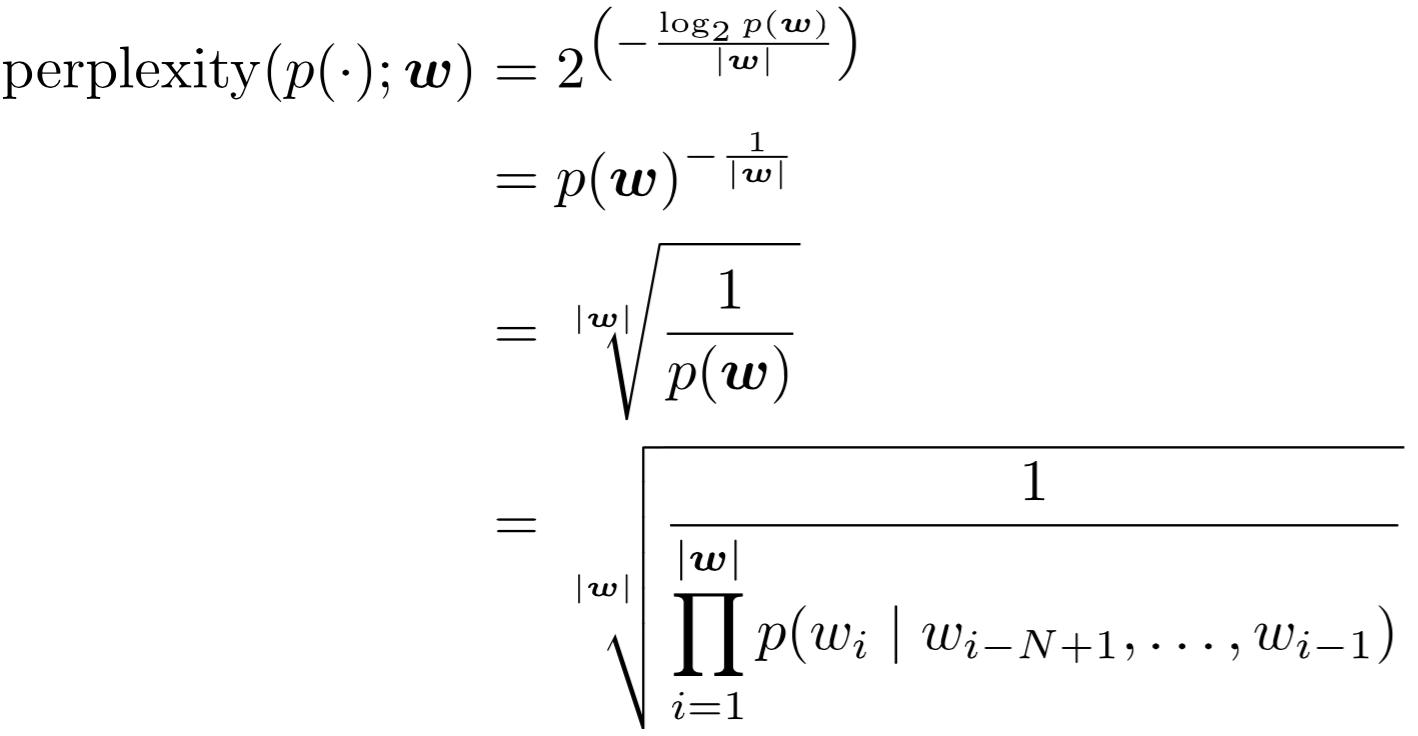
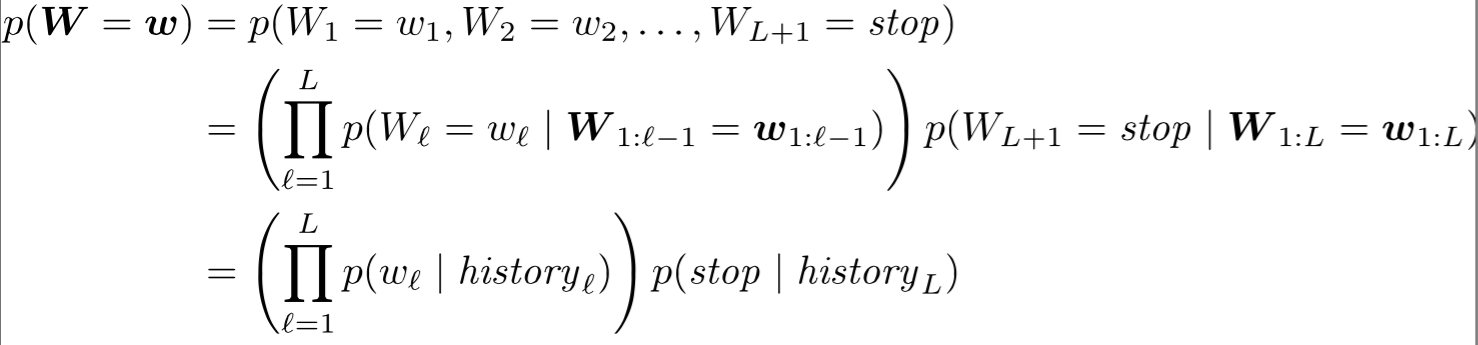
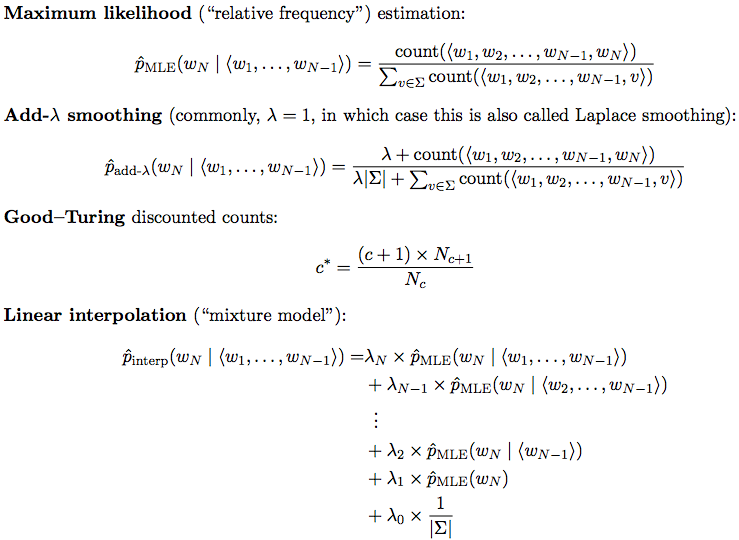
