MODULE 3

Module 3: Syntax Analysis

Syntax Analysis

- Syntax refers to the arrangement of words in a sentence such that they make grammatical sense.
- In NLP, Syntactic analysis is used to assess how the natural language aligns with the grammatical rules.
- Syntactic analysis helps us to understand the roles played by different words in a body of text.
- Example:
- Innocent peacefully children sleep little
- Vs.
- Innocent little children sleep peacefully.

1. Part-Of-Speech Tagging(POS)

- It is a process of assigning corresponding part of speech like noun, verb, adverb, adjective to each word in a sentence.
- It convert a sentence to forms-list of words, list of tuples (word, tag)
- A tag signifies whether the word is a noun, adjective , verb and so on.
- Automatic assignment of descriptors to the given tokens is called Tagging. And descriptor is called Tag.
- The Tag may indicate Part-of Speech or semantic information.
- POS tagging is applied to language grammatical rules to parse meanings of sentences and phrases.
- The main challenge in POS tagging is to resolving the ambiguity in possible POS tags for a word.

E.g. Our dog chased a brown cat away from the home.

Word	POS Category
Our	Pronoun
a, the	Determiners
dog, cat, home	Nouns
brown	Adjective
chased	verb (past tense)
away	Adverb
from	preposition

Text: **The boy put the toys in the bag.**

POS Category	Words
Noun (N)	boy, toys, bag
Verb (V)	put
Preposition(P)	in
Determiner (Det)	the

- Example:
- I am reading a book (noun).
- Book(verb) that flight.
- The word book in the above sentences have the same spelling but different meanings.
- Same is true for other languages, for example the hindi word “sona” may mean gold(noun) or sleep (verb), however only one possible is used at a time.
- In tagging we try to determine correct category of word in its context.

- **Part-of-Speech Categories:**
- 1. Closed class
- 2. Open class

- Closed classes are those that have relatively fixed set of words. For Example: prepositions are closed class because there is fixed set of them in English.
- Nouns and verbs are open classes because new nouns and verbs are continually added or borrowed from other languages.
- Closed classes differ more from language to language than open classes.
- Closed class examples:
- Prepositions: on, under, over, near, by, at, from, to ,with- occur before noun
- Determiners/ Articles: a, an, the – often begin noun phrase.
- Pronouns: She, who, I, Others- referring to some noun phrase or entity or event.
- Conjunctions: and, but, or, as, if, when- used to join two phrases, clauses and sentences.
- Auxiliary verbs: can, may, should, are - a type of verb that takes a supportive role in a sentence, second to the main verb.
- Particles: up, down, on, off, in, out, at, by- combine with verb to form a larger unit called phrasal verb.
- Numerals: one, two, three, first, second, third

- There are four major open classes that occur in languages of the world: Noun, verbs, adjectives, and adverbs
- A) Noun: Noun is the name given to the lexical class in which the words for most people, places, and things occur.
- 1. Proper Noun: are names of specific persons or entities.
- Proper nouns are usually capitalized and are not preceded by articles.
- Ex: Tina, Delhi.
- 2. Common Noun: Common nouns are those nouns that refer to a generic item, group or place.
- Common nouns are not capitalized unless they appear at the beginning of a sentence.
- Ex: pen, school
- Types: Singular Noun, Plural Noun, Countable noun, uncountable noun, collective noun, abstract noun.

- B) Verbs: words referring to actions and processes.
- 1. main verbs: draw, provide, eat, happen, etc.

- C) Adjective: words that describe the qualities or states of being of nouns: silly, yellow, fun, fast.
- They can also describe the quantity of nouns: many, few, millions, eleven.

- D) adverbs: An adverb is a word that modifies (describes) a verb (“he sings loudly”), an adjective (“very tall”), another adverb (“ended too quickly”), or even a whole sentence (“Fortunately, I had brought an umbrella.”).
- Adverbs often end in -ly,
- Directional and locative adverbs (home, here, downhill): specify direction or location od some action.

- Degree Adverbs (Extremely, very, somewhat): specify the extent of some action, process or property.
- Manner Adverbs (slowly, delicately): Describes the manner of some action or process.
- Temporal adverbs(Yesterday, Monday): Describe the time that some action or event took place.

