NLP
Introduction to NLP

Syntax
Syntax

• Is language more than just a “bag of words”?
  – Grammatical rules apply to categories and groups of words, not individual words.

• Example
  – a sentence includes a subject and a predicate. The subject is a noun phrase and the predicate is a verb phrase.
  – Noun phrase: The cat, Samantha, She
  – Verb phrase: arrived, went away, had dinner

• When people learn a new word, they learn its syntactic usage.
  – Examples: wug (n), cluvius (adj) – use them in sentences
  – Hard to come up with made up words: forkle, vleer, etc. all taken.
Defining Parts of Speech

• What do nouns typically have in common?
  – E.g., *can* be preceded by “the”.

• What about verbs?
  – Verbs can be preceded by “can’t”.

• Adjectives can come between “the” and a noun.
  – How is this different from grade school definitions?

• Determiners
  – a, the, many, no, five

• Prepositions
  – for, to, in, without, before
The Lexicon

• How do we think of words like cat, run, five?
  – pronunciation, part of speech, meaning
  – Five: /faɪv/, numeral, “5”

• Ambiguity
Constituents

• Constituents are continuous
• Constituents are non-crossing
  – if two constituents share one word, then one of them must completely contain the other.
• Each word is a constituent
Constituent Tests

• “coordination” test
  – She bought a bagel and three chocolate croissants

• “pronoun” test
  – A small dog is barking in the park.
  – It is barking in the park

• “question by repetition” test:
  – I have seen blue elephants
  – Blue elephants?
  – * Seen blue?
  – Seen blue elephants?

• “topicalization” test:
  – Blue elephants, I have seen.

• “question” test:
  – What have I seen?

• “deletion” test
  – Last year I saw a blue elephant in the zoo.

• “semantic” test
• “intuition” test
How to generate sentences

• One way: tree structure
  – Generate the tree structure first
  – Then fill the leaf nodes with terminals
A Simple Syntactic Rule

• The simplest rule for a sentence, e.g. “Birds fly”

\[ S \rightarrow N \ V \]
Simplest Grammar

S → N V
N → Samantha | Min | Jorge
V → left | sang | walked

Sample sentences:
  Samantha sang
  Jorge left
Syntax

• The verbs so far were intransitive (no direct object)
• What rules are needed next?
  – Transitive verbs and direct objects (“Jorge saw Samantha”)
  – Determiners (“the cats”)
• Combinatorial explosion (even for the simplest form of sentences)
  – Need for noun phrases
  – Ditto for verb phrases
Latest Grammar

\[
S \rightarrow NP \ VP \\
NP \rightarrow DT \ N \\
VP \rightarrow V \ NP \\
DT \rightarrow \text{the} \mid a \\
N \rightarrow \text{child} \mid \text{cat} \mid \text{dog} \\
V \rightarrow \text{took} \mid \text{saw} \mid \text{liked} \mid \text{scared} \mid \text{chased}
\]

Sample sentences:
- a dog chased the cat
- the child saw a dog
Alternatives

• Different expansions of a category are delineated with " | "

- NP $\rightarrow$ PN | DT CN

• One rule for proper nouns and another for common nouns
Latest Grammar

S → NP VP
NP → DT CN
NP → PN
VP → V NP
DT → the | a
CN → child | cat | dog
PN → Samantha | Jorge | Min
V → took | saw | liked | scared | chased

Sample sentences:

a child scared Jorge
Min took the child
Optional categories

• Wherever N is allowed in a sentence,
  – DT N
  – JJ N
  – DT JJ N

  are also allowed

• We can use the notation for alternatives
  – NP → N | DT N | JJ N | DT JJ N

• Optional categories can be also marked using parentheses:
  – NP → (DT) (JJ) N
Verb Phrases

- Samantha ran.
- Samantha ran to the park.
- Samantha ran away.
- Samantha bought a cookie.
- Samantha bought a cookie for John.

Overall structure

- \( VP \rightarrow V (NP) (P) (NP) \)
Latest Grammar

\[ S \rightarrow NP \ VP \]
\[ NP \rightarrow DT \ CN \]
\[ NP \rightarrow PN \]
\[ VP \rightarrow V \ (NP) \ (P) \ (NP) \]
\[ DT \rightarrow \text{the} | \text{a} \]
\[ CN \rightarrow \text{child} | \text{cat} | \text{dog} \]
\[ PN \rightarrow \text{Samantha} | \text{Jorge} | \text{Min} \]
\[ P \rightarrow \text{to} | \text{for} | \text{from} | \text{in} \]
\[ V \rightarrow \text{took} | \text{saw} | \text{liked} | \text{scared} | \text{chased} | \text{gave} \]

Sample sentences:

Samantha saw the cat
Jorge gave the cat to Min
**Prepositional Phrases**

- **Examples:**
  - Mary bought a book for John in a bookstore.
  - The bookstore sells magazines.
  - The bookstore on Main St. sells magazines.
  - Mary ran away.
  - Mary ran down the hill.

