Natural Language Processing

Chapter 3: Part-of-Speech Tagging

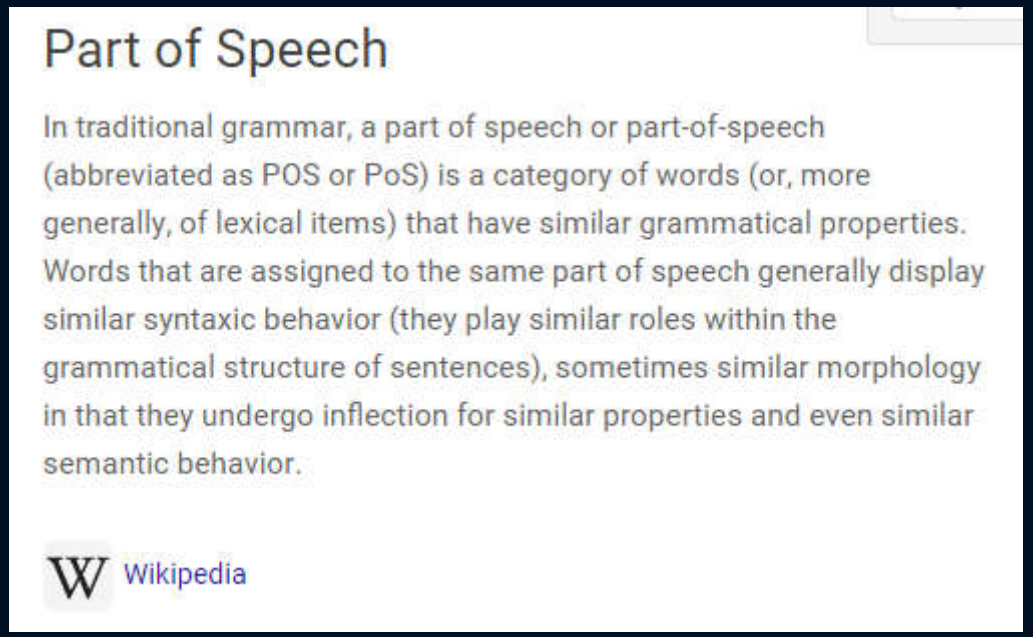
Dr Raymond Lee
Associate Professor, DST
BNU-HKBU United International College

Part-of-Speech Tagging

1. What is Part-of-Speech (POS) ?
2. POS in NLP
3. NLU - the BIG Picture
4. Different Types of POS
5. What is Tagset?
6. Ambiguous in POS Tags
7. Algorithms for POS Tagging
8. Rule-based POS Tagging
9. Stochastic POS Tagging
10. Hybrid POS Tagging - Brill Tagger
11. Evaluation of Taggers
12. Summary

What is Part-of-Speech (PoS)?

- Inflection refers to a process of word formation in which items are added to the base form of a word to express grammatical meanings.
- The word "inflection" comes from the Latin inflectere, meaning "to bend."
- For example, the inflection -s at the end of cats shows that the noun is plural.
- The same inflection -s at the end of the verb gets shows that the subject is in the third-person singular (e.g. He or She gets the book.).
- The inflection -ed is often used to indicate the past tense, changing take to took and hear to heard.
- In this way, inflections are used to show grammatical categories such as person, number and tense.
- To clarify how such inflection works, we need to know the different types of POS and their nature.



Nine Major POS in English Language

- Every sentence you write or speak in English includes words that fall into some of the nine parts of speech.
- In English Language, there are 9 major POS types.
- They are: nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions, determiners (articles), and interjections.
- Some linguists include only eight parts of speech and leave interjections in their own category.
- In terms of linguistics, POS is an important component to learn the categorization of word classes and usage.
- In terms of Use of English, POS is an vital component for learning and using grammars. Important component for speaking and written English.
- In terms of NLP, POS is an vital component for the "categorization" words (word types) and their functions inside the sentences and utterances.
- Of course, it is also the key component in POS Tagging.

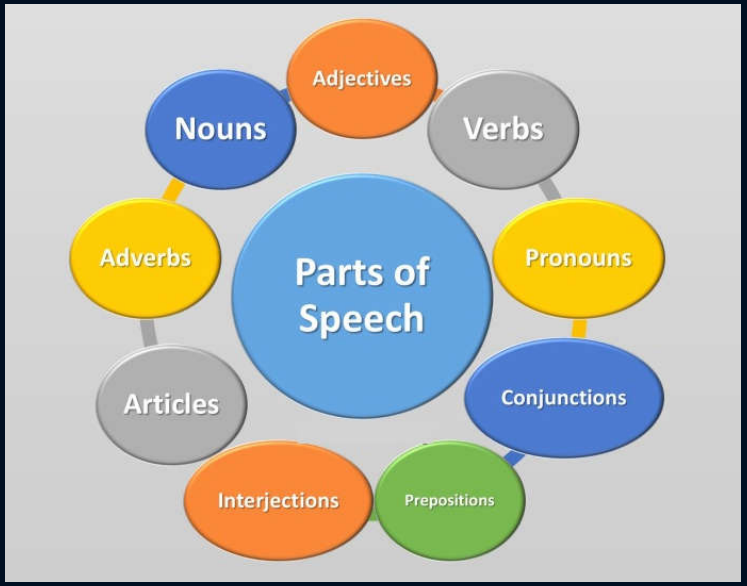
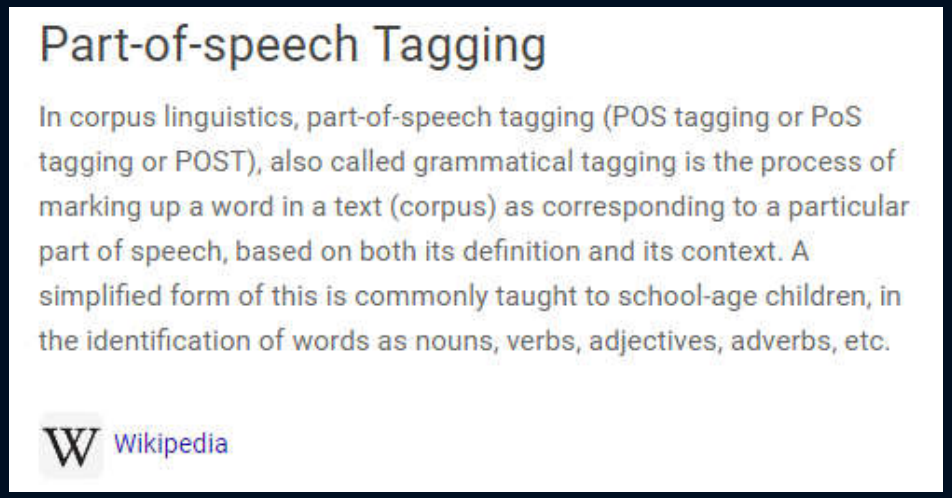


Fig. 3.1 Major POS in English Language

What is Part-of-Speech Tagging in Linguistics?

