CS4705
Part of Speech tagging

Some slides adapted from: Dan Jurafsky, Julia Hirschberg, Jim Martin

HW questions?

- Training files, question samples
  - /home/cs4705/corpora/wsj
  - /home/cs4705/corpora/wsj/wsj_2300questions.txt
  - CVN: will post on the CVN web site this afternoon

- Question and answer templates

- Not expected to use tools that we haven't gone over (e.g., named entity recognition)
    - Must allow paraphrases for indices
    - But company names will be provided in question exactly as they appear in article

- Any other questions?

Smoothing

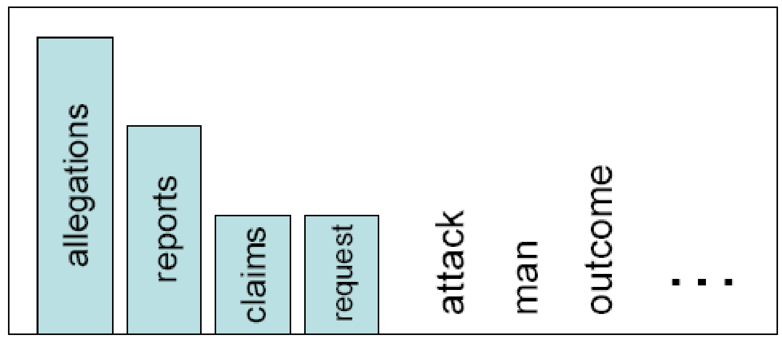
- Words follow a Zipfian distribution
  - Small number of words occur very frequently
  - A large number are seen only once
  - Zipf's law: a word's frequency is approximately inversely proportional to its rank in the word distribution list

- Zero probabilities on one bigram cause a zero probability on the entire sentence

Smoothing is like Robin Hood: Steal from the rich and give to the poor (in probability mass)

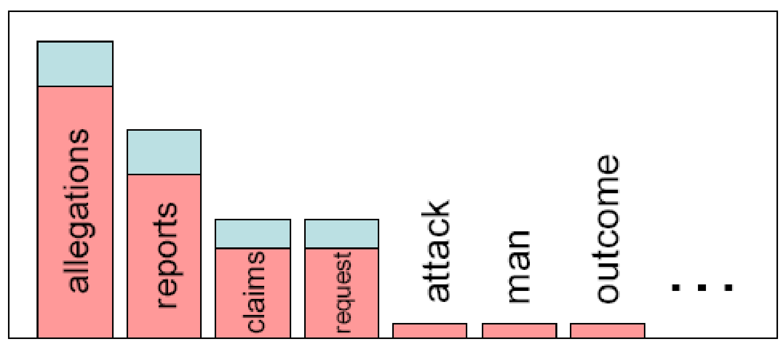
- We often want to make predictions from sparse statistics:

P(w | denied the)
3 allegations
2 reports
1 claims
1 request
7 total



- Smoothing flattens spiky distributions so they generalize better

P(w | denied the)
2.5 allegations
1.5 reports
0.5 claims
0.5 request
2 other
7 total



- Very important all over NLP, but easy to do badly!

Slide from Dan Klein

Smoothing Methods

- Add-one smoothing (easy, but inaccurate)
  - Add 1 to every word count (Note: this is type)
  - Increment normalization factor by Vocabulary size: N (tokens) + *V (types)* :

$$p_i^* = \frac{c_i + 1}{N + V}$$

- Backoff models
  - When a count for an n-gram is 0, back off to the count for the (n-1)-gram
  - These can be weighted
- Class-based smoothing
  - For certain types of n-grams, back off to the count of its syntactic class
  - E.g., Count ProperNouns in place of names (e.g., Obama)
- Good-Turing
  - Re-estimate amount of probability mass for zero (or low count) ngrams by looking at ngrams with higher counts
  - Estimate

$$c^* = (c+1)\frac{N_{c+1}}{N_c}$$

Garden path sentences

- The old dog the footsteps of the young.
- The cotton clothing is made of grows in Mississippi.
- The horse raced past the barn fell.

What is a word class?

