Announcements

- Readings sent out
  - Bayesian probability (Wasserman “All of Statistics”)
  - Part-of-Speech (Jurafsky and Martin)
  - Parsing (Jurafsky and Martin)
- Next two weeks: Parsing and machine translation
- After Spring break: review and midterm
- After that: Project

Parse Trees

- Central to the description of NL syntax
- Parts of speech were a first step
- Today:
  - Constituents
  - Dependencies
- Context-free grammars for English
Noun Phrases

- Examples:
  - the elephant arrived
  - it arrived
  - elephants arrived
  - the big ugly elephant arrived
  - the elephant I love to hate arrived
- (They all appear in the same context - before a verb.)

Other Kinds of Phrases

- Prepositional phrases
  - on Tuesday
  - in March
  - under the leaking roof
- Sentences (clauses)
  - John loves Mary
  - John loves the woman he thinks is Mary
  - sometimes John thinks he is Mary
- Verb phrases, adjective phrases, adverb phrases ...

What Makes A Phrase A Phrase?

- You can move it (fronting, passivizing, inversion to form a question)
  - she makes delicious cake → delicious cake she made.
- You can conjoin it with a similar thing
  - the cat died → the cat and the mouse died
- You can replace it with a pronoun, “do,” “there,” or “then”
  - the furry kittens lost their mittens → they lost them
  - the professor eats snacks ... and the student does (too)
- It can be an answer to a “Wh” question.
  - What did he do? Taught computer science.
Production Rules

- Alternative ways to build a particular kind of phrase
- NP $\rightarrow$ Determiner Noun
- NP $\rightarrow$ ProperNoun
- Determiner $\rightarrow$ an
- Determiner $\rightarrow$ the
- Noun $\rightarrow$ elephant
- ProperNoun $\rightarrow$ Smith

Note the use of parts of speech!

- Yes, you can write this in BNF if you’d like.

Building Noun Phrases

- NP $\rightarrow$ Determiner N' | ProperNoun
- N' $\rightarrow$ Noun | AP N' | N' PP
- AP $\rightarrow$ Adv AP | Adj
- PP $\rightarrow$ Preposition NP

Rules like “Determiner $\rightarrow$ the | an | a” are the kinds of part-of-speech rules you’d need for a POS tagger (e.g., HMM emissions). These rules - and generalizations of them - are sometimes called the “lexicon.” Can integrate morphology here.

A Complex NP

the very large man on the broken roof with a headache
Context-Free Grammars

- Vocabulary of terminal symbols $\Sigma$
- Set of nonterminal symbols (AKA variables) $N$
- Special start symbol $S \in N$
- Production rules of the form
  \[ X \rightarrow \gamma \]
  where $X \in N$ (a nonterminal symbol)
  and $\gamma \in (N \cup \Sigma)^*$ (a sequence of terminals and nonterminals)

Two Views of CFGs

- A system for generating sentences in the grammar's language
  - Start with an $S$ node.
  - While there are any nonterminal symbols, nondeterministically rewrite some nonterminal using a production rule.
  - At the end, you have a sequence of terminals.
- A set of rules for assigning structure to (parsing) a sentence

Definitions

- Grammatical: said of a sentence in the language
- Ungrammatical: said of a sentence not in the language
- Derivation: sequence of top-down production steps
- Parse tree: graphical representation of the derivation

- A string is grammatical iff there's a derivation for it.
Declarative Sentences

- $S \rightarrow NP\ VP$
- VP (verb phrase) is typically what you used to call a "predicate" - the verb and its right-side arguments, like object, indirect object, etc.

Questions

- Yes/no questions:
  - $S \rightarrow \text{AuxVerb} NP\ VP$
- Wh-as-subject:
  - $S \rightarrow \text{WhNP} VP$
- Wh-as-something else:
  - $S \rightarrow \text{WhNP} \text{Aux} NP\ VP$

High-Level Points

- The rules I/the book have given you are great in some cases.
- Some failures:
  - overgenerating (generate bad English)
  - ambiguity
  - undergenerating (trees or sentences)
- Remember: there's no spec! Getting "the right" grammar is a matter of research, not mere implementation.
- There's a difference between "ungrammatical as English" and "ungrammatical with respect to a given grammar"
Agreement

- John loves Mary
- *John love Mary
- These men are very smart
- *This clever little children want some books
- How do we make subjects agree with verbs, or determiners agree with nouns?