- In NLP, Tagging is a kind of classification that may be defined as the automatic assignment of description to the words (word tokens).
- We called them POS tags (or Tags in short), which may represent one of the part-of-speech, semantic information and so on.
- Now, if we talk about Part-of-Speech (PoS) tagging, then it may be defined as the process of assigning one of the parts of speech to the given word. It is generally called POS tagging.
- In simple words, we can say that POS tagging is a task of labelling each word in a sentence with its appropriate part of speech. We already know that parts of speech include nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.
- As seen, the same word might have different meaning as different roles in POS: E.g. Book can be used as "A Book" as noun or "Booking the table" as verb, which have totally different meaning in a sentence or utterance.



What is Part-of-Speech Tagging in NLP?

- In other word, POS Tagging is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form *(word, tag)*).
- The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.
- Parts of speech tags are defined by the relationship of words with the other words in the sentence.
- Fig 3.2 shows how POS tagging is apply to the sentence (utterance): She sells seashells on the seashore.
- So how can we assign POS tagging in NLP?
- We can apply machine learning models and rule-based models to obtain the parts of speech tags of a word.
- Most of the POS tagging falls under Rule Base POS tagging, Stochastic POS tagging and Hybrid POS Tagging using more advanced technology such as Transformation-based tagging.
- In this chapter, we will learnt the basic concepts and components of POS Tagging, for the workshop in the Part II of this book, we will study how it works by using NLTK and spaCy technologies.
- First, have a look on some realistic POS databank – PENN Treebank Tagset.

Fig. 3.2 POS example for utterance "She sells seashells on the seashore"

POS tags used in the Penn Treebank Project

- The most commonly used parts of speech tag databank are provided by the Penn Treebank corpus (Marcus et al., 1993).
- The English Penn Treebank tagset is used with English corpora annotated by the TreeTagger tool, developed by Helmut Schmid in the TC project at the Institute for Computational Linguistics of the University of Stuttgart.
- In which, a total of 45 P.O.S tags are defined according to their usage.
- The English Penn Treebank (PTB) corpus, and in particular the section of the corpus corresponding to the articles of Wall Street Journal (WSJ), is one of the most known and used corpus for the evaluation of models for sequence labelling.
- The corpus is also commonly used for character-level and word-level Language Modelling.
- Fig. 3.4 Shows the POS Tagging of the sentence "David has purchased a new laptop from Apple store".

No	POS Tag	Description	Example	No	POS Tag	Description	Example
1	CC	coordinating conjunction	and, but, or	24	SYM	Symbol	$ / [= *
2	CD	cardinal number	1, third	25	TO	infinitive 'to'	to
3	DT	determiner	a, the	26	UH	interjection	haha, oops
4	EX	existential there	there is	27	VB	verb - base form	drink
5	FW	foreign word	les	28	VBD	verb - past tense	drank
6	IN	preposition, sub-conj	in, of, by, like	29	VBG	verb - gerund	drinking
7	JJ	adjective	big, wide, green	30	VBN	verb - past participle	drunk
8	JJR	adjective, comparative	bigger, wider, greener	31	VBP	verb - non-3sg pres	drink
9	JJS	adjective, superlative	biggest, wildest, greenest	32	VBZ	verb - 3sg pres	drinks
10	LS	list marker	1), One, i	33	WDT	wh-determiner	which, that
11	MD	modal	can, could, shall, will	34	WP	wh-pronoun	who, what
12	NN	noun, singular or mass	table, shop	35	WP$	possessive wh-pronoun	whose, those
13	NNS	noun plural	tables, shops	36	WRB	wh-abverb	where, when, how
14	NNP	proper noun, singular	Samsung	37	#	#	#
15	NNPS	proper noun, plural	Vikings	38	$	$	$
16	PDT	predeterminer	all/both the students	39	"	Left quotation	' "
17	POS	possessive ending	friend's	40	``	right quotation	' "
18	PP	personal pronoun	I, he, it, you	41	(	Opening brackets	({
19	PPZ	possessive pronoun	my, his, your, one's	42	)	Closing brackets	) }
20	RB	adverb	however, quickly, here	43	,	Comma	,
21	RBR	adverb, comparative	better, quicker	44	:	Sent-final punc	. ! ?
22	RBS	adverb, superlative	best, quickest	45	:	Mid-sentence punc	: ; ... -
23	RP	particle	of, up (e.g. give up)				

Fig. 3.3 Penn Treebank POS Tags (with punctuation)

Fig. 3.4 POS example for "David has purchased a new laptop from Apple Store"

M. Marcus, B. Santorini and M.A. Marcinkiewicz (1993). Building a large annotated corpus of English: The Penn Treebank. In Computational Linguistics, volume 19, number 2, pp. 313–330.

Why Do We Care about Parts of Speech in NLP?

- Pronunciation (e.g. Same word has different pronunciation in different POS classes)

  Here are the student records vs. The teacher records his lecture.

- Predicting what words can be expected next

  Personal pronoun (e.g., *I*, he, *she, they*) ? will/ would/ think …

- Stemming 词干

  E.g. comput for computer

- As the basis for syntactic parsing and then meaning extraction

  Give me back my money or I will call the police.

