Outline

- Part-of-Speech tags
- Part-of-Speech tagging
  - Rule-Based Tagging
  - HMM Tagging
  - Transformation-Based Tagging
- Evaluation
Outline

● Part-of-Speech tags
● Part-of-Speech tagging
  - Rule-Based Tagging
  - HMM Tagging
  - Transformation-Based Tagging
● Evaluation
Part-of-Speech (POS) Tags

- Also known as:
  - Part-of-speech tags, lexical categories, word classes, morphological classes, lexical tags

Plays [VERB] well [ADVERB] with [PREPOSITION] others [NOUN]

Plays [VBZ] well [RB] with [IN] others [NNS]
Examples of POS tags

- **Noun**: book/books, nature, Germany, Sony
- **Verb**: eat, wrote
- **Auxiliary**: can, should, have
- **Adjective**: new, newer, newest
- **Adverb**: well, urgently
- **Number**: 872, two, first
- **Article/Determiner**: the, some
- **Conjunction**: and, or
- **Pronoun**: he, my
- **Preposition**: to, in
- **Particle**: off, up
- **Interjection**: Ow, Eh
Motivation: Speech Synthesis

• Word „content“
  - „Eggs have a high protein content."
  - „She was content to step down after four years as chief executive."
Motivation: Machine Translation

- e.g., translation from English to German:
  - „I like ...“
    - „Ich mag ....“ (verb)
    - „Ich wie ...“ (preposition)
Motivation: Syntactic parsing

Your query

I saw the man on the roof

Tagging

I/PRP saw/VBD the/DT man/NN on/IN the/DT roof/NN

Parse

(ROOT
  (S
    (NP (PRP I))
    (VP (VBD saw)
      (NP (DT the) (NN man))
      (PP (IN on)
        (NP (DT the) (NN roof))))))

(http://nlp.stanford.edu:8080/parser/index.jsp)
Motivation: Information extraction

- Named-entity recognition (usually nouns)

```bash
> echo "Inhibition of NF-kappaB activation reversed the anti-apoptotic effect of isochamaejasmin." | ./geniatagger

Inhibition Inhibition NN B-NP 0
of of IN B-PP 0
NF-kappaB NF-kappaB NN B-NP B-protein
activation activation NN I-NP 0
reversed reverse VBD B-VP 0
the the DT B-NP 0
anti-apoptotic anti-apoptotic JJ I-NP 0
effect effect NN I-NP 0
of of IN B-PP 0
isochamaejasmin isochamaejasmin NN B-NP 0
." 
```

(http://www.nactem.ac.uk/tsujii/GENIA/tagger/)
Motivation: Information extraction

- Relation extraction (usually verbs)

```bash
> echo "Inhibition of NF-kappaB activation reversed the anti-apoptotic effect of isochamaejasmin." | ./geniatagger
```

```
Inhibition  Inhibition  NN  B-NP  0
of           of           IN  B-PP  0
NF-kappaB   NF-kappaB   NN  B-NP  B-protein
activation  activation  NN  I-NP  0
reversed    reverse      VBD B-VP  0
the          the          DT  B-NP  0
anti-apoptotic anti-apoptotic JJ  I-NP  0
effect      effect       NN  I-NP  0
of           of           IN  B-PP  0
isochamaejasmin isochamaejasmin NN  B-NP  0
.          .          .      .    0  0
```
Open vs. Closed Classes

- **Closed**
  - limited number of words, do not grow usually
  - e.g., Auxiliary, Article, Determiner, Conjunction, Pronoun, Preposition, Particle, Interjection

- **Open**
  - unlimited number of words
  - e.g., Noun, Verb, Adverb, Adjective
POS Tagsets

• There are many parts of speech tagsets

• Tag types
  – Coarse-grained
    • Noun, verb, adjective, ...
  – Fine-grained
    • noun-proper-singular, noun-proper-plural, noun-common-mass, ..
    • verb-past, verb-present-3rd, verb-base, ...
    • adjective-simple, adjective-comparative, ...
POS Tagsets

- Brown tagset (87 tags)
  - Brown corpus

- C5 tagset (61 tags)

- C7 tagset (146 tags!)

