

Mieleni minun tefewi,  
Niwoni ajattelewī,  
Mieli ruweta runoille,  
Laatiua iaulamahan.  
5. Weli kulta weikkoseni,  
Kaunis kielikumppalini!

# Natural Language Processing

Info 159/259

Lecture 9: Parts of speech (Sept 21, 2017)

David Bamman, UC Berkeley

# Announcements

- My office hours: next Monday 9/25 10am-noon (not tomorrow!)

# Announcements

- NLP Seminar (talks by NLP researchers every 3 or so weeks). 4pm Monday, 202 South Hall  
<http://nlp.berkeley.edu>



# The Viral Texts Project

Mapping Networks of Reprinting in 19th-Century  
Newspapers and Magazines

- Next Monday 9/25, 4pm, 202 South Hall  
David Smith, Northeastern

“...In our Viral Texts project, for example, we have built models of reprinting for noisily-OCR’d nineteenth-century newspapers to trace the flow of news, literature, jokes, and anecdotes throughout the United States. ...”

# NLP Seminar

- For any talk in the NLP seminar this semester, feel free to write up 500-word review of the talk + ideas for how it can inspire your future work
- You can swap that grade for your lowest quiz/homework grade.

context

everyone likes \_\_\_\_\_

a bottle of \_\_\_\_\_

is on the table

\_\_\_\_\_ makes you drunk

a cocktail with \_\_\_\_\_

and seltzer

from last time

# Distribution

- Words that appear in similar contexts have similar representations (and similar meanings, by the distributional hypothesis).

# Parts of speech

- Parts of speech are categories of words defined **distributionally** by the morphological and syntactic contexts a word appears in.

# Morphological distribution

POS often defined by distributional properties; verbs  
= the class of words that each combine with the  
same set of affixes

|         | -s       | -ed      | -ing      |
|---------|----------|----------|-----------|
| walk    | walks    | walked   | walking   |
| slice   | slices   | sliced   | slicing   |
| believe | believes | believed | believing |
| of      | *ofs     | *ofed    | *ofing    |
| red     | *reds    | *redded  | *reding   |

# Morphological distribution

We can look to the function of the affix (denoting past tense) to include irregular inflections.

|       | -s     | -ed    | -ing     |
|-------|--------|--------|----------|
| walk  | walks  | walked | walking  |
| sleep | sleeps | slept  | sleeping |
| eat   | eats   | ate    | eating   |
| give  | gives  | gave   | giving   |

# Syntactic distribution

- Substitution test: if a word is replaced by another word, does the sentence remain **grammatical**?

Kim saw the

elephant

before we did

dog

idea

\*of

\*goes

# Syntactic distribution

- These can often be too strict; some contexts admit substitutability for some pairs but not others.

Kim saw the

elephant

before we did

\*Sandy

both nouns but  
common vs. proper

Kim \*arrived the

elephant

before we did

both verbs but  
transitive vs. intransitive

|              |                                                                                                       |
|--------------|-------------------------------------------------------------------------------------------------------|
| Nouns        | People, places, things, actions-made-nouns (“I like <b>swimming</b> ”). Inflected for singular/plural |
| Verbs        | Actions, processes. Inflected for tense, aspect, number, person                                       |
| Adjectives   | Properties, qualities. Usually modify nouns                                                           |
| Adverbs      | Qualify the manner of verbs (“She ran <b>downhill extremely quickly yesteray</b> ”)                   |
| Determiner   | Mark the beginning of a noun phrase (“ <b>a</b> dog”)                                                 |
| Pronouns     | Refer to a noun phrase (he, she, it)                                                                  |
| Prepositions | Indicate spatial/temporal relationships ( <b>on</b> the table)                                        |
| Conjunctions | Conjoin two phrases, clauses, sentences (and, or)                                                     |

