



HOW TO TAG A TEXT SEQUENCE?

Sandeep Soni

10/29/2024

WORDS

They can fish

WORDS

They can fish

Is this sentence about the
ability to catch fish?

WORDS

They can fish

... Or is the sentence about
storing fish?

How do you differentiate between these two meanings?

PARTS OF SPEECH

Categories of words that have similar morphological and syntactic properties

PARTS OF SPEECH

Categories of words that have similar morphological and syntactic properties

They can fish

PARTS OF SPEECH

Categories of words that have similar morphological and syntactic properties

Noun Verb Noun

They can fish

PARTS OF SPEECH

Categories of words that have similar morphological and syntactic properties

Noun Verb Noun

They can fish

Noun Verb Verb

PARTS OF SPEECH

Categories of words that have similar morphological and syntactic properties

Noun Verb Noun

Store fish

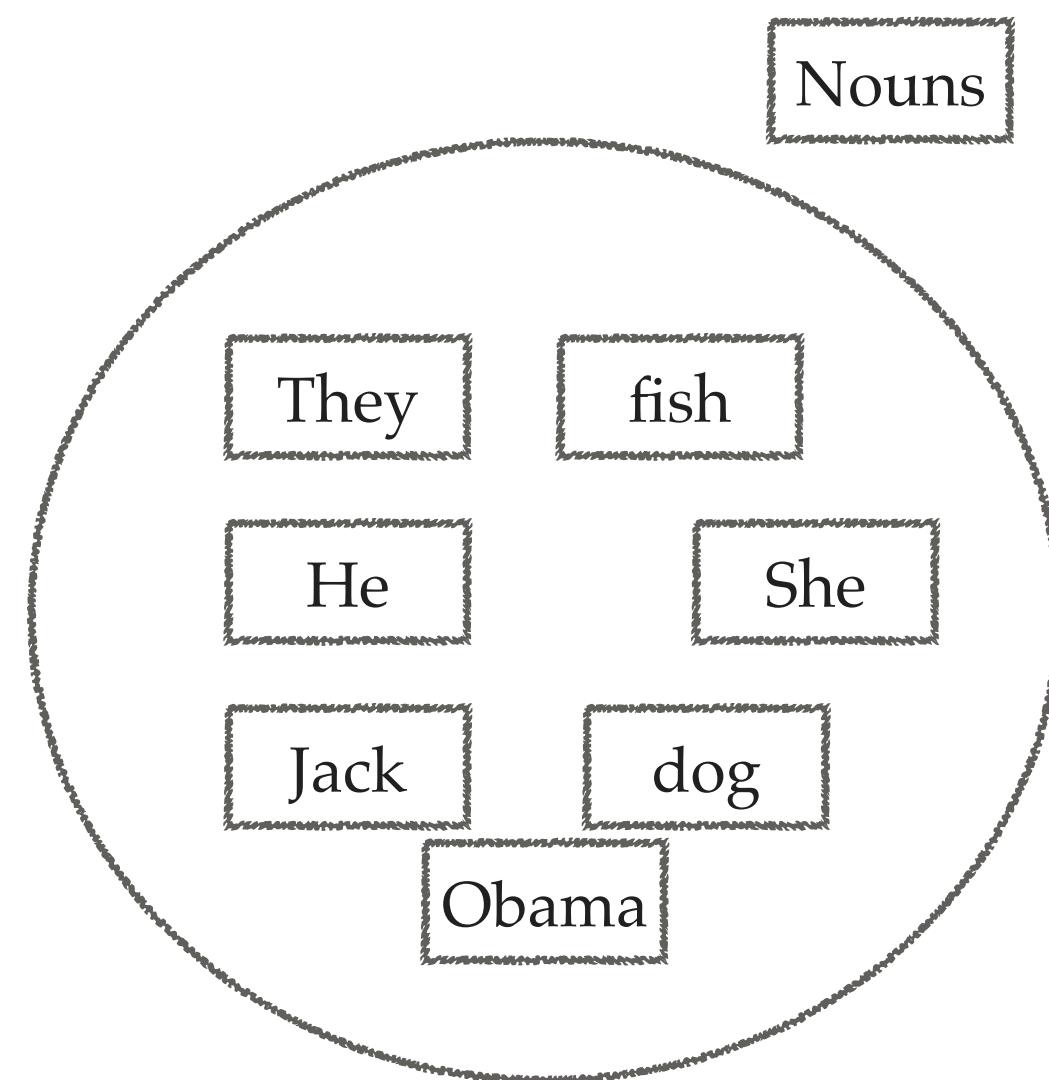
They can fish