- Penn TreeBank (45 tags)
  - A large annotated corpus of English tagset
Penn Treebank Tagset

<table>
<thead>
<tr>
<th>Number</th>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CC</td>
<td>Coordinating conjunction</td>
</tr>
<tr>
<td>2.</td>
<td>CD</td>
<td>Cardinal number</td>
</tr>
<tr>
<td>3.</td>
<td>DT</td>
<td>Determiner</td>
</tr>
<tr>
<td>4.</td>
<td>EX</td>
<td>Existential <em>there</em></td>
</tr>
<tr>
<td>5.</td>
<td>FW</td>
<td>Foreign word</td>
</tr>
<tr>
<td>6.</td>
<td>IN</td>
<td>Preposition or subordinating conjunction</td>
</tr>
<tr>
<td>7.</td>
<td>JJ</td>
<td>Adjective</td>
</tr>
<tr>
<td>8.</td>
<td>JJR</td>
<td>Adjective, comparative</td>
</tr>
<tr>
<td>9.</td>
<td>JJS</td>
<td>Adjective, superlative</td>
</tr>
<tr>
<td>10.</td>
<td>LS</td>
<td>List item marker</td>
</tr>
<tr>
<td>11.</td>
<td>MD</td>
<td>Modal</td>
</tr>
<tr>
<td>12.</td>
<td>NN</td>
<td>Noun, singular or mass</td>
</tr>
<tr>
<td>13.</td>
<td>NNS</td>
<td>Noun, plural</td>
</tr>
<tr>
<td>14.</td>
<td>NNP</td>
<td>Proper noun, singular</td>
</tr>
<tr>
<td>15.</td>
<td>NNPS</td>
<td>Proper noun, plural</td>
</tr>
<tr>
<td>16.</td>
<td>PDT</td>
<td>Predeterminer</td>
</tr>
<tr>
<td>17.</td>
<td>POS</td>
<td>Possessive ending</td>
</tr>
<tr>
<td>18.</td>
<td>PRP</td>
<td>Personal pronoun</td>
</tr>
<tr>
<td>19.</td>
<td>PRP$</td>
<td>Possessive pronoun</td>
</tr>
<tr>
<td>20.</td>
<td>RB</td>
<td>Adverb</td>
</tr>
<tr>
<td>21.</td>
<td>RBR</td>
<td>Adverb, comparative</td>
</tr>
<tr>
<td>22.</td>
<td>RBS</td>
<td>Adverb, superlative</td>
</tr>
<tr>
<td>23.</td>
<td>RP</td>
<td>Particle</td>
</tr>
<tr>
<td>24.</td>
<td>SYM</td>
<td>Symbol</td>
</tr>
<tr>
<td>25.</td>
<td>TO</td>
<td><em>to</em></td>
</tr>
<tr>
<td>26.</td>
<td>UH</td>
<td>Interjection</td>
</tr>
<tr>
<td>27.</td>
<td>VB</td>
<td>Verb, base form</td>
</tr>
<tr>
<td>28.</td>
<td>VBD</td>
<td>Verb, past tense</td>
</tr>
<tr>
<td>29.</td>
<td>VBG</td>
<td>Verb, gerund or present participle</td>
</tr>
<tr>
<td>30.</td>
<td>VBN</td>
<td>Verb, past participle</td>
</tr>
<tr>
<td>31.</td>
<td>VBP</td>
<td>Verb, non-3rd person singular present</td>
</tr>
<tr>
<td>32.</td>
<td>VBZ</td>
<td>Verb, 3rd person singular present</td>
</tr>
<tr>
<td>33.</td>
<td>WDT</td>
<td>Wh-determiner</td>
</tr>
<tr>
<td>34.</td>
<td>WP</td>
<td>Wh-pronoun</td>
</tr>
<tr>
<td>35.</td>
<td>WP$</td>
<td>Possessive wh-pronoun</td>
</tr>
<tr>
<td>36.</td>
<td>WRB</td>
<td>Wh-adverb</td>
</tr>
</tbody>
</table>

(https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)
Outline

• Part-of-Speech tags
• Part-of-Speech tagging
  – Rule-Based Tagging
  – HMM Tagging
  – Transformation-Based Tagging
• Evaluation
POS Tagging

- The process of assigning a part of speech to each word in a text

- Challenge: words often have more than one POS
  - On my back\_[NN] (noun)
  - The back\_[JJ] door (adjective)
  - Win the voters back\_[RB] (adverb)
  - Promised to back\_[VB] the bill (verb)
Ambiguity in POS tags

