ECE365: Introduction to NLP

Spring 2021

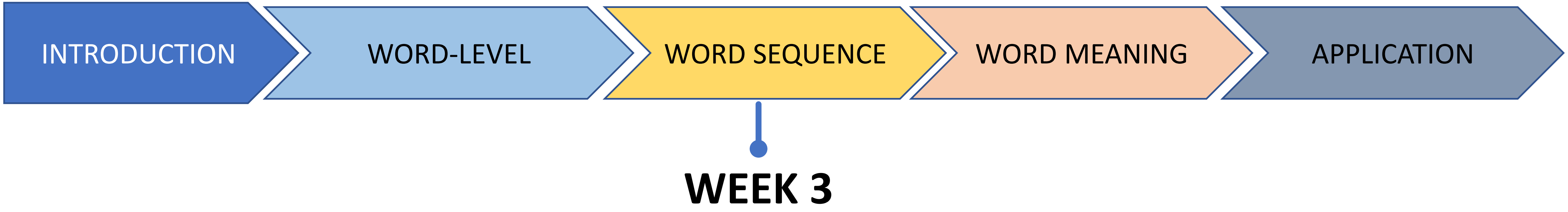
Lecture 5: Words in a sequence – Sequence labeling

[Reading J&M Chapter 8 (up to and including 8.4)]

Logistics

- Lab 3 is up

Course Progress



INTRODUCTION

WORD-LEVEL

WORD SEQUENCE

WORD MEANING

APPLICATION

WEEK 3

What is the nature of understanding we can get considering words as sequences?

- Mr. Forever who lives dangerously thinks he has nine lives.

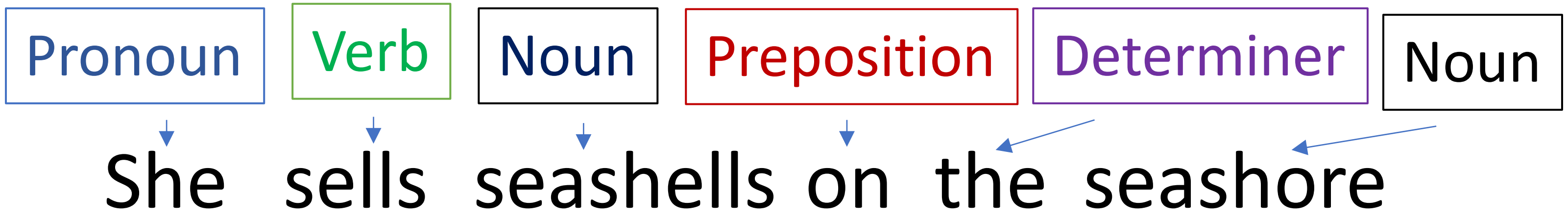
- We ate in the afternoon and went on to have an afternoon tea.

What is sequence labeling?

- Input: a sequence of word tokens **w**

- Output: a sequence of tags **t**, one per word (t∈ T)

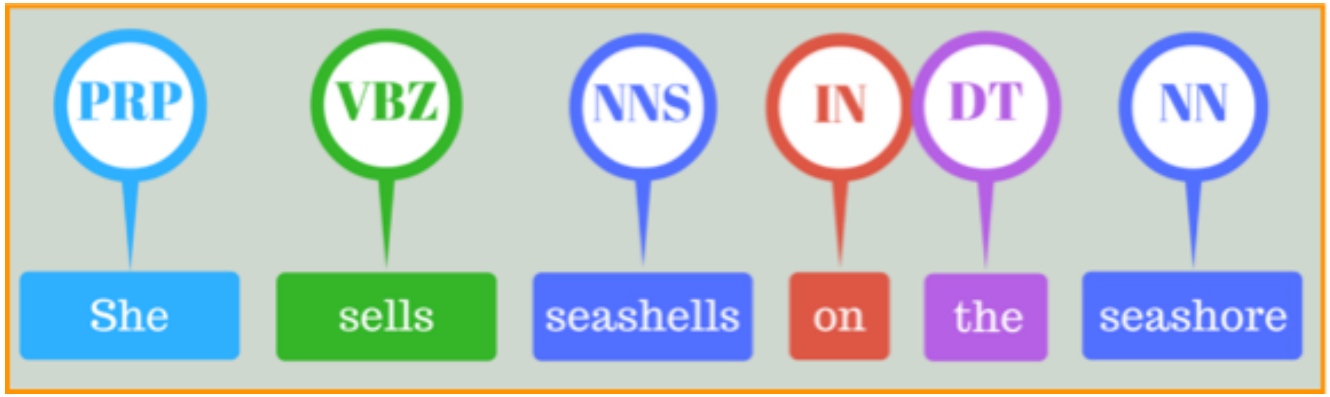
What is sequence labeling?



