Linguistic Essentials

Berlin Chen

Department of Computer Science & Information Engineering
National Taiwan Normal University

References:

1. D. Jurafsky & J. H. Martin, *Speech and Language Processing,* Chapter 3 (2nd)
2. C.Manning & H. Schütze, *Foundations of Statistical Natural Language Processing*, Chapter 3
3. J. Allen, *Natural Language Understanding,* Chapter 2

Introduction

- Basic linguistic concepts
  - Word
    - Morphology
    - Part-of-speech (POS; 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)

- Pen Treebank tag set (45 tags)

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

https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Important Syntactic Classes of Words (1/7)

- 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 (2/7)

- 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

7

Important Syntactic Classes of Words (3/7)

- Pronoun Forms in English

主格代名詞　受格代名詞　所有格代名詞　反身代名詞

	Nominative	Accusative	Possessive	2nd Possessive	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

a friend of mine / yours / hers / ....

- Relative pronouns (關係代名詞)
  - who, which, that

Important Syntactic Classes of Words (4/7)

- English **pronouns** have another possessive form, often called the 'second' possessive personal pronoun, used when the object of the preposition of describes the possessor
  - E.g., a friend of mine.

- There are **reflexive pronouns**, which are used similarly to ordinary (personal) pronouns except that they always refer to a nearby antecedent in the same sentence, normally the subject of the sentence.

Mary saw her in the mirror.
Mary saw herself in the mirror.

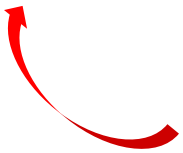
We expected something of **each other** then.
We expected something of **one another** then.

- Reflexive pronouns (and certain other expressions like *each other)* are often referred to as **anaphors*,*** and must refer to something very nearby
- in the text

Important Syntactic Classes of Words (5/7)

- 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.



look up
look over
look out
take off
….

  - How to pronounce phrasal verbs ? (stress patterns)

I **look up** the table.
A **look-up** table is used.

Important Syntactic Classes of Words (6/7)

- 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, that, although, before, …)

husband ***and*** wife

She bought ***or*** leased a car.

I won't wait ***if*** he is late.

She left ***before*** he arrived.

Important Syntactic Classes of Words (7/7)

- 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
    - Demonstratives are used to specify the distance of something in space or time in relation to the speaker

This apple looks ripe.
I'll call round this afternoon.

That car is speedy.
He was very poor at that time.

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?
What book will you read?

13

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)

- Adverbial Nouns
  - n\Nouns or noun phrases that function grammatically as adverbs to modify verbs and complement certain adjectives
  - E.g., home, west, tomorrow

14

An Example Sentence

Children ate sweet candy.

noun verb adjective 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 (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
        (green is ambiguous between being an adjective or a noun)

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 (Taiwner ?)
Hong Kong → Hong Kongese ? (Hong Konger ?)

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"
    - "weak" → "weaken"
    - "contextual" → "contextualized"
  - Less systematic, quite complex in English
- Compounding
  - Refer to merging of two or more words into a new word
    - White house, blueprint, mad cow disease etc.

Inflectional Morphology (1/5)

- Only nouns, verbs, and sometimes adjectives can be inflected in English
  - In English, adjectives only take two inflections: the comparative and superlative (thin thinner; thinnest)
- Nominal inflection
  - Inflections for nouns: number, case, gender (English)
  - 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	mouse	ox
Plural	cats	thrushes	butterflies	mice	oxen

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)

Quirk et al., 1985

- Verbal inflection
  - More complicated than nominal inflection
    - Three kinds of verbs
      - Main Verbs (eat, sleep, impeach, …)
      - Primary Verbs (be, have, do) 基本(助)動詞
      - Modal Verbs (can, will, should, …) 情態(助)動詞

Auxiliary Verbs

  - Main verbs (can be regular or irregular)
    - Regular verbs: with three predictable endings

24

Inflectional Morphology (4/5)

Morphological Form Classes	Regularly Inflected Verbs			
stem	walk	merge	try	map
-s form	walks	merges	tries	maps
-ing participle(分詞)	walking	merging	trying	mapping
past form or -ed participle	walked	merged	tried	mapped

- The regular class is **productive**: new words can be automatically included, e.g., fax (n: fax → faxes; v: fax → faxed → faxed → faxing)

25

Inflectional Morphology (5/5)

"be"

- 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

Morphological Form Classes	Irregularly Inflected Verbs		
stem	eat	catch	cut
-s form	eats	catches	cuts
-ing participle (also for a gerund 動名詞)	eating	catching	cutting
past form (preterite)	ate	caught	cut
-ed participle (prefect construction, passive construction)	eaten	caught	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)	computerization
-ee	appoint (V)	appointee
-er	kill (V)	killer
-ness	fuzzy (A)	fuzziness

- Adjectives derived from nouns and verbs

Suffix	Base Noun/Verb	Derived Adjective
-al	computation (N)	computational
-able	embrace (V)	embraceable
-less	clue (N)	clueless

- 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

  (The way we form comparative adjectives is based on the number of syllables in the adjective and whether or not the adjective ends with the letter 'y'.)