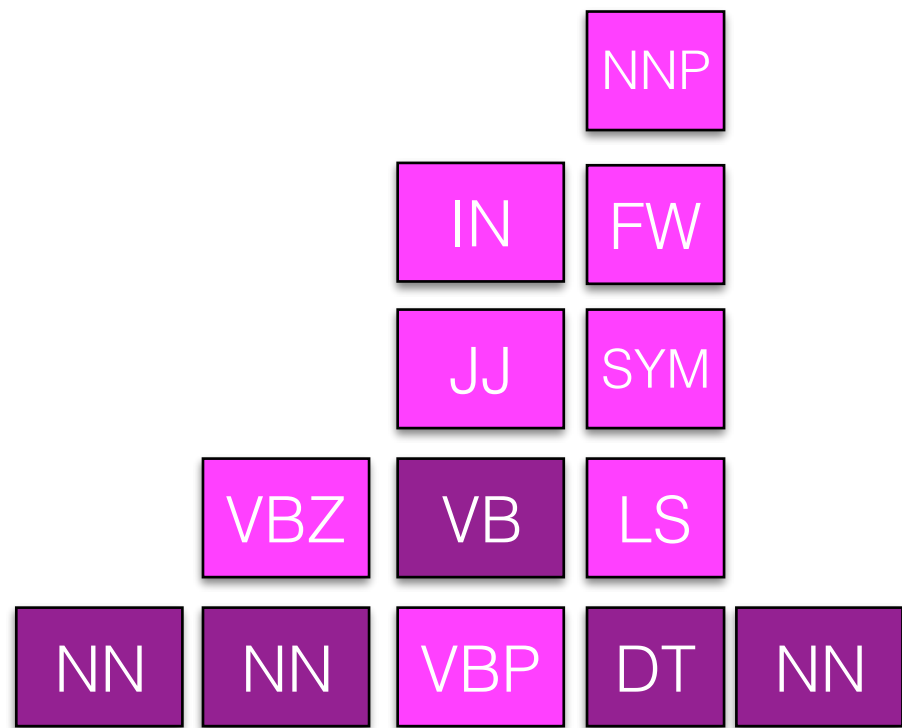
Open class

|              |                                                                                     |
|--------------|-------------------------------------------------------------------------------------|
| Nouns        | fax, affluenza, subtweet, bitcoin, cronut, emoji, listicle, mocktail, selfie, skort |
| Verbs        | text, chillax, manspreading, photobomb, unfollow, google                            |
| Adjectives   | crunk, amazeballs, post-truth, woke                                                 |
| Adverbs      | hella, wicked                                                                       |
| Determiner   | OOV? Guess Noun                                                                     |
| Pronouns     |                                                                                     |
| Prepositions | English has a new preposition, because internet<br>[Garber 2013; Pullum 2014]       |
| Conjunctions |                                                                                     |

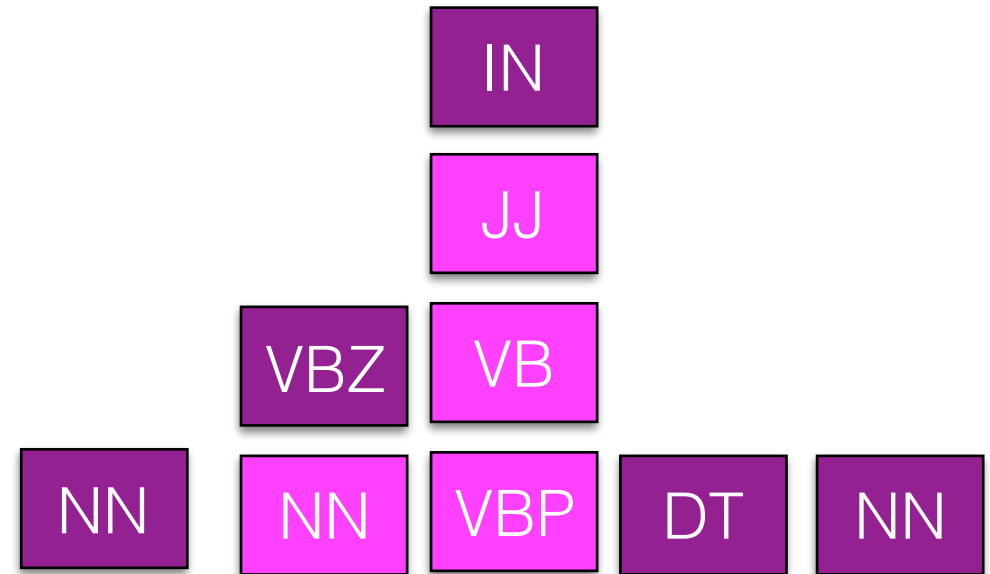
Closed class

# POS tagging

Labeling the tag that's correct  
for the context.