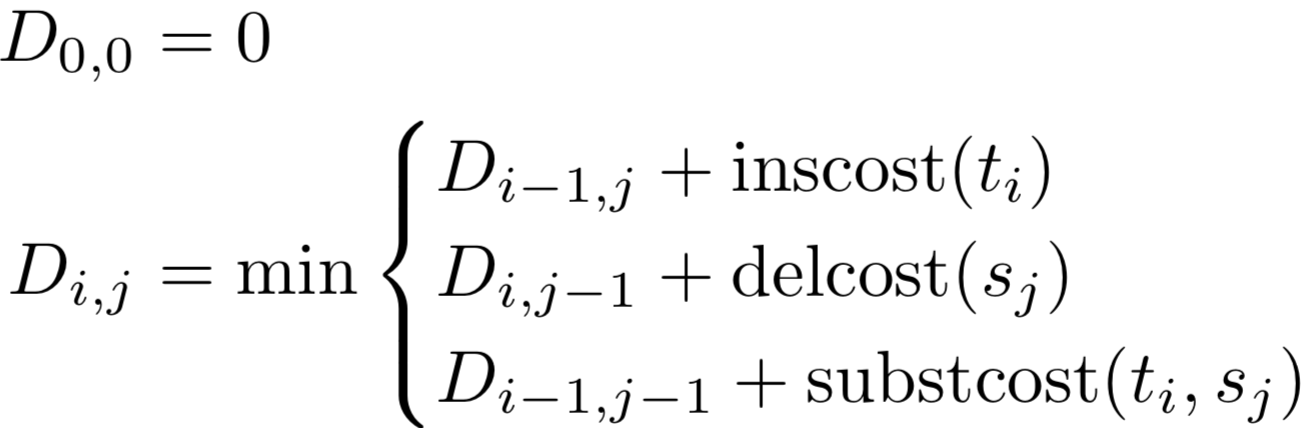
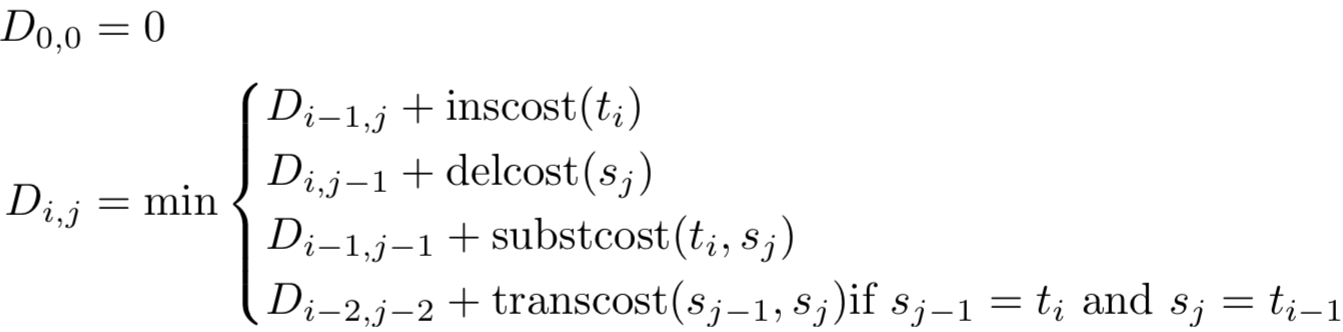
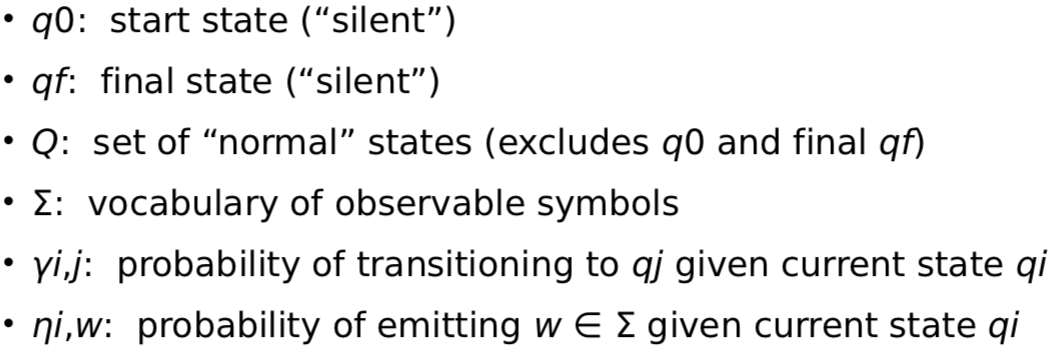
**# Language Models & Smoothing**