- Machine translation (Same word with different POS classes most likely have totally difference translation in other language)
  - (E) *book* +*N* ⇒ (F) *acheter* +*N* *(e.g. Buy a book -> Achète un livre)*
  - (E) *book* + *VB* ⇒ (F) *réserver* +*VB* (e.g. Book a room -> réserver une chambre)

Same words with difference stress accents as difference POS scenarios

- One important function of POS is the different stress accents of the same word (word type) as different POS inside a sentence or utterance.
- Commonly occurs
  - Noun vs Verb confusion
  - Adjective vs Verb confusion
  - Adjective vs Noun confusion
- Fig 3.5 shows the common examples of same English word with different stress accents and meaning.
- Commonly occurs when dealing with a noisy channel
- Probabilistic techniques we can use for various subproblem
- Corpora we can analyze to collect our facts
- POS tagging is the first step.

Noun	Verb	Noun	Verb	Noun	Verb
ABstract	abstRACT	ENvelope	enVELope	REBel	reBEL
ACcent	acCENT	EScort	esCORT	REcap	reCAP
ADdict	adDICT	EXploit	exPLOIT	REcall	reCALL
ADdress	adDRESS	EXport	exPORT	REcord	reCORD
ANnex	anNEX	EXtract	exTRACT	REfill	reFILL
ALly	alLY	FInance	fiNANCE	REfund	reFUND
ATtribute	atTRIBute	FRAgment	fragMENT	REfuse	refUSE
COMbat	comBAT	IMpact	imPACT	REject	reJECT
COMmune	comMUNE	IMprint	imPRINT	REplay	rePLAY
COMpact	comPACT	INcrease	inCREASE	SUBject	subJECT
COMpound	comPOUND	INsert	inSERT	SURvey	surVEY
COMpress	comPRESS	INsult	inSULT	SUSpect	susPECT
CONduct	conDUCT	MANdate	manDATE	TORment	torMENT
CONfines	conFINES	OBject	obJECT	TRANSfer	transFER
CONflict	conFLICT	OVERcharge	overCHARGE	TRANSplant	transPLANT
CONscript	conSCRIPT	OVERwork	overWORK	TRANSport	transPORT
CONsort	conSORT	PERmit	perMIT	UPset	upSET
CONtract	conTRACT	PERvert	perVERT		
CONtrast	conTRAST	PREfix	preFIX	**Adjective**	**Verb**
CONverse	conVERSE	PREsent	preSENT	ABsent	abSENT
CONvert	conVERT	PROceeds	proCEEDS	FREquent	freQUENT
CONvict	conVICT	PROcess	proCESS	PERfect	perFECT
DEcrease	deCREASE	PROduce	proDUCE		
DEsert	deSERT	PROgress	proGRESS	**Adjective**	**Noun**
DEtail	deTAIL	PROject	proJECT	inVALid	INvalid
DIScard	disCARD	PROtest	proTEST	miNUTE (my noot)	MInute (min it)
DIScharge	disCHARGE	RAMpage	ramPAGE	comPLEX	COMplex

Fig. 3.5 Common example of same English word with different stress accents

Natural Language Understanding (NLU) The Big Picture

- Although the NLP consists of many modules and components.
- One of the most important aspect is NLU – Natural Language Understanding (NLU)
- Essential for almost ALL NLP applications: Text summarization, Sentiment analysis, Information Retrievals (IR) to Q&A Chatbot systems.
- Fig 3.6 shows major processes of NLU.
- POS Tagging is the key process to provide "basic function" and "category" (word class) of words
- NLG (Natural Language Generation) goes backwards.
- For this reason, we generally want declarative representations of the facts.

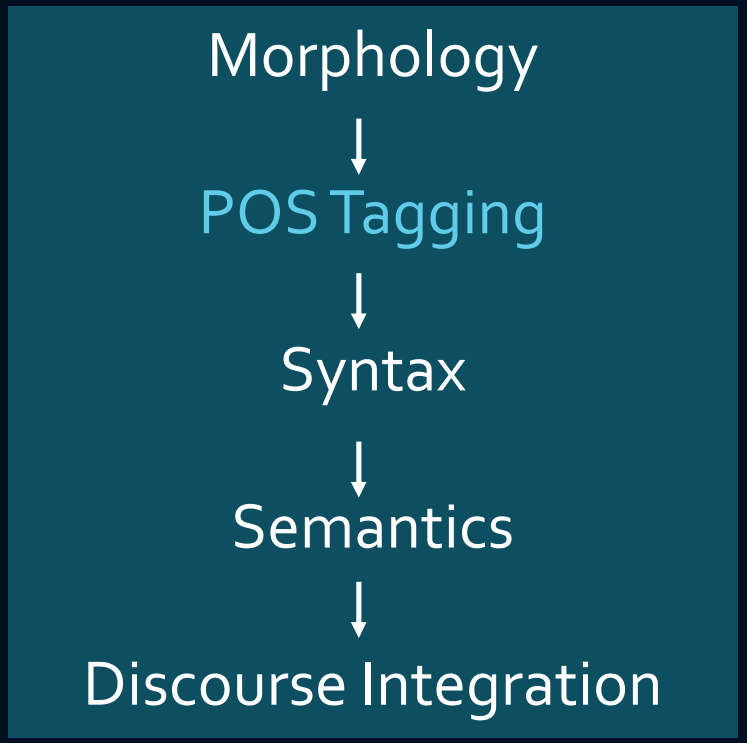


Fig. 3.6 Major Components in NLU

Computational Linguistics and POS

Computational Linguistics and POS

- Computational linguistics (CL) is the scientific and engineering discipline concerned with understanding written and spoken language from a computational aspect.
- CL aims at the building artifacts that usefully process and produce language, either in bulk or in a dialogue setting.
- To the extent that language is a mirror of mind, a computational understanding of language also provides insight into thinking and intelligence.
- Since language is our most natural and most versatile means of communication, linguistically competent computers would greatly facilitate our interaction with machines and software of all sorts, and put at our fingertips, in ways that truly meet our needs, the vast textual and other resources of the internet.
- Two major issues we need to tackle with Computational Linguistics
  - Linguistic – what are the facts about language?
  - Algorithmic – what are effective computational procedures for dealing with those facts?
- The theoretical goals of computational linguistics include:
  - Formulation of grammatical and semantic frameworks for characterizing languages in ways enabling computationally tractable implementations of syntactic and semantic analysis
  - Discovery of processing techniques and learning principles that exploit both the structural and distributional (statistical) properties of language
  - Development of cognitively and neuroscientifically plausible computational models of how language processing and learning might occur in the brain.
- Part-of-Speech can be considered as the fundamental process in Computational Linguistics for the understanding and modelling of human languages.