Fruit flies like a banana



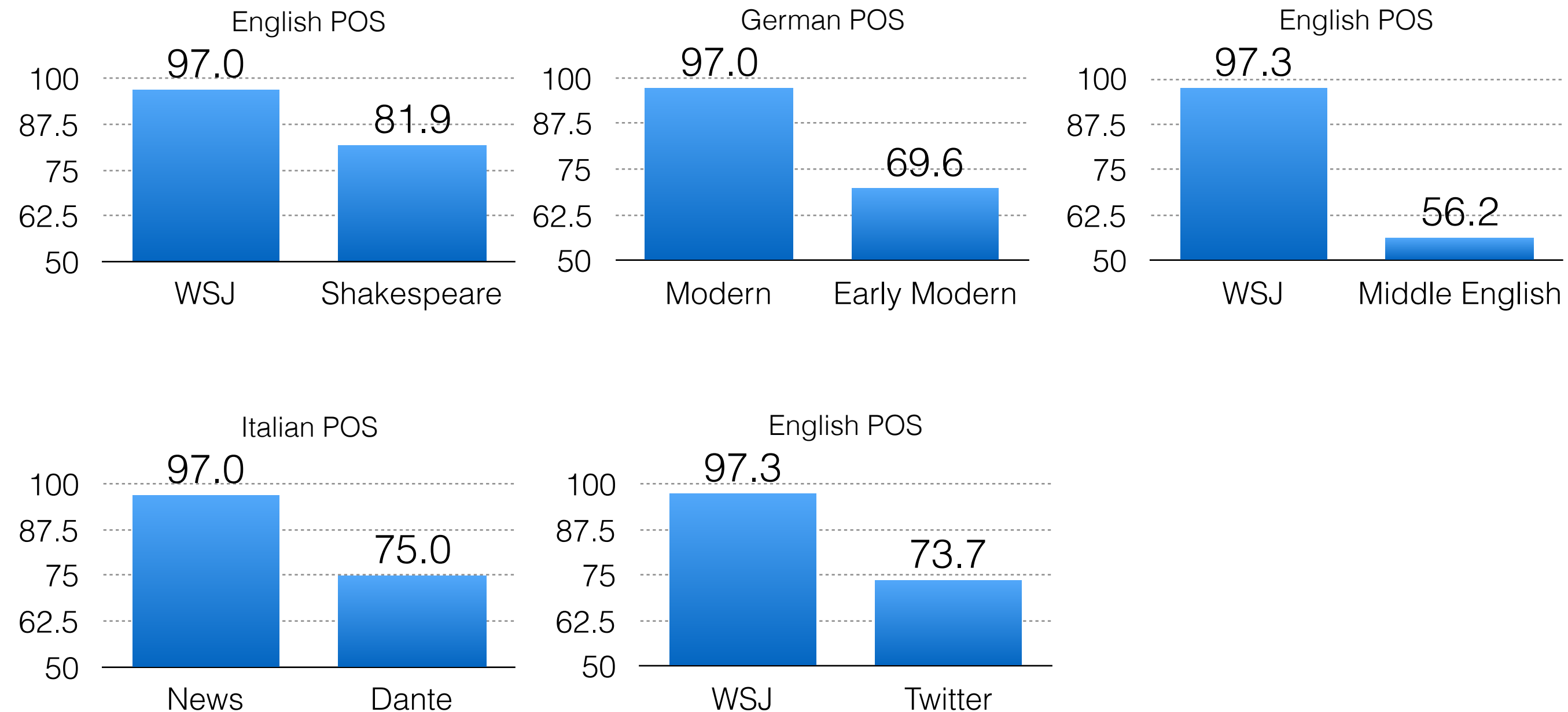
# Time flies like an arrow

(Just tags in evidence within the Penn Treebank — more are possible!)

# State of the art

- Baseline: Most frequent class = 92.34%
- Token accuracy: 97% (English news)  
[Toutanova et al. 2003; Søgaard 2010]
  - Optimistic: includes punctuation, words with only one tag (deterministic tagging)
  - Substantial drop across domains (e.g., train on news, test on literature)
- Whole sentence accuracy: 55%

# Domain difference

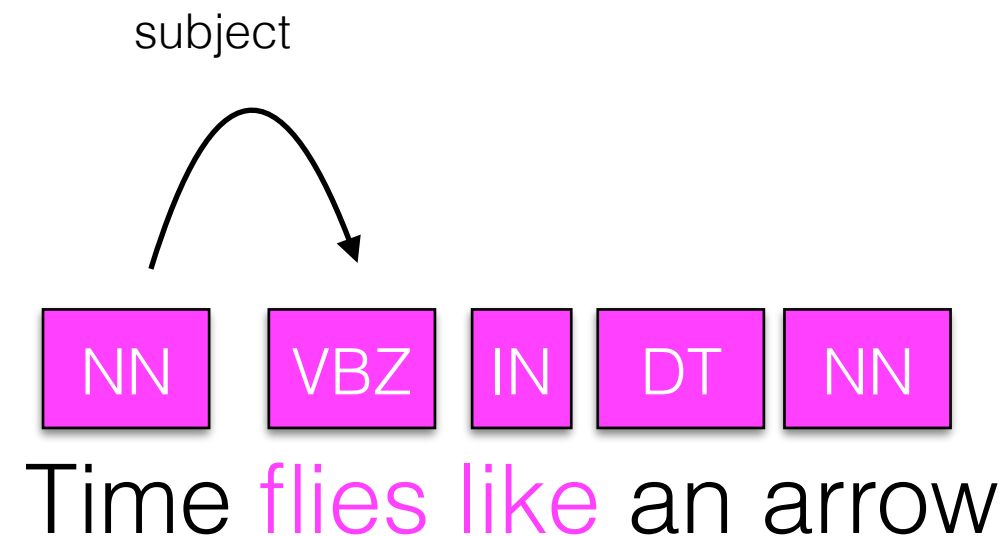
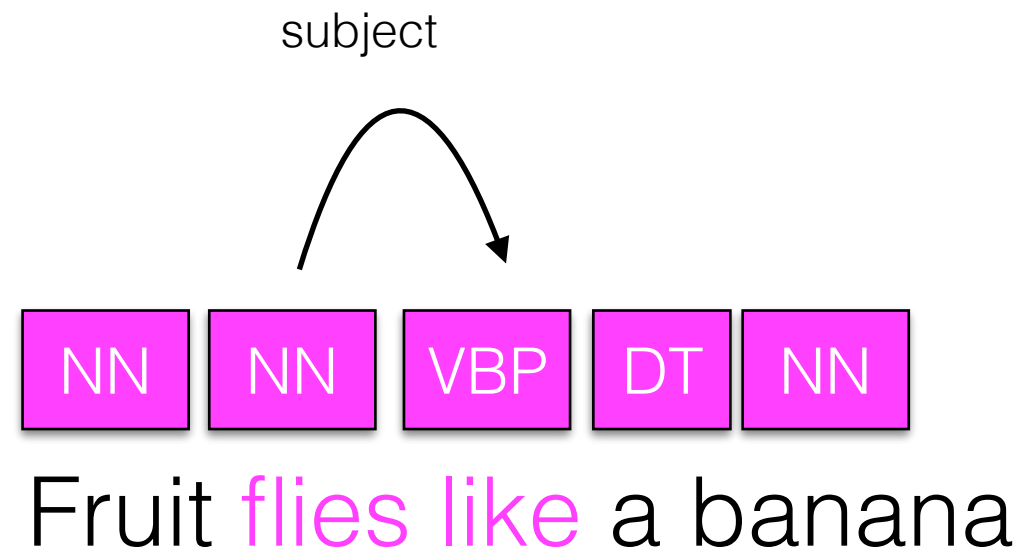