Pronoun Verb Noun Preposition Determiner Noun

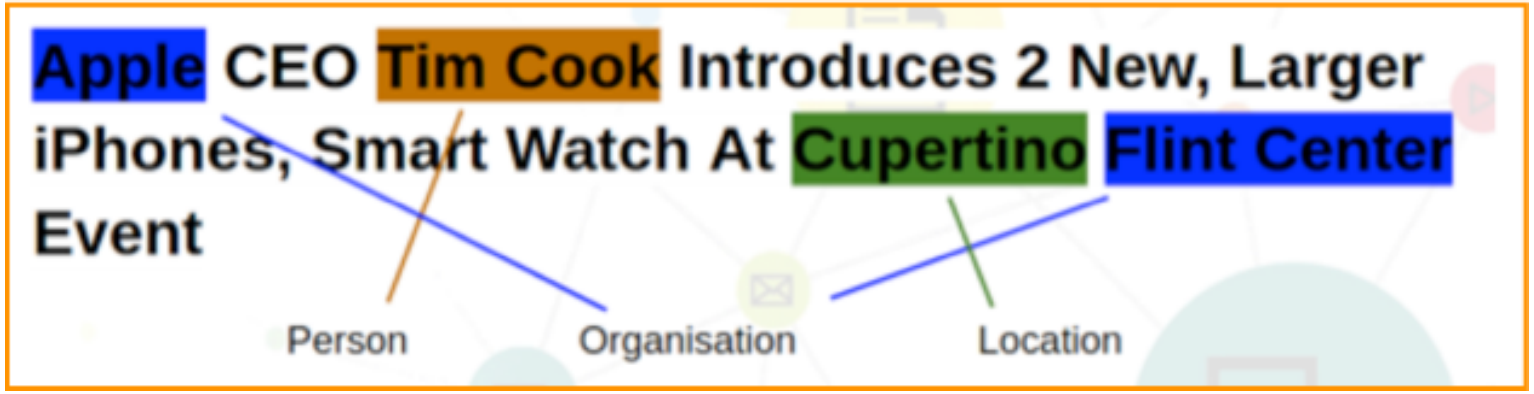
She sells seashells on the seashore

Sequence labeling tasks



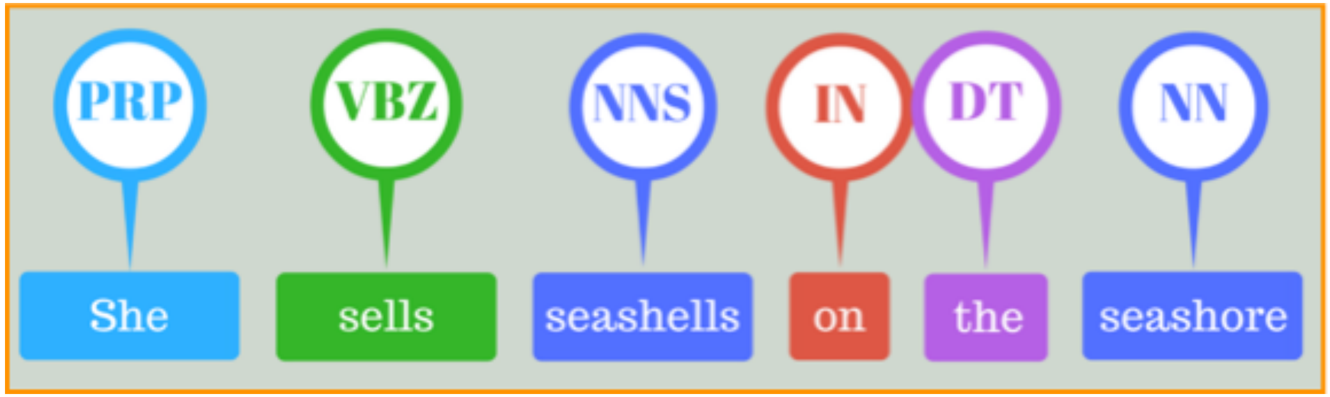
Part of Speech (POS) Tagging

Named Entity Recognition
(NER)


Apple CEO Tim Cook Introduces 2 New, Larger iPhones, Smart Watch At Cupertino Flint Center Event

Overview

POS tagging



Hidden Markov Model (HMM)

Viterbi Algorithm

What is sequence labeling?

- Input: a sequence of word tokens **w**

- Output: a sequence of POS tags **t**, one per word (t∈ T)

Part Of Speech Tagging

What are POS tags?

- Word classes
  - Nouns, Verbs, Adjectives, Adverbs

  - Prepositions, Conjunctions, Auxiliary verbs, Pronouns, determiners, numerals

POS tags

Open Classes

Nouns

Verbs

Adjectives

Adverbs

Closed Classes

Pronouns

Determiners

Auxiliary verbs

Prepositions

Conjunctions

Particles

numerals

Google Universal POS Tags

ADJ: adjective

ADP: adposition (preposition or postposition)

ADV: adverb

AUX: auxiliary

CCONJ: coordinating conjunction

DET: determiner

INTJ: interjection

NOUN: noun

NUM: numeral

PART: particle

PRON: pronoun

PROPN: proper noun

PUNCT: punctuation

SCONJ: subordinating conjunction

SYM: symbol

VERB: verb

X: other

Why do POS tagging?

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

Pronoun verb adjective noun! pronoun verb adjective, conjunction preposition adjective noun. Determiner….

POS tagging permits abstraction, allowing models to be more general

Why do POS tagging?

Text-to-Speech (how to pronounce the following words?)

English

- Transport
- Object. (She did not object to taking the object with her.)
- Discount
- Address
- Content