What is a Part of Speech and Semantic Meaning

Is this a semantic distinction?

For example, maybe Noun is the class of words for people, places and things. Maybe Adjective is the class of words for properties of nouns.

Consider: *green book*

book is a Noun

green is an Adjective

Now consider: *book worm ??? (which one is noun/adjective?)*

This green is very soothing. (which is which?)

Morphological and Syntactic Definition of POS

An Adjective is a word that can fill the blank in:

It's so _____.

A Noun is a word that can be marked as plural.

A Noun is a word that can fill the blank in:

the _____ is

What is *green*?

It's so green.

Both greens could work for the walls.

The green is a little much given the red rug.

NINE Key Part-of-Speech in English
Meaning and Example

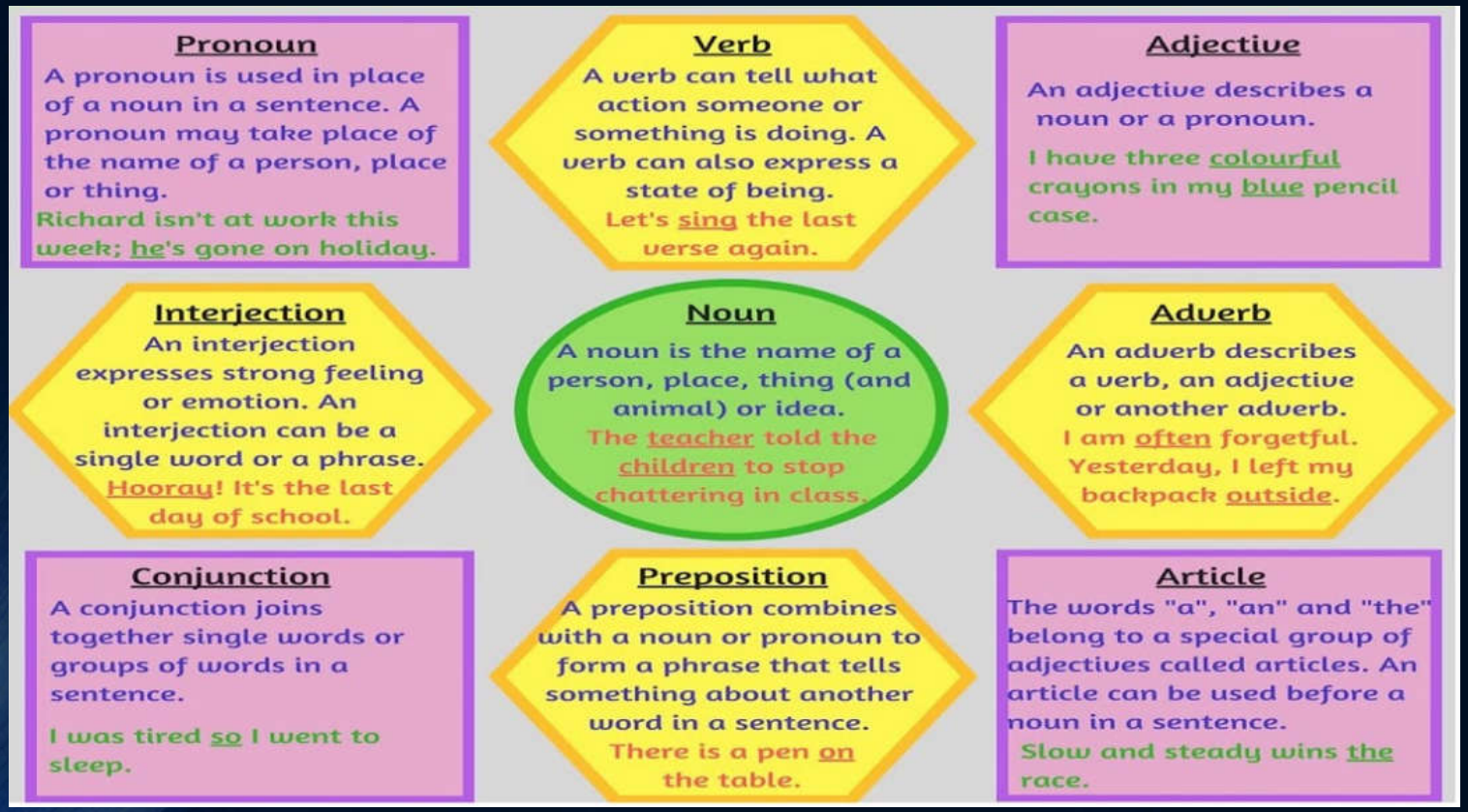
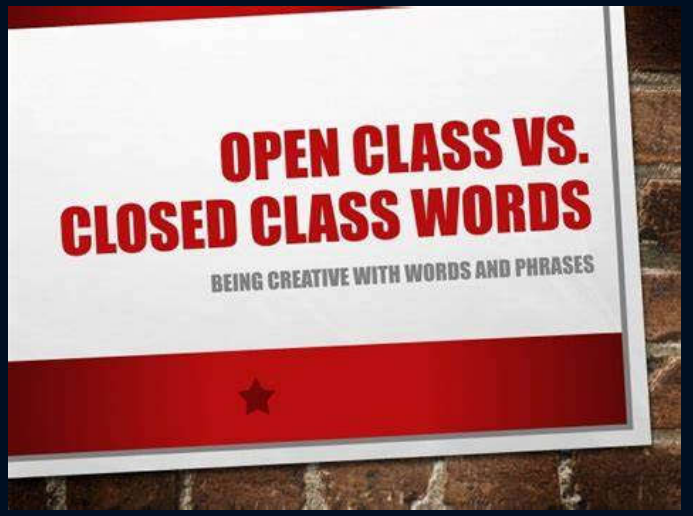


Fig. 3.7 8 Major POS in English Language (with description)

English Word Classes

Open vs Closed Class Types

- All English words can be classified as belonging to a "closed class" or an "open class."
- Closed-class words are also called "function words" or "grammar words"
- Open-class words are also called "lexical words" or "content words."
- Closed classes are called "closed" because new words are seldom added to them.
- By contrast, new items are regularly added to the open classes.
- The entire noun ,verb, adjective and adverbs classes are made up of open-class items as is the entire sub-class of full verbs.
- Example: Fax, Telex, Internet, Hub, Bitcoin, …
- All determiners, pronouns, conjunctions, prepositions are closed class.
- The meaning of an open-class item such as a noun or a verb can be given in a dictionary where words are defined in isolation.
- By contrast, the meaning of closed-class words cannot be explained in isolation because it is tied up with the grammatical structures that they are part of.
- For example, in the utterance "the style of this painting", both "the" and "this" don't have a special meaning as compared with the word "painting", which have a specific meaning in our usual knowledge.



