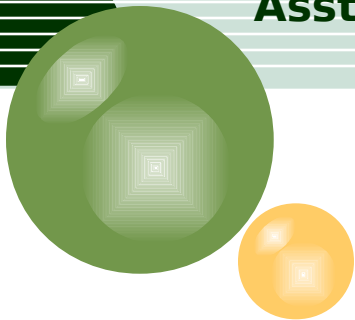


# *Word Classes [English]*

**By:**

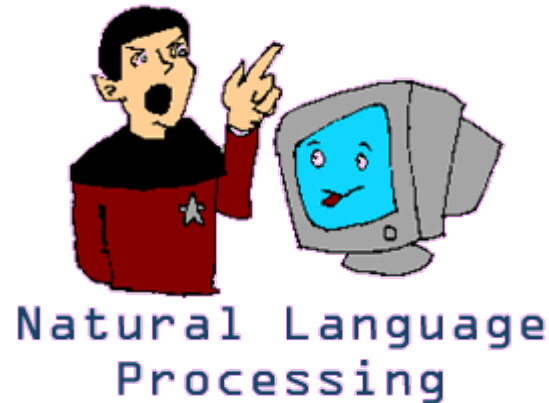
**B. SENTHIL KUMAR**

**Asst. Prof / CSE, SSNCE**



# Overview

- ❖ Background
- ❖ English word classes
  - Noun
  - Verb
  - Adjectives
  - Adverbs
  - Preposition
  - Participles
  - Conjunctions
  - Pronouns
  - Auxiliary verbs
  - Others



# Background

- **Part of speech:**
  - Noun, verb, pronoun, preposition, adverb, conjunction, participle, and article
- Recent lists of POS (also know as *word classes*, *morphological class*, or *lexical tags*) have much larger numbers of word classes.
  - 45 for Penn Treebank
  - 87 for the Brown corpus, and
  - 146 for the C7 tagset
- The significance of POS for language processing is that it gives a significant amount of information about the word and its neighbours.

# Background

- Possessive pronouns (*my, your, his, her*) are likely to be followed by a noun
- Personal pronouns (*I, you, he, me*) are likely to be followed by a verb
- A word's part-of-speech can tell us something about how the word is pronounced
  - The noun is pronounced *CONtent* and the adjective *conTENT*
  - *OBject* (noun) and *obJECT*(verb)
- POS can be used in stemming for IR, since
  - Knowing a word's POS can help tell us which morphological affixes it can take.

# English Word Classes

- To look more complete definition of the classes of POS.
  - Traditionally, the definition of POS has been based on morphological and syntactic function.
    - words that function similarly with respect to the affixes they take
    - words that function similarly with respect to what can occur nearby
  - While, it has tendencies toward semantic coherence (e.g., nouns describe “*people, places, or things*” and adjectives describe properties), this is not necessarily the case.
- Two broad subcategories of POS:
  1. Closed class
  2. Open class

# English Word Classes

- Closed class
  - Have relatively fixed membership, e.g., prepositions
  - Function words like *of*, *it*, *and*, or *you*, which tend to be very short, occur frequently.
- Open class
  - Four major open classes occurring in the languages of the world: nouns, verbs, adjectives, and adverbs.
    - Nouns and verbs are continually coined or borrowed from other languages (eg: *to fax*)
    - Many languages have no adjectives, e.g., the native American language Lakota, and Chinese

# Open Class: Noun

- Noun
  - The name given to the lexical class in which the words for most people, places, or things occur
  - Since lexical classes like noun are defined functionally (morphological and syntactically) rather than semantically,
    - some words for people, places, or things may not be nouns, and
    - some nouns may not be words for people, places, or things.
  - Thus, nouns include
    - Concrete terms, like *ship*, and *chair*,
    - Abstractions like *bandwidth* and *relationship*, and
    - Verb-like terms like *pacing* (*His pacing to and fro became quite annoying*)

# Open Class: Noun

- Noun
  - Define noun in English
    - Ability to occur with determiners (*a goat, its bandwidth, Plato's Republic*),
    - To take possessives (*IBM's annual revenue*), and
    - To occur in the plural form (*goats, abaci*) – but not all nouns



# Open Class: Noun

- Nouns are grouped into **proper nouns** and **common nouns**
  - **Proper nouns:**
    - *Rama, Chennai, and IBM* (names of specific persons or entities)
    - Not preceded by articles, e.g., *the book is upstairs*, but *Rama is upstairs*
    - Usually capitalized
  - Common nouns are divided into count nouns and mass nouns

# Open Class: Noun

- Nouns are grouped into **proper nouns** and **common nouns**
  - **Common nouns**
    - **Count nouns:**
      - Allow grammatical enumeration, i.e., both singular and plural (*goat/goats*), and can be counted (*one goat / two goats*)
    - **Mass nouns:**
      - Something is conceptualized as a homogeneous group, *snow, salt, and communism* not counted (\*two snow)
      - Appear without articles where singular count nouns cannot (*Snow is white* but not \**Goat is white*)