- 45-tags Brown corpus (word types)
  - Unambiguous (1 tag): 38,857
  - Ambiguous: 8,844
    - 2 tags: 6,731
    - 3 tags: 1,621
    - 4 tags: 357
    - 5 tags: 90
    - 6 tags: 32
    - 7 tags: 6 (well, set, round, open, fit, down)
    - 8 tags: 4 ('s, half, back, a)
    - 9 tags: 3 (that, more, in)
Baseline method

1. Tagging unambiguous words with the correct label
2. Tagging ambiguous words with their most frequent label
3. Tagging unknown words as a noun

- This method performs around 90% precision
POS Tagging

- Plays well with others
  - Plays (NNS/VBZ)
  - well (UH/JJ/NN/RB)
  - with (IN)
  - others (NNS)

- Plays_[VBZ] well_[RB] with_[IN] others_[NNS]
Outline

• Part-of-Speech tags
• Part-of-Speech tagging
  – Rule-Based Tagging
  – HMM Tagging
  – Transformation-Based Tagging
• Evaluation
Rule-Based Tagging

- Standard approach (two steps):
  1. Dictionaries to assign a list of potential tags
     - Plays (NNS/VBZ)
     - well (UH/JJ/NN/RB)
     - with (IN)
     - others (NNS)
  2. Hand-written rules to restrict to a POS tag
     - Plays (VBZ)
     - well (RB)
     - with (IN)
     - others (NNS)
Rule-Based Tagging

- Some approaches rely on morphological parsing
  - e.g., EngCG Tagger
Outline

- Part-of-Speech tags
- Part-of-Speech tagging
  - Rule-Based Tagging
  - HMM Tagging
  - Transformation-Based Tagging
- Evaluation
Sequential modeling

- Many of the NLP techniques should deal with data represented as sequence of items
  - Characters, Words, Phrases, Lines, ...

- Part-of-speech tagging

- Named-entity recognition
Sequential modeling

• Making a decision based on:
  – Current Observation:
    • Word ($W_0$): „35-years-old“
    • Prefix, Suffix: „computation“ → „comp“, „ation“
    • Lowercased word: „New“ → „new“
    • Word shape: „35-years-old“ → „d-a-a“
  – Surrounding observations
    • Words ($W_{+1}, W_{-1}$)
  – Previous decisions
    • POS tags ($T_{-1}, T_{-2}$)
Learning Model (classification)
Sequential modeling

- Greedy inference
  - Start in the beginning of the sequence
  - Assign a label to each item using the classifier
  - Using previous decisions as well as the observed data
Sequential modeling

- Beam inference
  - Keeping the top k labels in each position
  - Extending each sequence in each local way
  - Finding the best k labels for the next position
Hidden Markov Model (HMM)

- Finding the best sequence of tags \((t_1 \ldots t_n)\) that corresponds to the sequence of observations \((w_1 \ldots w_n)\)

- Probabilistic View
  - Considering all possible sequences of tags
  - Choosing the tag sequence from this universe of sequences, which is most probable given the observation sequence

\[
\hat{t}_1^n = \arg \max_{t_1^n} P(t_1^n | w_1^n)
\]
Using the Bayes Rule

\[ \hat{t}_1^n = \arg\max_{t_1^n} P(t_1^n | w_1^n) \]

\[ P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \]

\[ P(t_1^n | w_1^n) = \frac{P(w_1^n | t_1^n) \cdot P(t_1^n)}{P(w_1^n)} \]

\[ \hat{t}_1^n = \arg\max_{t_1^n} P(w_1^n | t_1^n) \cdot P(t_1^n) \]

\text{likelihood} \quad \text{prior probability}
Using Markov Assumption

\[ \hat{t}_1^n = \arg\max_{t_1^n} P(w_1^n|t_1^n) \cdot P(t_1^n) \]

\[ P(w_1^n|t_1^n) \approx \prod_{i=1}^n P(w_i|t_i) \quad \text{(it depends only on its POS tag and independent of other words)} \]

\[ P(t_1^n) \approx \prod_{i=1}^n P(t_i|t_{i-1}) \quad \text{(it depends only on the previous POS tag, thus, bigram)} \]

\[ \hat{t}_1^n = \arg\max_{t_1^n} \prod_{i=1}^n P(w_i|t_i) \cdot P(t_i|t_{i-1}) \]
Two Probabilities