- 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, ,,,,

Only for pronoun in English

29

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*.

market street

street market

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

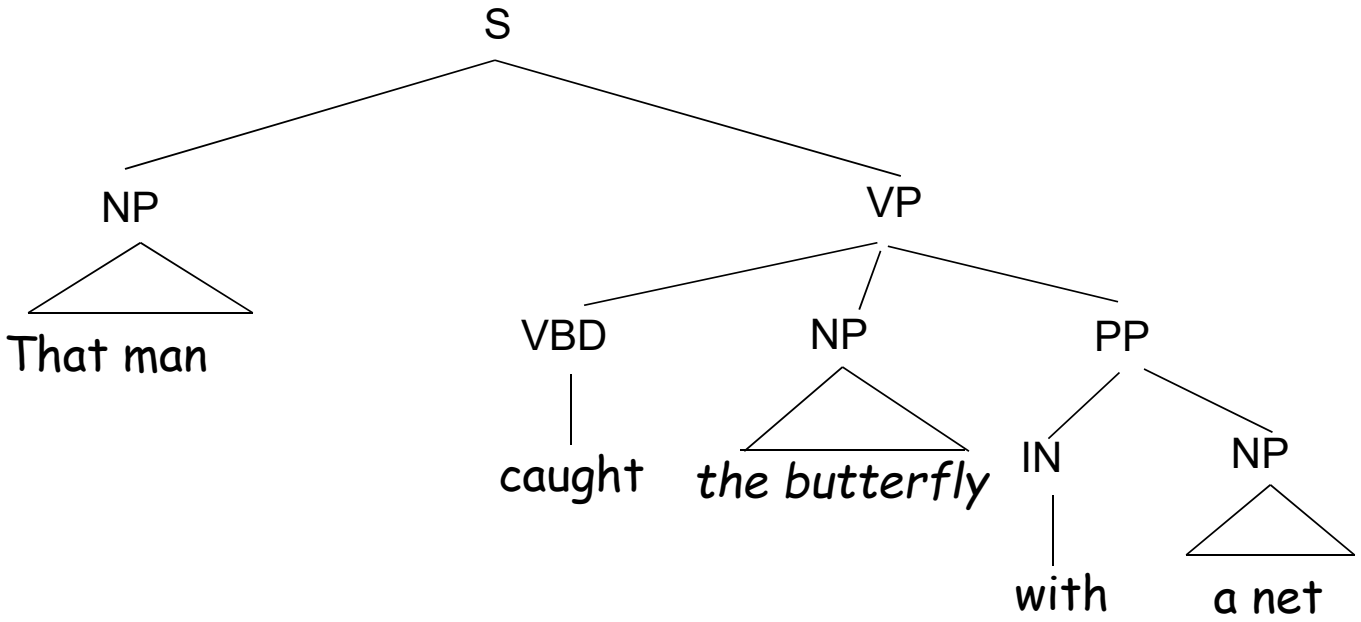
She			him	
The woman	saw		the man	
The woman with sad eyes			the man with red hair	
….			…	

- Syntagmatic Relationship (結構體關係)
  - Two words bear a syntagmatic relationship if they can form a phrase (or syntagma) like sewed clothes or sewed a dress
    - *sewed wood?*
  - An important class of syntagmatically related words are *collocations*

strong tea	strong computer (?)
powerful tea (?)	powerful computer

Typical English Sentence Structure

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



主詞片語

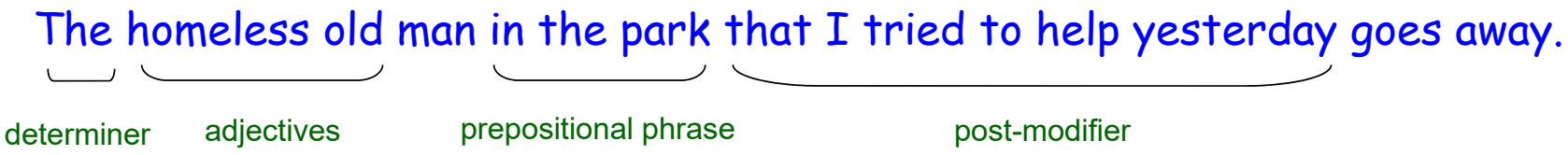
動詞片語

Noun Phrases (NP) (1/2)