# Open Class: Verb

- **Verbs**

- Most of the words referring to actions and processes including main verbs like *draw, provide, differ, go*.
- A number of morphological forms: non-3rd-person-sg (*eat*), 3rd-person-sg (*eats*), progressive (*eating*), past participle (*eaten*)
- A subclass: **auxiliaries** (discussed in closed class)

# Open class: Adjectives

- **Adjectives**

- Terms describing properties or qualities
- Adjectives are said to modify nouns
  - large elephant
  - yellow cab
  - pale face
  - old documents
  - friendly teacher
- Most languages have adjectives for the concepts of color (*white, black*), age (*old, young*), and value (*good, bad*), but
- There are languages without adjectives, e.g., Chinese.

# Open class: Adverbs

- **Adverbs:** Words viewed as modifying something (often verbs)
  - **Directional (or locative) adverbs:** specify the direction or location of some action - *home, here, downhill, there*
  - **Degree adverbs:** specify the extent of some action, process, or property - *extremely, very, somewhat, farthest*
  - **Manner adverb:** describe the manner of some action or process - *slowly, slinkily, delicately*
    - Adverbs can also modify adjectives / adverbs
    - *delicately* designed dress, *badly* trained dogs
  - **Temporal adverbs:** describe the time that some action or event took place - *yesterday, now, daily*

# Closed Classes

- Some important closed classes in English:
  - **Prepositions:** on, under, over, near, by, at, from, to, with
  - **Determiners:** a, an, the
  - **Pronouns:** she, who, I, others
  - **Conjunctions:** and, but, or, as, if, when
  - **Auxiliary verbs:** can, may, should, are
  - **Particles:** up, down, on, off, in, out, at, by
  - **Numerals:** one, two, three, first, second, third

# Closed Classes: Prepositions

- **Prepositions** occur before nouns, semantically they are relational
  - Indicating spatial or temporal relations, whether literal (*on it, before then, by the house*) or metaphorical (*on time, with guest, beside herself*)
  - Other relations as well [*Hamlet was written by Shakespeare*]

of	540,085	through	14,964	worth	1,563	pace	12
in	331,235	after	13,670	toward	1,390	nigh	9
for	142,421	between	13,275	plus	750	re	4
to	125,691	under	9,525	till	686	mid	3
with	124,965	per	6,515	amongst	525	o'er	2
on	109,129	among	5,090	via	351	but	0
at	100,169	within	5,030	amid	222	ere	0
by	77,794	towards	4,700	underneath	164	less	0
from	74,843	above	3,056	versus	113	midst	0
about	38,428	near	2,026	amidst	67	o'	0
than	20,210	off	1,695	sans	20	thru	0
over	18,071	past	1,575	circa	14	vice	0

*Preposition (and particles) of English from CELEX*

# Closed Classes: Particles

- A **particle** is a word that resembles a preposition or an adverb
- Often combines with a verb to form a larger unit call a **phrasal verb**

So I *went on* for some days cutting and hewing timber ...

Moral reform is the effort to *throw off* sleep ...

- These are particles rather than prepositions
- *on* is followed by a true preposition phrase, not by a noun phrase
- *off* is a particle which appear after their objects (*throw sleep off*, *throw off sleep*), but not possible for prepositions (*The horse went off its track*, but \**The horse went its track off*)



# Closed Classes: Particles

- Extremely hard to automatically distinguish particles from prepositions
- Some tagsets in *CELEX* do not distinguish them

aboard	aside	besides	forward(s)	opposite	through
about	astray	between	home	out	throughout
above	away	beyond	in	outside	together
across	back	by	inside	over	under
ahead	before	close	instead	overhead	underneath
alongside	behind	down	near	past	up
apart	below	east, etc.	off	round	within
around	beneath	eastward(s),etc.	on	since	without

*English single-word particles from Quirk, et al (1985)*

# Closed Classes: Articles & Conjunctions

- English has three articles: *a*, *an*, and *the*
  - Articles begin a noun phrase.
  - *a* and *an* mark a noun phrase as indefinite, *the* can mark it as definite
- ❖ **Conjunctions** are used to join two phrases, clauses, or sentences.
  - Coordinating conjunctions like *and*, *or*, *but* join two elements of equal status
  - Subordinating conjunctions are used when one of the elements is of some sort of embedded status. *I thought **that** you might like some milk...***complementizers** (linking verb to its argument)

