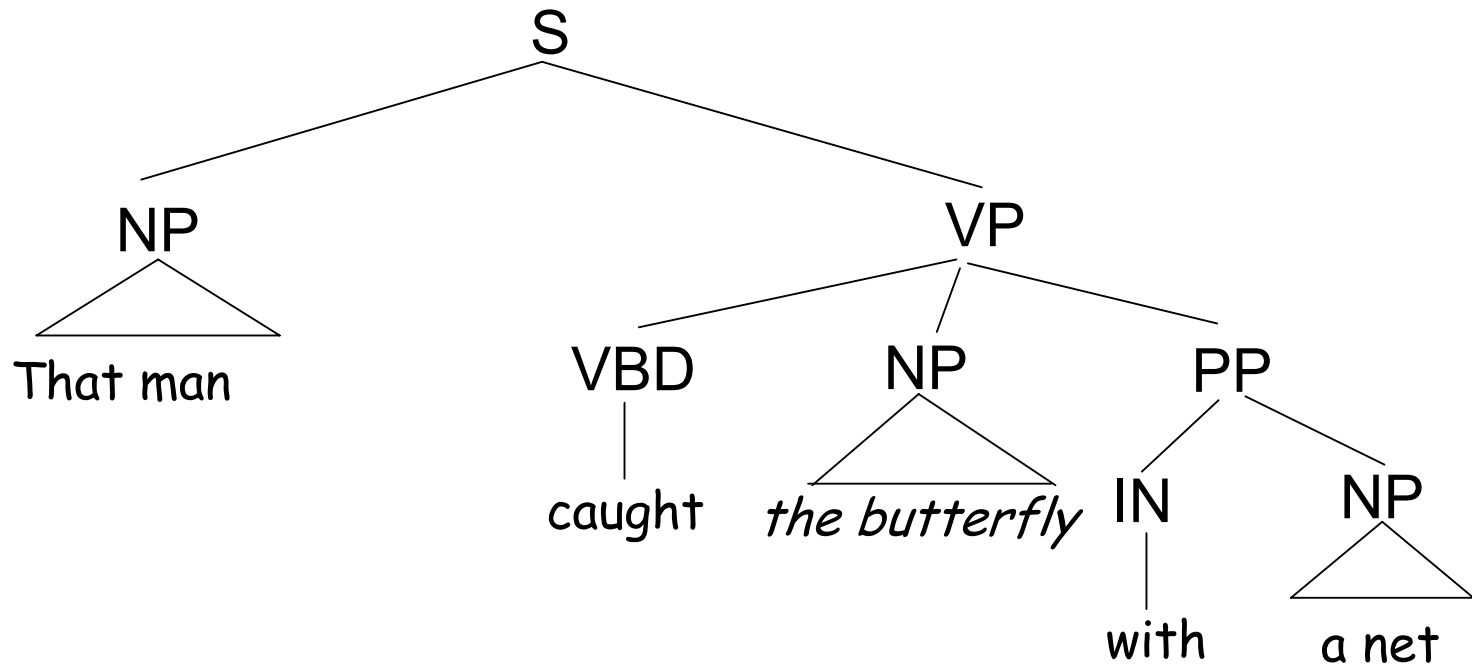
- Syntagmatic Relationship (結構體關係)

- Two words are syntagmatic relationship in 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



# Noun Phrases (NP)

- Noun phrase is a syntactic unit of the sentence in which information about the noun is gathered
  - 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 occur in noun phrases

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



# Verb Phrases (VP)

- 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)

- 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** (補語)

- Subject (NP) and (direct/indirect) objects (NP), PP, etc., are arguments of a verb *We deprived him of food.*

- Centrally involved in the action of the verb

- All non-subject arguments are complements

- **Adjuncts** (附加語)

- Phrases that have a less tight link to the verb

- Specify time, place, manner of the action

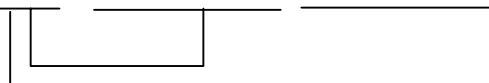
- She saw a Woody Allen movie in Paris.*

- She saw a Woody Allen movie with a couple of friends.*

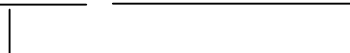
# Prepositional Phrases (PP)

- 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 put the book inside the box.



Jack gave the book inside the box to me.



