

# Statistical Natural Language Processing

## Part of speech tagging

Çağrı Çöltekin

University of Tübingen  
Seminar für Sprachwissenschaft

Summer Semester 2020

## Part of speech tagging

Time flies like an arrow .

NOUN VERB ADP DET NOUN PUNC

- Part of speech (POS or PoS) tags are morphosyntactic classes of words
- The words belonging to the same POS class share some syntactic and morphological properties

## Traditional POS tags

what you learn in (primary?) school

- noun apple, chair, book
- verb go, read, eat
- adjective blue, happy, nice
- adverb well, fast, nicely
- pronoun I, they, mine
- determiner a, the, some
- preposition in, since, past, ago (?)
- conjunction and, or, since
- interjection uh, ouch, hey

With minor differences, this list of categories has been around for a long time.

## When we say 'traditional' ...



- POS tags in modern linguistics are based on Greek/Latin linguistic traditions
- But others, e.g., Sanskrit linguists, also proposed POS tags

## What are the POS tags good for

- Linguistic theory
- Parsing
- Speech synthesis: pronounce *lead*, *wind*, *object*, *insult* differently based on their POS tag
- The same goes for machine translation
- Information retrieval: if *wug* is a noun, also search for *wugs*
- Text classification: improves some tasks
- As a back-off strategy for some language models

## POS tagsets in practice

example: Penn treebank tagset

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	eats
JJS	Adj., superlative	widest	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	WPS	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	( [, (, {, <
PRPS	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			

## POS tagsets in practice

example 2: STTS tagset

POS	description	examples
...	...	...
KOUI	subordinating conjunction	um [zu leben], anstatt [zu fragen]
KOUS	subordinating conjunction	weil, daß, damit, wenn, ob
KON	coordinative conjunction	und, oder, aber
KOKOM	particle of comparison, no clause	als, wie
NN	noun	Tisch, Herr, [das] Reisen
NE	proper noun	Hans, Hamburg, HSV
PDS	substituting demonstrative	dieser, jener
PIS	substituting indefinite pronoun	keiner, viele, man, niemand
PIAT	attributive indefinite	kein [Mensch], irgendein [Glas]
PIDAT	attributive indefinite	[ein] wenig [Wasser],
PPER	irreflexive personal pronoun	ich, er, ihm, mich, dir
PPOSS	substituting possessive pronoun	meins, deiner
PPOSAT	attributive possessive pronoun	mein [Buch], deine [Mutter]
PRELS	substituting relative pronoun	[der Hund,] der
PRELAT	attributive relative pronoun	[der Mann,] dessen [Hund]
...	...	...

## POS tagset choices

- The choice of tagsets depends on the language and application
- Example tag set sizes (for English)
  - Brown corpus, 87 tags
  - Penn treebank 45 tags
  - BNC, 61 tags
- Differences can be large, for Chinese Penn treebank has 34 tags, but tagsets with about 300 tags exist
- For other languages, the choice varies roughly between about 10 to a few hundred

## Shift towards more 'universal' tag sets

- The variation in POS tagset choices often makes it difficult to
  - compare alternative approaches
  - use the same tools on different languages or data sets
- There has been a recent trend for 'universal' tag sets
- The result is a smaller POS tag set (back to the tradition)
- But often supplemented with *morphological features*

## POS tagsets in recent practice

example: Universal Dependencies tag set

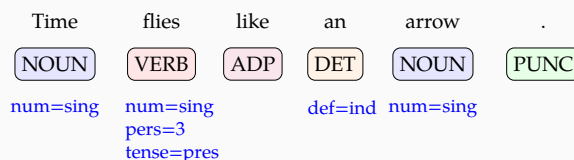
ADJ	adjective	PART	particle
ADP	adposition	PRON	pronoun
ADV	adverb	PROPN	proper noun
AUX	auxiliary	PUNCT	punctuation
CCONJ	coordinating conjunction	SCONJ	subordinating conjunction
DET	determiner	SYM	symbol
INTJ	interjection	VERB	verb
NOUN	noun		
NUM	numeral	X	other

## Morphological features

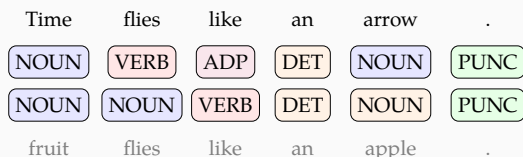
- Annotating words with morphological features has been common in (non-English) NLP
- Morphological features give additional sub-categorization information for the word
- For example
  - nouns typically have *number* and *case* feature
  - verbs typically have *tense*, *aspect*, *modality* *voice* features
  - adjectives typically have *degree*
- Morphological feature sets change depending on the language (typology)

## Morphological features

an example



## POS tags are ambiguous



Part of speech tagging is essentially an ambiguity resolution problem.