- 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 occurred in noun phrases



Noun Phrases (NP) (2/2)

- More specifically, noun phrases normally consist of an optional determiner, zero or more adjective phrases, a noun head, and then perhaps some post-modifiers, such as prepositional phrases or clausal modifiers, with the constituents appearing in that order

- Clausal modifiers of nouns are referred to as relative clauses

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)

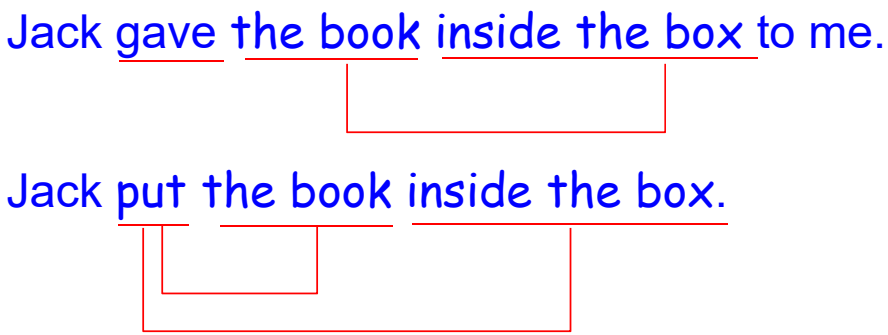
- Sub-categorization
  - **Transitive and Intransitive**
    - Transitive: the verb with a following noun phrase (or a complement) John loves Mary.
    - Intransitive: the verb may stand alone The women walked.
  - Arguments and Complements (補語)
    - Subject (NP) and (direct/indirect) objects (NP), 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 (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 (名詞片語補語)
  - Express spatial and temporal locations and other attributes
- Can appear within other major phrase types
  - Nominal modifier prepositional phrases
  - Verbal modifier prepositional phrases



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)

    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

41

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:

  Category →category*
  - The symbol on the left side can be rewritten as the sequence of symbols on the right side

Rewrite Rules (2/3)

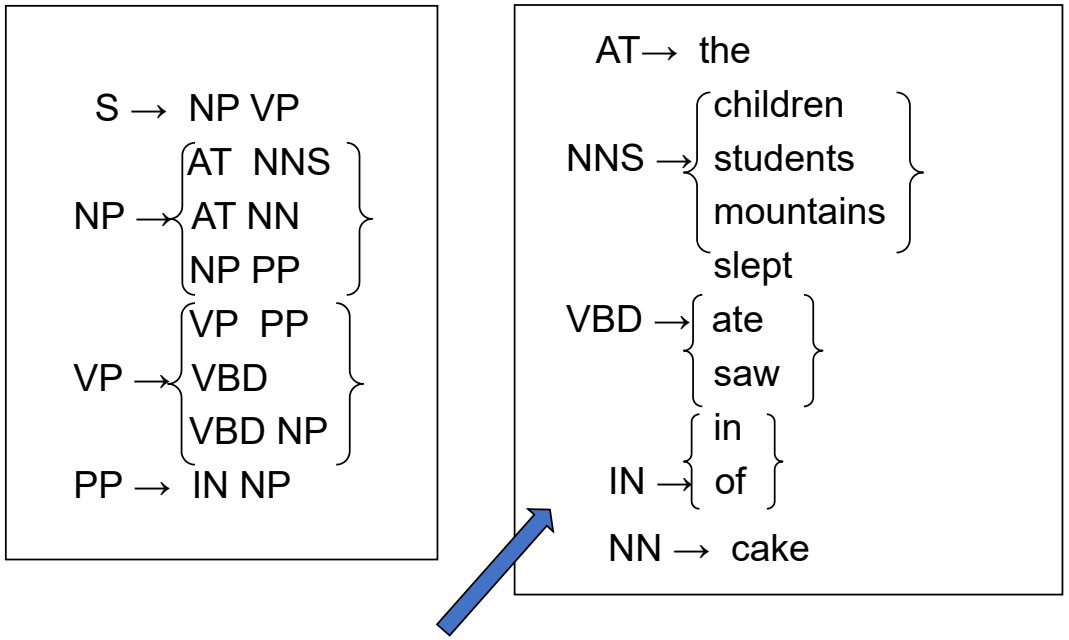
- To produce a language
  - We can start with the start symbol 'S' (for a sentence)