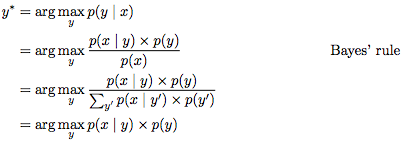
****

**# Noisy Channel & Edit Distance**

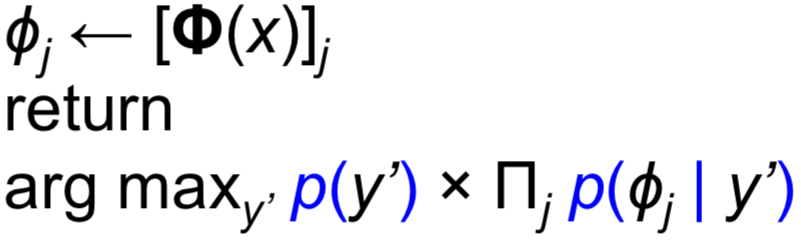
****

**+ Noisy Channel**

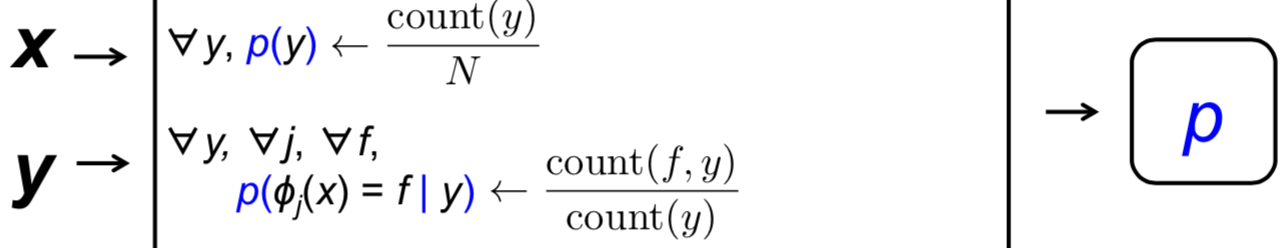
**The Noisy Channel Model assumes we want to recover the most likely system output y for an observed input x by modeling the probability distribution p(y | x)**