French: est, president

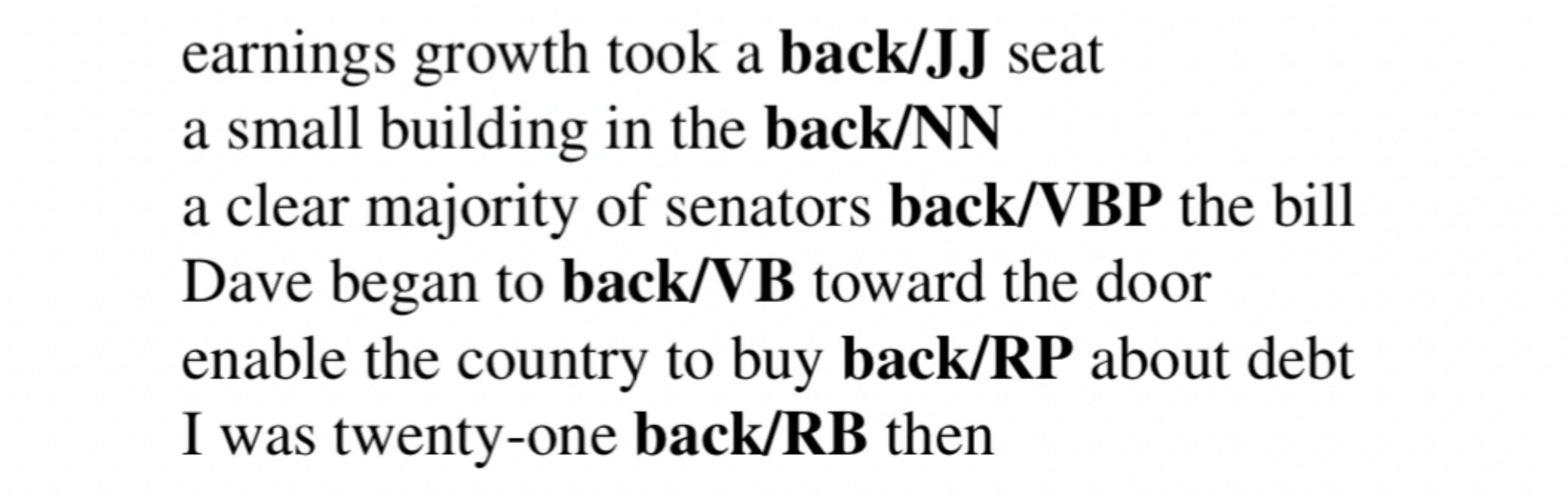
Useful for machine translation

How do humans assign tags?

- Jabberwocky (by Lewis Carroll 1872)

‘Twas brillig, and the slithy toves
  Did gyre and gimble in the wabe:
  All mimsy were the borogoves,
  And the mome raths outgrabe.

Why is POS tagging hard?