- The tag transition probabilities: $P(t_i|t_{i-1})$
  - Finding the likelihood of a tag to proceed by another tag
  - Similar to the normal bigram model

\[ P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})} \]
Two Probabilities

- The word likelihood probabilities: $P(w_i|t_i)$
  - Finding the likelihood of a word to appear given a tag

$$P(w_i|t_i) = \frac{C(t_i, w_i)}{C(t_i)}$$
Two Probabilities

I saw the man on the roof.

\[ P([NN][DT]) = \frac{C([DT],[NN])}{C([DT])} \]

\[ P(man|[NN]) = \frac{C([NN],man)}{C([NN])} \]
Ambiguity

Secretariat is expected to race tomorrow.

People inquire the reason for the race.
Secretariat is expected to race tomorrow.

Secretariat is expected to race tomorrow.
Secretariat is expected to race tomorrow.

\[
P(VB|TO) = 0.83 \\
P(\text{race}|VB) = 0.00012 \\
P(NR|VB) = 0.0027 \\
P(VB|TO).P(NR|VB).P(\text{race}|VB) = 0.00000027
\]
Secretariat is expected to race tomorrow.
Formalization of Hidden Markov Model (HMM)

- Extension of Finite-State Automata

- Finite-State Automata
  - set of states
  - set of transitions between states
Formalization of Hidden Markov Model (HMM)

• Weighted finite-state automaton
  - Each arc is associated with a probability
  - The probabilities leaving any arc must sum to one
Hidden Markov Model (HMM)

- POS tagging
  - Ambiguous
  - We observe the words, not the POS tags

- HMM
  - Observed events: words
  - Hidden events: POS tags
Hidden Markov Model (HMM)

- Transition probabilities: $P(t_i|t_{i-1})$
Hidden Markov Model (HMM)

- Word likelihood probabilities: $P(w_i|t_i)$

```
B_2
P("aardvark"|TO)
.. P("race"|TO)
.. P("the"|TO)
.. P("to"|TO)
.. P("zebra"|TO)
```

```
B_1
P("aardvark"|VB)
.. P("race"|VB)
.. P("the"|VB)
.. P("to"|VB)
.. P("zebra"|VB)
```

```
B_3
P("aardvark"|NN)
.. P("race"|NN)
.. P("the"|NN)
.. P("to"|NN)
.. P("zebra"|NN)
```
Viterbi algorithm

- Decoding algorithm for HMM
  - Determine the best sequence of POS tags

- Probability matrix
  - Columns corresponding to inputs (words)
  - Rows corresponding to possible states (POS tags)
Viterbi algorithm

1. Move through the matrix in one pass filling the columns left to right using the transition probabilities and observation probabilities

2. Store the max probability path to each cell (not all paths) using dynamic programming
I want to race.
$V_0(0) = 1.0$

$v_{t-1}$: previous Viterbi path probability (from the previous time step)
\( q_0 \): start
\( q_1 \): PPSS
\( q_2 \): VB
\( q_3 \): TO
\( q_4 \): NN

\( v_0(0) = 1.0 \)

\[ a_{ij} \]: transition probability (from previous state \( q_i \) to current state \( q_j \))

\[ b_j(o_t) \]: state observation likelihood (observation \( o_t \) given the current state \( j \))
\[ v_t(j) = \max_{i=1}^{N} v_{t-1}(i) \cdot a_{ij} \cdot b_j(o_t) \]

\[ v_1(1) = P(\text{PPSS}|\text{start}) \cdot P(\text{start}) \cdot P(\text{i}|\text{PPSS}) = 0.067 \cdot 0.37 = 0.025 \]

\[ v_1(2) = P(\text{PPSS}|\text{start}) \cdot P(\text{start}) \cdot P(\text{i}|\text{VB}) = 0.019 \cdot 0 = 0 \]

\[ v_1(3) = P(\text{PPSS}|\text{start}) \cdot P(\text{start}) \cdot P(\text{i}|\text{TO}) = 0.043 \cdot 0 = 0 \]

\[ v_1(4) = P(\text{PPSS}|\text{start}) \cdot P(\text{start}) \cdot P(\text{i}|\text{NN}) = 0.041 \cdot 0 = 0 \]
I want to race.