**# Classification**

**+ Naïve Bayes Classifier**

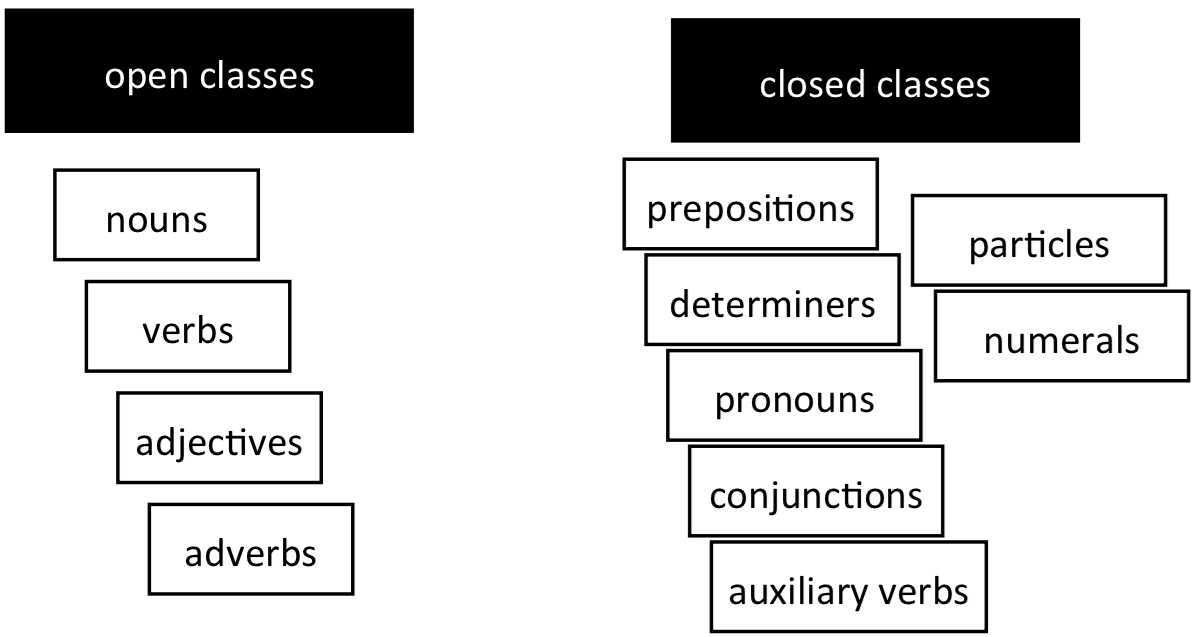
****

**+ Naïve Bayes Learner**

****

**# Sentiment Analysis**

**Social Media (tweets), product reviews, discussion forum posts, and blog posts**

**# Part of Speech Tags**

|  |  |  |  |
| --- | --- | --- | --- |
| **CC** | **Conj.** | **PRP** | **Pronoun, personal** |
| **CD** | **Numeral, Cardinal** | **RB** | **Adverb** |
| **DT** | **Determiner** | **RBR** | **Adverb, comparative** |
| **IN** | **Prep., Conj.** | **RP** | **Particle** |
| **JJ** | **Adj. or Numeral** | **VB** | **Verb, base form** |
| **NN** | **Noun, common, s.** | **VBD** | **Verb, past tense** |
| **NNP** | **Noun, proper, s.** | **VBG** | **V., present participle** |
| **NNPS** | **Noun, proper, pl.** | **VBN** | **V., past participle** |
| **NNS** | **Noun, common, pl.** | **VBP** | **V., present, not 3rd** |