- A property of the most formalizations of natural language in terms of rewrite rules is recursive

43

Rewrite Rules (3/3)

- A simple set of rewrite rules



- 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:

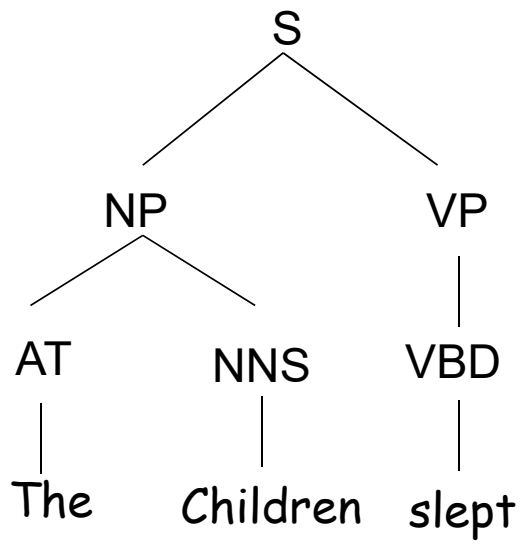
S
→ NP VP
→ AT NNS VBD
→ The children slept.

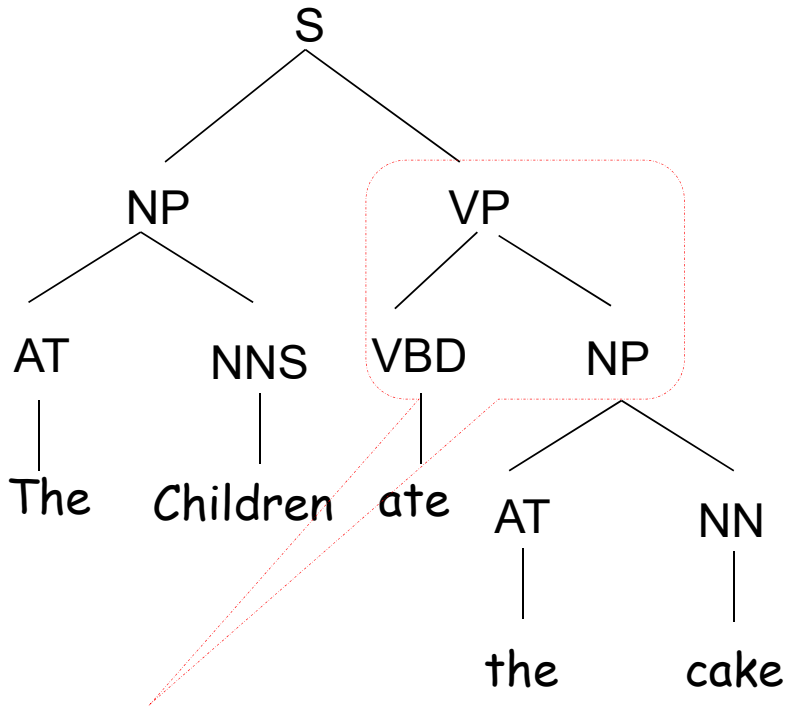
S
→ NP VP
→ AT NNS VBD NP
→ AT NNS VBD AT NN
→ The children ate the cake.

- Context-free grammar
  - The possibilities for rewriting depend solely on the category, not the surrounding context
  - E.g., we could expand VP to a natural phrase like ***sewed clothes***, but we can as easily choose a nonsensical expansion like ***sewed wood blocks***

Representing Phrase Structures as a Tree (1/3)

- 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 (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

