Three Models for the Description of Language

written by Noam Chomsky

Matt Willsey

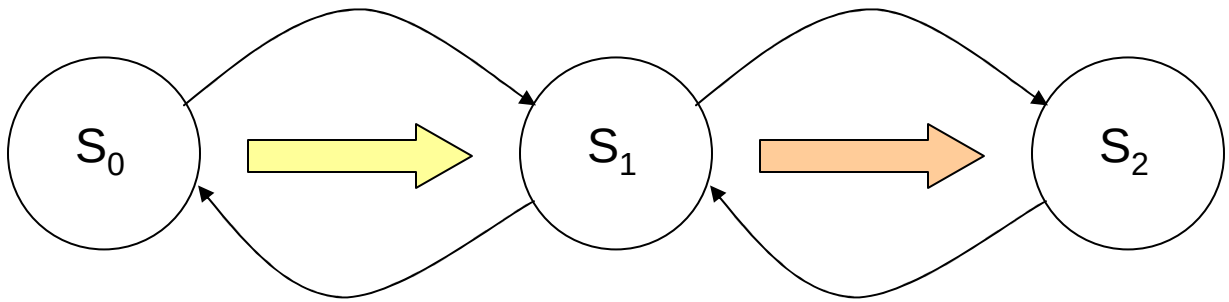
6.441: Transmission of Information

May 12, 2006

Background

- **Goal:** Develop simple structural model for the English language
  - Provides a useful method for generating/classifying grammatical English sentences
  - “Reveals” underlying structure of the English Language and how ideas are communicated
- **Three Models**
  - Finite-State Markov Processes
  - [∑, F] Grammars
  - Transformational Grammars
- **Disclaimer:** Sweeping details under the rug

Finite-State Markov Processes



Example

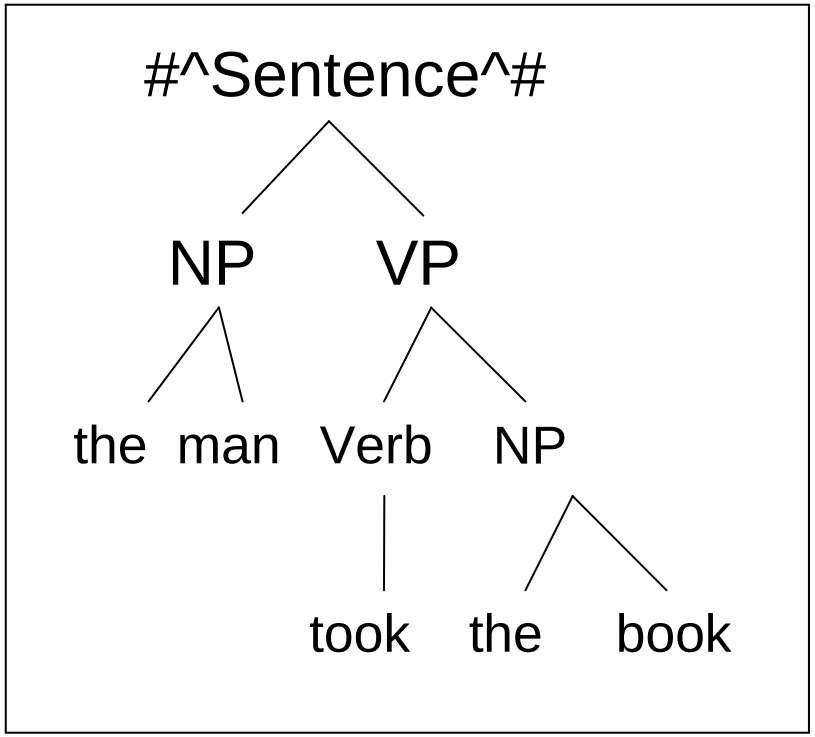
$$a_{0,1,5} \quad a_{1,2,3} \quad \cdots$$

Finite-State Markov Processes

- Formal System Description
  - finite number of states: $S_0, S_1, \ldots, S_q$
  - set of transition symbols:
    $$A = \{a_{ijk} \mid 0 \le i, j \le q; 1 \le k \le N_{ij} \ \forall\, i, j\}$$
  - pairs of connected states: $C = \{(S_i, S_j)\}$
- English cannot be modeled with a finite-state Markov process.
- Fails with dependencies: if-then, either-or

$$a \wedge a \wedge \ldots \wedge b \wedge b$$

[∑, F] Grammars



(5) ∑: # ^ Sentence ^ #
F: Sentence → NP ^ VP
VP → Verb ^ NP
NP → the ^ man, the ^ book
Verb → took

(6) D: # ^ Sentence ^ #
^ NP ^ VP ^
^ the ^ man ^ VP ^
^ the ^ man ^ Verb ^ NP ^
^ the ^ man ^ took ^ NP ^
^ the ^ man ^ took ^ the ^ book ^

[∑, F] Grammars are MORE powerful than finite-state Markov models because they can describe $a \wedge a \wedge \ldots \wedge b \wedge b$.