**Two difficulties:**

**1) What tags? Unknown words?**

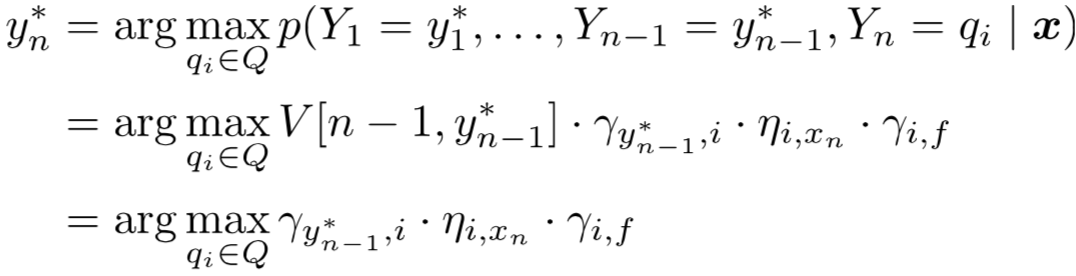
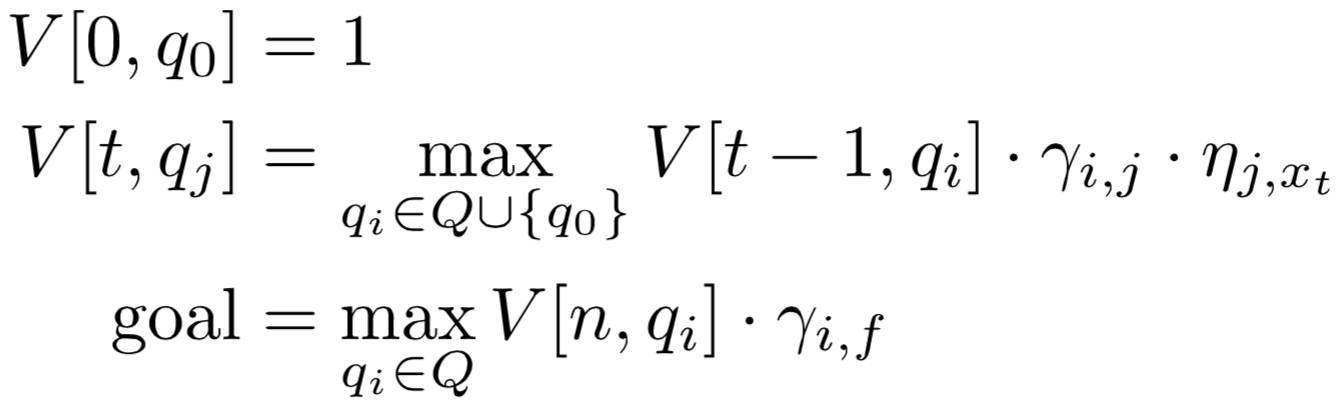
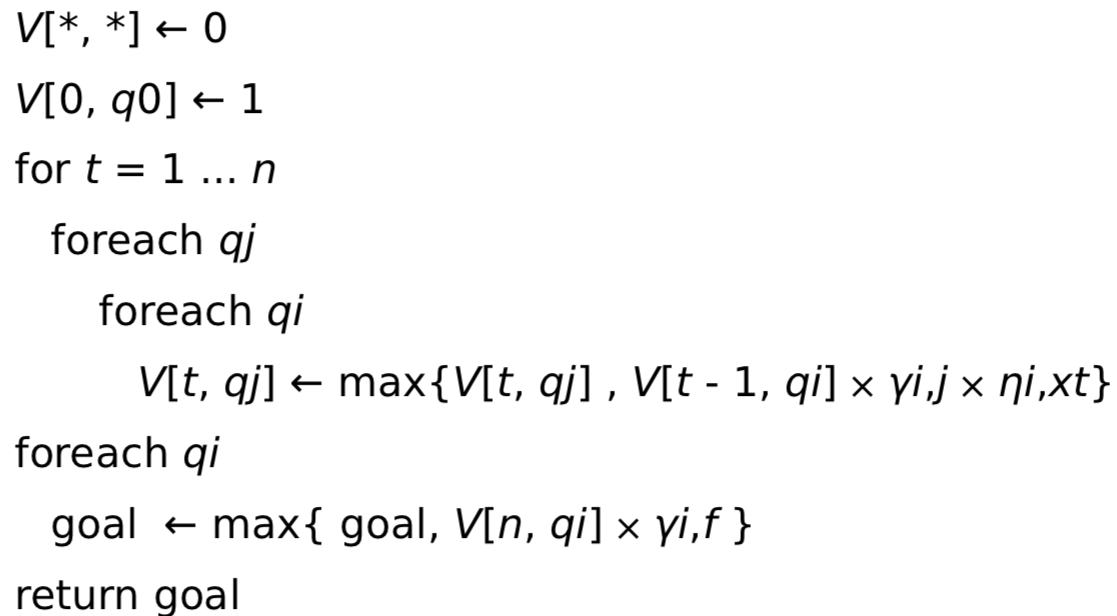
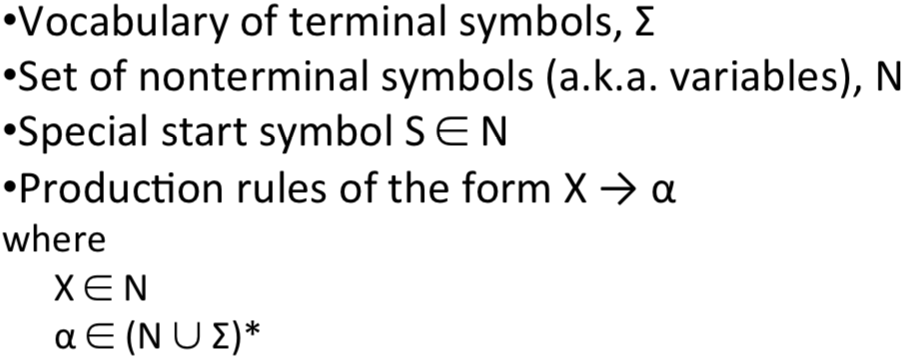
**2) Disambiguating words (more than one tag)**