- Words that somehow ‘behave’ alike:
  - Appear in similar contexts
  - Perform similar functions in sentences
  - Undergo similar transformations

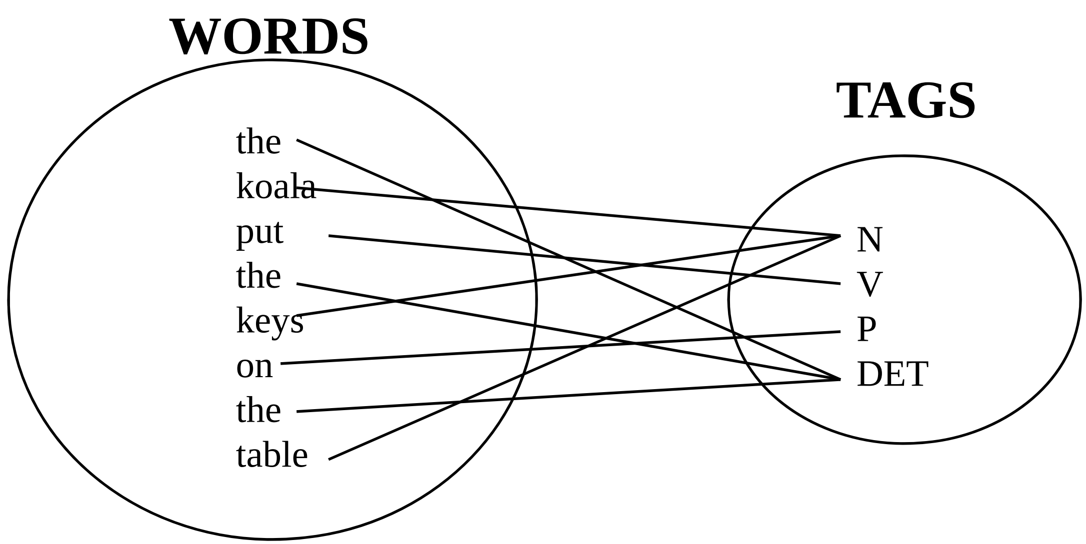
- 9 (or so) traditional parts of speech
  - Noun, verb, adjective, preposition, adverb, article, interjection, pronoun, conjunction,

POS examples

N	noun	chair, bandwidth, pacing
V	verb	study, debate, munch
ADJ	adjective	purple, tall, ridiculous
ADV	adverb	unfortunately, slowly,
P	preposition	of, by, to
PRO	pronoun	I, me, mine
DET	determiner	the, a, that, those

POS Tagging: Definition

- The process of assigning a part-of-speech or lexical class marker to each word in a corpus:



What is POS tagging good for?

- Is the first step of a vast number of Comp Ling tasks
- Speech synthesis:
  - How to pronounce "lead"?
  - INsult inSULT
  - OBject obJECT
  - OVERflow overFLOW
  - DIScount disCOUNT
  - CONtent conTENT
- Parsing
  - Need to know if a word is an N or V before you can parse
- Word prediction in speech recognition
  - Possessive pronouns (my, your, her) followed by nouns
  - Personal pronouns (I, you, he) likely to be followed by verbs
- Machine Translation

Open and closed class words

- Closed class: a relatively fixed membership
  - Prepositions: of, in, by, …
  - Auxiliaries: may, can, will had, been, …
  - Pronouns: I, you, she, mine, his, them, …
  - Usually function words (short common words which play a role in grammar)
- Open class: new ones can be created all the time
  - English has 4: Nouns, Verbs, Adjectives, Adverbs
  - Many languages have all 4, but not all!
  - In Lakhota and possibly Chinese, what English treats as adjectives act more like verbs.

Open class words

- Nouns
  - Proper nouns (Columbia University, New York City, Arthi Ramachandran, Metropolitan Transit Center). English capitalizes these.
  - Common nouns (the rest). German capitalizes these.
  - Count nouns and mass nouns
    - Count: have plurals, get counted: goat/goats, one goat, two goats
    - Mass: don't get counted (fish, salt, communism) (*two fishes)
- Adverbs: tend to modify things
  - Unfortunately, John walked home extremely slowly yesterday
  - Directional/locative adverbs (here, home, downhill)
  - Degree adverbs (extremely, very, somewhat)
  - Manner adverbs (slowly, slinkily, delicately)
- Verbs:
  - In English, have morphological affixes (eat/eats/eaten)
  - Actions (walk, ate) and states (be, exude)

- Many subclasses, e.g.
  - eats/V $\Rightarrow$ eat/VB, eat/VBP, eats/VBZ, ate/VBD, eaten/VBN, eating/VBG, ...
  - Reflect morphological form & syntactic function

How do we decide which words go in which classes?

- Nouns denote people, places and things and can be preceded by articles? But…
  - My typing is very bad.
  - *The Mary loves John.
- Verbs are used to refer to actions, processes, states
  - But some are closed class and some are open

I will have emailed everyone by noon.

  - Adverbs modify actions
  - Is Monday a temporal adverb or a noun? Some others?