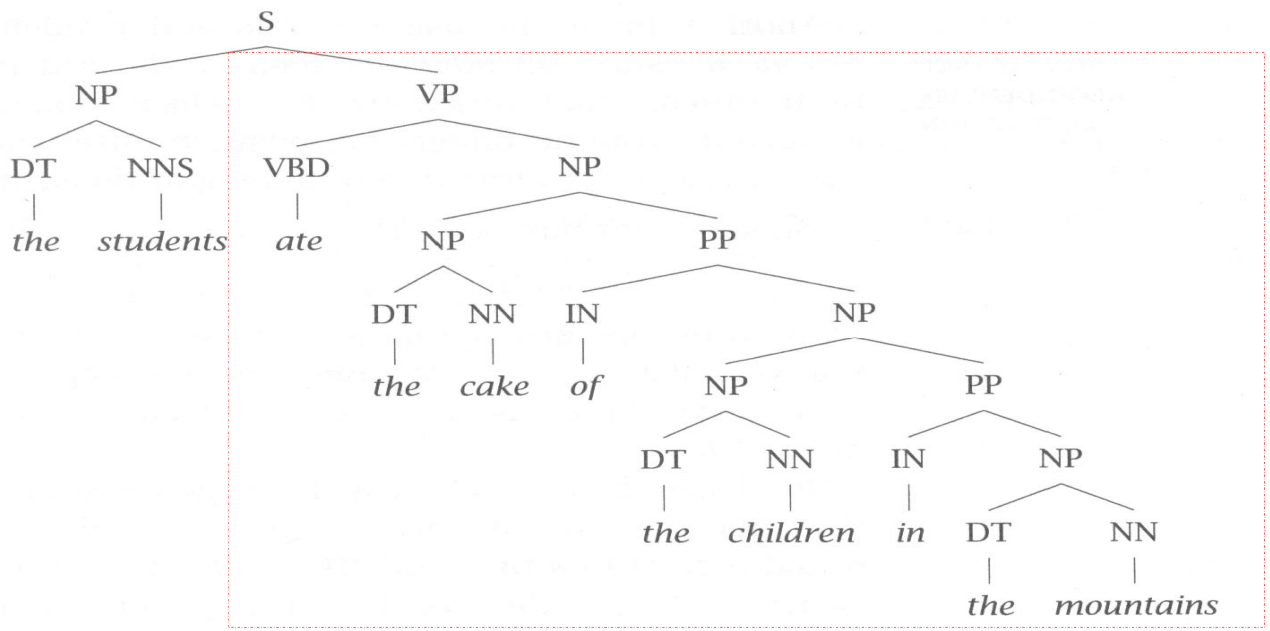


Figure 3.1 An example of recursive phrase structure expansion.

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

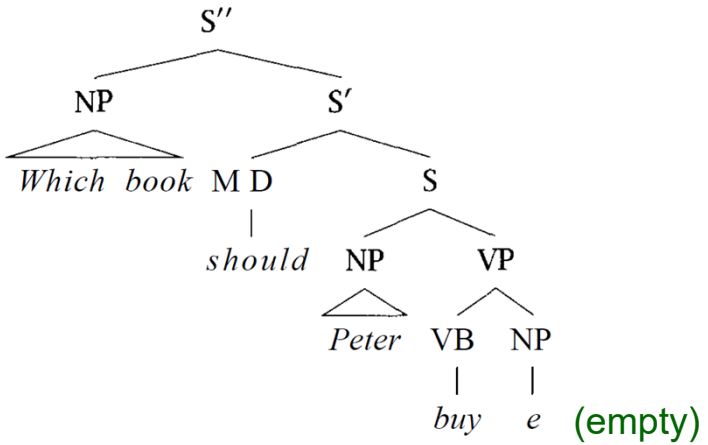
The **women** who found the wallet **were** given a reward.

(Non-local phenomena are a challenge for some Statistical NLP approaches like *n*-grams that model local dependencies. An *n*-gram model would predict that the word after wallet in is was, not were.)

    - Long-distance dependence

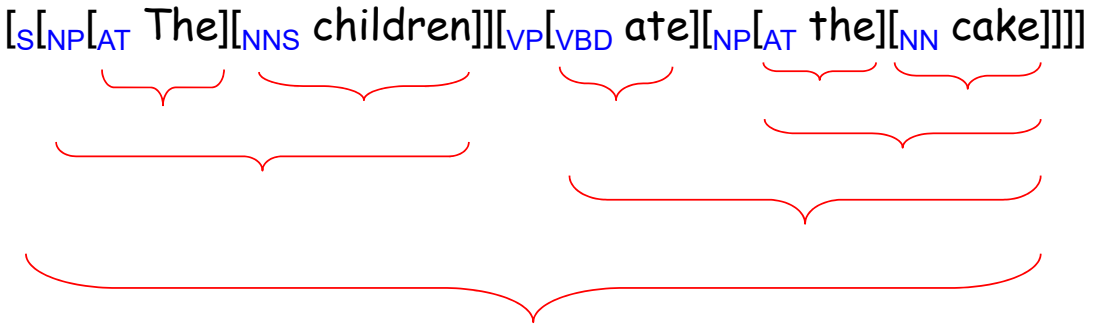
Should Peter **buy a book**?

Which book should Peter **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



Dependency: Arguments and Adjuncts (1/2)

- Most commonly, **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

Sue watched the man at next table.

Sue and **the mm** are dependents of a watching event. We will say that they are the two arguments of the verb **watch**. The PP **at the next table** is a dependent of **man**, which modifies man.

