Brief Recap

- We covered the Viterbi Algorithm for POS tagging
- We know what CFGs and Syntax
- Questions?

Syntax in NLP applications

- Language modeling
  - Is "The girl I met wore a hat" a valid sentence in the language?
- Grammar checking
  - What's wrong with this sentence: "She wear of a hat"?
- Information extraction/Question Answering
  - In this sentence: "John worked at Pitt for two years, since the winter of 2014" when did John start working at Pitt?
    - Identify temporal expression noun phrase "the winter of 2014"
- Compositional semantics
  - Who did what to whom in this sentence: "The helpful man gave the crying child a coloring book about dinosaurs"
    - Identify subject, verb, direct object, indirect object

Syntax in NLP applications

-
- Sentiment analysis
  - In this sentence: "It is a shame that the expensive renovation drove out the long term residents of the neighborhood" how does the writer feel about various entities mentioned in the sentence?
    - Identify embedded sentence (renovation drove out residents) as well as the relationship between entities in the embedded sentence (renovation, residents)
- Framing
  - "The ball broke the window" vs. "I broke the window with the ball"
- Machine translation
  - Need to know how lanuages have different ways of organizing sentences (e.g., typical adjectives come after noun in French)

An English Grammar Fragment

- Sentences
- Noun phrases
  - Agreement
- Verb phrases
  - Subcategorization

Speech and Language Processing - Jurafsky and Martin

Sentence Types

- Declaratives: *A plane left.*
  - *S → NP VP*
- Imperatives: *Leave!*
  - *S → VP*
- Yes-No Questions: *Did the plane leave?*
  - *S → Aux NP VP*
- WH Questions: *When did the plane leave?*
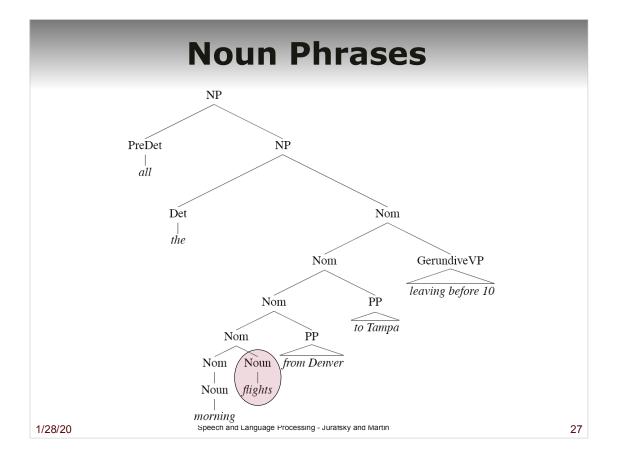 - *S → WH-NP Aux NP VP*

Noun Phrases

- Let's consider the following rule in more detail...
  - *NP → Det Nominal*
- Most of the complexity of English noun phrases is hidden in this rule.
- Consider the derivation for the following example
  - *All the morning flights from Denver to Tampa leaving before 10*

Noun Phrases

NP Structure

- Clearly this NP is really about *flights.* That's the central critical noun in this NP. Let's call that the *head*.

- We can dissect this kind of NP into the stuff that can come before the head, and the stuff that can come after it.

Determiners

- Noun phrases can start with determiners…
- Determiners can be
  - Simple lexical items: *the, this, a, an*, etc.
    - A car
  - Or simple possessives
    - John's car
  - Or complex recursive versions of that
    - John's sister's husband's son's car

Nominals

- Contains the head and any pre- and post- modifiers of the head.
  - Pre-
    - Quantifiers, cardinals, ordinals…
      - Three cars
    - Adjectives
      - large cars
    - Ordering constraints
      - Three large cars
      - ?large three cars

Postmodifiers

- Three kinds
  - Prepositional phrases
    - From Seattle
  - Non-finite clauses
    - Arriving before noon
  - Relative clauses
    - That serve breakfast
- Same general (recursive) rule to handle these
  - *Nominal → Nominal PP*
  - *Nominal → Nominal GerundVP*
  - *Nominal → Nominal RelClause*

Agreement

- By **agreement**, we have in mind constraints that hold among various constituents that take part in a rule or set of rules
- For example, in English, determiners and the head nouns in NPs have to agree in their number.