Closed Class Words

- Idiosyncratic
- Closed class words (Prep, Det, Pron, Conj, Aux, Part, Num) are easier, since we can enumerate them….but
  - Part vs. Prep
    - George eats up his dinner/George eats his dinner up.
    - George eats up the street/*George eats the street up.
  - Articles come in 2 flavors: definite (the) and indefinite (a, an)

POS tagging: Choosing a tagset

- To do POS tagging, need to choose a standard set of tags to work with
- Could pick very coarse tagsets
  - N, V, Adj, Adv.
- Brown Corpus (Francis & Kucera ‘82), 1M words, 87 tags
- Penn Treebank: hand-annotated corpus of *Wall Street Journal*, 1M words, 45-46 tags
  - Commonly used
  - set is finer grained,
- Even more fine-grained tagsets exist

Penn TreeBank POS Tag set

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	*and, but, or*	SYM	Symbol	+,%, &
CD	Cardinal number	*one, two, three*	TO	"to"	*to*
DT	Determiner	*a, the*	UH	Interjection	*ah, oops*
EX	Existential 'there'	*there*	VB	Verb, base form	*eat*
FW	Foreign word	*mea culpa*	VBD	Verb, past tense	*ate*
IN	Preposition/sub-conj	*of, in, by*	VBG	Verb, gerund	*eating*
JJ	Adjective	*yellow*	VBN	Verb, past participle	*eaten*
JJR	Adj., comparative	*bigger*	VBP	Verb, non-3sg pres	*eat*
JJS	Adj., superlative	*wildest*	VBZ	Verb, 3sg pres	*eats*
LS	List item marker	*1, 2, One*	WDT	Wh-determiner	*which, that*
MD	Modal	*can, should*	WP	Wh-pronoun	*what, who*
NN	Noun, sing. or mass	*llama*	WP$	Possessive wh-	*whose*
NNS	Noun, plural	*llamas*	WRB	Wh-adverb	*how, where*
NNP	Proper noun, singular	*IBM*	$	Dollar sign	$
NNPS	Proper noun, plural	*Carolinas*	#	Pound sign	#
PDT	Predeterminer	*all, both*	"	Left quote	(' or ")
POS	Possessive ending	*'s*	"	Right quote	(' or ")
PRP	Personal pronoun	*I, you, he*	(	Left parenthesis	([, (, {, <)
PRP$	Possessive pronoun	*your, one's*	)	Right parenthesis	( ],), }, >)
RB	Adverb	*quickly, never*	,	Comma	,
RBR	Adverb, comparative	*faster*	.	Sentence-final punc	(. ! ?)
RBS	Adverb, superlative	*fastest*	:	Mid-sentence punc	(: ; ... – -)
RP	Particle	*up, off*			

Using the UPenn tagset

- The/DT grand/JJ jury/NN commmented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS ./.
- Prepositions and subordinating conjunctions marked IN (“although/IN I/PRP..”)
- Except the preposition/complementizer “to” is just marked “to”.

POS Tagging

- Words often have more than one POS: *back*
  - The *back* door = JJ
  - On my *back* = NN
  - Win the voters *back* = RB
  - Promised to *back* the bill = VB
- The POS tagging problem is to determine the POS tag for a particular instance of a word.

These examples from Dekang Lin

How do we assign POS tags to words in a sentence?

- Time flies like an arrow.
- Time/[V,N] flies/[V,N] like/[V,Prep] an/Det arrow/N
- Time/N flies/V like/Prep an/Det arrow/N
- Fruit/N flies/N like/V a/DET banana/N
- Fruit/N flies/V like/Prep a/DET banana/N
- The/Det flies/N like/V a/DET banana/N

How hard is POS tagging? Measuring ambiguity

		Original 87-tag corpus		Treebank 45-tag corpus	
Unambiguous (1 tag)		**44,019**		**38,857**	
Ambiguous (2–7 tags)		**5,490**		**8844**	
Details:	2 tags	4,967		6,731	
	3 tags	411		1621	
	4 tags	91		357	
	5 tags	17		90	
	6 tags	2	(*well*, *beat*)	32	
	7 tags	2	(*still*, *down*)	6	(*well*, *set*, *round*, *open*, *fit*, *down*)
	8 tags			4	(*'s*, *half*, *back*, *a*)
	9 tags			3	(*that*, *more*, *in*)

Potential Sources of Disambiguation

- Many words have only one POS tag (e.g. is, Mary, very, smallest)
- Others have a single most likely tag (e.g. a, dog)
- But tags also tend to co-occur regularly with other tags (e.g. Det, N)
- In addition to conditional probabilities of words $P(w_1|w_{n-1})$, we can look at POS likelihoods $P(t_1|t_{n-1})$ to disambiguate sentences and to assess sentence likelihoods

3 methods for POS tagging

1. Rule-based tagging
   - (ENGTWOL)
2. Transformation-based tagging
   1. Learned rules (statistic and linguistic)
   - Brill tagger
3. Stochastic (=Probabilistic) tagging
   - HMM (Hidden Markov Model) tagging

Rule-based tagging

- Start with a dictionary
- Assign all possible tags to words from the dictionary
- Write rules by hand to selectively remove tags
- Leaving the correct tag for each word.

