ANTONIS ANASTASOPOULOS
CS499 INTRODUCTION TO NLP

PART-OF-SPEECH TAGGING

https://cs.gmu.edu/~antonis/course/cs499-spring21/

With adapted slides by David Mortensen and Alan Black
1. The POS-tagging task
2. Parts of speech
3. POS-tagging with count-based models
4. Neural POS Tagging
POS TAGGING
POS TAGGING

My cat who lives dangerously no longer has nine lives.
POS TAGGING

My cat who `lives` dangerously no longer has nine `lives`.

`lives`: noun /lajvz/
`lives`: verb /livz/

The task is:

Input: a sequence of word tokens \textit{w}
Output: a sequence of part-of-speech tags \textit{t} (one per word)

Note: the linguistic facts are considerably more complicated that the assumptions of the structure of this task, but there are good reasons for keeping it simple.
EXAMPLE

<table>
<thead>
<tr>
<th>POS</th>
<th>Charlie</th>
<th>Brown</th>
<th>received</th>
<th>a</th>
<th>valentine</th>
<th>.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WHY HAVE PARTS OF SPEECH

There are too many words
- You’d need a lot of data to train rules
- Due to sparsity, rules would be very specific

PoS tags allow models to generalize

Give useful reduction in model sizes

There are many different tag sets
- You want the right one for your task
PARTS OF SPEECH
SO YOU WORK ON POS TAGGING. WHAT’S A PART OF SPEECH?

— David Kaplan
WHAT ARE PARTS OF SPEECH

The lexicon (collection of words of a language) is not some amorphous soup

It is somewhat soup-like, but it is a chunky soup:

- Small, finite number of categories
- Structured subcategories within these categories
- These categories are soft

If you ignore the structured nature of the lexicon, you are making life hard for yourself!
WHAT ARE PARTS OF SPEECH

A limited number of tags for word “class”

Distributional
- has the same contexts
- has the same syntactic functions (subj, obj, mod)
- occurs in same positions in syntactic structure

Morphological
- allows the same suffixes, prefixes

Not about meaning
- your English teacher lied to you — get used to it
OPEN-CLASS
PARTS OF SPEECH
Can be subjects and objects of verbs
- This book is about geography.
- I read a good book.

Can be objects of prepositions
- I’m mad about book

Can be plural or singular (book, books)

Can have determiners (the book)

Can be modified by adjectives (blue book)

Can have possessors (my books, John’s book)
ENGLISH VERBS

Take noun phrases as arguments
- At least a subject
  - Dr. A parsed aggressively.
- Sometimes one or two objects
  - Dr. A parsed the data.
  - Dr. A passed [the function] [an argument].

Can take tense morphology (past/non-past)

Can be modified by adverbs
ENGLISH ADJECTIVES

Modify nouns (restrict their reference)
- *his* heavy book.
- *His* book *is* heavy.

Can take comparative/superlative suffixes when allowed by prosody
- heavy, heavier, heaviest
- But pitiful, more pitiful, most pitiful

Not all languages have adjectives
Some languages (e.g. Korean, Hmong, Vietnamese) use verbs to modify nouns in this way
ENGLISH ADVERBS

Modify verbs, adjectives, and other adverbs (restrict their reference)
- He *erroneously concluded* that PHP is a real programming language *simply* because it is Turing complete.
- The design of PHP is *exceptionally poor*. 
CLOSED CLASS
PARTS OF SPEECH
ENGLISH PREPOSITIONS

Occur before noun phrases

Relate noun phrase to some higher-level constituent
- *He lingered in the depths of despair.*

It is not difficult to characterize prepositions *formally*, but they are very difficult to characterize *semantically*
(a good argument to not introduce semantic considerations into PoS categories)

They are often identical in spelling and pronunciation to *particles.*
Determiners are words that come at the beginning of English noun phrases

**Articles** like *the*, *a*, and *an*

- *The interpreter choked on an unknown identifier.*

Other determiners include some demonstratives like *this* and *that*.

- *That version of Python really chaps my hide.*
ENGLISH PRONOUNS

Pronouns can replace noun phrases, acting as a short hand for them
- You code like a wizard.
- Who knows Haskell, really?.
Conjunctions join phrases, clauses, or sentences.

Typically, the conjuncts joined by a conjunction are of the same type

Coordinating conjunctions:
- and, or, but, ...

Subordinating conjunctions:
- if, because, though, while, ...
ENGLISH AUXILIARY VERBS

“Helping verbs” that occur before main verbs

Some can occur as main verbs as well
- Be
  - I am the type system. (main)
  - I am working on my project. (aux verb)
- Have

Others (e.g. modals) occur only as auxiliary verbs
- must, might, would, will, could, can, ...
ENGLISH PARTICLES

Particle is sometimes used as a grab-bag category for closed-class items that do not fit in other categories.

Most often, in English, these resemble prepositions or adverbs and are used in combination with a verb:
- He tore off his shirt.
- He tore his shirt off.
- I want to leave.
Numerals have properties of both nouns and adjectives

They can be the subject and object of verbs:
- Two will enter but only one will leave.
- I bought twenty.

They can function both attributively and predicatively:
- Two variables were left undeclared.
- We are three.