Dependency: Arguments and Adjuncts (2/2)

- Examples

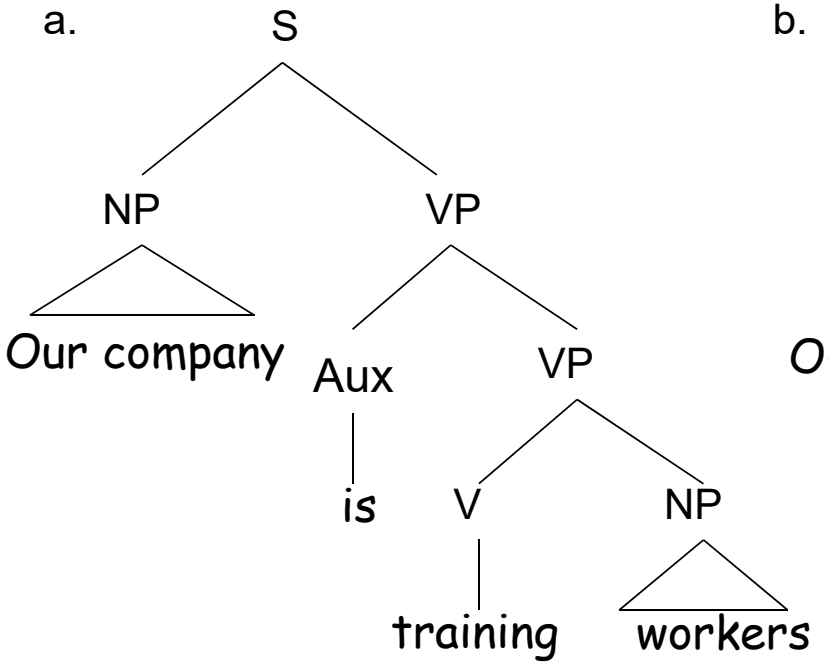
Phrase Structure Ambiguity (1/4)

- 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)

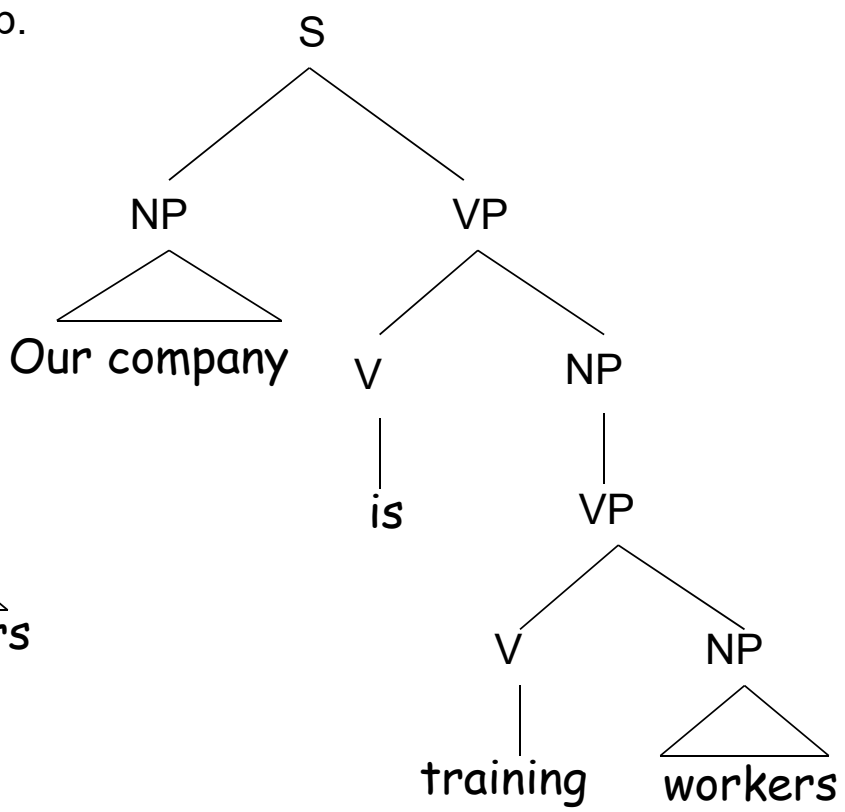
- Therefore, a practical NLP system must be good at making disambiguation decisions of word sense, word category, syntactic structure, and semantic scope

Phrase Structure Ambiguity (2/4)

a.