What's a Preposition?

Preposition (abbreviated PP)

- Part-of-Speech with a word or group of words used before a noun, pronoun.
- Or a noun phrase to show direction, time, place, location, spatial relationships or to introduce an object.
- Common examples of prepositions are words like "in," "at," "on," "of," and "to."
- There are approximately 80 to 100 prepositions in the English language.
- Prepositions are words that introduce information to the recipient.
- This information can include where something takes place (such as 'at' the store), when or why something takes place (such as 'before' dinner), or general descriptive information (such as the girl 'with' the cool hair style).
- The object of the preposition is the noun that follows the preposition.
- It is also the stopping point for each prepositional phrase.
- For instance, we might say, 'to the supermarket.' The word 'to' is the preposition and 'supermarket' is the object of the preposition. Here's another example, 'over the rainbow.' The word 'over' is the preposition and 'rainbow' is the object of the preposition.
- Fig 3.8 shows the TOP 40 prepositions from CELEX on-line dictionary from COBUILD 16 million word corpus

Rank	PP	Freq.	Rank	PP	Freq.
1	of	540,085	21	above	3056
2	in	331235	22	near	2,026
3	for	142,421	23	off	1695
4	to	125691	24	past	1,575
5	with	124,965	25	worth	1563
6	on	109129	26	toward	1,390
7	at	100,169	27	plus	750
8	by	77794	28	till	686
9	from	74,843	29	amongst	525
10	about	38428	30	via	351
11	than	20,210	31	amid	222
12	over	18071	32	underneath	164
13	through	14,964	33	versus	113
14	after	13670	34	amidst	67
15	between	13,275	35	sans	20
16	under	9525	36	circa	14
17	per	6,515	37	pace	12
18	among	5090	38	nigh	9
19	within	5,030	39	re	4
20	towards	4700	40	mid	3

Fig. 3.8 TOP 40 prepositions from CELEX on-line dictionary from COBUILD 16 million word corpus

What's a Conjunction?

A conjunction (abbreviated CONJ or CNJ)

- Part-of-Speech that connects words, phrases, or clauses that are called the conjuncts of the conjunctions.
- This definition may overlap with that of other parts of speech, so what constitutes a "conjunction" must be defined for each language.
- In English, a given word may have several senses, being either a preposition or a conjunction depending on the syntax of the sentence.
- For example, after is a preposition in "she left after the show.", but it is a conjunction in "she left after she finished her homework".
- Coordinating conjunctions allow you to join words, phrases, and clauses of equal grammatical rank in a sentence. The most common coordinating conjunctions are for, and, nor, but, or, yet, etc.
- Subordinating conjunctions join independent and dependent clauses. A subordinating conjunction can signal a cause-and-effect relationship, or some other kind of relationship between the clauses. Common subordinating conjunctions are because, since, as, although, though, while, and whereas.
- In general, a conjunction is an invariable (non-inflected) grammatical particle and it may or may not stand between the items conjoined.
- Fig. 3.9 shows the TOP 50 co-ordinating and subordinating conjunctions from CELEX on-line dictionary from COBUILD 16 million word corpus.

Rank	CONJ.	Freq.	Rank	CONJ.	Freq.
1	and	514,946	26	now	1,290
2	that	134,773	27	neither	1,120
3	but	96,889	28	whenever	913
4	or	76,563	29	whereas	867
5	as	54,608	30	except	864
6	if	53,917	31	till	686
7	when	37,975	32	provided	594
8	because	23,626	33	whilst	351
9	so	12,933	34	suppose	281
10	before	10,720	35	cos	188
11	though	10,329	36	supposing	185
12	than	9,511	37	considering	174
13	while	8,144	38	lest	131
14	after	7,042	39	albeit	104
15	whether	5,978	40	providing	96
16	for	5,935	41	whereupon	85
17	although	5,424	42	seeing	63
18	until	5,072	43	directly	26
19	yet	5,040	44	ere	12
20	since	4,843	45	notwithstanding	3
21	where	3,952	46	according as	0
22	nor	3,078	47	as if	0
23	once	2,826	48	as long as	0
24	unless	2,205	49	as though	0
25	why	1,333	50	both and	0

Fig. 3.9 TOP 50 co-ordinating and subordinating conjunctions from CELEX on-line dictionary from COBUILD 16 million word corpus

What's a Pronoun?

Pronoun (abbreviated PRN or PN)

- Part-of-Speech defined as a word or phrase that is used as a substitution for a noun or noun phrase
- Which is also known as the pronoun's antecedent.
- Pronouns are short words and can do everything that nouns can do and are one of the building blocks of a sentence.
- Common pronouns are he, she, you, me, I, we, us, this, them, that.
- A pronoun can act as a subject, direct object, indirect object, object of the preposition, and more and takes the place of any person, place, animal or thing.
- By using pronouns, we can he to replace John in an utterance e.g.
  John is sick today, he cannot attend the evening seminar.
- Pronoun is a powerful tool for us to simplify the content in our dialogue and conversation by replace the pronouns with simple "token".
- Fig. 3.10 shows the TOP 50 pronouns from CELEX on-line dictionary from COBUILD 16 million word corpus.
- Without pronouns, we'd constantly have to repeat nouns, and that would make our speech and writing repetitive, not to mention cumbersome.
- However, the use of pronouns will cause some ambiguous problem:
  E.g. John blame Jim for the investment loss, he felt sorry for that. So, who "feel sorry", John or Jim?