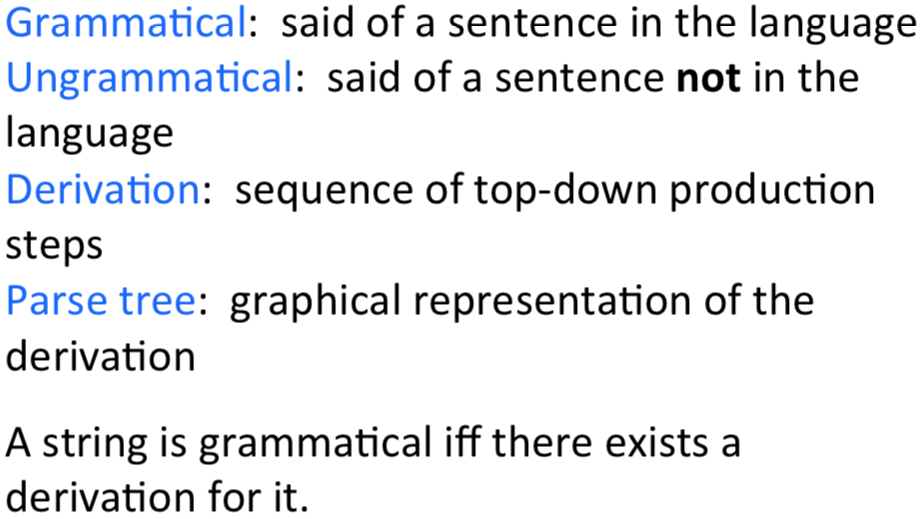
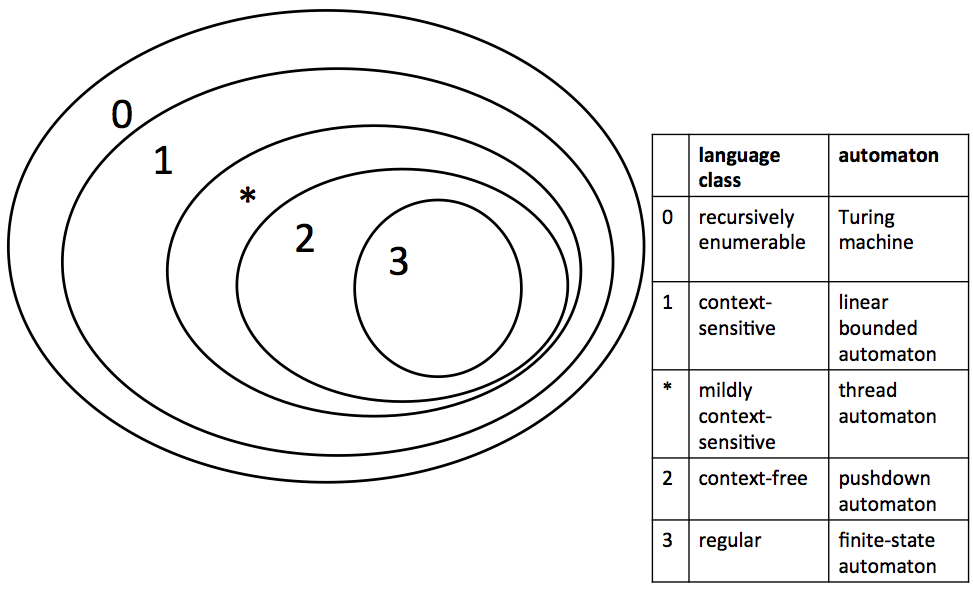
**# Hidden Markov Models**