# 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

- sentence moods
- Declaratives(直述句): Subject -Verb -Object  
The children (subject) should (auxiliary verb) eat spinach (object).
  - Interrogatives (詢問句): (question) Yes/No question  
Did he cry? 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

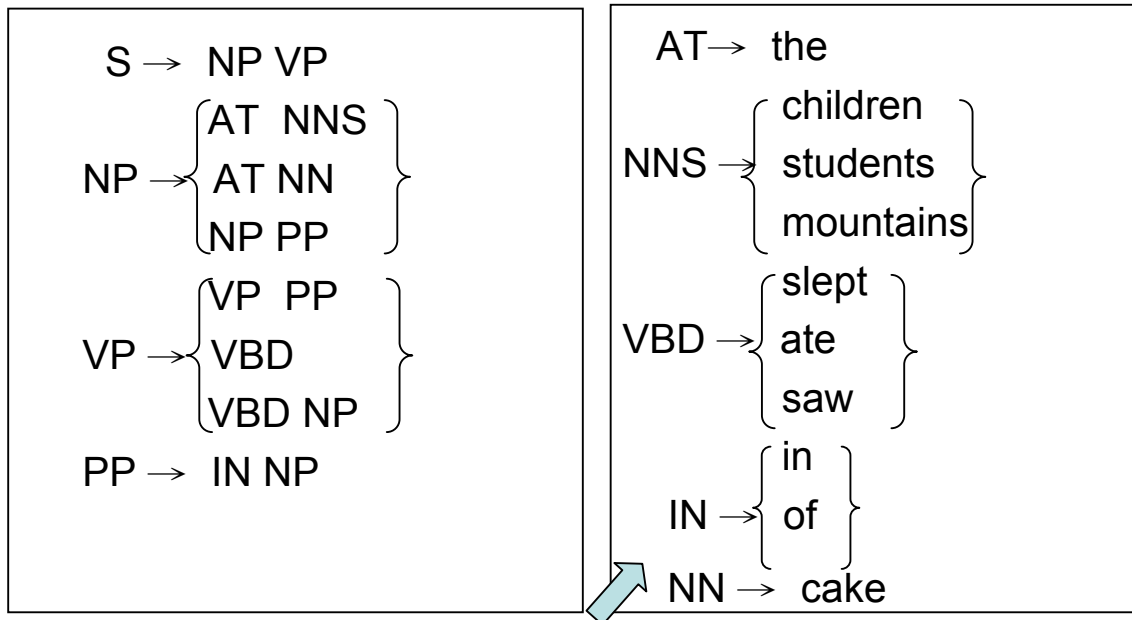
- The regularities of word order are often captured by means of *rewrite rules*
  - **Generate sentences**
  - **Parsing**: the process of reconstructing the derivation(s) or phrase structure tree(s) for a particular sequence of words
    - A phrase structure tree is called a “**parse**”
    - Multiple parses → “**syntactic ambiguity**”
- A rewrite rule has the form:  
 $Category \rightarrow category^*$ 
  - The symbol on the left side can be rewritten as the sequence of symbols on the right side

# Rewrite Rules

- 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

- A simple set of rewrite rules



- The rules on the righthand 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:

S

→ NP VP

→ AT NNS VBD

→ *The children slept.*

S

→ NP VP

→ AT NNS VBD NP

→ AT NNS VBD AT NN

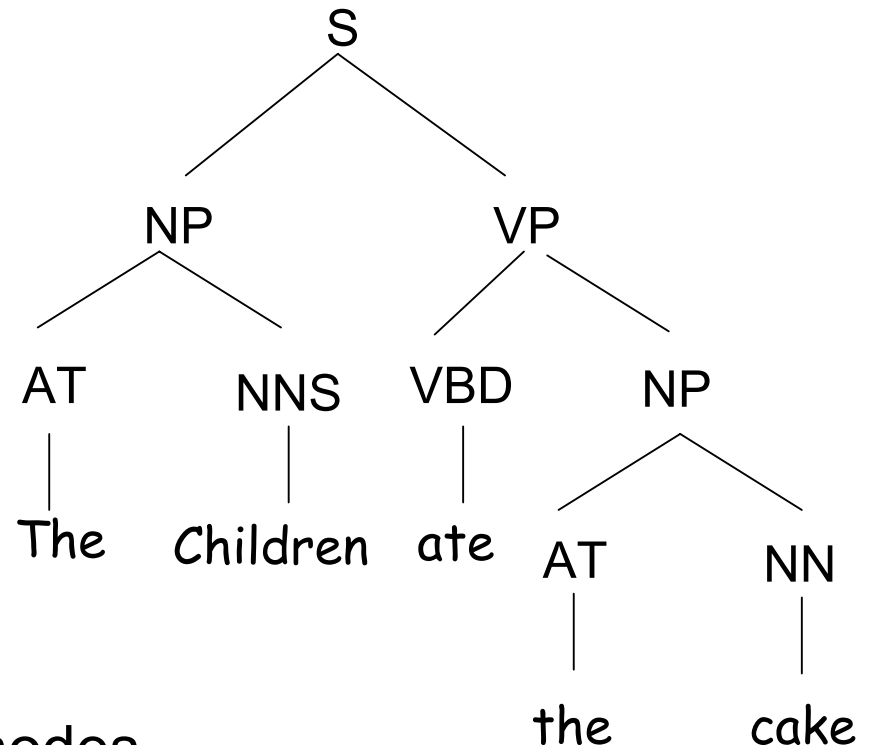
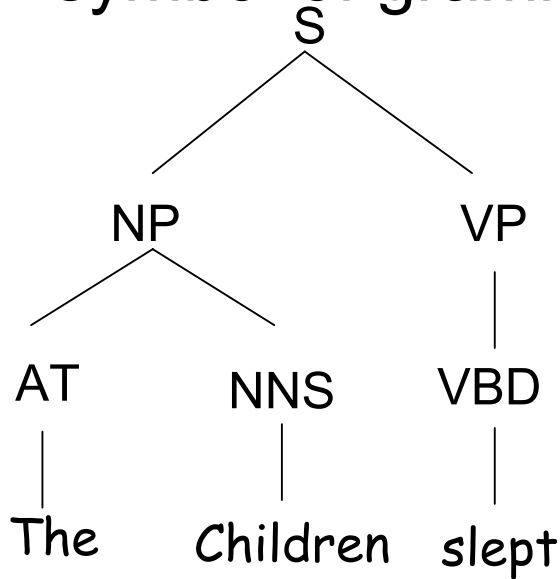
→ *The Children ate the cake.*

- Context-free grammar

- The possibilities for rewriting depend solely on the category, not the surrounding context

# Representing Phrase Structures as a Tree

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



– 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

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

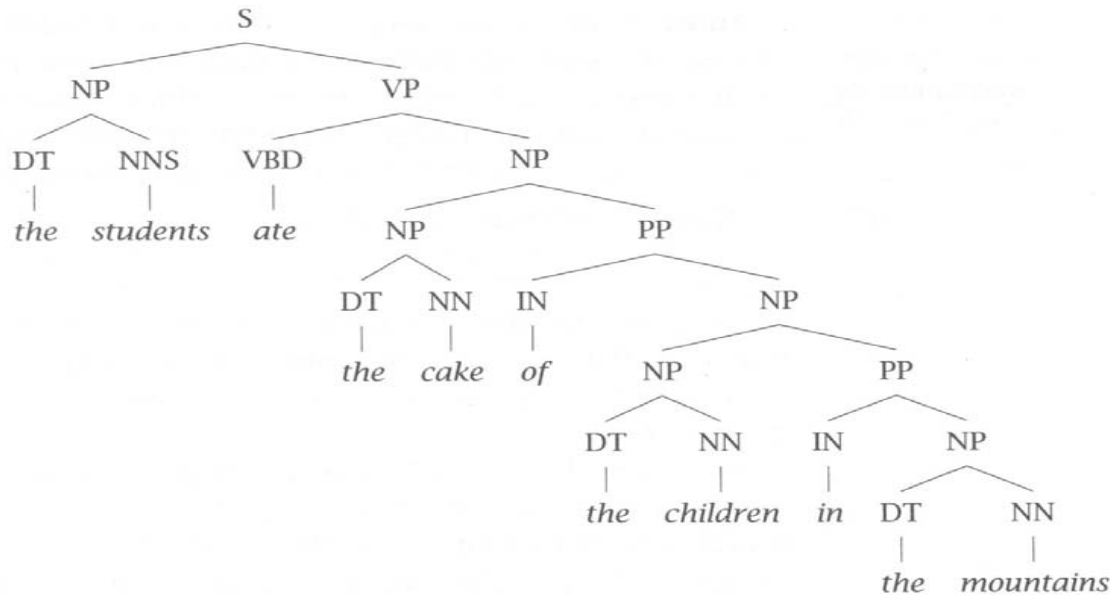



Figure 3.1 An example of recursive phrase structure expansion.

# Representing Phrase Structures as a Tree

- 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.