Rank	PRN	Freq.	Rank	PRN	Freq.
1	it	199,920	26	our	23,029
2	I	198,139	27	these	22,697
3	he	158,366	28	any	22666
4	you	128,688	29	more	21,873
5	his	99,820	30	many	17343
6	they	88,416	31	such	16,880
7	this	84,927	32	those	15819
8	that	82,603	33	own	15,741
9	she	73,966	34	us	15724
10	her	69,004	35	how	13,137
11	we	64,846	36	another	12,551
12	all	61,767	37	where	11,857
13	which	61,399	38	same	11,841
14	their	51,922	39	something	11,754
15	what	50,116	40	each	11,320
16	my	46,791	41	both	10,930
17	him	45,024	42	last	10,816
18	me	43,071	43	every	9,788
19	who	42,881	44	himself	9,113
20	them	42,099	45	nothing	9,026
21	no	33,458	46	when	8,336
22	some	32,863	47	one	7,423
23	other	29,391	48	much	7,237
24	your	28,923	49	anything	6,937
25	its	27,783	50	next	6,047

Fig. 3.10 TOP 50 pronouns from CELEX on-line dictionary from COBUILD 16 million word corpus

What's a verb?

Verb (abbreviated VB)

- A verb is a word that in syntax generally conveys an action, a process, an occurrence, or a state of being.
- In the usual description of English, the basic form, with or without the particle to, is the infinitive.
- In many languages, verbs are inflected to encode tense, aspect, mood, and voice.
- In some languages, such as Chinese, verbs and nouns of a word is often interchangeable.
- A verb may also agree with the person, gender or number of some of its arguments, such as its subject, or object.
- Verbs have tenses: present, to indicate that an action is being carried out; past, to indicate that an action has been done; future, to indicate that an action happen in future and future perfect to indicate an action that will be completed I in future.
- A modal verb is a type of verb that contextually indicates a modality such as a likelihood, ability, permission, request, capacity, suggestion, order, obligation, or advice.
- Modal verbs always accompany the base (infinitive) form of another verb having semantic content. In English, the modal verbs commonly used are can, could, may, might, shall, should, will, would, and must.
- Fig. 3.11 shows the TOP 25 verbs from CELEX on-line dictionary from CO BUILD 16 million word corpus.

Rank	VB	Freq.	Rank	VB	Freq.
1	can	70,930	14	won't	3,100
2	will	69,206	15	'd	2,299
3	may	25,802	16	ought	1,845
4	would	18,448	17	will	862
5	should	17,760	18	shouldn't	858
6	must	16,520	19	mustn't	332
7	need	9,955	20	'll	175
8	can't	6,375	21	needn't	148
9	have	6,320	22	mightn't	68
10	might	5,580	23	oughtn't	44
11	couldn't	4,265	24	mayn't	3
12	shall	4,118	25	dare	3
13	wouldn't	3,548			

Fig. 3.11 TOP 25 modal verbs from CELEX on-line dictionary from CO BUILD 16 million word corpus

What is Tagset?

Tagsets

- In primary school, we are commonly teach that there are 9 parts of speech in English: noun, verb, article, adjective, preposition, pronoun, adverb, conjunction, and interjection.
- However, there are clearly many more categories and sub-categories.
- For example in nouns, the plural, possessive, and singular forms can be distinguished.
- A tagset is a list of part-of-speech tags (POS tags for short), i.e. labels used to indicate the part of speech and sometimes also other grammatical categories (case, tense etc.) of each token in a text corpus.
- Commonly used tagsets include: Brown Corpus Tagset, PENN Treebank Tagset and CLAWS Tagset.
- The first major corpus of English for computer analysis was the Brown Corpus developed at Brown University by Henry Kučera and W. Nelson Francis, in the mid-1960s. It consists of about 1,000,000 words of running English prose text, made up of 500 samples from randomly chosen publications. Each sample is 2,000 or more words (ending at the first sentence-end after 2,000 words, so that the corpus contains only complete sentences).
- The English Penn Treebank tagset is used with English corpora annotated by the TreeTagger tool, developed by Helmut Schmid in the TC project at the Institute for Computational Linguistics of the University of Stuttgart.
- English CLAWS part-of-speech tagset version 7 is available in English corpora annotated by the tool using CLAWS (the Constituent Likelihood Automatic Word-tagging System) developed by University Centre for Computer Corpus Research on Language at Lancaster University.
- CLAWS uses a Hidden Markov model to determine the likelihood of sequences of words in anticipating each part-of-speech label.

Brown corpus tagset (87 tags):

https://web.archive.org/web/20080706074336/http://www.scs.leeds.ac.uk/ccalas/tagsets/brown.html

Penn Treebank tagset (45 tags):

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

C7 tagset (146 tags)

https://ucrel.lancs.ac.uk/claws7tags.html

Ambiguous in POS Tags

- One might wonder, to check for the POS, why can't we just look them up in a dictionary rather than using the Tagset databank?
- Reason #1: In reality, there are many ambiguous in POS Tags for many words:
  1. Noun-verb Amb: E.g. Record : Records the lecture vs Play the CD records.
  2. Adj-Verb Amb: E.g. Perfect: A perfect plan vs John perfects the invention.
  3. Adj-Noun Amb: E.g. Complex: A complex case vs A shopping complex.
- Ambiguity – In the Brown corpus, 10.40% of the word types are ambiguous (Fig 3.12)
- In which over 40% ambiguous words are easy to disambiguate (WHY?)
- 3-Tag ambiguous e.g. Green. 1. Color Green (NN) 2 A green apple (ADJ) 3 The roof was greening with leaves.
- Which one got 7 POS Tags ambiguous – "Still"
  - The first 4 usage of STILL. ADJ: The still status, NN: The still of the night, ADV: She still lives in here, VB: still can also be using as act of still. Exercise: Figure out the other THREE POS Tag usage.

Unambiguous (1 tag)	**35,340**
2 tags	3,760
3 tags	264
4 tags	61
5 tags	12
6 tags	2
7 tags	1
Ambiguous (2-7 tags)	**4,100**
Ambiguous %	**10.40%**

Fig. 3.12 Classification of word types in Brown corpus by the level of ambiguity

Algorithms for POS Tagging

Why can't we just look them up in a dictionary?