- **Changes are needed to both NP and VP to accommodate prepositional phrases**
  - Wherever a preposition is allowed, it can be followed by a noun phrase.
  - Run up
  - NP can contain any number of PPs but only up to two NPs.

- **How do we revise the grammar accordingly?**
The Rules So Far

- $S \rightarrow NP\ VP$
- $NP \rightarrow (DT)\ (JJ)\ N\ (PP)$
- $VP \rightarrow V\ (NP)\ (PP)$
- $PP \rightarrow P\ (NP)$
PP Ambiguity

• The boy saw the woman with the telescope.

  PP \rightarrow \text{PREP NP}
  VP \rightarrow \text{V NP PP}
  VP \rightarrow \text{V NP}
  NP \rightarrow \text{DT N}
  NP \rightarrow \text{DT N PP}
Repetition (*)

• \((JJ^*)\) = a sequence of zero or more JJ
• Are all sequences of adjectives allowed?
  – a big red house
  – * a red big house
• Adjective ordering in English depends on semantics!
Exercise

- The Little Red Riding Hood
- Three Little Pigs
- The Three Musketeers
- The Steadfast Tin Soldier
- The French Connection
- Old Macdonald
- Five Golden Rings
- The Ancient Mariner
Adjective ordering

- **Det**
- Number
- Strength
- Size
- Age
- Shape
- Color
- Origin
- Material
- Purpose
- **Noun**

- det < number < size < color < purpose < noun
- strength < material < noun
- origin < noun
Nested Sentences

• Examples:
  – I don’t recall whether I took the dog out.
  – Do you know if the mall is still open?
• \( VP \rightarrow V \ (NP) \ (NP) \ (C \ S) \ (PP^*) \)
• Can \( (C \ S) \) appear inside an NP?
  – Whether he will win the elections remains to be seen.
Recursion

• S can generate VP, VP can generate S
• NP can generate PP, PP can generate NP
• What does recursion allow?
• Is there a longest sentence in English?
• Conjunction of NPs:
  \[ \text{NP} \rightarrow \text{NP} \text{ and } \text{NP} \]
• Conjunction of PPs:
  \[ \text{PP} \rightarrow \text{PP} \text{ and } \text{PP} \]
• Conjunction of VPs:
  \[ \text{VP} \rightarrow \text{VP} \text{ and } \text{VP} \]
Meta-patterns

• $S \rightarrow NP\ VP$
  – $NP \rightarrow (DT)\ (JJ)\ N\ (PP)$
  – $VP \rightarrow V\ (NP)\ (PP)$
  – $PP \rightarrow P\ (NP)$

• Is there a meta-pattern here?
  – $XP \rightarrow (\text{specifier})\ X'$
  – $X' \rightarrow X\ (\text{complement})$

• Example: $NP \rightarrow DT\ N'$

• X-bar Theory
Meta-rules for Conjunctions

• Conjunction
  – $X \rightarrow X \text{ and } X$

• This kind of rule even covers entire sentences
  – $S \rightarrow S \text{ and } S$
**Auxiliaries**

- Is “Aux V” a constituent?
  - I have seen blue elephants and will remember them forever.

- Recursion:
  - VP $\rightarrow$ Aux VP
  - Raj may have been sleeping.

- Is such recursion unlimited?
Exercise

• Grammar:
  - \( S \rightarrow NP \ VP \ | \ CP \ VP \)
  - \( NP \rightarrow (DT) \ (JJ^*) \ N \ (CP) \ (PP^*) \)
  - \( VP \rightarrow V \ (NP) \ (NP) \ (PP^*) \ | \ V \ (NP) \ (CP) \ (PP^*) \)
  - \( PP \rightarrow P \ NP \)
  - \( CP \rightarrow C \ S \)

• What rules are needed to generate these three sentences:
  - 1. The small dog of the neighbors brought me an old tennis ball.
  - 2. That wugs have three eyes is unproven by scientists.
  - 3. I saw the gift that the old man gave me at the meeting.
Notes

- Syntax helps with sentences like “* The milk drank the cat”, “The milk is drunk by the cat”
- Overgeneration “The girl saw”
- Undergeneration
- Grammar – between the two
Arguments vs. Adjuncts

• Arguments
  – Mandatory (e.g., “* Romeo likes”, “*likes Juliet”)
  – Cannot be repeated (e.g., “* Juliet likes Romeo John”)
  – Verbs can have more than one subcategorization frame

• Adjuncts
  – Optional
  – Typically prepositional phrases or adverbs
  – Can be repeated (e.g., “Apparently Candace ate pizza yesterday at the restaurant with pleasure”)