Agreement, Using More Detailed Rules

- $S \rightarrow NP \ VP$
- $S_{3sg} \rightarrow NP_{3sg} \ VP_{3sg}$
- $SOther \rightarrow NPOther \ VPOther$
- $NP_{3sg} \rightarrow Det \ N'_{3sg} \mid ProperNoun_{3sg}$
- $N'_{3sg} \rightarrow N_{3sg} \mid AP \ N'_{3sg} \mid N'_{3sg} \ PP$
- $VP_{3sg} \rightarrow TransitiveVerb_{3sg} \ NP \mid ...

Verb Arguments

- A related problem: some verbs require certain constellations of arguments.
  - $VP \rightarrow TransitiveVerb \ NP$
  - $VP \rightarrow IntransitiveVerb$
  - $VP \rightarrow DitransitiveVerb \ NP \ PP \mid DitransitiveVerb \ NP \ NP$
  - $VP \rightarrow STakingVerb \ that \ S$
  - $VP \rightarrow VPTaking \ Verb \ to \ VP$
  - $VP \rightarrow TransitiveVerb \ kill \mid love$
  - $IntransitiveVerb \rightarrow eat \mid sleep$
  - $DitransitiveVerb \rightarrow show \mid give$
  - $STakingVerb \rightarrow know \mid believe$
  - $VPTakingVerb \rightarrow want \mid need$
Dependencies

- A somewhat different view of English grammar.
- The words are the vertices in a graph.
- Every word has a parent (except the root), forming a tree.
- The edges may be labeled to denote grammatical relations:
  - subject, object, indirect object of a verb
  - complement of a preposition or copula
  - temporal adverbial

Dependency Tree

I gave him my address on Tuesday

Context-Free Dependency Grammars

- gave → I (subject) gave
- gave → gave (indirect object) him
- gave → gave (object) address
- address → my (attributive) address
- gave → gave (temporal) on
- on → on (preposition complement) Tuesday
Food For Thought

• How are we going to find the structures?
• How are we going to decide among competing parses?
• Where are the rules going to come from?

Parsing

• Given a grammar $G$ and a sentence $x = (x_1, x_2, ..., x_n)$, find the best parse tree.

• We’re not going to simply build it step by step; we need to entertain many partial possibilities in parallel.

First View: Parsing as Search

Trees break into pieces (partial trees), which can be used to define a search space.
Top-Down Parsing (Recursive Descent)

x = “Book that flight”

Top-Down Parsing (Recursive Descent)

x = “Book that flight”

Top-Down Parsing (Recursive Descent)

x = “Book that flight”
Top-Down Parsing (Recursive Descent)

x = “Book that flight”

Top-Down Parsing (Recursive Descent)

x = “Book that flight”

Top-Down Parsing (Recursive Descent)

- Never wastes time exploring ungrammatical trees!
- Inefficiency: most search states (partial trees) could never lead to a derivation of our sentence.
Bottom-Up Parsing

book that flight

(Verb book) (Det that) (Noun flight)

(Noun book) (Det that) (Noun flight)

book that flight

(Nominal (Noun book)) (Det that) (Nominal (Noun flight))

(Nominal (Noun book)) (Det that) (Nominal (Noun flight))

book that flight

(Nominal (Noun book)) (Det that) (Nominal (Noun flight))

(Nominal (Noun book)) (Det that) (Nominal (Noun flight))

book that flight
Bottom-Up Parsing

- Never generates trees that are inconsistent with the sentence.
- Generates partial trees that have no hope of getting to S.
**Ambiguity Redux**

- A sentence may have many parses.
- Even if a sentence has only one parse, finding it may be difficult, because there are many misleading paths you could follow.
- Bottom-up: fragments that can never have a home in any S
- Top-down: fragments that never get you to x
- What to do when there are many parses ... how to choose? Return them all?