Noun Verb Verb

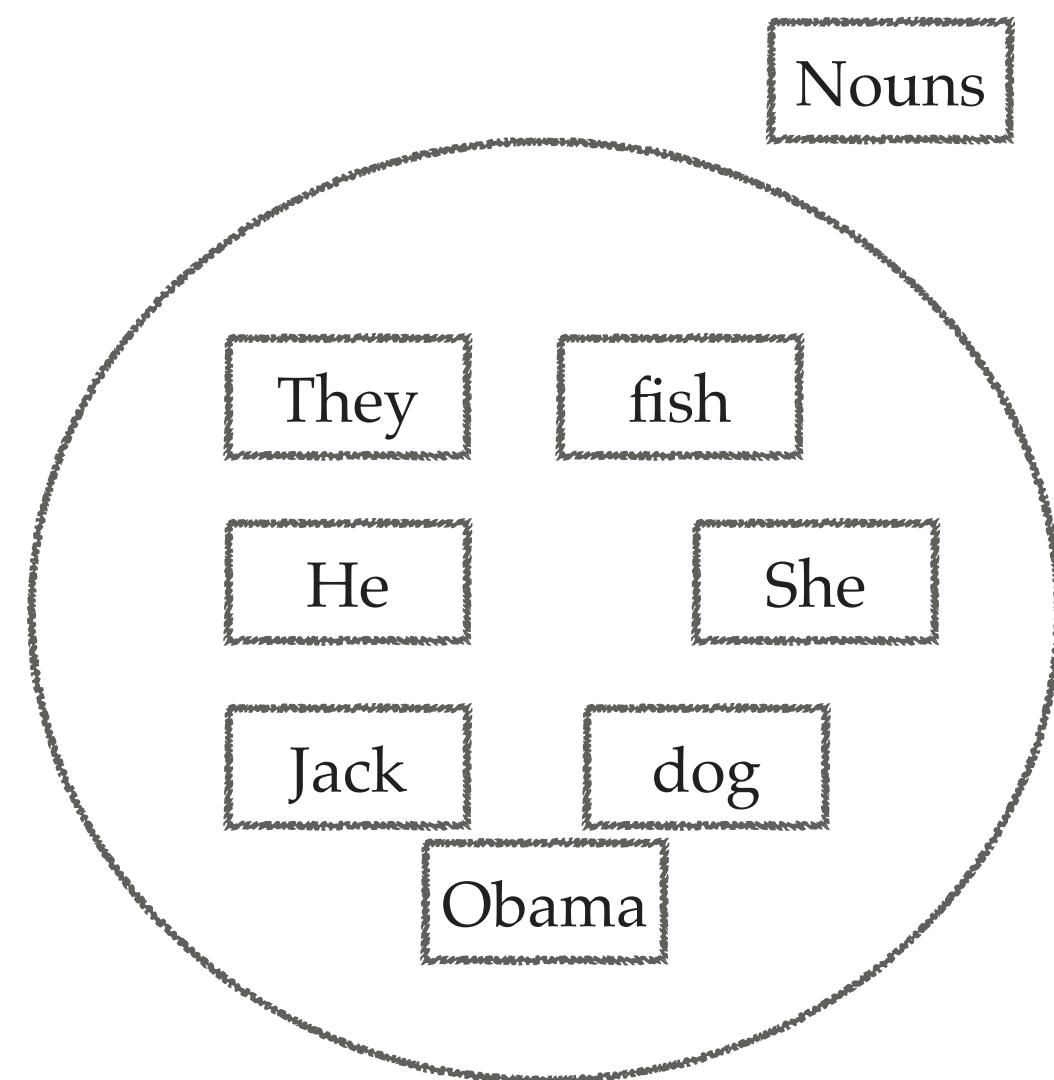
Catch fish

PoS AS DIMENSIONALITY REDUCTION

PoS AS DIMENSIONALITY REDUCTION



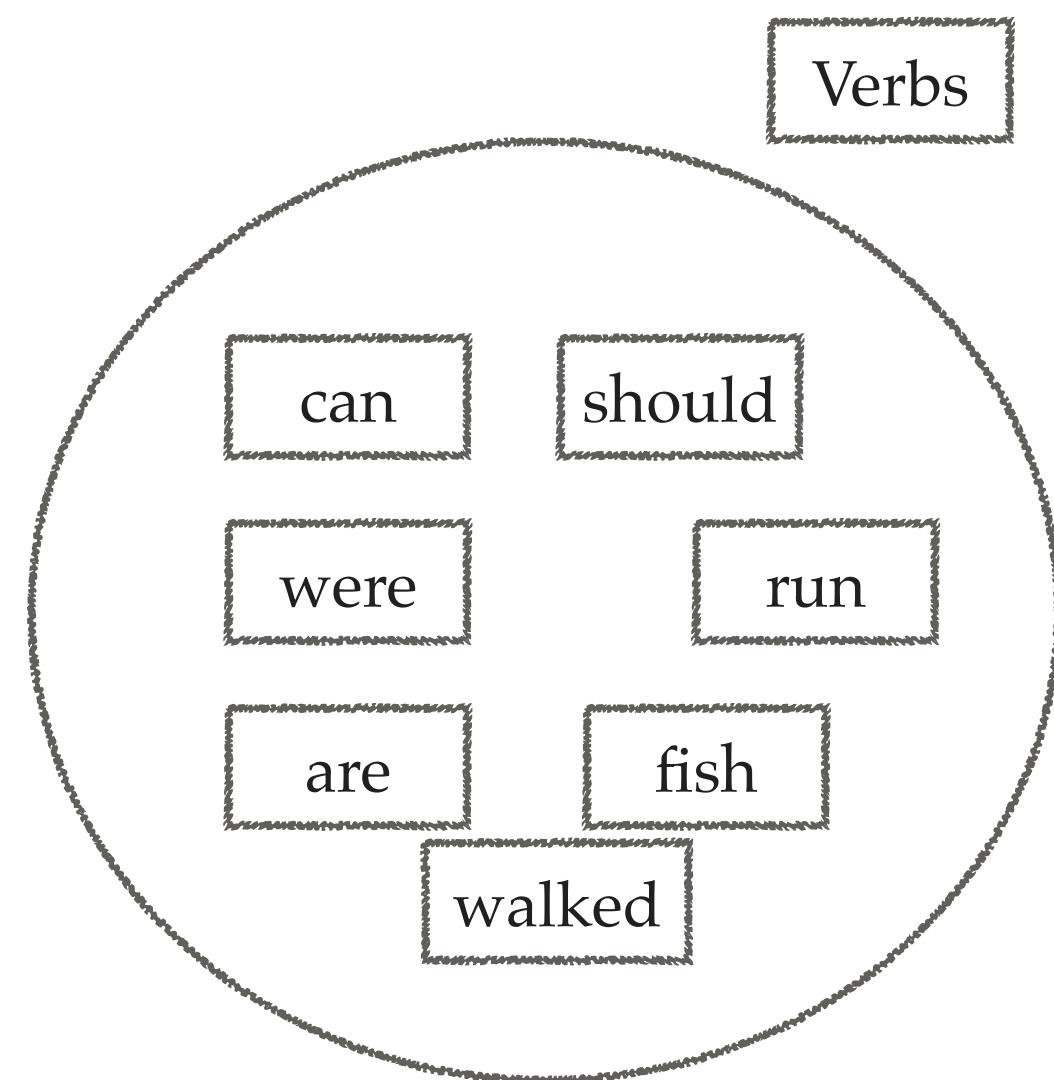
PoS AS DIMENSIONALITY REDUCTION



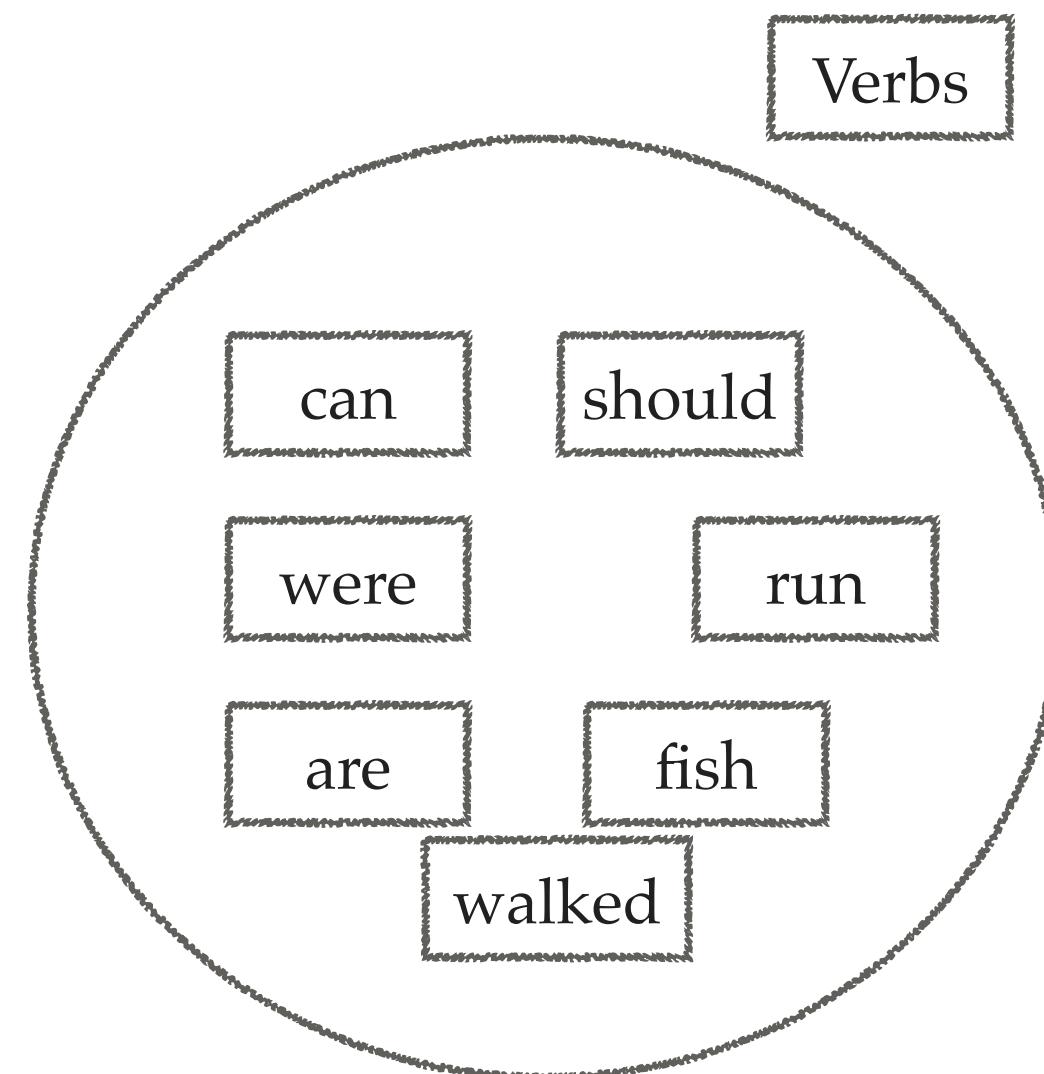
They

PoS AS DIMENSIONALITY REDUCTION

PoS AS DIMENSIONALITY REDUCTION



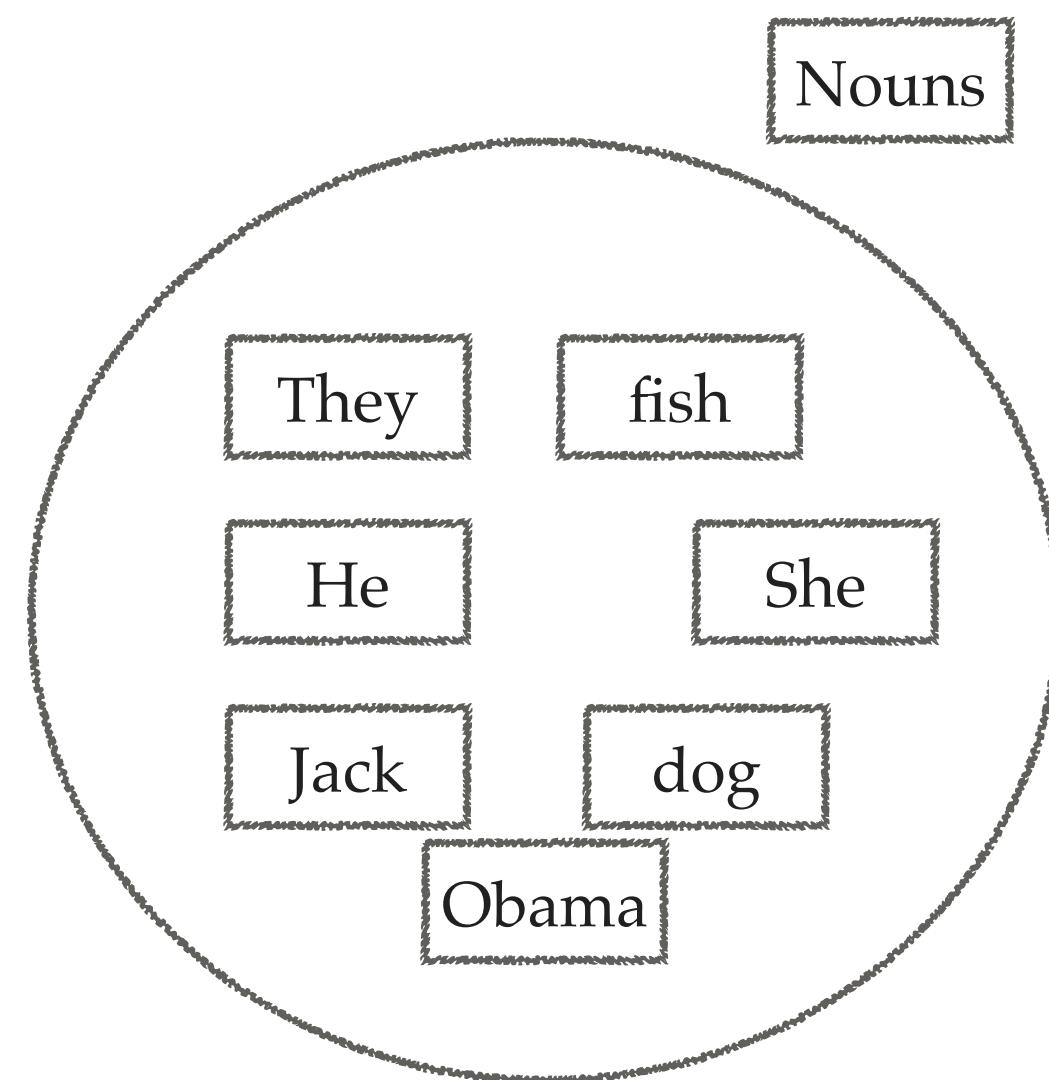
PoS AS DIMENSIONALITY REDUCTION



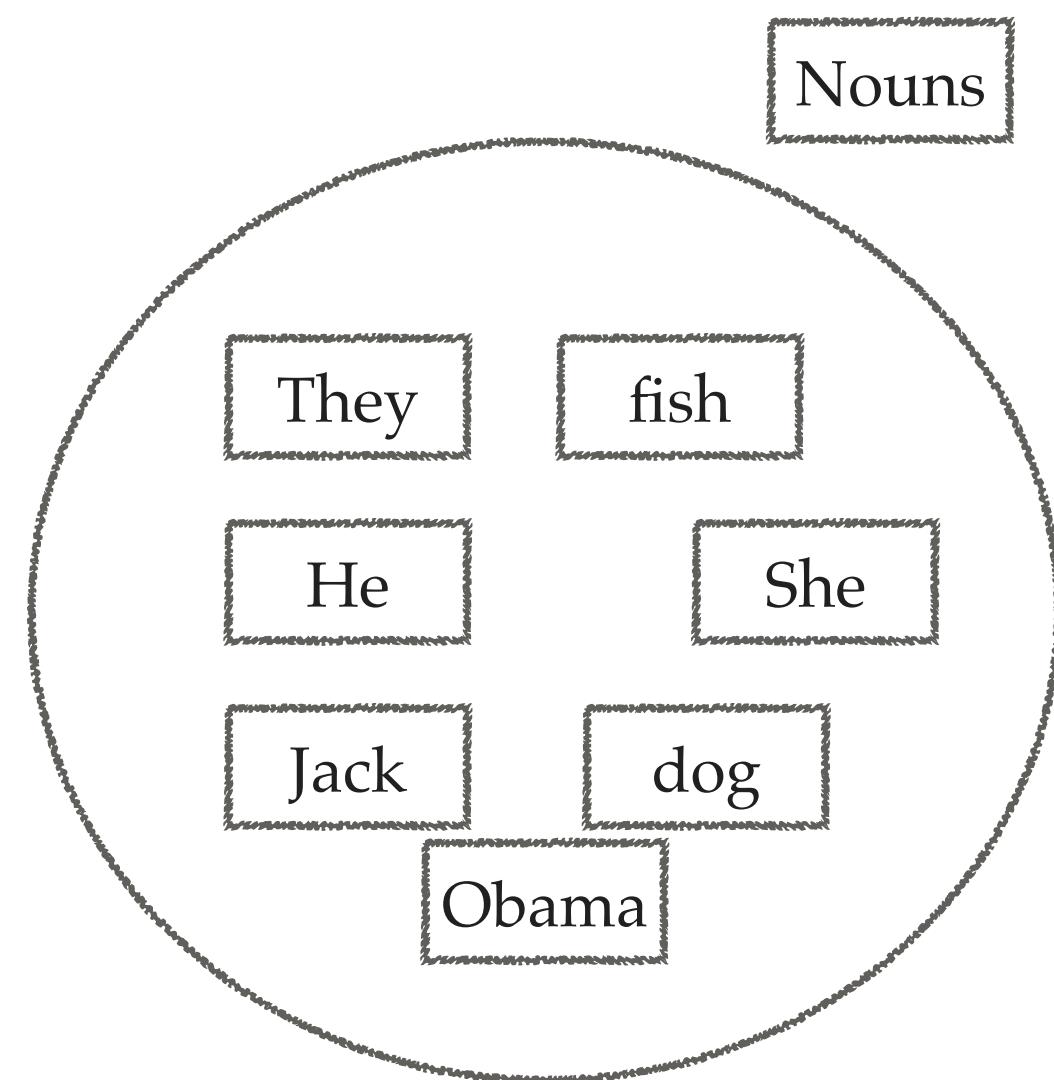
They can