# Sources of error

|                           |       |                                                                               |
|---------------------------|-------|-------------------------------------------------------------------------------|
| Lexicon gap               | 4.5%  | a 60% slash/NN the common stock dividend                                      |
| Unknown word              | 4.5%  | blaming the disaster on substandard/JJ construction                           |
| Could plausibly get right | 16.0% | market players overnight/RB in Tokyo began bidding up oil prices              |
| Difficult linguistics     | 19.5% | They set/VBP up absurd situations, detached from reality                      |
| Underspecified/unclear    | 12.0% | it will take a \$ 10 million fourth-quarter charge against/IN discontinued/JJ |
| Inconsistent/no standard  | 28.0% | Orson Welles 's Mercury Theater in the '30s/NNS                               |
| Gold standard wrong       | 15.5% | Our market got hit/VB a lot harder on Monday than the listed market           |

Why is part of speech tagging useful?

# POS indicative of syntax



# POS indicative of MWE

at least one adjective/noun or noun phrase

and definitely  
one noun

$$((A \mid N)^+ \mid ((A \mid N)^*(NP))(A \mid N)^*)N$$

*AN*: linear function; lexical ambiguity; mobile phase

*NN*: regression coefficients; word sense; surface area

*AAN*: Gaussian random variable; lexical conceptual paradigm; aqueous mobile phase

*ANN*: cumulative distribution function; lexical ambiguity resolution; accessible surface area

*NAN*: mean squared error; domain independent set; silica based packing

*NNN*: class probability function; text analysis system; gradient elution chromatography

*NPN*: degrees of freedom; [*no example*]; energy of adsorption

# POS is indicative of pronunciation

| Noun                  | Verb                  |
|-----------------------|-----------------------|
| My conduct is great   | I conduct myself well |
| She won the contest   | I contest the ticket  |
| He is my escort       | He escorted me        |
| That is an insult     | Don't insult me       |
| Rebel without a cause | He likes to rebel     |
| He is a suspect       | I suspect him         |

# Tagsets

- Penn Treebank
- Universal Dependencies
- Twitter POS

# Homework 3

- Annotate ~1000 words of text using the Penn Treebank tags
- You'll be correcting the output of a tagger with ~92% accuracy (→ you should be making ~80 corrections)

| Tag   | Description          | Example               | Tag  | Description          | Example              |
|-------|----------------------|-----------------------|------|----------------------|----------------------|
| CC    | coordin. conjunction | <i>and, but, or</i>   | SYM  | symbol               | <i>+, %, &amp;</i>   |
| CD    | cardinal number      | <i>one, two</i>       | TO   | "to"                 | <i>to</i>            |
| DT    | determiner           | <i>a, the</i>         | UH   | interjection         | <i>ah, oops</i>      |
| EX    | existential 'there'  | <i>there</i>          | VB   | verb base form       | <i>eat</i>           |
| FW    | foreign word         | <i>mea culpa</i>      | VBD  | verb past tense      | <i>ate</i>           |
| IN    | preposition/sub-conj | <i>of, in, by</i>     | VBG  | verb gerund          | <i>eating</i>        |
| JJ    | adjective            | <i>yellow</i>         | VBN  | verb past participle | <i>eaten</i>         |
| JJR   | adj., comparative    | <i>bigger</i>         | VBP  | verb non-3sg pres    | <i>eat</i>           |
| JJS   | adj., superlative    | <i>wildest</i>        | VBZ  | verb 3sg pres        | <i>eats</i>          |
| LS    | list item marker     | <i>1, 2, One</i>      | WDT  | wh-determiner        | <i>which, that</i>   |
| MD    | modal                | <i>can, should</i>    | WP   | wh-pronoun           | <i>what, who</i>     |
| NN    | noun, sing. or mass  | <i>llama</i>          | WP\$ | possessive wh-       | <i>whose</i>         |
| NNS   | noun, plural         | <i>llamas</i>         | WRB  | wh-adverb            | <i>how, where</i>    |
| NNP   | proper noun, sing.   | <i>IBM</i>            | \$   | dollar sign          | <i>\$</i>            |
| NNPS  | proper noun, plural  | <i>Carolinas</i>      | #    | pound sign           | <i>#</i>             |
| PDT   | predeterminer        | <i>all, both</i>      | "    | left quote           | <i>' or "</i>        |
| POS   | possessive ending    | <i>'s</i>             | "    | right quote          | <i>' or "</i>        |
| PRP   | personal pronoun     | <i>I, you, he</i>     | (    | left parenthesis     | <i>[, (, {, &lt;</i> |
| PRP\$ | possessive pronoun   | <i>your, one's</i>    | )    | right parenthesis    | <i>], ), }, &gt;</i> |
| RB    | adverb               | <i>quickly, never</i> | ,    | comma                | <i>,</i>             |
| RBR   | adverb, comparative  | <i>faster</i>         | .    | sentence-final punc  | <i>. ! ?</i>         |
| RBS   | adverb, superlative  | <i>fastest</i>        | :    | mid-sentence punc    | <i>: ; ... --</i>    |
| RP    | particle             | <i>up, off</i>        |      |                      |                      |