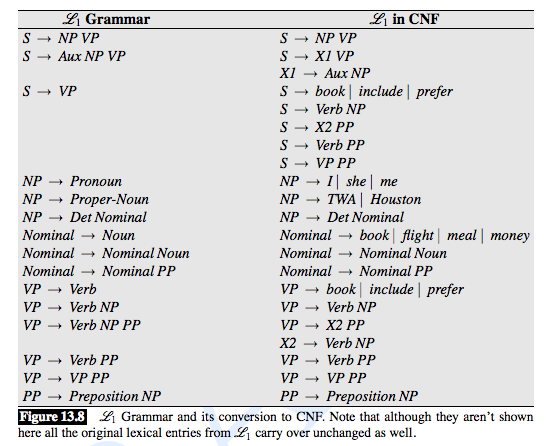
**+ Markov Property: the state we move to at time t depends only on the state we were at time t-1**

**+ Symbols are observed, state sequence is hidden**

**+ Viterbi Algorithm**

****

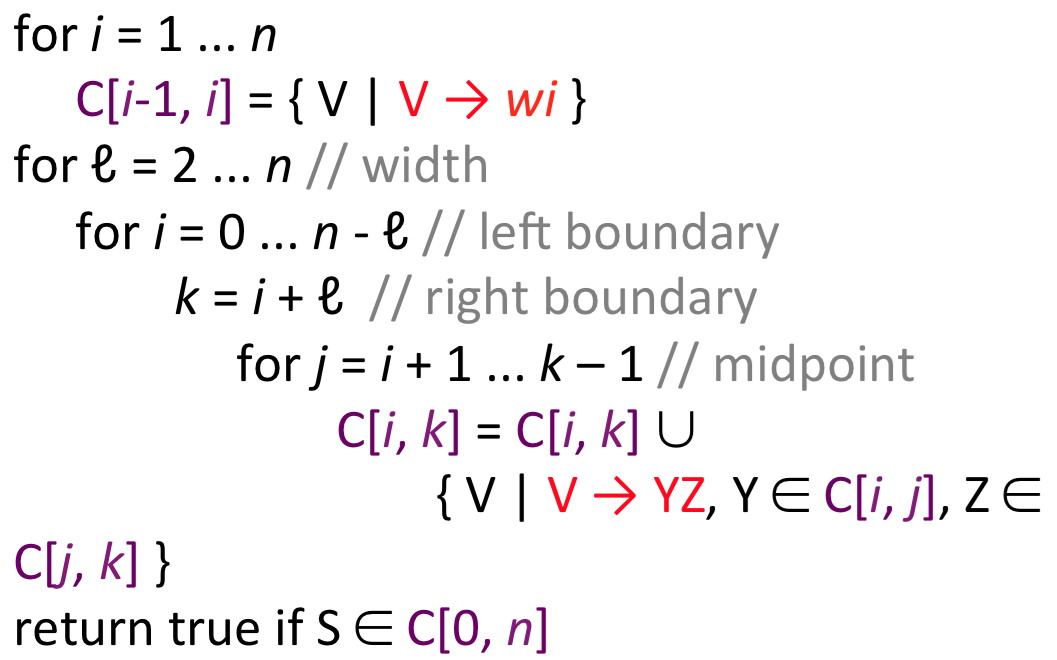
**# Syntactic Representation**

**# Chomsky Hierarchy**

**+ Pumping Lemma**

**If L is an infinite regular language, then there are strings x, y, and z such that y is not empty, xynz is in L, for all n >= 0**

**# Context-free Recognition & CKY Algorithm**

**Requires the grammar be in Chomsky normal form**