Reason #2: Words that aren't in the dictionary

- One idea: $P(t_i \mid w_i)$ = the probability that a random hapax legomenon in the corpus has tag t_i.
- Nouns are more likely than verbs, which are more likely than pronouns.
- Another idea: use morphology, e.g. Bitcoin.

Hapax Legomenon

- Hapax Legomenon (plural legomena) is a Greek term that literally means "being said once."
- It has come to refer to a term that is used only once in a given context, whether it is the works of a particular author, a particular work of literature, or even within all the known writings of a particular language.
- The works of Shakespeare are estimated to contain thousands of hapax legomena, as about 6,500 words appear only once in all of Shakespeare's poems and plays.
- The Bible contains hundreds of hapax legomena, that is, words that only appear once in the Bible. However, the meaning of most of these terms is not in question because they are used in other Greek and Hebrew/Ancient Semitic literature.
- As always, context is important in Hapex Legomenon.
- If the meaning of the word is known from other ancient literature and that meaning fits the context of the biblical passage, then the word can be translated with a great deal of confidence.

Algorithms for POS Tagging Using Knowledge

1 Dictionary

2 Morphological rules, e.g.,

- _____-ed
- _____-ing
- _____-tion
- _____-ly
- capitalization

3 N-gram frequencies – next word prediction

- to _____
- DET _____ N
- But what about rare words, e.g, *smelt* (two verb forms, melt and past tense of smell, and one noun form, a small fish)

4 Combining these

- V _____-ing

I could almost hear the two-tone foghorns knelling my demise
vs. 'The huntsman's horn sounded the final knell when the last traditional hunt by the Tedworth came to end.'

Approaches of POS Tagging

1. Rule-based Approach
2. Stochastic-based Approach
3. Hybrid Approach

Rule-based Approach of POS Tagging

- Rule-based Approaches
  - This is one of the oldest approaches to POS tagging.
  - It involves a TWO-stage process.
  - Stage 1: Using a dictionary consisting of all the possible POS tags for a given word.
  - Stage 2: If any of the words have more than one tag, hand-written rules are used to assign the correct tag based on the tags of surrounding words.
  - The accuracy of the obtained rule set directly affects the tagging results.
  - The lexicon is first used for basic segmentation and tagging of the corpus, listing all possible lexical properties of the object, and then combining the rule base with the contextual information to disambiguate and finally retain the only suitable lexical property.
  - The rule-based approach is clear and has a wide range of applications
  - The rule generation can be achieved by:
    1. Hand Creation
    2. Training from a corpus by Machine Learning
  - Advantages of hand creation: based on sound linguistic principles, sensible to people, explainable
  - Advantages of training from a corpus: less work, extensible to new languages, customizable for specific domains.

Rule-based Approach of POS Tagging

- Rule-based Approaches
  - But it is not easy to obtain the rules automatically through machine learning, and manual construction is a difficult and time-consuming task, and if the rules are described in too much detail, the coverage of the rules (Oliver et al 2003) will be greatly reduced, and it is difficult to adjust them according to the actual situation.
  - If the rules are not based on the context but only on the lexical nature of the rules, ambiguity may arise.
  - Sample rule: If the preceding of a word an article, then the word has to be a noun.
    - Consider the words: A Book
    - Get all the possible POS tags for individual words:
      - A – Article; Book – Noun or Verb
    - Use the rules to assign the correct POS tag:
      - As per the possible tags, "A" is an Article and we can assign it directly.
      - But, a book can either be a Noun or a Verb.
      - However, if we consider "A Book", A is an article and following our rule above, Book has to be a Noun.
      - Thus, we assign the tag of Noun to book.
    - POS Tag: [("A", "Article"), ("Book", "Noun")]

Example of Rule-Based POS Tagging

Step 1: Using a dictionary, assign to each word a list of possible tags.

Step 2: Figure out what to do about words that are unknown or ambiguous. Two approaches:

- Rules that specify what to do.
- Rules that specify what not to do:

Fig. 3.13 shows the sample Adverbial "that" rule:

- The first two statements of this rule check to see that the that directly precedes a sentence-final adjective, adverb, or quantifier.
- In all other cases the adverb reading is eliminated.
- The last clause eliminates cases preceded by verbs like consider or believe which can take a noun and an adjective.
- The main logic is to avoid tagging the following instance of that as an adverb such as "It isn't that odd".
- The other rule is used to check if the previous word is a verb which expects a complement (like think or hope), and if the that is followed by the beginning of a noun phrase, and a finite verb such as "I consider that a win" or more complex structure such as "I hope that she is ignorance".

Example: Adverbial "that" rule
Given input: "that"
If
(+1 A/ADV/QUANT)
(+2 SENT-LIM)
(NOT -1 SVOC/A)
Then eliminate non-ADV tags
Else eliminate ADV

It isn't that odd . vs

I consider that a win. vs

I hope that she is ignorance.

Fig. 3.13 Sample Rule for Adverbial "that" rule

Stochastic-based Approach of POS Tagging

- Stochastic-based Approach
  - Different from the Rule-based approach, the Stochastic-based approach of POS tagging is a supervised model, involves using with frequencies or probabilities of the tags in the given training corpus to assign a tag to a new word.
  - These taggers entirely rely on statistics of the tag occurrence, i.e. probability of the tags.
  - Based on the words used for determining a tag, Stochastic Taggers are divided into 2 parts:
    - Word Frequency: In this approach, we find the tag that is most assigned to the word. For example: Given a training corpus, the word "check" occurs 10 times – 6 times as Noun, 4 times as a Verb; the word book will always be assigned as "Noun" since it occurs the most in the training set. Hence, a Word Frequency Approach is not very reliable.
    - Tag Sequence Frequency: Here, the best tag for a word is determined using the probability the tags of N previous words, i.e. it considers the tags for the words preceding book. Although this approach provides better results than a Word Frequency Approach, it may still not provide accurate results for rare structures. Tag Sequence Frequency is also referred to as the N-gram approach.
  - The stochastic POS tagging model allows features to be non-independent and allows for the addition of features of various granularities.
  - Commonly used stochastic-based approach with using of Hidden Markov Model (HMM) Tagger.
  - The Maximum Entropy Markov Model (MEMM) (Al-Taani and Al-Rub, 2009) is a stochastic POS tagging model that defines an exponential algorithm for each state as the conditional probability of the next state given the current state, which has the advantages of a stochastic POS tagging model.
  - However, this model suffers from label bias problem.
  - Unlike the MEMM model, the Conditional Random Field (CRF) model uses only one finger model as the joint probability of the entire label sequence given the sequence of observations.
  - Lafferty et. al. (2001) verified that this model can effectively solve the tagging bias problem.