When used attributively, they come before any adjectives:
- The two undeclared variables were the cause of much consternation.
- *The undeclared two variables were the cause of much consternation.
1. Substitution test

The ADJ cat sat on the mat
The blue NOUN sits on the NOUN
The blue cat VERB on the mat
The blue cat sat PREP the mat
FINE-GRAINED CLASSES

- Nouns
  - Proper
  - Common
    - Mass
    - Countable

- Verbs

- Adjectives
  - Directional
    - Degree
      - manner
      - Temporal
I will call up my friend

I will call my friend up

I will call my friend up in the treehouse

Gerunds
   I like walking
   I like apples
   His walking kept him fit
   His apples kept him fit
   His walking slowly kept him fit
   His apples slowly kept him fit

Potential Additional Classes

Interjections
Negatives
Politeness markers
Greetings
Existential there
Numbers, Symbols, Money, ....
Emojis
URLs
Hashtags
GOOGLE’S UNIVERSAL TAGS

<table>
<thead>
<tr>
<th>Tag</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>adjective</td>
</tr>
<tr>
<td>ADP</td>
<td>adposition (preposition or postposition)</td>
</tr>
<tr>
<td>ADV</td>
<td>adverb</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
</tr>
<tr>
<td>CCONJ</td>
<td>coordinating conjunction</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>INTJ</td>
<td>interjection</td>
</tr>
<tr>
<td>NOUN</td>
<td>noun</td>
</tr>
<tr>
<td>NUM</td>
<td>numeral</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>PRON</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROPN</td>
<td>proper noun</td>
</tr>
<tr>
<td>PUNCT</td>
<td>punctuation</td>
</tr>
<tr>
<td>SCONJ</td>
<td>subordinating conjunction</td>
</tr>
<tr>
<td>SYM</td>
<td>symbol</td>
</tr>
<tr>
<td>VERB</td>
<td>verb</td>
</tr>
<tr>
<td>X</td>
<td>other</td>
</tr>
</tbody>
</table>

**Warning**

Don’t tell a linguist these are *truly* universal. They will be very offended — and they will be right to do so.

But, they can be very useful!
WHY DO WE NEED MODELS? IS POS TAGGING HARD?

If every “word” could only be associated with a single tag, PoS tagging would be trivial
- How would you do it?
- Do you foresee any problems?

But, this won’t always work
- *lives* can be a noun or a verb
- *black* can be an adjective, verb, proper noun, common noun, etc...

How bad is this problem, really?

<table>
<thead>
<tr>
<th>PoS tags per orthographic word in Penn Treebank</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 down</td>
</tr>
<tr>
<td>6 that</td>
</tr>
<tr>
<td>6 set</td>
</tr>
<tr>
<td>6 put</td>
</tr>
<tr>
<td>6 open</td>
</tr>
<tr>
<td>6 hurt</td>
</tr>
<tr>
<td>6 cut</td>
</tr>
<tr>
<td>6 bet</td>
</tr>
<tr>
<td>6 back</td>
</tr>
<tr>
<td>5 vs.</td>
</tr>
<tr>
<td>5 the</td>
</tr>
<tr>
<td>5 spread</td>
</tr>
<tr>
<td>5 split</td>
</tr>
<tr>
<td>5 say</td>
</tr>
<tr>
<td>5 ’s</td>
</tr>
<tr>
<td>5 run</td>
</tr>
<tr>
<td>5 repurchase</td>
</tr>
<tr>
<td>5 read</td>
</tr>
<tr>
<td>5 present</td>
</tr>
<tr>
<td>5 out</td>
</tr>
<tr>
<td>5 many</td>
</tr>
<tr>
<td>5 less</td>
</tr>
<tr>
<td>5 left</td>
</tr>
<tr>
<td>5 Japanese</td>
</tr>
<tr>
<td>5 in</td>
</tr>
<tr>
<td>5 hit</td>
</tr>
<tr>
<td>5 half</td>
</tr>
<tr>
<td>5 further</td>
</tr>
<tr>
<td>5 forecast</td>
</tr>
<tr>
<td>5 fit</td>
</tr>
<tr>
<td>5 first</td>
</tr>
<tr>
<td>5 East</td>
</tr>
<tr>
<td>5 counter</td>
</tr>
<tr>
<td>5 cost</td>
</tr>
<tr>
<td>5 close</td>
</tr>
<tr>
<td>5 bid</td>
</tr>
<tr>
<td>5 beat</td>
</tr>
<tr>
<td>5 a</td>
</tr>
</tbody>
</table>

317 down RB
200 down RP
138 down IN
10 down JJ
1 down VBP
1 down RBR
1 down NN
MODELING APPROACH

1. Pick the most frequent tag per word
   - What accuracy do you think it would give you over average English text?

2. Look at the context
   - Preceding (and succeeding) words
   - Preceding (and succeeding) tags
   - The ...
   - To ...
   - John’s blue ...
THE OUT-OF-VOCABULARY PROBLEM

How do you handle cases where your dictionary does not include all of the words?

Proper names?
Borrowed words?
Neologisms?

As a language user, these are not a problem for you.

How would you give a POS-tagger the same superpower?

Stay tuned for the CRF class!
NEURAL POS TAGGING
TRAINING A NEURAL POS TAGGER

I hate this movie.
TRAINING A NEURAL POS TAGGER

Calculating the loss

I hate this movie

RNN RNN RNN RNN

predict predict predict predict

prediction 1 prediction 2 prediction 3 prediction 4

loss 1 loss 2 loss 3 loss 4

label 1 label 2 label 3 label 4

sum

total loss

34
CODE EXAMPLE
Classification 101