**Classical NLP: Parsing**

- **Fed raises interest rates 0.5 percent**
- Write symbolic or logical rules:
- **Grammar (CFG)**
  - ROOT \(\rightarrow\) S
  - S \(\rightarrow\) NP VP
  - NP \(\rightarrow\) DT NN
  - NP \(\rightarrow\) NN NNS
  - VP \(\rightarrow\) VBP NP
  - PP \(\rightarrow\) IN NP
- **Lexicon**
  - NN \(\rightarrow\) interest
  - NNS \(\rightarrow\) raise
  - VBP \(\rightarrow\) raises
  - VBZ \(\rightarrow\) raises

- Use deduction systems to prove parses from words
  - Minimal grammar on "Fed raises" sentence: 36 parses
  - Simple 10-rule grammar: 592 parses
  - Real-size grammar: many millions of parses

- This scaled very badly, didn't yield broad-coverage tools

**Ambiguities: PP Attachment**

The board approved its acquisition by Royal Truxton [at its monthly meeting].

For $27 a share.
Attachments

- I cleaned the dishes from dinner
- I cleaned the dishes with detergent
- I cleaned the dishes in my pajamas
- I cleaned the dishes in the sink

PP Attachment

<table>
<thead>
<tr>
<th>V</th>
<th>N1</th>
<th>P</th>
<th>N2</th>
<th>Attachment</th>
</tr>
</thead>
<tbody>
<tr>
<td>join</td>
<td>board</td>
<td>as</td>
<td>director V</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td>chairman</td>
<td>of</td>
<td>N.V.</td>
<td>N</td>
</tr>
<tr>
<td>using</td>
<td>crocitchite</td>
<td>in</td>
<td>filters</td>
<td>V</td>
</tr>
<tr>
<td>bring</td>
<td>attention</td>
<td>to</td>
<td>problem</td>
<td>V</td>
</tr>
<tr>
<td>is</td>
<td>asbestos</td>
<td>in</td>
<td>products</td>
<td>N</td>
</tr>
<tr>
<td>making</td>
<td>paper</td>
<td>for</td>
<td>filters</td>
<td>N</td>
</tr>
<tr>
<td>including</td>
<td>three</td>
<td>with</td>
<td>cancer</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always noun attachment</td>
<td>59.0</td>
</tr>
<tr>
<td>Most likely for each preposition</td>
<td>72.2</td>
</tr>
<tr>
<td>Average Human (4 head words only)</td>
<td>88.2</td>
</tr>
<tr>
<td>Average Human (whole sentence)</td>
<td>93.2</td>
</tr>
</tbody>
</table>

Syntactic Ambiguities I

- Prepositional phrases:
  They cooked the beans in the pot on the stove with handles.

- Particle vs. preposition:
  The puppy tore up the staircase.

- Complement structures
  The tourists objected to the guide that they couldn't hear.
  She knows you like the back of her hand.

- Gerund vs. participial adjective
  Visiting relatives can be boring.
  Changing schedules frequently confused passengers.
Syntactic Ambiguities II

- Modifier scope within NPs
  - impractical design requirements
  - plastic cup holder

- Multiple gap constructions
  - The chicken is ready to eat.
  - The contractors are rich enough to sue.

- Coordination scope:
  - Small rats and mice can squeeze into holes or cracks in the wall.

Dark Ambiguities

- Dark ambiguities: most analyses are shockingly bad (meaning, they don’t have an interpretation you can get your mind around)

- Unknown words and new usages

- Solution: We need mechanisms to focus attention on the best ones, probabilistic techniques do this

Human Processing

- Garden pathing:
  - the man who hunts ducks out on weekends
  - the cotton shirts are made from grows in Mississippi
  - the daughter of the king’s son loves himself

- Ambiguity maintenance

  - Have the police . . . eaten their supper?
  - come in and look around.
  - taken out and shot.