Tag Set for English (Penn Treebank)

- Examples:
- 1. Brown Corpus tagset (87 tags)
- 2. Penn Treebank Tagset (45 tags)
- 3. C7 tagset (147 tags)

- A tagset is a list of part-of-speech tags, i.e used to indicate the part of speech and often also other grammatical categories of each token in a text corpus.
- The English Penn Treebank tagset is used with English corpus.
- Example: The grand jury commented on a number of other topics.
- The/DT grand/JJ jury/NN commented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS

Noun Types

POS Type	Explanation	Examples
NN	Singular Common Noun	Woman, Orange, Table
NNS	Plural Common Noun	Women, Oranges, Tables
NNP	Singular Proper Noun	Priya, Zenith, Jack
NNPS	Plural Proper Noun	Indians, Americas

Verb Types

POS Type	Explanation	Examples
VB	Base form of a Verb	walk, play, eat, read
VBD	Past Tense of a Verb	Walked, played, ate, read
VBN	Past Participle of a Verb	Walked, played, eaten, read
VBG	Gerund form of a Verb	fishing, walking, reading
VBZ	3rd person verb on present tense	Walks, plays, eats, read, is
VBP	Non 3rd person verb on present tense	Walk, play, eat, read, am, are
MD	Modal Verb	Can, may, should

Adjective & Adverb Types

POS Type	Explanation	Examples
JJ	Adjective	Intelligent, small, fast
JJR	Comparative Adjective	Better, smaller, faster
JJS	Superlative Adjective	Best, smallest, fastest
RB	Adverb	Back, behind, fast, slow
RBR	Comparative Adverb	Slower, faster
RBS	Superlative Adverb	Slowest, fastest

Pronoun, Determiner, Preposition Types

POS Type	Explanation	Examples
PRP	Pronoun	He, she, they, I, we
PRP$	Possessive Pronoun	His, her, your, our
POS	Possessive Marker	India's, Asian's
DT	Determiner	The, a
CC	Coordinating Conjunction	And, or, also, but
IN	Preposition	In, under, of, from, with
CD	Cardinal Number	20, two

Statement: **I am a girl.**

Word	POS Type	Explanation
I	PRP	Pronoun
am	VBP	Non 3rd person verb on present tense
a	DT	Determiner
girl	NN	Singular Common Noun

Statement: **Kavya is a intelligent girl.**

Word	POS Type	Explanation
Kavya	NNP	Singular Proper Noun
is	VBZ (Aux. verb)	3rd person verb on present tense
a	DT	Determiner
intelligent	JJ	Adjective
girl	NN	Singular Common Noun

Statement: **She plays tennis.**

Word	POS Type	Explanation
She	PRP	Pronoun
plays	VBZ	3rd person verb on present tense
tennis	NN	Singular Common Noun

Statement: **They play football.**

Word	POS Type	Explanation
They	PRP	Pronoun
play	VBP	Non 3rd person verb on present tense
football	NN	Singular Common Noun

Penn Treebank Part-of-Speech 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	*eat*
JJS	Adj., superlative	*wildest*	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*	WP$	Possessive wh-	*whose*
NNS	Noun, plural	*llamas*	WRB	Wh-adverb	*how, where*
NNP	Proper noun, singular	*IBM*	$	Dollar sign	*$*
NNPS	Proper noun, plural	*Carolinas*	#	Pound sign	#
PDT	Predeterminer	*all, both*	“	Left quote	(‘ or “)
POS	Possessive ending	*’s*	”	Right quote	(’ or ”)
PRP	Personal pronoun	*I, you, he*	(	Left parenthesis	([, (, {, <)
PRP$	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*			

- The input to a tagging algorithm is a string of words and specified tagset.

  eg: Book/VB that/DT flight/NN
- Tagging algorithms automatically choose multiple tags for single word and select only one best appropriate tag for that word.
- Eg: word book can be consider as verb for book that flight and can be noun for please give me a book.

- Challenges in POS tagging:
- The problem with POS tagging is ambiguity.
- In English, many common words have multiple meanings hence multiple POS.
- The job of POS tagger is to resolve this ambiguity accurately based on the context of use.

- Examples of ambiguities in POS tagging
- The attack/NN was brutal.
- King was planning to attack/VB neighboring states.
- Tigers usually attack/VBP their prey in a group.

- On Sunday, I read two book/NNS.
- During winter season, he books/VBZ a flight a ticket to avail discount.
- They will book/VBP a flight on Sunday.