## POS tag ambiguity

More examples

- Some words are highly ambiguous
  - ADJ the *back* door
  - NOUN on our *back*
  - ADV take it *back*
  - VERB we will *back* them
- The garden-path sentences are often POS ambiguities
  - The *old man* the boats
  - The complex *houses* married and single soldiers and their families

## POS tagging: strategies

POS tagging can be solved in a number of different methods

- Rule-based methods: 'constraint grammar' (CG)
- Transformation based: Brill tagger
- Machine-learning approaches
  - Typical statistical approaches involve *sequence learning* methods:
    - Hidden Markov models
    - Conditional random fields
    - (Recurrent) neural networks

## Rule-based POS tagging

typical approach

- Using a tag lexicon, start with assigning all possible tags to each word
- Eliminate tags based on hand-crafted rules
- Rules typically rely on the words and (potential) tags of the words in the context
- Result is not always full disambiguation, some ambiguity may remain
- Some probabilistic constraints may also be applied

## Rule-based POS tagging

an example

- Among others, the word *that* can be  
 SCONJ we know *that* it is bad  
 ADV it is not *that* bad

An example rule for disambiguation (simplified):

```

1 if the next word is ADJ
2 and the next word is sentence final
3 and the previous word is not a verb like 'consider'
4 then eliminate SCONJ
5 else eliminate ADV
    
```

- The rules above prefer SCONJ for cases like *I consider that funny*.

## Transformation based tagging (TBL)

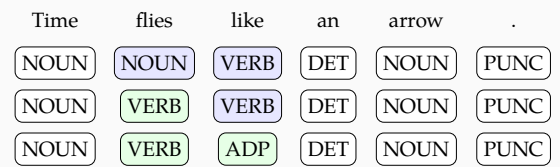
- The idea: learn a sequence of rules (similar to CG) using a tagged corpus
- The rules transform an initial POS assignment to (approximately) the POS tag assignment in the training corpus
- During test time apply the rules in the same order

## Learning in TBL

1. Start with most likely tags for each word
  2. Find the best rule that improves the tagging accuracy,
  3. Repeat step 2 for all possible rules
- Rules need to be restricted, often templates are used. For example:  
 Change tag x to tag y if
    - the preceding/following word is tagged z
    - the preceding word tagged v and the following word is tagged z
    - the preceding word tagged v and the following word is tagged z and two words before is tagged t

## Transformation based learning

An example

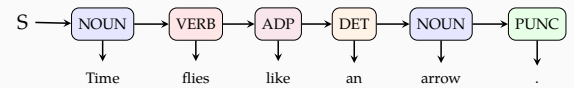


- Start with most likely POS tags
- Apply: change NOUN to VERB if preceding word is NOUN and ...
- Apply: change VERB to ADP if preceding word is tagged as VERB
- Stop when none of the rules improve the result

## ML methods for POS tagging

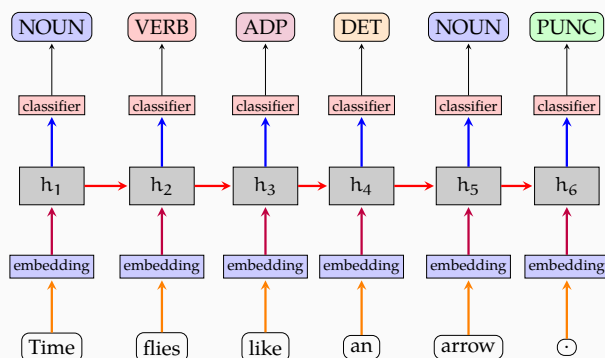
- POS tagging is a typical example of 'sequence labeling'
- Many of the ML methods introduced earlier can be used for POS tagging
- Sequence learning methods are more suitable, since the tags depend on the neighboring tags
  - Hidden Markov models (HMMs)
  - Hidden Markov max-ent models (HMMEMs)
  - Conditional random fields (CRFs)
  - Recurrent neural networks (RNNs)

## POS tagging using Hidden Markov models (HMM)



- The tags are hidden
- Probability of a tag depends on the previous tag
- Probability of a word at a given state depends only on the current tag
- Parameters of the model can be learned
  - supervised from a tagged corpus (e.g., MLE)
  - unsupervised using EM (Baum-Welch algorithm)

## RNNs for POS tagging



## POS tagging accuracy

- Tagging each word with the most probable tag gives around 90% accuracy
- State-of-the-art POS taggers (for English) achieve 95%–97%
- Human agreement on annotation also seems to be around 97%: not a lot of space for improvement
  - at least for well-studied resource-rich languages

## Summary

- POS is an old idea in linguistics
- POS tags have uses in both linguistics, and practical applications
- Common methods for automatic POS tagging include
  - rule-based
  - transformation-based
  - statistical (more on this next week)

methods

Next:

- Vector representations
- Parsing
- Text classification

## Open vs. closed class words

Open class words (e.g., nouns) are productive

- new words coined are often in these classes
- we often cannot rely on a fixed lexicon
- they are typically ‘content’ words

Closed class words (e.g., determiners) are generally static

- the lexicon does not grow
- they are typically ‘function’ words

- This distinction is often language dependent,

## Some issues with traditional POS tags

- Not all POS tags are observed in (or theorized for) all languages
- Often finer granularity is necessary
  - *book*, *water* and *Mary* are all nouns, but
    - The book is here
    - \* The Mary is here
    - We have water
    - \* We have book