q_{end} (end) → end

q_4 (NN) → NN

q_3 (TO) → TO

q_2 (VB) → VB

q_1 (PPSS) → PPSS

q_0 (start) → start

\[
v_1(4) \cdot P(VB|NN) = 0.0040
\]

\[
v_1(3) \cdot P(VB|TO) = 0.83
\]

\[
v_1(2) \cdot P(VB|VB) = 0.0038
\]

\[
v_1(1) \cdot P(VB|PPSS) = 0.025 \cdot 0.23 = 0.0055
\]

Q_1 \quad Q_2 \quad Q_3 \quad Q_4
\( q_{\text{end}} \rightarrow \text{end} \)

\( q_4 \rightarrow \text{NN} \)

\( q_3 \rightarrow \text{TO} \)

\( q_2 \rightarrow \text{VB} \)

\( q_1 \rightarrow \text{PPSS} \)

\( q_0 \rightarrow \text{start} \)

\( v_2(2) = \max(0, 0, 0.0055) \cdot 0.0093 = 0.000051 \)
I want to race.
I want to race.
I want to race.
Outline

- Part-of-Speech tags
- Part-of-Speech tagging
  - Rule-Based Tagging
  - HMM Tagging
  - Transformation-Based Tagging
- Evaluation
Transformation-Based Tagging

- Also called Brill Tagging
- Uses Transformation-Based Learning (TBL)
  - Rules are automatically induced from training data
Transformation-Based Tagging

- Brill's Tagger

  1. Assign every word with the most likely tag:
     - Secretariat/NNP is/VBZ expected/VBN to/TO race/NN tomorrow/NN
     - The/DT race/NN for/IN outer/JJ space/NN

  2. Apply transformation rules:
     - e.g., change „NN“ to „VB“ when previous tag is „TO“
       - Secretariat/NNP is/VBZ expected/VBN to/TO race/VB tomorrow/NN
Transformation-Based Tagging

- TBL learning process
  - Based on a set of templates (abstracted transformations)
    - „the word two before (after) is tagged z“
    - „the preceding word is tagged z and the following word is tagged w“
    - etc.
Outline

• Part-of-Speech tags
• Part-of-Speech tagging
  – Rule-Based Tagging
  – HMM Tagging
  – Transformation-Based Tagging
• Evaluation
Evaluation

• Corpus
  - Training and test, and optionally also development set
  - Training (cross-validation) and test set

• Evaluation
  - Comparison of gold standard (GS) and predicted tags
  - Evaluation in terms of Precision, Recall and F-Measure
Precision and Recall

- **Precision:**
  - Amount of labeled items which are correct

\[
\text{Precision} = \frac{tp}{tp + fp}
\]

- **Recall:**
  - Amount of correct items which have been labeled

\[
\text{Recall} = \frac{tp}{tp + fn}
\]
F-Measure

- There is a strong anti-correlation between precision and recall
- Having a trade off between these two metrics
- Using F-measure to consider both metrics together
- F-measure is a weighted harmonic mean of precision and recall

\[ F = \frac{(\beta^2 + 1) PR}{\beta^2 P + R} \]
Error Analysis

- Confusion matrix or contingency table
  - Percentage of overall tagging error

<table>
<thead>
<tr>
<th></th>
<th>IN</th>
<th>JJ</th>
<th>NN</th>
<th>NNP</th>
<th>RB</th>
<th>VBD</th>
<th>VBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN</td>
<td>-</td>
<td>.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.7</td>
</tr>
<tr>
<td>JJ</td>
<td>.2</td>
<td>-</td>
<td>3.3</td>
<td>2.1</td>
<td>1.7</td>
<td>.2</td>
<td>2.7</td>
</tr>
<tr>
<td>NN</td>
<td>8.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.2</td>
</tr>
<tr>
<td>NNP</td>
<td>.2</td>
<td>3.3</td>
<td>4.1</td>
<td>-</td>
<td>.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RB</td>
<td>2.2</td>
<td>2.0</td>
<td>.5</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>VBD</td>
<td>.3</td>
<td>.5</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>4.4</td>
</tr>
<tr>
<td>VBN</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
</tr>
</tbody>
</table>
Further reading and tools

- Book Jurafski & Martin
  - Chapter 5

- Tools
  - Stanford POS parser
  - OpenNLP
  - TreeTagger
  - and many others...
Project update

• Integration of POS tagging
  − Presentation in the next lecture (May 23rd, 2016)