CSCE 771: Computer Processing of Natural Language
Lecture 8: Review Recent Topics, Quiz

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

13TH SEPTEMBER, 2022

Carolinian Creed: "I will practice personal and academic integrity."

Acknowledgement: Used materials by Jurafsky & Martin,

Organization of Lecture 8

- Opening Segment
  - Announcements
- Main Lecture
- Concluding Segment
  - About Next Lecture – Lecture 9

Main Section
- Review quiz
- Review parsing
- Introduce evaluation metrics in NLP context
- Review projects

6	Sep 8 (Th)	Statistical Parsing, QUIZ
7	Sep 13 (Tu)	Review Parsing, Quiz review, Review Project, Introduce Evaluation
8	Sep 15 (Th)	Semantics
9	Sep 20 (Tu)	Review: Machine Learning for NLP, Evaluation – Metrics
10	Sep 22 (Th)	Language Model – Vector embeddings, CNN/ RNN
11	Sep 27 (Tu)	Guest Lecture – Dr. Amitava Das: Glove, Word2Vec, Transformer Review: Reasoning for NLP
12	Sep 29 (Th)	Representation: Ontology, Knowledge Graph, QUIZ
13	Oct 4 (Tu)	Representation: Embeddings, Language Models
14	Oct 6 (Th)	Entity extraction
15	Oct 11 (Tu)	Guest Lecture – Dr. Amitava Das: Using lang models to solve NLP tasks

Announcements

GUEST LECTURES ON LANGUAGE MODELS BY DR. AMITAVA DAS

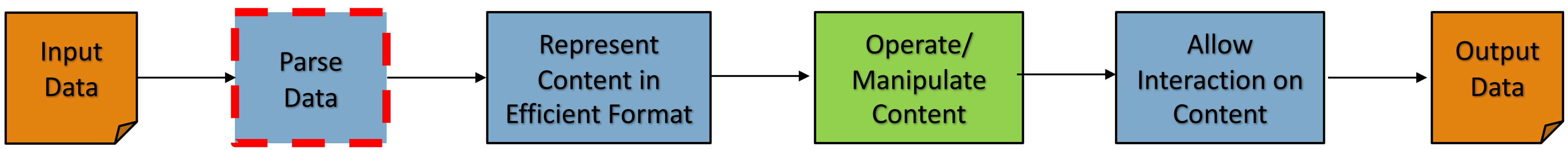
Recap of Lecture 7

- We discussed statistical parsing - Probabilistic grammars
  - assign a probability to a sentence or string of words
  - In a probabilistic context-free grammar (PCFG), every rule is annotated with the probability of that rule being chosen assuming conditional independence.
  - The probability of a sentence is computed by multiplying the probabilities of each rule in the parse of the sentence.
  - We looked at Stanford parser
- We had Quiz 1