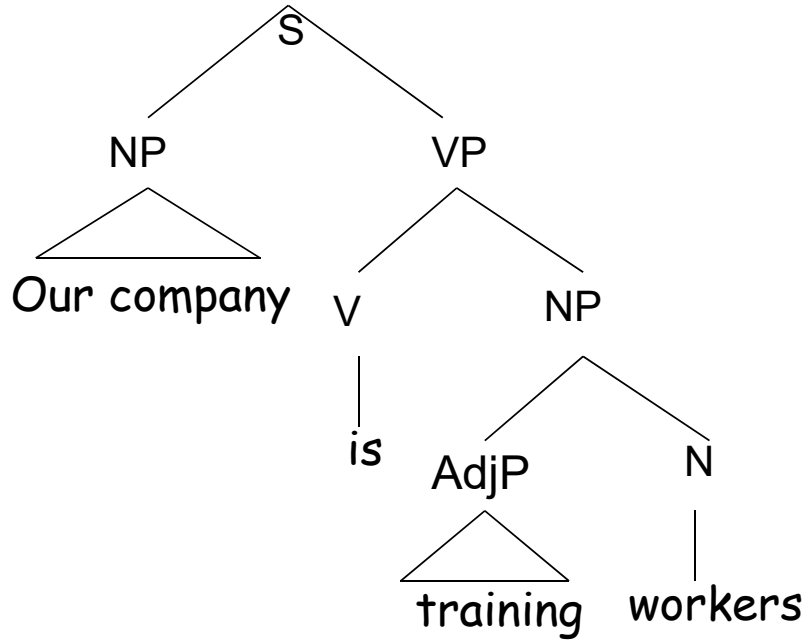
b.



(Cf. Our problem is training workers.)

Phrase Structure Ambiguity (3/4)

c.



(Cf. Those are training wheels.)

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

Phrase Structure Ambiguity (4/4)

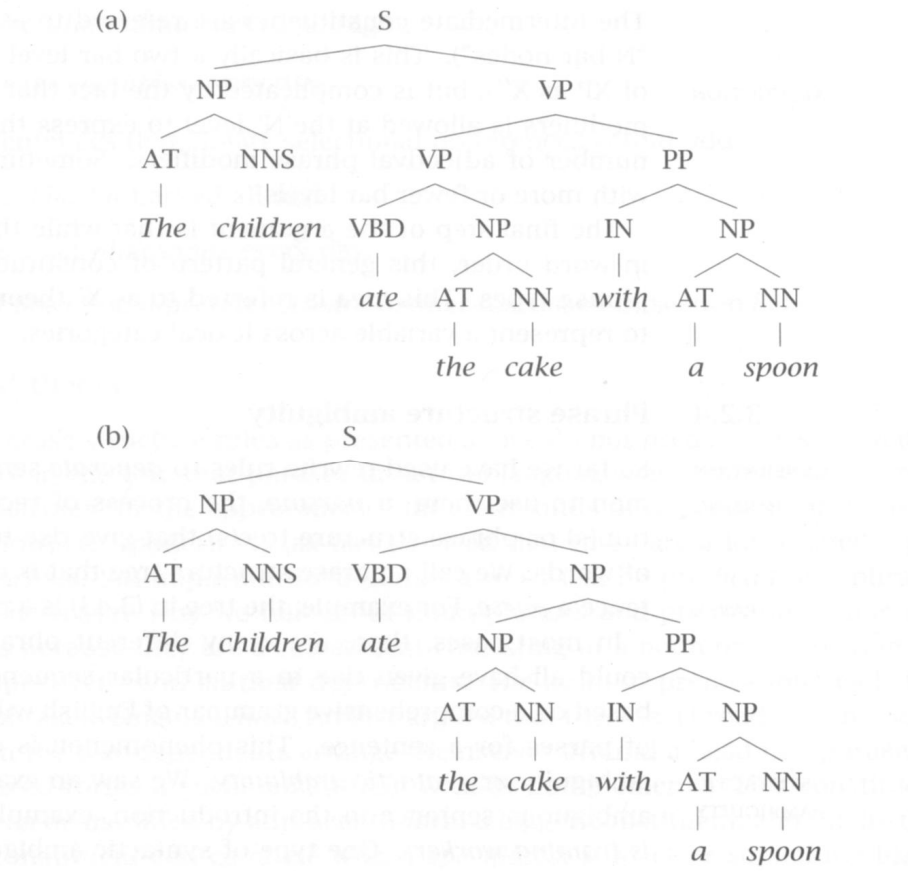
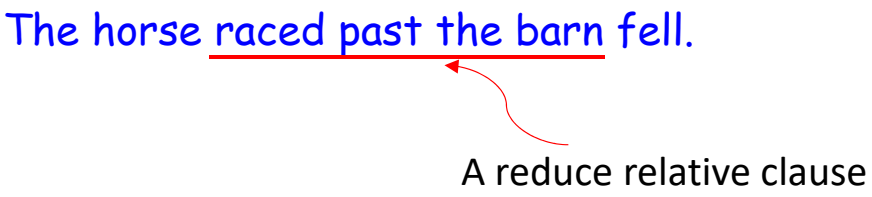


Figure 3.2 An example of a prepositional phrase attachment ambiguity.