A horizontal line with vertical end caps connects the word "women" to the word "were", illustrating a subject-verb agreement dependency.

- Long-distance dependence

Should Peter **buy a book**?

**Which book** should Peter **buy**?



An argument of “buy”

# Representing Phrase Structures via a Bracketing

- Bracketing
  - 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]]]]

The diagram shows the sentence "The children ate the cake" with red curly braces indicating its hierarchical structure. The outermost brace spans the entire sentence. Inside, there are two main branches: one for "The children" and one for "ate the cake". The "The children" branch has two sub-brackets for "The" and "children". The "ate the cake" branch has three sub-brackets: one for "ate", one for "the", and one for "cake".

# Dependency: Arguments and Adjuncts

- 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

- Examples

Children eat sweet candy.

agent/subject                  patient/object

She gave him the book.

recipient/                  patient/  
indirect object      direct object

She gave the book to him .

patient                  recipient  
(accusative case)      (dative case)  
受格                          與格

Bill received a package from the mailman .

patient/subject                          agent/indirect object

Candy is eaten by children.

patient/subject                          agent/prepositional by-phrase

Active voice

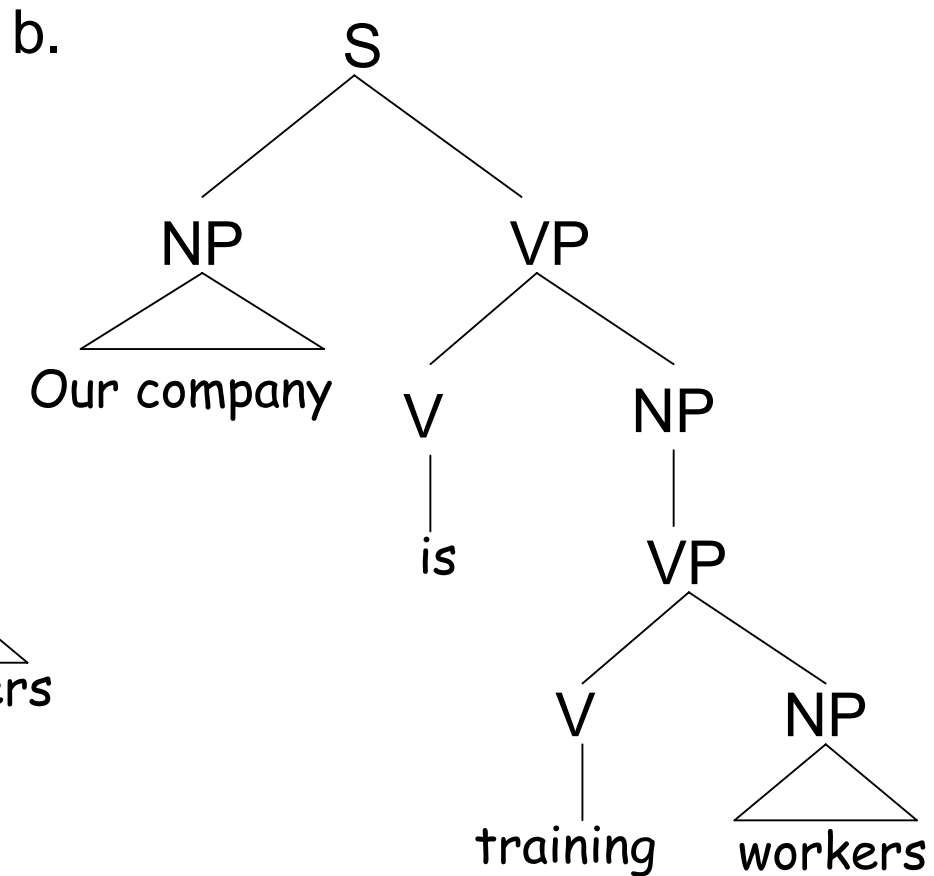
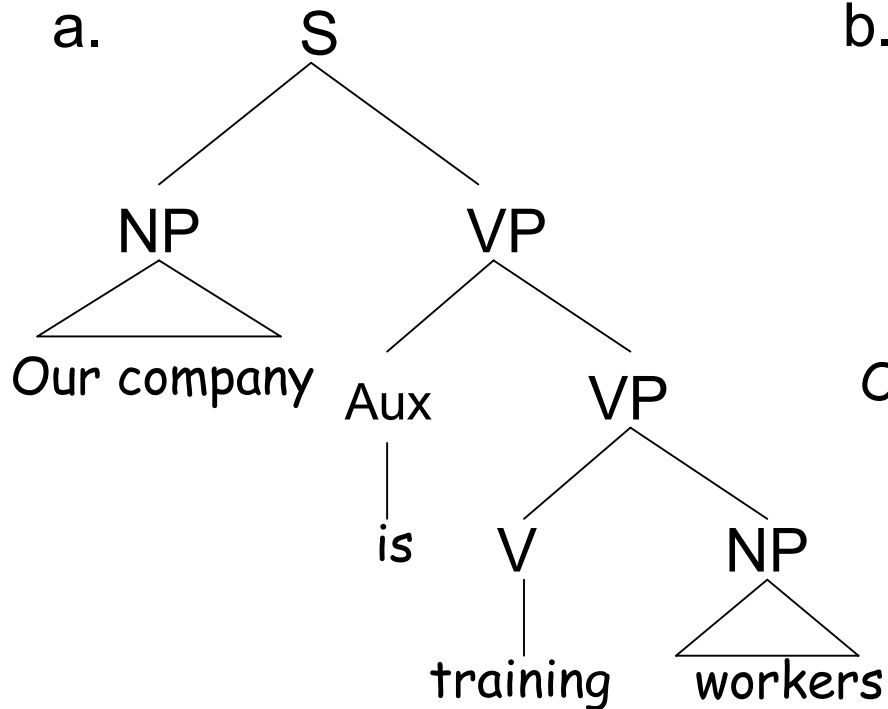
Passive voice