# Homework 3

- What features are you using as as human to assign the correct tag?

| Tag   | Description          | Example               | Tag  | Description          | Example              |
|-------|----------------------|-----------------------|------|----------------------|----------------------|
| CC    | coordin. conjunction | <i>and, but, or</i>   | SYM  | symbol               | <i>+, %, &amp;</i>   |
| CD    | cardinal number      | <i>one, two</i>       | TO   | “to”                 | <i>to</i>            |
| DT    | determiner           | <i>a, the</i>         | UH   | interjection         | <i>ah, oops</i>      |
| EX    | existential ‘there’  | <i>there</i>          | VB   | verb base form       | <i>eat</i>           |
| FW    | foreign word         | <i>mea culpa</i>      | VBD  | verb past tense      | <i>ate</i>           |
| IN    | preposition/sub-conj | <i>of, in, by</i>     | VBG  | verb gerund          | <i>eating</i>        |
| JJ    | adjective            | <i>yellow</i>         | VCN  | verb past participle | <i>eaten</i>         |
| JJR   | adj., comparative    | <i>bigger</i>         | VBP  | verb non-3sg pres    | <i>eat</i>           |
| JJS   | adj., superlative    | <i>wildest</i>        | VBZ  | verb 3sg pres        | <i>eats</i>          |
| LS    | list item marker     | <i>1, 2, One</i>      | WDT  | wh-determiner        | <i>which, that</i>   |
| MD    | modal                | <i>can, should</i>    | WP   | wh-pronoun           | <i>what, who</i>     |
| NN    | noun, sing. or mass  | <i>llama</i>          | WP\$ | possessive wh-       | <i>whose</i>         |
| NNS   | noun, plural         | <i>llamas</i>         | WRB  | wh-adverb            | <i>how, where</i>    |
| NNP   | proper noun, sing.   | <i>IBM</i>            | \$   | dollar sign          | <i>\$</i>            |
| NNPS  | proper noun, plural  | <i>Carolinas</i>      | #    | pound sign           | <i>#</i>             |
| PDT   | predeterminer        | <i>all, both</i>      | “    | left quote           | <i>‘ or “</i>        |
| POS   | possessive ending    | <i>’s</i>             | ”    | right quote          | <i>’ or ”</i>        |
| PRP   | personal pronoun     | <i>I, you, he</i>     | (    | left parenthesis     | <i>[, (, {, &lt;</i> |
| PRP\$ | possessive pronoun   | <i>your, one’s</i>    | )    | right parenthesis    | <i>], ), }, &gt;</i> |
| RB    | adverb               | <i>quickly, never</i> | ,    | comma                | <i>,</i>             |
| RBR   | adverb, comparative  | <i>faster</i>         | .    | sentence-final punc  | <i>. ! ?</i>         |
| RBS   | adverb, superlative  | <i>fastest</i>        | :    | mid-sentence punc    | <i>: ; ... --</i>    |
| RP    | particle             | <i>up, off</i>        |      |                      |                      |

# Verbs

| tag | description            | example               |
|-----|------------------------|-----------------------|
| VB  | base form              | I want to like        |
| VBD | past tense             | I/we/he/she/you liked |
| VBG | present participle     | He was liking it      |
| VBN | past participle        | I had liked it        |
| VBP | present (non 3rd-sing) | I like it             |
| VBZ | present (3rd-sing)     | He likes it           |
| MD  | modal verbs            | He can go             |

# VB

## (verb, base form)

- The base form of verbs, found in imperatives, infinitives and subjunctives

- Just do it
- You should do it
- He wants to do it

|      |         |
|------|---------|
| 5031 | be/vb   |
| 1491 | have/vb |
| 669  | make/vb |
| 558  | sell/vb |
| 554  | buy/vb  |
| 534  | get/vb  |
| 518  | take/vb |
| 458  | do/vb   |
| 372  | pay/vb  |
| 325  | see/vb  |

# VBD

## (verb, past tense)

- Verbs used in the past tense

- He ate the food

|      |              |
|------|--------------|
| 7806 | said/vbd     |
| 5456 | was/vbd      |
| 2682 | were/vbd     |
| 2367 | had/vbd      |
| 876  | rose/vbd     |
| 834  | did/vbd      |
| 594  | fell/vbd     |
| 394  | reported/vbd |
| 392  | closed/vbd   |
| 384  | added/vbd    |

# VBG

## (verb, gerund)

- Verb forms in the gerund or present participle; generally end in **-ing**.
  - He was **going** to the store

|     |               |
|-----|---------------|
| 573 | including/vbg |
| 545 | being/vbg     |
| 543 | according/vbg |
| 412 | going/vbg     |
| 302 | making/vbg    |
| 268 | trying/vbg    |
| 250 | selling/vbg   |
| 236 | buying/vbg    |
| 213 | getting/vbg   |
| 205 | operating/vbg |

# VCN

## (verb, past participle)

- Verb form in the past participle

- The apple was eaten
- He had expected to go

|      |              |
|------|--------------|
| 2156 | been/vbn     |
| 643  | expected/vbn |
| 435  | made/vbn     |
| 435  | based/vbn    |
| 367  | compared/vbn |
| 356  | used/vbn     |
| 344  | so1d/vbn     |
| 267  | priced/vbn   |
| 229  | named/vbn    |
| 211  | held/vbn     |