Start with a dictionary

- she: PRP
- promised: VBN,VBD
- to TO
- back: VB, JJ, RB, NN
- the: DT
- bill: NN, VB

- Etc… for the ~100,000 words of English

Use the dictionary to assign every possible tag

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	**promised**	**to**	**back**	**the**	**bill**

Write rules to eliminate tags

Eliminate VBN if VBD is an option when VBN|VBD follows “<start> PRP”

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	**promised**	**to**	**back**	**the**	**bill**

Sample ENGTWOL Lexicon

Word	POS	Additional POS features
smaller	ADJ	COMPARATIVE
entire	ADJ	ABSOLUTE ATTRIBUTIVE
fast	ADV	SUPERLATIVE
that	DET	CENTRAL DEMONSTRATIVE SG
all	DET	PREDETERMINER SG/PL QUANTIFIER
dog's	N	GENITIVE SG
furniture	N	NOMINATIVE SG NOINDEFDETERMINER
one-third	NUM	SG
she	PRON	PERSONAL FEMININE NOMINATIVE SG3
show	V	IMPERATIVE VFIN
show	V	PRESENT -SG3 VFIN
show	N	NOMINATIVE SG
shown	PCP2	SVOO SVO SV
occurred	PCP2	SV
occurred	V	PAST VFIN SV

Stage 1 of ENGTWOL Tagging

- First Stage: Run words through FST morphological analyzer to get all parts of speech.
- Example: *Pavlov had shown that salivation …*

Pavlov	PAVLOV N NOM SG PROPER
had	HAVE V PAST VFIN SVO
	HAVE PCP2 SVO
shown	SHOW PCP2 SVOO SVO SV
that	ADV
	PRON DEM SG
	DET CENTRAL DEM SG
	CS
salivation	N NOM SG

Stage 2 of ENGTWOL Tagging

- Second Stage: Apply NEGATIVE constraints.
- Example: Adverbial “that” rule
  - Eliminates all readings of “that” except the one in
    - “It isn’t *that* odd”

Given **input**: “that”

If

(+1 A/ADV/QUANT) **;if next word is adj/adv/quantifier**

(+2 SENT-LIM) **;following which is E-O-S**

(NOT -1 SVOC/A) **; and the previous word is not a**

; verb like “consider” which

; allows adjective complements

; in “I consider that odd”

Then eliminate non-ADV tags

Else eliminate ADV

Transformation-Based Tagging (Brill Tagging)

- Combination of Rule-based and stochastic tagging methodologies
  - Like rule-based because rules are used to specify tags in a certain environment
  - Like stochastic approach because machine learning is used—with tagged corpus as input
    - *Rules are learned*
- Input:
  - tagged corpus
  - dictionary (*with most frequent tags*)