# Phrase Structure Ambiguity

- 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



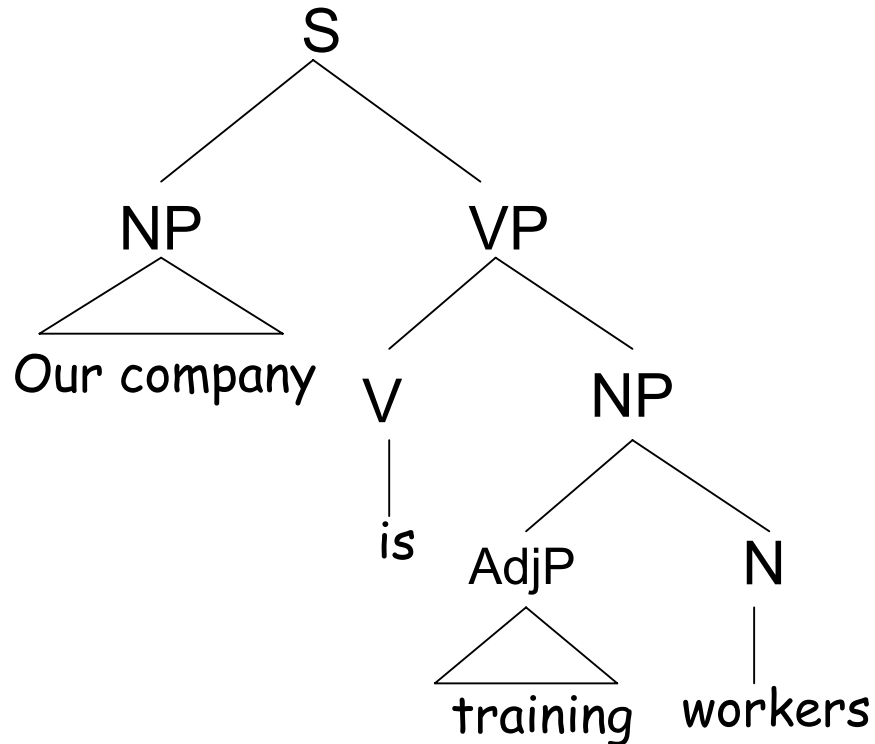
# Phrase Structure Ambiguity



(Cf. Our problem is training workers.)

# Phrase Structure Ambiguity

c.



(Cf. Those are training wheels.)

\* The last two parses (b. and c.) are semantic anomalous!