  This flight   *This flights

  Those flights   *Those flight

Problem

- Our earlier NP rules are clearly deficient since they don't capture this constraint
  - *NP → Det Nominal*
    - Accepts, and assigns correct structures, to grammatical examples (*this flight*)
    - But its also happy with incorrect examples (*these flight)
  - Such a rule is said to *overgenerate.*
  - We'll come back to this in a bit
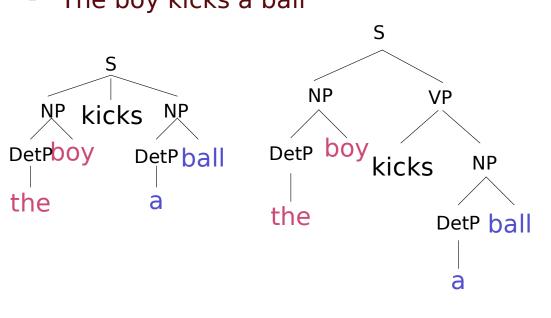
NP Constituency: Review

- NPs can all appear before a verb:
  - Some big dogs and some little dogs are going around in cars...
  - Big dogs, little dogs, red dogs, blue dogs, yellow dogs, green dogs, black dogs, and white dogs are all at a dog party!
  - I do not
- But individual words can't always appear before verbs:
  - *little are going...
  - *blue are...
  - *and are
- Must be able to state generalizations like:
  - Noun phrases occur before verbs

PP Constituency

- There is some structure:
  - Under a tree is a yellow dog.
  - A yellow dog is under a tree.
  - Called <u>Pre-posing</u> and <u>Post-posing</u>
- But not:
  - *Under, is a yellow dog a tree.
  - *Under a is a yellow dog tree.
- Consider the following sentence
  - I saw a man on a hill with a telescope.

VP Constituency

- "The boy kicks a ball"

VP Constituency

- Existence of VP is a linguistic (i.e., empirical) claim, not a methodological claim
- Syntactic evidence
  - VP-fronting (Bob was told to clean the carpet and quickly clean the carpet he did!)
  - VP-ellipsis (He cleaned the carpet quickly, and so did she)
  - Adjuncts can occur before and after VP, but not *in* VP (
    - He often eats beans ✓
    - He eats beans often ✓
    - He eats often beans ✗

Verb Phrases

- English *VP*s consist of a head verb along with 0 or more following constituents which we'll call *arguments*.

$$VP \rightarrow Verb \quad \text{disappear}$$
$$VP \rightarrow Verb\ NP \quad \text{prefer a morning flight}$$
$$VP \rightarrow Verb\ NP\ PP \quad \text{leave Boston in the morning}$$
$$VP \rightarrow Verb\ PP \quad \text{leaving on Thursday}$$

Subcategorization

- But, even though there are many valid VP rules in English, not all verbs are allowed to participate in all those VP rules.
- We can subcategorize the verbs in a language according to the sets of VP rules that they participate in.
- This is a modern take on the traditional notion of transitive/intransitive.
- Modern grammars may have 100s or such classes.

Subcategorization

- Sneeze: John sneezed
- Find: Please find [a flight to NY]$_{NP}$
- Help: Can you help [me]$_{NP}$[with a flight]$_{PP}$
- Prefer: I prefer [to leave earlier]$_{TO-VP}$
- …

Subcategorization

- *John sneezed the book
- *I prefer United has a flight
- *Give with a flight

- As with agreement phenomena, we need a way to formally express the constraints

Why?

- Right now, the various rules for VPs *overgenerate*.
  - They permit the presence of strings containing verbs and arguments that don't go together
  - For example
  - VP -> V NP therefore

    Sneezed the book is a VP since "sneeze" is a verb and "the book" is a valid NP

Let's try to solve this

- Ordering constraint on Adjectives:
  - *Opinion-size-age-shape-color-origin-material-purpose*
  - You can have a lovely little old rectangular green French silver whittling knife.
  - But not a green great dragon.

  - What rules would we need to satisfy this ordering.

Possible CFG Solution

- Possible solution for agreement.
- Can use the same trick for all the verb/VP classes.

- SgS -> SgNP SgVP
- PlS -> PlNp PlVP
- SgNP -> SgDet SgNom
- PlNP -> PlDet PlNom
- PlVP -> PlV NP
- SgVP ->SgV Np
- …

CFG Solution for Agreement

- It works and stays within the power of CFGs
- But its ugly
- And it doesn't scale all that well because of the interaction among the various constraints explodes the number of rules in our grammar.

The Point

- CFGs appear to be just about what we need to account for a lot of basic syntactic structure in English.
- But there are problems
  - That can be dealt with adequately, although not elegantly, by staying within the CFG framework.
- There are simpler, more elegant, solutions that take us out of the CFG framework (beyond its formal power)
  - LFG, HPSG, Construction grammar, XTAG, etc.
  - Prior edition explores the unification approach