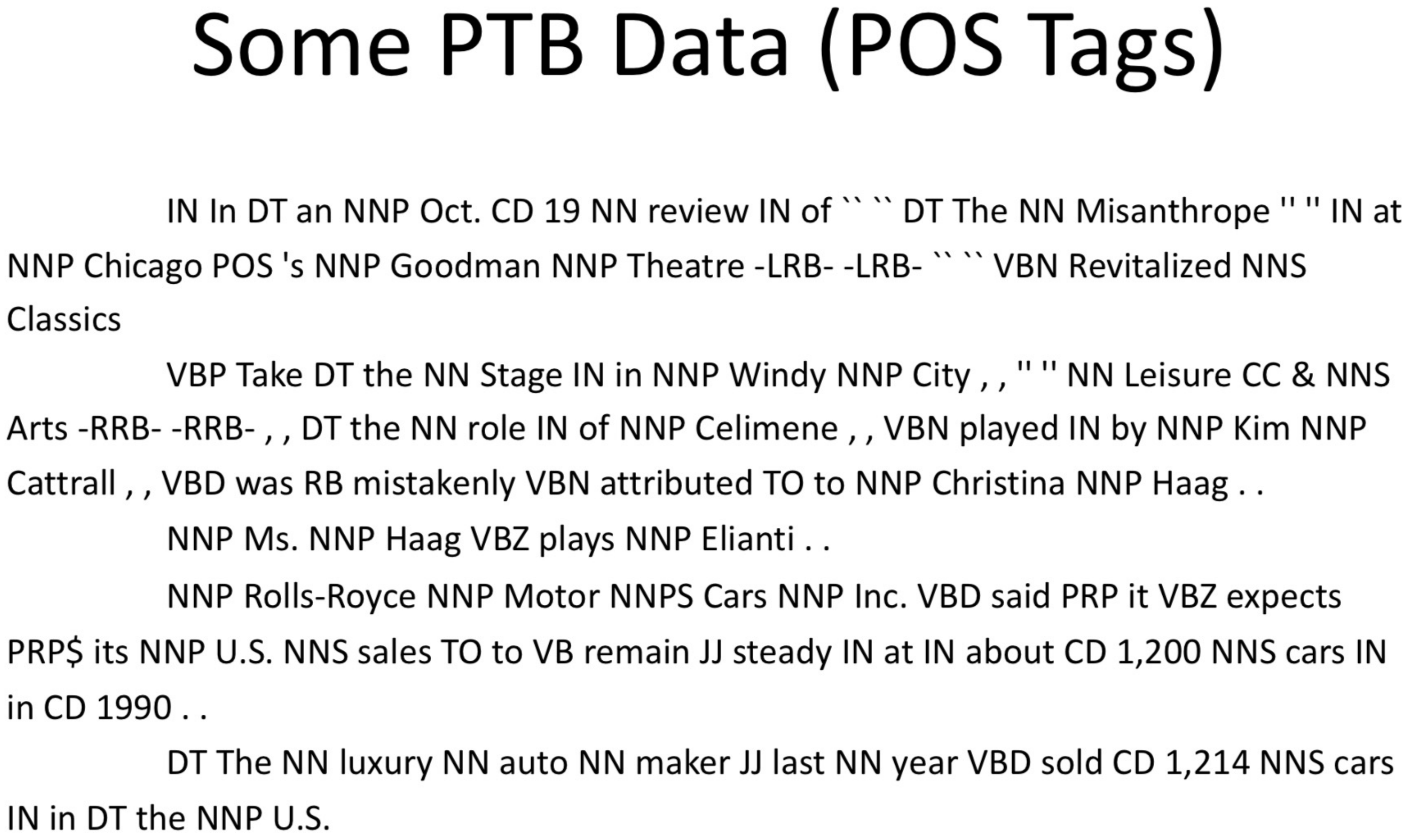


earnings growth took a **back/JJ** seat
a small building in the **back/NN**
a clear majority of senators **back/VBP** the bill
Dave began to **back/VB** toward the door
enable the country to buy **back/RP** about debt
I was twenty-one **back/RB** then

Tag ambiguity: each word may have multiple POS tags

11% of all word types or 40% of all word tokens in Brown corpus (1M words) are ambiguous

What resources are available?



Some PTB Data (POS Tags)

IN In DT an NNP Oct. CD 19 NN review IN of `` `` DT The NN Misanthrope '' '' IN at NNP Chicago POS 's NNP Goodman NNP Theatre -LRB- -LRB- `` `` VBN Revitalized NNS Classics

VBP Take DT the NN Stage IN in NNP Windy NNP City , , '' '' NN Leisure CC & NNS Arts -RRB- -RRB- , , DT the NN role IN of NNP Celimene , , VBN played IN by NNP Kim NNP Cattrall , , VBD was RB mistakenly VBN attributed TO to NNP Christina NNP Haag . .

NNP Ms. NNP Haag VBZ plays NNP Elianti . .