PoS AS DIMENSIONALITY REDUCTION

PoS AS DIMENSIONALITY REDUCTION



PoS AS DIMENSIONALITY REDUCTION



They can fish

CONTEXTUAL DEPENDENCE

To find the syntactic role that a word has in a sentence,
more context is helpful

Verb Verb

Noun Noun Noun

They can fish

Noun Verb Noun

They can fish in a cold room

PART OF SPEECH TAGGING

The task of assigning a tag (category) to every word in a sequence.

PRP VB NN IN DT JJ NN

They can fish in a cold room

Classification

+ve

They can fish in a cold room

Classification

+ve

They can fish in a cold room

Part of Speech Tagging

PRP

VB

NN

IN

DT

JJ

NN

They can fish in a cold room

TAG SET

The number of categories depends on language, genre, etc although there are some core categories

PRP VB NN IN DT JJ NN

They can fish in a cold room

Nouns	People, places, things, dates, etc. Depend on quantity (singular Vs Plural)
Adjectives	Properties or qualities of noun phrases
Verbs	Actions, processes. Depends on tense, number, person, etc
Adverbs	Modifiers of verbs; qualify the action (e.g., quickly stopped)
Determiner	Beginning of a noun phrase; emphasizes specificity (e.g., a film)
Prepositions	Indicates spatial/direction/temporal relationship with a noun phrase (e.g., in a room)
Conjunctions	Connectives between phrases, clauses, sentences (e.g., Jack and Jill)
Pronouns	References to noun phrases (e.g., he, she, it)