Garden Pathing (1/2)

- A garden path sentence leads you along a path that suddenly turns out not to work
  - E.g., "The horse raced past the barn fell."
  - By the time most people get to the word ***barn,*** they have constructed a parse that roughly corresponds to the meaning (i.e., "The horse ran past the barn.")
  - But then there is an additional word fell that cannot be incrementally added to this parse
  - We have to backtrack to ***raced*** and construct a completely different parse, corresponding to the meaning "*The horse fell after it had been raced past the burn*"



Garden Pathing (2/2)

- Garden-path sentences are rarely a problem in spoken language. Semantic preferences, since intonational patterns all usually prevent us from garden pathing
  - We can see this in "The horse raced past the barn fell.", where an intonational break between horse and ***raced*** would tip the hearer off that ***raced*** introduces a reduced relative clause, not the verb of the main clause
- However, garden-pathing can be a real problem when reading complex sentences of written English

Semantics

- The meaning of words, constructions and utterances. We can divide semantics into two parts:
  - 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 traditional NLU systems (developed before the era of deep learning), semantic analysis is done only after syntactic analysis!

Lexical Semantics (1/2)

- We can organize word in to a lexical hierarchy, for example WordNet
  - Thus, 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 hot vs. cold
    - Synonyms (同義詞): Words have the same (very similar) meanings
    - Homonyms (同形異議詞): Words are written the same way but have different meanings which seems unrelated (e.g.: bank, suit, bass)

bass { instrument /'bes/ ; fish /'bæs/ ; tree } suit { case ; dress ; suite }

    - Homophones (同形同音異議詞) is a subcase of homonyms : two word are not only written the same way but also same pronunciation (bank, suit)

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.

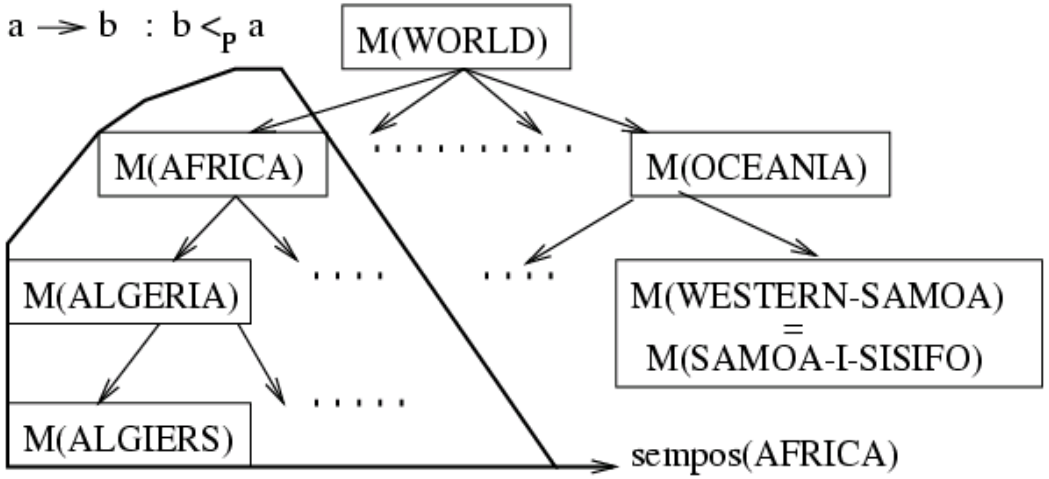


Figure 1: Example of Geographic Semantic Poset

Compositionality

- Once we have the meanings of individual words, we need to assemble them into the meaning of the whole sentence
- That is a hard problem because natural language often does not obey the principle of compositionality (by which the meaning of the whole can be strictly predicted from the meaning of the parts)
- E.g., The word *white* refers to very different colors in the following expressions:

  white paper, white hair, white skin, white wine

  - White hair is grey, a white skin really has a rosy color, and white wine is actually yellow (but yellow wine doesn't sound very appealing)
  - The groupings *white hair, white skin, and white wine* are examples of **collocations**
- Idioms in English are non-compositional (e.g., *kick the bucket*)

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

- 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 Related 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