# VBP

## (verb, non-3sg pres)

- Present tense of verbs, excluding the 3rd-person

- I **am** tall
- You **are** tall
- We **are** tall
- I **like** ice cream
- You **like** ice cream
- We **like** ice cream

|      |             |
|------|-------------|
| 4920 | are/vbp     |
| 2621 | have/vbp    |
| 838  | do/vbp      |
| 722  | say/vbp     |
| 460  | 're/vbp     |
| 272  | think/vbp   |
| 243  | want/vbp    |
| 227  | 've/vbp     |
| 170  | include/vbp |
| 166  | expect/vbp  |

# VBZ

## (verb 3sg pres)

- Present tense of verbs, only the 3rd-person
  - he **is** tall
  - he **likes** ice cream

|      |             |
|------|-------------|
| 9328 | is/vbz      |
| 4368 | has/vbz     |
| 2675 | says/vbz    |
| 1623 | 's/vbz      |
| 663  | does/vbz    |
| 341  | expects/vbz |
| 225  | plans/vbz   |
| 225  | makes/vbz   |
| 178  | remains/vbz |
| 167  | owns/vbz    |

# MD

## (Modal verb)

- All verbs that don't take -s ending in third-person singular present
- can, could, dare, may, might, must, ought, shall, should, will, would

|      |           |
|------|-----------|
| 4057 | will/md   |
| 2973 | would/md  |
| 1483 | could/md  |
| 1233 | can/md    |
| 1066 | may/md    |
| 598  | should/md |
| 459  | might/md  |
| 332  | must/md   |
| 326  | wo/md     |
| 246  | ca/md     |

# RP (particle)

- Used in combination with a verb
  - she turned the paper over
- verb + particle = phrasal verb, often non-compositional
  - turn down, rule out, find out, go on

|     |           |
|-----|-----------|
| 774 | up/rp     |
| 487 | out/rp    |
| 301 | off/rp    |
| 209 | down/rp   |
| 124 | in/rp     |
| 98  | over/rp   |
| 81  | on/rp     |
| 72  | back/rp   |
| 46  | around/rp |
| 25  | away/rp   |

# Nouns

non-proper

proper

| tag  | description                     | example       |
|------|---------------------------------|---------------|
| NN   | non-proper, singular<br>or mass | the company   |
| NNS  | non-proper, plural              | the companies |
| NNP  | proper, singular                | Carolina      |
| NNPS | proper, plural                  | Carolinas     |

# DT (Article)

- Articles (a, the, every, no)
- Indefinite determiners (another, any, some, each)
- That, these, this, those when preceding noun
- All, both when not preceding another determiner or possessive pronoun

|       |          |
|-------|----------|
| 65548 | the/dt   |
| 26970 | a/dt     |
| 4405  | an/dt    |
| 3115  | this/dt  |
| 2117  | some/dt  |
| 2102  | that/dt  |
| 1274  | all/dt   |
| 1085  | any/dt   |
| 953   | no/dt    |
| 778   | those/dt |

# PDT (Predeterminer)

- Determiner-like words that precede an article or possessive pronoun

- **all** his marbles
- **both** the girls
- **such** a good time

|     |           |
|-----|-----------|
| 263 | all/pdt   |
| 114 | such/pdt  |
| 84  | half/pdt  |
| 24  | both/pdt  |
| 7   | quite/pdt |
| 2   | many/pdt  |
| 1   | nary/pdt  |

# PRP

## (Personal pronouns)

- Personal pronouns (I, me, you, he, him, it, etc.)
- Reflexive pronouns (ending in -self): himself, herself
- Nominal possessive pronouns: mine, yours

|      |           |
|------|-----------|
| 7854 | it/prp    |
| 4601 | he/prp    |
| 3260 | they/prp  |
| 2323 | his/prp\$ |
| 1792 | we/prp    |
| 1584 | i/prp     |
| 1001 | you/prp   |
| 874  | them/prp  |
| 694  | she/prp   |
| 438  | him/prp   |

# PRP\$

## (Possessive pronouns)

- Adjectival possessive forms
  - my car

|      |             |
|------|-------------|
| 5013 | its/prp\$   |
| 2364 | their/prp\$ |
| 2323 | his/prp\$   |
| 521  | our/prp\$   |
| 430  | her/prp\$   |
| 328  | my/prp\$    |
| 269  | your/prp\$  |

# JJ

## (Adjectives)

- General adjectives

- *happy person*
- *new mail*

- Ordinal numbers

- *fourth person*

|      |              |
|------|--------------|
| 2002 | other/jj     |
| 1925 | new/jj       |
| 1563 | last/jj      |
| 1174 | many/jj      |
| 1142 | such/jj      |
| 1058 | first/jj     |
| 824  | major/jj     |
| 715  | federal/jj   |
| 698  | next/jj      |
| 644  | financial/jj |

# JJR

## (Comparative adjectives)

- Adjectives with a comparative ending *-er* and comparative meaning.
  - *happier* person
- More and less (when used as adjectives)
  - *more* mail
- Comparative meaning but no comparative ending (*superior*) = JJ

|      |             |
|------|-------------|
| 1498 | more/jjr    |
| 518  | higher/jjr  |
| 432  | lower/jjr   |
| 285  | less/jjr    |
| 158  | better/jjr  |
| 136  | smaller/jjr |
| 122  | earlier/jjr |
| 112  | greater/jjr |
| 93   | larger/jjr  |
| 75   | bigger/jjr  |

# JJS

## (Superlative adjectives)

- Adjectives with a superlative ending *-est* and superlative meaning.