Tag	Description	Examples	%
Nominal, Nominal + Verbal			
N	common noun (NN, NNS)	books someone	13.7
O	pronoun (personal/WH; not possessive; PRP, WP)	it you u meeee	6.8
S	nominal + possessive	books' someone's	0.1
[^]	proper noun (NNP, NNPS)	lebron usa iPad	6.4
Z	proper noun + possessive	America's	0.2
L	nominal + verbal (= <i>I don't know</i>)	he's book'll iono	1.6
M	proper noun + verbal	Mark'll	0.0
Other open-class words			
V	verb incl. copula, auxiliaries (V*, MD)	might gonna ought couldn't is eats	15.1
A	adjective (J*)	good fav lil	5.1
R	adverb (R*, WRB)	2 (i.e., <i>too</i>)	4.6
!	interjection (UH)	lol haha FTW yea right	2.6
Other closed-class words			
D	determiner (WDT, DT, WP\$, PRP\$)	the teh its it's	6.5
P	pre- or postposition, or subordinating conjunction (IN, TO)	while to for 2 (i.e., to) 4 (i.e., <i>for</i>)	8.7
&	coordinating conjunction (CC)	and n & + BUT	1.7
T	verb particle (RP)	out off Up UP	0.6
X	existential <i>there</i> , predeterminers (EX, PDT)	both	0.1
Y	X + verbal	there's all's	0.0
Twitter/online-specific			
#	hashtag (indicates topic/category for tweet)	#acl	1.0
@	at-mention (indicates another user as a recipient of a tweet)	@BarackObama	4.9
~	discourse marker, indications of continuation of a message across multiple tweets	RT and : in retweet construction RT @user : hello	3.4
U	URL or email address	http://bit.ly/xyz	1.6
E	emoticon	:-) :b (: <3 o_O	1.0

Miscellaneous			
\$	numeral (CD)	2010 four 9:30	1.5
,	punctuation (#, \$, ' ', (,)) , , . , :, ` `)	!!! ?!?	11.6
G	other abbreviations, foreign words, possessive endings, symbols, garbage (FW, POS, SYM, LS)	ily (<i>I love you</i>) wby (<i>what about you</i>) 's ♪ --> awesome...I'm	1.1

SEQUENCE LABELING

- Classification: Independently predict each word tag

The old man the boat

SEQUENCE LABELING

- Structured prediction: Predict the entire tag sequence

$$f([x_1, x_2, \dots, x_n]) \rightarrow [y_1, y_2, \dots, y_n]$$

Where can part of speech tagging be used?

Sarcasm as Contrast between a Positive Sentiment and Negative Situation

**Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva,
Nathan Gilbert, Ruihong Huang**

School Of Computing

University of Utah

Salt Lake City, UT 84112

{riloff,asheq,alnds,ngilbert,huangrh}@cs.utah.edu, prafulla.surve@gmail.com

- (a) Oh how I love *being ignored*. #sarcasm
- (b) Thoroughly enjoyed *shoveling the driveway* today! :) #sarcasm
- (c) Absolutely adore it when *my bus is late* #sarcasm
- (d) I'm so pleased mom *woke me up* with *vacuuming my room* this morning. :) #sarcasm

Identified positive sentiment phrases and negative situation phrases using part of speech tags to build a sarcasm recognizer

ANOTHER SEQUENCE LABELING TASK

Apple was founded on April 1, 1976, by Steve Jobs in California.

ANOTHER SEQUENCE LABELING TASK

Apple was founded on April 1, 1976, by Steve Jobs in California

NAMED ENTITIES

Apple was founded on April 1, 1976, by Steve Jobs in California

- Such proper nouns are called named entities
- Named entities are references to entities that can be grounded in reality

NAMED ENTITIES

Apple **ORG** was founded on April 1, 1976 **DATE**, by Steve Jobs **PERSON** in California **GPE**.

- Named entity recognition (NER) is the task of identifying spans of text that are named entities and categorizing them

How to cast NER into a sequence labeling task?