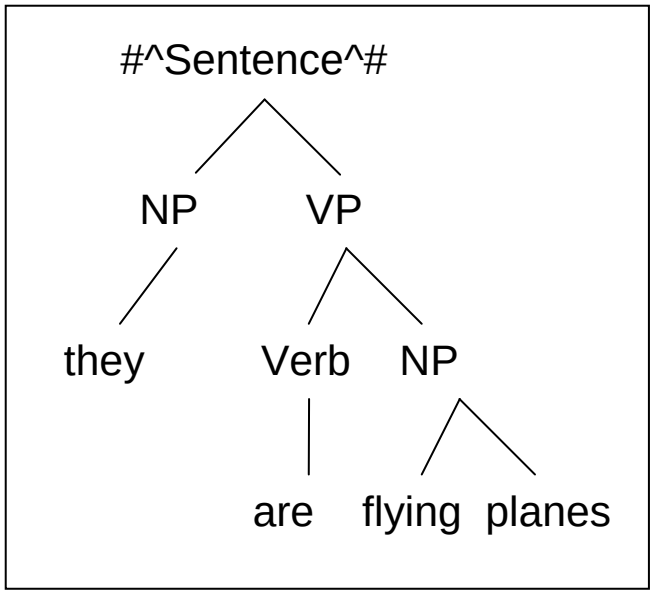
Z → a ^ Z ^ b

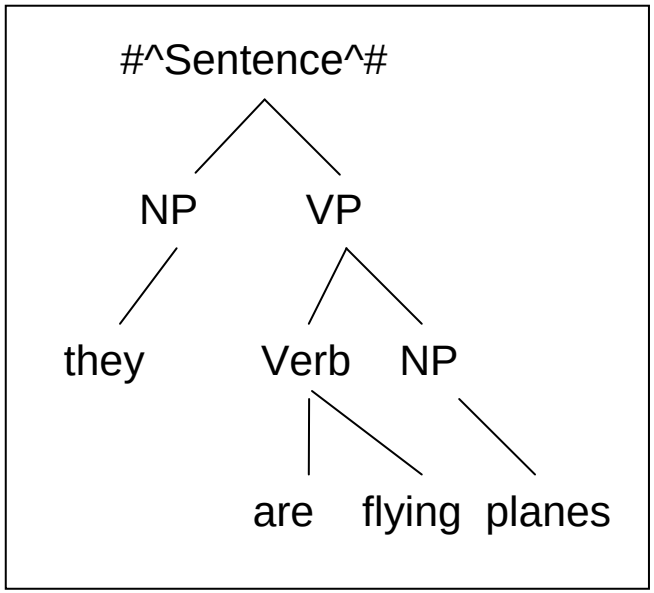
Transformation Grammar

- Still [∑, F] grammars are too limited for English
- How do we generate “have taken” from the string “Verb”?
- Define the rules:
  - Verb → Auxiliary ^ V
  - Auxiliary → have ^ en
  - V → take
- Output is the kernel string “have ^ en ^ take”
- Now define the transformation: Af ^ v → v ^ Af ^ #
- Applying the transformation to the kernel string yields “have ^ take ^ en”

What Did We Do?

- Developed a simple model to generate and classify grammatical sentences
  - “Chung asked a hard question.” => Grammatical
  - “Chung a asked question hard.” => Not Grammatical
- Developed a clue how ideas in the brain are converted to sentences, for example, “they are flying planes.”





What else?

- To compress English strings, Lempel-Ziv assumes English is derived from a Markovian source
- Chomsky proved that English cannot be modeled by a finite-state Markov model because of dependencies.

  "If either if the candle falls, then the floor catches on fire, or the toaster smokes, then the fire alarm will sound."

- English sentences used in practice still can be modeled by a finite Markov source when using Lempel-Ziv.

Questions