# Phrase Structure Ambiguity

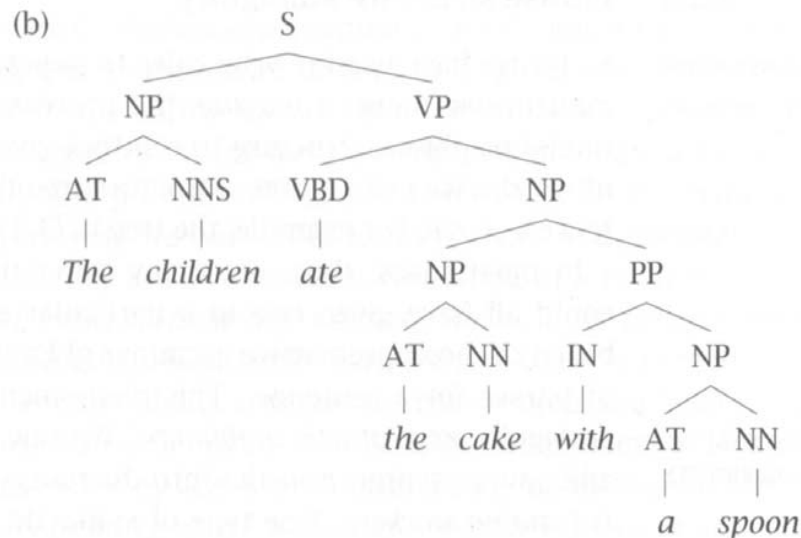
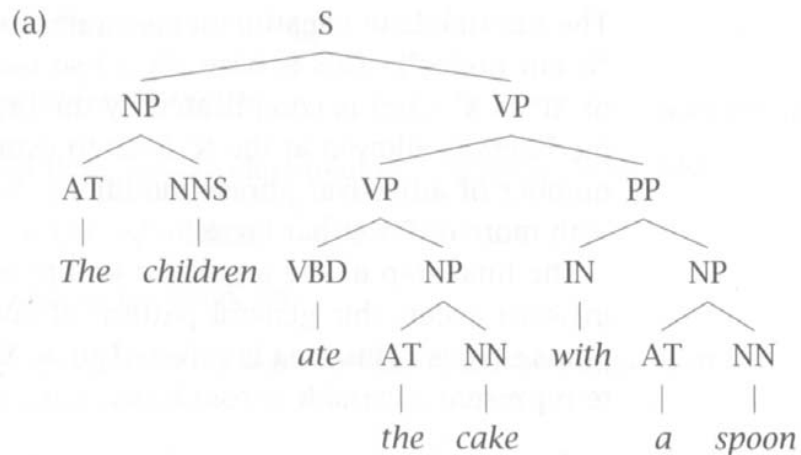


Figure 3.2 An example of a prepositional phrase attachment ambiguity.

# 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

- WordNet defines the lexical hierarchy thesaurus
  - **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, ...)

# Lexical Semantics

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

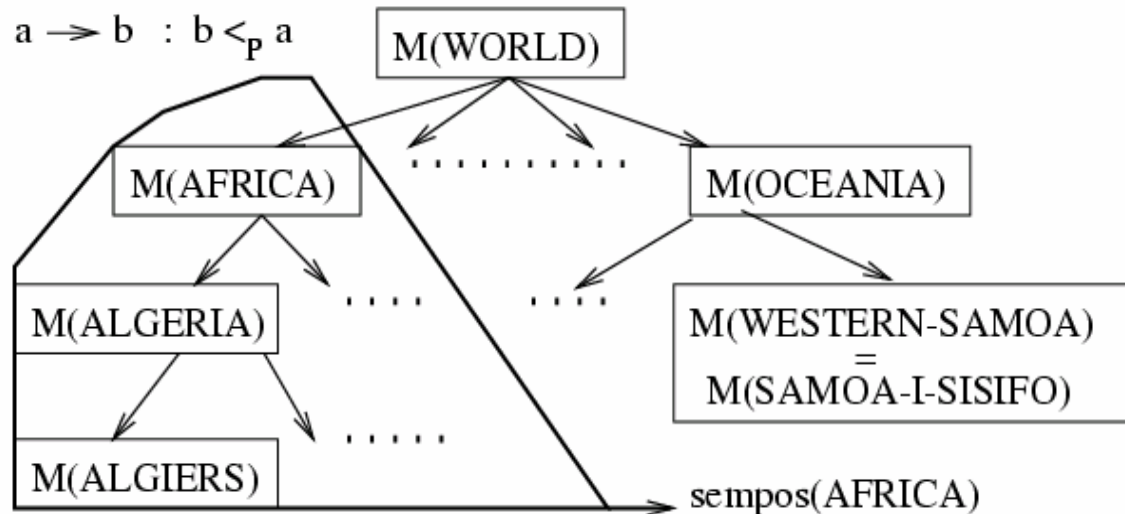


Figure 1: Example of Geographic Semantic Poset

# Discourse

- 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
    - 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

- 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



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