BIO NOTATION

BIO NOTATION

B Beginning of a named entity

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

BIO NOTATION

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O

Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O O
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O O B
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O O B I
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O O B I O
Apple was founded on April 1976, by Steve Jobs in California

BIO NOTATION

B	Beginning of a named entity
I	Inside of a named entity
O	Outside of the named entity

B O O O B I O O B I O B
Apple was founded on April 1976, by Steve Jobs in California

EXTENDED BIO NOTATION

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG
Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O O B-PER

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O O B-PER I-PER

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O O B-PER I-PER O

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

EXTENDED BIO NOTATION

B-CAT	Beginning of a named entity of category <i>CAT</i>
I-CAT	Inside of a named entity of category <i>CAT</i>
O	Outside of the named entity

B-ORG O O O B-DATE I-DATE O O B-PER I-PER O B-LOC

Apple was founded on April 1976, by Steve Jobs in California

- If there are n categories, then there are $2n+1$ BIO tags

If there are n PoS categories, how many PoS tags in the extended BIO notation?

NER TAGS

Person	Proper nouns refer to people or characters (e.g., Harry Potter, Sandeep Soni)	PER
Location	Refer to natural geographical regions (e.g., mountains) or structures built by humans (e.g., building)	LOC
GPE	Geo political entities such as cities, states, countries,etc (e.g., London)	GPE
Organization	Organization names, company names, sports teams, etc (e.g., The United Nations)	ORG

SPACY TAGS

CARDINAL	LAW	PERCENT
DATE	LOC	PERSON
EVENT	MONEY	PRODUCT
FAC	NORP	QUANTITY
GPE	ORDINAL	TIME
LANGUAGE	ORG	WORK_OF_ART

PERSON:	People, including fictional.
NORP:	Nationalities or religious or political groups.
FAC:	Buildings, airports, highways, bridges, etc.
ORG:	Companies, agencies, institutions, etc.
GPE:	Countries, cities, states.
LOC:	Non-GPE locations, mountain ranges, bodies of water.
PRODUCT:	Objects, vehicles, foods, etc. (Not services.)
EVENT:	Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART:	Titles of books, songs, etc.
LAW:	Named documents made into laws.
LANGUAGE:	Any named language.
DATE:	Absolute or relative dates or periods.
TIME:	Times smaller than a day.
PERCENT:	Percentage, including "%".
MONEY:	Monetary values, including unit.
QUANTITY:	Measurements, as of weight or distance.
ORDINAL:	"first", "second", etc.
CARDINAL:	Numerals that do not fall under another type.

FINE GRAINED NER

- Expanding the number of categories for named entities gives us more ways to distinguish between them

person	doctor engineer monarch musician politician religious_leader soldier terrorist	organization		terrorist_organization government_agency government political_party educational_department military news_agency
location	body_of_water city country county province railway road bridge	product	camera mobile_phone computer software game spacecraft instrument train	art film play
	island mountain glacier astral_body cemetery park			written_work newspaper music
				event military_conflict attack natural_disaster election sports_event protest terrorist_attack
building	time color award educational_degree title symptom drug body_part living_thing animal food	chemical_thing biological_thing medical_treatment disease symptom drug body_part living_thing animal food	website broadcast_network broadcast_program tv_channel currency stock_exchange algorithm programming_language transit_system transit_line	
airport dam hospital hotel library power_station restaurant sports_facility theater				

DOMAIN SPECIFIC NER

- The categories are highly dependent on the domain of interest
- Of course, the **black King** is out in the open, but the **white Bishop** is hemmed in by the **pawns**.

Tag	Meaning
Hu	Human
Tu	Turn
Po	Position
Pi	Piece
Ps	Piece specifier
Mc	Move compliment
Pa	Piece attribute
Pq	Piece quantity
Re	Region
Ph	Phase
St	Strategy
Ca	Castle
Me	Move eval.
Mn	Move name
Ee	Eval. element
Ev	Evaluation
Ti	Time
Ac	Player action
Ap	Piece action
Ao	Other action
Ot	Other notion

NESTED NER

- Typically NER assumes all categories are in a flat structure.
 - E.g., [Emory University]_{ORG}
- But many named entities have a nested structure
 - E.g., [[Emory]_{PER} University]_{ORG}

NAMED ENTITY RECOGNITION

- Many different ways to identify named entities
 - Max Entropy Models (e.g., Multinomial Logistic Regression)
 - Max Entropy Markov Models
 - Recurrent Neural networks
 - Transformer models

IN CLASS

- Sequence labeling demo