NNP Rolls-Royce NNP Motor NNPS Cars NNP Inc. VBD said PRP it VBZ expects PRP$ its NNP U.S. NNS sales TO to VB remain JJ steady IN at IN about CD 1,200 NNS cars IN in CD 1990 . .

DT The NN luxury NN auto NN maker JJ last NN year VBD sold CD 1,214 NNS cars IN in DT the NNP U.S.

Is POS tagging a solved problem?

- **Most frequent class:** Assign each word token to the tag with which it occurred most in the training set. (e.g. back/NN) gives 90% accuracy

- **State of the art:** 97% accuracy at word level

- Average English sentence ~ 14 words

  Sentence level accuracies: 0.97^14 = **65%**

POS tagging not solved yet!

Techniques of POS tagging

- Rule based approaches

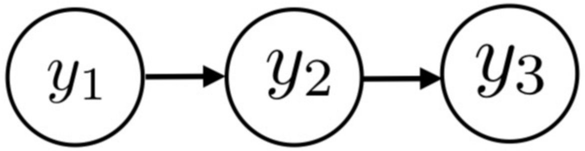
- Machine-learning methods – Hidden Markov Model

Classic Solution: Hidden Markov Model

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$

- Model the sequence of tags ***y*** over words ***x*** as a Markov process

- Markov property: future is conditionally independent of the past given the present