- *happiest* person

- Most and least (when used as adjectives)

- *most* mail

- Comparative meaning but no comparative ending (*unsurpassed*)  
= JJ

|     |              |
|-----|--------------|
| 695 | most/jjs     |
| 428 | least/jjs    |
| 315 | largest/jjs  |
| 299 | latest/jjs   |
| 209 | biggest/jjs  |
| 194 | best/jjs     |
| 76  | highest/jjs  |
| 63  | worst/jjs    |
| 31  | lowest/jjs   |
| 30  | greatest/jjs |

# RB (Adverb)

- Most words that end in -ly
- Degree words (quite, too, very)
- Negative markers: not, n't, never

|      |          |
|------|----------|
| 4410 | n't/rb   |
| 2071 | also/rb  |
| 1858 | not/rb   |
| 1109 | now/rb   |
| 1070 | only/rb  |
| 1027 | as/rb    |
| 961  | even/rb  |
| 839  | so/rb    |
| 810  | about/rb |
| 804  | still/rb |

# RBR

## (Comparative Adverb)

- Adverbs with a comparative ending *-er* and comparative meaning.
- *More/less*

|      |             |
|------|-------------|
| 1121 | more/rbr    |
| 516  | earlier/rbr |
| 192  | less/rbr    |
| 88   | further/rbr |
| 82   | lower/rbr   |
| 75   | better/rbr  |
| 65   | higher/rbr  |
| 57   | longer/rbr  |
| 53   | later/rbr   |
| 34   | faster/rbr  |

# RBS

## (Comparative Adverb)

- Adverbs with a superlative ending *-est* and superlative meaning.

- *Most/least*

|     |              |
|-----|--------------|
| 549 | most/rbs     |
| 21  | best/rbs     |
| 9   | least/rbs    |
| 8   | hardest/rbs  |
| 2   | most/rbs jjs |
| 1   | worst/rbs    |
| 1   | rbs/nnp      |
| 1   | highest/rbs  |
| 1   | earliest/rbs |

# IN (preposition, subordinating conjunction)

- All prepositions (except *to*) and subordinating conjunctions
- He jumped **on** the table **because** he was excited

|       |         |
|-------|---------|
| 31111 | of/in   |
| 22967 | in/in   |
| 11425 | for/in  |
| 7181  | on/in   |
| 6684  | that/in |
| 6399  | at/in   |
| 6229  | by/in   |
| 5940  | from/in |
| 5874  | with/in |
| 5239  | as/in   |

# CC

## (Coordinating conjunction)

- And, but, not, or
- Math operators (plus, minor, less, times)
- For (meaning “because”)  
[he asked to be transferred, for he was unhappy]

|       |            |
|-------|------------|
| 22362 | and/cc     |
| 4604  | but/cc     |
| 3436  | or/cc      |
| 1410  | &/cc       |
| 94    | nor/cc     |
| 68    | either/cc  |
| 53    | yet/cc     |
| 53    | plus/cc    |
| 37    | both/cc    |
| 32    | neither/cc |

# EX

## (Existential “there”)

- There was a party in progress
- There ensued a melee

1176 there/ex

# FW

## (Foreign word)

- Words in a foreign language (here, non-English) that haven't been incorporated into the language yet.
- e.g., *persona non grata*
- Words that are also in the English lexicon (e.g., *yoga*) should be tagged with their function in the sentence (as any English word)

|    |                |
|----|----------------|
| 39 | de/fw          |
| 15 | vs./fw         |
| 15 | perestroika/fw |
| 13 | pro/fw         |
| 13 | glasnost/fw    |
| 9  | bono/fw        |
| 8  | a/fw           |
| 7  | la/fw          |
| 7  | etc/fw         |
| 6  | naczelnik/fw   |

# UH (Interjection)

- oh, uh, um
- yes, no
- please, well

|    |           |
|----|-----------|
| 22 | yes/uh    |
| 19 | no/uh     |
| 13 | well/uh   |
| 11 | oh/uh     |
| 5  | quack/uh  |
| 5  | ok/uh     |
| 3  | please/uh |
| 3  | indeed/uh |
| 3  | hello/uh  |
| 3  | ah/uh     |

# WP

## (wh-pronoun)

- Who, what, whom

|      |            |
|------|------------|
| 2101 | who/wp     |
| 973  | what/wp    |
| 77   | whom/wp    |
| 5    | whoever/wp |

# WP\$

(possessive wh- word)

- Whose (that's it)

243 whose/wp\$

# TO

(“to”)

- Any instance of to as either a preposition (“to the river”) or an infinitive (“want to go”)

30190 to/to

# WRB

## (wh- adverb)

- A wh-term that functions as an adverb (modifying a verb rather than acting like a pronoun/noun)

- How did it go?
- Where was it?
- Why did you go?

|      |              |
|------|--------------|
| 1659 | when/wrb     |
| 506  | where/wrb    |
| 501  | how/wrb      |
| 197  | why/wrb      |
| 18   | whenever/wrb |
| 6    | wherever/wrb |
| 5    | whereby/wrb  |
| 3    | however/wrb  |
| 1    | wherein/wrb  |