HMM Tagger

- Let's use HMM Tagger as example.
- The rationale of HMM Tagger is: Using N-gram frequencies to choose the most-likely tag for a given word.
- Mathematically, all we need is to maximum conditional probability: The conditional probability w_i is tag t_i in the context given w_i :

$$P(t_i \, in \, context \mid w_i) = \frac{P(w_i \mid t_i \, in \, context) P(t_i \, in \, context)}{P(w_i)}$$

- In other word, given a sentence or word sequence, HMM taggers choose the tag sequence that maximizes the following formula:

$$P(\text{word}|\text{tag}) * P(\text{tag}|\text{previous n tags})$$

- For a bigram-HMM tagger of this kind chooses the tag t_i for word w_i that is most probable given the previous tag t_{i-1} and the current word w_i:

$$t_i = \underset{j}{\text{argmax}} P(t_j|t_{i-1}, w_i)$$

- Through some simplifying Markov assumptions that we will give below, we the previous equation to give the basic HMM equation for a single tag as follows:

$$t_i = \underset{j}{\text{argmax}} P(t_j|t_{i-1}) P(w_i|t_j)$$

Transformation-based Tagging using Brill Taggers
Analog – Oil Painting

Transformation-based Tagging (also called "Brill Taggers") was invented by Brill (1995) which is the direct implementation of so-called Transformation-Based Learning (TBL) which is based on the integration of both rule-based and stochastic-based method.

What is Transformation-based Learning?

Oil Painting usually using "Layering-and-Refinement" method painting approach:

1. Start with the background "theme", such as sky or household background
2. Paint the background first
3. Paint the "main theme" or "object" over the background
4. Refine the "main theme" or "object" to make it more precise
5. Further "refine" the "objects" or "main theme" until perfect.

For example:

1. Human portrait painting such as the famous "Mona Lisa" or the "Last Supper" from Leonardo da Vinci.
2. Figure 3.13 for the oil paint of a sunflower in a vase.

Fig. 3.14 Oil Painting Analog to Brill Tagger Transformation Technique

Hybrids – the Brill Tagger

Learning rules stochastically: Transformation Based Learning

Step 1: Assign each word the tag that is most likely given no contextual information.

Race example: P(NN|*race*) = .98 P(VB|*race*) = .02

Step 2: Apply transformation rules that use the context that was just established.

Race example: Change NN to VB when the previous tag is TO.

Example:

Secretariat is expected to race tomorrow. *-- Change the tag of race from NN to VB*
The race is already over. *-- No change, race remains as NN*

Learning Brill Tagger Transformations

Three major stages:

1. Label every word with its most-likely tag.
2. Examine every possible transformation and select the one with the most improved tagging.
3. Retag the data according to this rule.

These three stages are repeated until some stopping point is reached.

In terms of POS tagging, the output of TBL is an ordered list of transformations, which constitute a tagging procedure that can be applied to a new corpus.

Fig. 3.15 shows sample rules by using Brill's TBL scheme.

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**.

Fig. 3.15 Sample rules used in Brill's TBL Scheme

Evaluation of POS Taggers
Confusion Matrix for Error Analysis

Major consideration for the evaluation of POS Taggers:

- Given an algorithm, how good is it?
- What is causing the errors?
- Can anything be done about them?

The confusion matrix suggests that some major problems facing current taggers are:

1. NN versus NNP versus JJ:
   These are hard to distinguish. Distinguishing proper nouns is especially important for information extraction and machine translation.
2. RP versus RB versus IN:
   All of these can appear in sequences of satellites immediately following the verb.
3. VBD versus VBN versus JJ:
   Distinguishing these is important for partial parsing (participles are used to find passives),and for correctly labeling the edges of noun-phrases.

Fig. 3.16 shows the confusion matrix from the HMM error analysis of Adventures of Sherlock Holmes. For example, 7.56% of by mis-tagging a NN by JJ, which is commonly occurred in many texts.

	IN	JJ	NN	NNP	RB	VBD	VBN
IN		0.18			0.56		
JJ	0.32		4.35	3.21	2.25	0.31	2.54
NN		7.56					0.35
NNP	0.31	3.12	5.23		0.15		
RB	2.45	3.21	0.43				
VBD		0.56	0.52				4.31
VBN		3.21				2.12	

Fig 3.16 Confusion matrix from the HMM of Adventures of Sherlock Holmes

How Good is An POS Tagging Algorithm?

- How good is the POS Tagging algorithm?
- What's the maximum performance we have any reason to believe is achievable? (How well can people do?)
- How good is good enough?

Is 97% good enough?

- Example 1: A speech dialogue system correctly assigns a meaning to a user's input 97% of the time.
- Example 2: An OCR systems correctly determines letters 97% of the time.

Summary

- Part-of-speech tagging is the process of assigning a part-of-speech label to each of a sequence of words. Taggers can be characterized as rule-based or stochastic.
- Rule-based taggers use hand-written rules to distinguish tag ambiguity.
- Stochastic taggers are either HMM-based, choosing the tag sequence which maximizes the product of word likelihood and tag sequence probability, or cue-based, using decision trees or maximum entropy models to combine probabilistic features.
- Brill's Taggers provides a hybrid POS Tagging solution with the integration of both rule-based and stochastic-based tagging scheme, such machine learning scheme based on transformation learning not only useful for POS Tagging, but also useful for other NLP process and application such as the training of Q&A chatbots.
- Taggers are often evaluated by comparing their output from a test-set to human labels for that test set.
- Error analysis can help pinpoint areas where a tagger doesn't perform well.

Next

Syntax & Parsing