# Closed Classes: Conjunctions

and	514,946	yet	5,040	considering	174	forasmuch as	0
that	134,773	since	4,843	lest	131	however	0
but	96,889	where	3,952	albeit	104	immediately	0
or	76,563	nor	3,078	providing	96	in as far as	0
as	54,608	once	2,826	whereupon	85	in so far as	0
if	53,917	unless	2,205	seeing	63	inasmuch as	0
when	37,975	why	1,333	directly	26	insomuch as	0
because	23,626	now	1,290	ere	12	insomuch that	0
so	12,933	neither	1,120	notwithstanding	3	like	0
before	10,720	whenever	913	according as	0	neither nor	0
though	10,329	whereas	867	as if	0	now that	0
than	9,511	except	864	as long as	0	only	0
while	8,144	till	686	as though	0	provided that	0
after	7,042	provided	594	both and	0	providing that	0
whether	5,978	whilst	351	but that	0	seeing as	0
for	5,935	suppose	281	but then	0	seeing as how	0
although	5,424	cos	188	but then again	0	seeing that	0
until	5,072	supposing	185	either or	0	without	0

*Coordinating and subordinating conjunctions of English  
From the CELEX on-line dictionary.*

# Closed Classes: Pronouns

- **Pronouns** act as a kind of shorthand for referring to some noun phrase or entity or event.
  - **Personal pronouns:** persons or entities (*you, she, I, it, me*)
  - **Possessive pronouns:** forms of personal pronouns indicating actual possession or just an abstract relation between the person and some objects (*my, your, his, her, its, one's, our*)
  - **Wh-pronouns:** (*what, who, whom, whoever*) used in certain question forms, or may also act as complementizers (*Ram, who I met five years ago...*)

it	199,920	how	13,137	yourself	2,437	no one	106
I	198,139	another	12,551	why	2,220	wherein	58
he	158,366	where	11,857	little	2,089	double	39
you	128,688	same	11,841	none	1,992	thine	30
his	99,820	something	11,754	nobody	1,684	summat	22
they	88,416	each	11,320	further	1,666	suchlike	18
this	84,927	both	10,930	everybody	1,474	fewest	15
that	82,603	last	10,816	ourselves	1,428	thyslf	14
she	73,966	every	9,788	mine	1,426	whomever	11
her	69,004	himself	9,113	somebody	1,322	whosoever	10
we	64,846	nothing	9,026	former	1,177	whomsoever	8
all	61,767	when	8,336	past	984	wherefore	6
which	61,399	one	7,423	plenty	940	whereat	5
their	51,922	much	7,237	either	848	whatsoever	4
what	50,116	anything	6,937	yours	826	whereon	2
my	46,791	next	6,047	neither	618	whoso	2
him	45,024	themselves	5,990	fewer	536	aught	1
me	43,071	most	5,115	hers	482	howsoever	1
who	42,881	itself	5,032	ours	458	thrice	1
them	42,099	myself	4,819	whoever	391	wheresoever	1
no	33,458	everything	4,662	least	386	you-all	1
some	32,863	several	4,306	twice	382	additional	0
other	29,391	less	4,278	theirs	303	anybody	0
your	28,923	herself	4,016	wherever	289	each other	0
its	27,783	whose	4,005	oneself	239	once	0
our	23,029	someone	3,755	thou	229	one another	0
these	22,697	certain	3,345	'un	227	overmuch	0
any	22,666	anyone	3,318	ye	192	such and such	0
more	21,873	whom	3,229	thy	191	whate'er	0
many	17,343	enough	3,197	whereby	176	whenever	0
such	16,880	half	3,065	thee	166	whereof	0
those	15,819	few	2,933	yourselves	148	whereto	0
own	15,741	everyone	2,812	latter	142	whereunto	0
us	15,724	whatever	2,571	whichever	121	whichsoever	0

*Pronouns of English from the  
CELEX on-line dictionary.*

# Closed Classes: Auxiliary verbs

- **Auxiliary verbs:** mark certain semantic feature of a main verb, including
  - whether an action takes place in the present, past or future (tense),
  - whether it is completed (aspect),
  - whether it is negated (polarity), and
  - whether an action is necessary, possible, suggested, desired (mood).
  - Including **copula** verb *be*, the two verbs *do* and *have* along with their inflection forms, as well as a class of **modal verbs**.
  - **Verb** *have* used to mark the perfect tenses – *I have gone, I had gone*
  - *Be* is used as part of the passive construction – *We were robbed* or part of progressive – *We are leaving*

# Closed Classes: Auxiliary verbs

- **Modals** used to mark the mood associated with event or action depicted by main verb.
  - *can* indicates ability or possibility
  - *may* indicates permission or possibility
  - *must* indicates necessity

can	70,930	might	5,580	shouldn't	858
will	69,206	couldn't	4,265	mustn't	332
may	25,802	shall	4,118	'll	175
would	18,448	wouldn't	3,548	needn't	148
should	17,760	won't	3,100	mightn't	68
must	16,520	'd	2,299	oughtn't	44
need	9,955	ought	1,845	mayn't	3
can't	6,375	will	862	dare	??
have	???				

*English modal verbs from the CELEX on-line dictionary.*

# Closed Classes: Others

- **Interjections:** *oh, ah, hey, phew, alas*
- Negatives: *no, not*
- Politeness markers: *please, thank you*
- Greetings: *hello, goodbye*
- Existential there: *there are two on the table*



# References

- *Speech and Language Processing*, Jurafsky and H.Martin

**Thank you!**