Transformation-Based Tagging

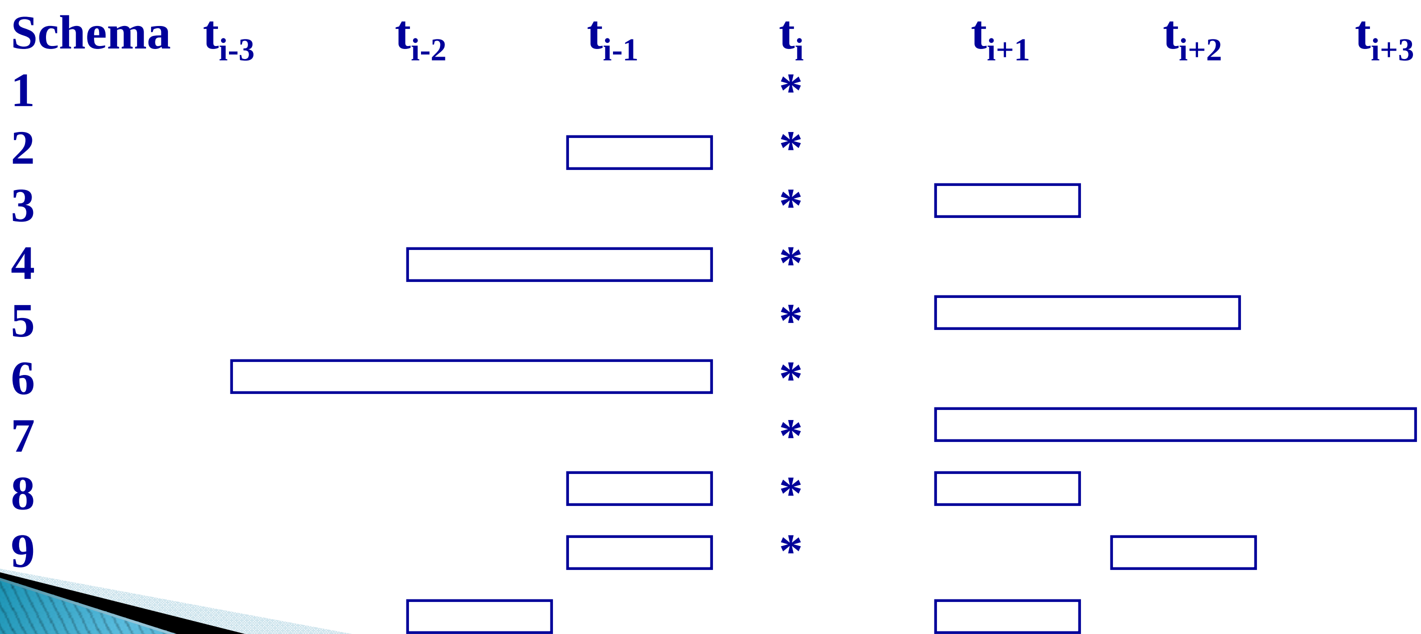
- Basic Idea:
  - Set the most probable tag for each word as a start value
  - Change tags according to rules of type “if word-1 is a determiner and word is a verb then change the tag to noun” in a specific order
- Training is done on tagged corpus:
  - Use a set of rule templates
  - Among the set of rules, find one with highest score
  - Continue finding rules until lowest score threshold is passed
  - Keep the ordered set of rules
- Rules make errors that are corrected by later rules

TBL Rule Application

- Tagger labels every word with its most-likely tag
  - For example: *race* has the following probabilities in the Brown corpus:
    - *P(NN|race) = .98*
    - *P(VB|race)= .02*
- Transformation rules make changes to tags
  - "Change NN to VB when previous tag is TO"
    … is/VBZ expected/VBN to/TO race/NN tomorrow/NN
    becomes
    … is/VBZ expected/VBN to/TO race/VB tomorrow/NN

TBL: Rule Learning

- 2 parts to a rule
  - Triggering environment
  - Rewrite rule
- The range of triggering environments of templates *(from Manning & Schutze 1999:363)*



TBL: The Tagging Algorithm

- Step 1: Label every word with most likely tag (from dictionary)
- Step 2: Check every possible transformation & select one which most improves tagging
- Step 3: Re-tag corpus applying the rules
- Repeat 2-3 until some criterion is reached, e.g., X% correct with respect to training corpus
- RESULT: Sequence of transformation rules

TBL: Rule Learning (cont.)

- Problem: Could apply transformations ad infinitum!
- Constrain the set of transformations with “templates”:
  - Replace tag X with tag Y, provided tag Z or word Z’ appears in some position
- Rules are learned in ordered sequence
- Rules may interact.
- Rules are compact and can be inspected by humans

Templates for TBL

The preceding (following) word is tagged **z**.
The word two before (after) is tagged **z**.
One of the two preceding (following) words is tagged **z**.
One of the three preceding (following) words is tagged **z**.
The preceding word is tagged **z** and the following word is tagged **w**.
The preceding (following) word is tagged **z** and the word
two before (after) is tagged **w**.

	Change tags			
#	From	To	Condition	Example
1	NN	VB	Previous tag is TO	to/TO race/NN → VB
2	VBP	VB	One of the previous 3 tags is MD	might/MD vanish/VBP → VB
3	NN	VB	One of the previous 2 tags is MD	might/MD not reply/NN → VB
4	VB	NN	One of the previous 2 tags is DT	
5	VBD	VBN	One of the previous 3 tags is VBZ	

Comparison of two approaches

- Accuracy
- Coverage
- Ease of building such a system
  - What is needed?
- Ease in porting to a new genre/new domain
- Baseline?

Summary

Parts of speech

- What's POS tagging good for anyhow?
- Tag sets
- Rule-based tagging
- Learning rules: statistical and linguistic
- Next time:
  - HMM Tagging