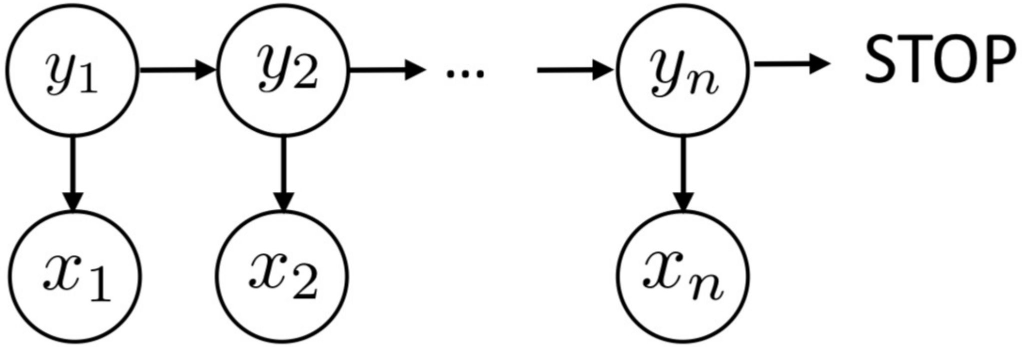
$$P(y_3|y_1, y_2) = P(y_3|y_2)$$

- If ***y*** are tags, this roughly corresponds to assuming that the next tag only depends on the current tag, not anything before

Classic Solution: Hidden Markov Model

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$

y ∈ T = set of possible tags (including STOP);
x ∈ V = vocab of words

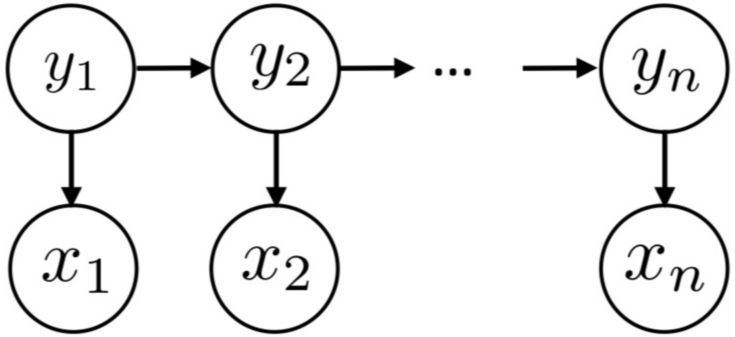


$$P(\mathbf{y},\mathbf{x}) = \underbrace{P(y_1)}_{\text{Initial distribution}} \underbrace{\prod_{i=2}^{n} P(y_i|y_{i-1})}_{\text{Transition probabilities}} \underbrace{\prod_{i=1}^{n} P(x_i|y_i)}_{\text{Emission probabilities}}$$

- Observation (*x*) depends only on current state (*y*)

Hidden Markov Model: Parameters

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$



$$P(\mathbf{y}, \mathbf{x}) = P(y_1) \prod_{i=2}^{n} P(y_i | y_{i-1}) \prod_{i=1}^{n} P(x_i | y_i)$$

- Initial distribution: |T| x 1 vector (distribution over initial states)
- Emission distribution: |T| x |V| matrix (distribution over words per tag)
- Transition distribution: |T| x |T| matrix (distribution over next tags per tag)

Learning

Learning

- Transitions
  - Count up all pairs (y_i, y_{i+1}) in the training data
  - Count up occurrences of what tag *T* can transition to
  - Normalize to get a distribution for P(next tag | *T*)
  - Need to *smooth* this distribution, won't discuss here
- Emissions: similar scheme, but trickier smoothing!

Decoding using Viterbi algorithm

- Dynamic Programming algorithm
- Intuition:
  - If I knew the best state sequence for words <o_1, …o_{n-1}>, then I can figure out the last state.
  - Decision would depend on s_{n-1}
  - So I only need the score of the best sequence up to n-1, ending in each possible state at n-1.
  - Ditto at every time step n-2, n-3, … 1

Decoding using Viterbi algorithm

- Given an HMM (A, B, P) and a sequence of observations, find the most probable sequence of tags

Summary

- Input: a sequence of word tokens **w**

- Output: a sequence of tags **t**, one per word (t∈ T)

- Why do POS tagging

- How to do POS tagging