# WDT

## (wh- determiner)

- Which, that when used as a relative pronoun
  - The car that was speeding
  - The car, which was speeding, stopped.

|      |               |
|------|---------------|
| 3014 | which/wdt     |
| 2718 | that/wdt      |
| 48   | whatever/wdt  |
| 35   | what/wdt      |
| 6    | whichever/wdt |

# SYM (symbol)

- Mathematical, technical, scientific symbols that aren't words in the language (here, English)

|    |         |
|----|---------|
| 13 | a/sym   |
| 9  | b/sym   |
| 8  | c/sym   |
| 4  | f/sym   |
| 3  | x/sym   |
| 3  | ffr/sym |
| 3  | e/sym   |
| 2  | z/sym   |
| 2  | d/sym   |
| 1  | r/sym   |

# CD

## (cardinal number)

- Any cardinal number (either written out or numerical)

- 4
- one million

|      |            |
|------|------------|
| 5742 | million/cd |
| 2327 | billion/cd |
| 2014 | one/cd     |
| 1525 | two/cd     |
| 814  | 1/cd       |
| 812  | three/cd   |
| 727  | 10/cd      |
| 668  | 30/cd      |
| 554  | 8/cd       |
| 546  | 1988/cd    |

# POS

## (possessive ending)

- Just the 's possessive ending

11032 's/pos

# LS

## (List item marker)

- Words used as item markers in lists:
  - ... for the following reasons: 1.  
because ... 2. ...

|    |           |
|----|-----------|
| 13 | 2/ls      |
| 13 | 1/ls      |
| 12 | 3/ls      |
| 7  | 4/ls      |
| 3  | second/ls |
| 3  | first/ls  |
| 3  | 5/ls      |
| 2  | third/ls  |
| 2  | b/ls      |
| 2  | a/ls      |

# Punctuation

\$

#

``

''

opening parenthesis

(

[ ( { <

closing parenthesis

)

] ) } >

,

sentence-final

.

. ! ?

mid-sentence

:

:: ; ... — -

# CD or NN

Can it be modified like an adjective?

CD

One of the best reasons

NN

The only (good) one of its kind

# DT or PDT

When articles precede another article, they are pre-determiners

DT

All girls

PDT

All the girls

# IN or RB

- Prepositions usually precede noun phrases (to form a prepositional phrase) but don't have to

IN

The credit car you won't want to do without

RB

We'll just have to do without

# IN or RB

- A preposition may precede another preposition

Blaze out into space

Come out of the woodwork

# IN or RP

She told off her friends

- If it can precede *or follow* the noun phrase = **RP**
  - She told off her friends  
She told her friends off
- If it must precede the noun phrase = **IN**
  - She stepped off the train  
\*She stepped the train off

# IN or WDT

When *that* introduces a complement, it is a subordinating conjunction; when introducing relative clause, a wh-pronoun

IN

the claim *that* angels have wings

[cf. he claimed *that* angels have wings]

WDT

a man *that* I know

# NN or JJ

- Nouns used as modifiers = NN
  - wool sweater
  - life insurance company
- Substantive adjectives = JJ if they can be modified by an adverb
  - The (very) rich pay far too few taxes

# JJ or NP/NPS

- Proper names can be adjectives or nouns

JJ

French cuisine is delicious

NNPS

The French tend to be inspired cooks

# JJ or RB

- If a word modifies a noun, it's usually an adjective (JJ); if it modifies a non-noun it's typically an adverb (RB)

JJ

rapid growth

RB

rapid growing plants

# JJ or VBG

- JJ if it precedes a noun and the corresponding verb is intransitive or does not have the same meaning

an appealing/JJ face;  
an appetizing/JJ dish;  
a revolving/JJ fund;  
a winning/JJ smile;

\*a face that appeals  
\*a dish that appetizes  
\*a fund that revolves  
\*a smile that wins

But:

the existing/VBG safeguards;  
a holding/VBG company;  
a managing/VBG director;  
the ruling/VBG class;

safeguards that exist  
a company that holds another one  
a director who manages  
the class that rules

# JJ or VBN

- If it's gradable (can insert *very*) = JJ

JJ

- He was very surprised

- If can be followed by a *by* phrase = VBN. If that conflicts with #1 above, then = JJ

VBN

- He was invited by some friends of her

JJ

- He was very surprised by her remarks

# NN or VBG

- Only nouns can be modified by adjectives; only gerunds can be modified by adverbs

NN

Good **cooking** is something to enjoy

VBG

**Cooking** well is a useful skill

# WDT or WH

- If a wh-word precedes a noun, it's a wh- determiner (WDT)

WDT

What kind do you want?

WH

What do you want?

<http://bit.ly/wshtags>

Ok

UH

,

,

one

CD

good

JJ

experience

NN

...

:

fine

UH

<http://bit.ly/wshtags>

|         |     |
|---------|-----|
| The     | DT  |
| station | NN  |
| wagons  | NN  |
| arrived | VBD |
| at      | IN  |
| noon    | NN  |
| ,       | ,   |
| a       | DT  |
| long    | JJ  |
| shining | VBG |
| line    | NN  |
| that    | WDT |
| coursed | VBD |
| through | IN  |
| the     | DT  |
| west    | JJ  |
| campus  | NN  |
| .       | .   |