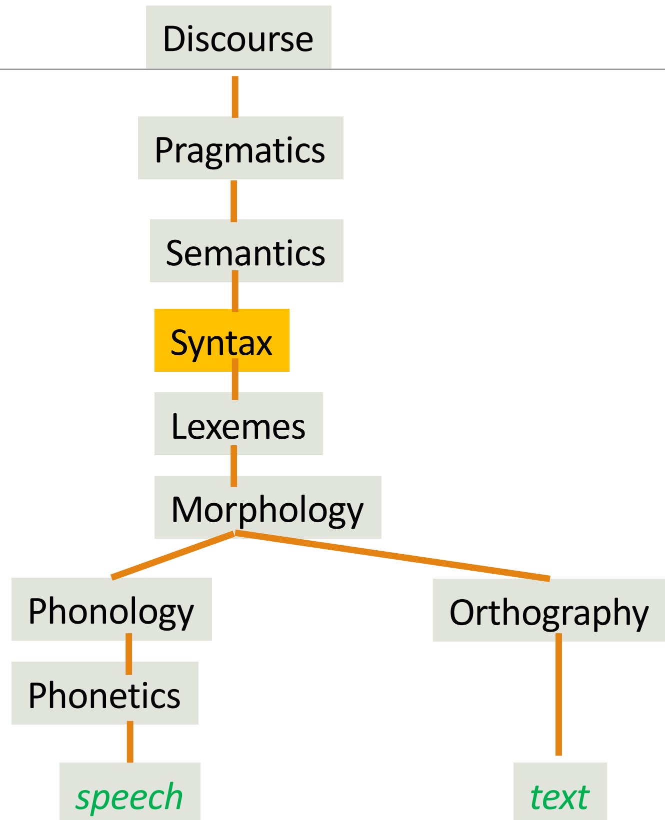
Review of Quiz 1

Main Lecture

Review Parsing



Levels of Linguistic Studies



- **Discourse:** study of group of sentences
- **Pragmatics:** how context contributes to meaning of sentences
- **Semantics:** meaning of words and combinations of words
- **Syntax:** rules for combining and using words/ phonemes.
- **Lexemes:** a set of words that are related through inflection (fly: verb, fly: noun)
- **Morphology**—rules that govern morphemes - the minimal meaningful units
  of language (lemmas and affixes)
- **Orthography**: convention for writing a language. E.g., spelling
- **Phonology:** organization of speech sound (i.e., phoneme)
- **Phonetics**: study of how sound is made and received

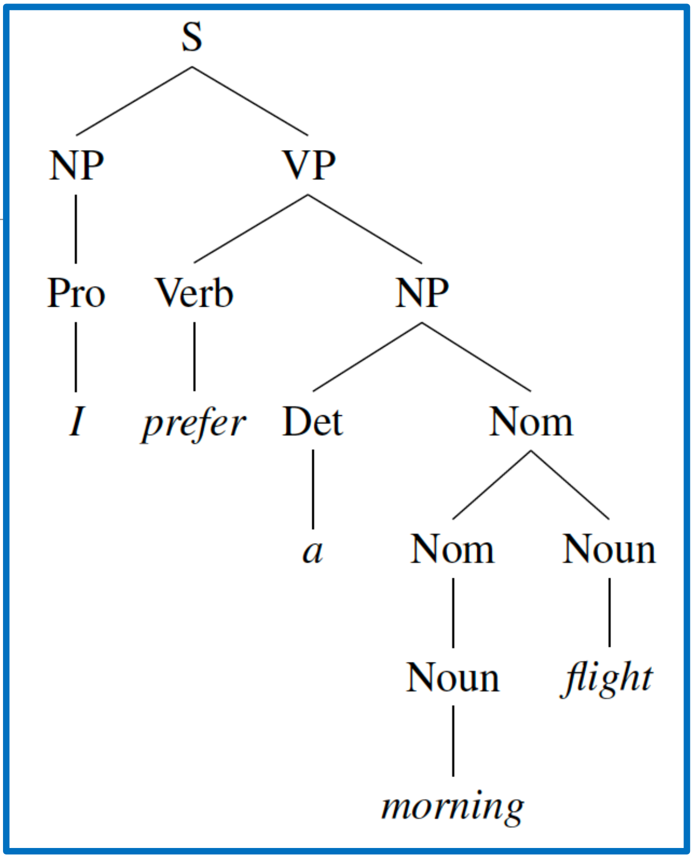
Why Parsing

- Recognizing legal inputs from illegal
- Usage of parse representation - parse tree
  - Grammar checking
  - Semantic analysis
  - Machine translation
  - Question answering
  - Information extraction
  - Speech recognition
  - …

Adapted from material by Robert C. Berwick

An Example Using CFGs

Grammar Rules			Examples
S	→	*NP VP*	I + want a morning flight
NP	→	*Pronoun*	I
	|	*Proper-Noun*	Los Angeles
	|	*Det Nominal*	a + flight
Nominal	→	*Nominal Noun*	morning + flight
	|	*Noun*	flights
VP	→	*Verb*	do
	|	*Verb NP*	want + a flight
	|	*Verb NP PP*	leave + Boston + in the morning
	|	*Verb PP*	leaving + on Thursday
PP	→	*Preposition NP*	from + Los Angeles



From Jurafsky & Martin

$[_S\ [_{NP}\ [_{Pro}\ \text{I}]]\ [_{VP}\ [_V\ \text{prefer}]\ [_{NP}\ [_{Det}\ \text{a}]\ [_{Nom}\ [_N\ \text{morning}]\ [_{Nom}\ [_N\ \text{flight}]]]]]]$

Bracketed Notation

Example: Larger English CFG

Grammar
$S \to NP\ VP\ .$
$S \to NP\ VP$
$S \to$ " S " $,\ NP\ VP\ .$
$S \to$ -NONE-
$NP \to DT\ NN$
$NP \to DT\ NNS$
$NP \to NN\ CC\ NN$
$NP \to CD\ RB$
$NP \to DT\ JJ\ ,\ JJ\ NN$
$NP \to PRP$
$NP \to$ -NONE-
$VP \to MD\ VP$
$VP \to VBD\ ADJP$
$VP \to VBD\ S$
$VP \to VBN\ PP$
$VP \to VB\ S$
$VP \to VB\ SBAR$
$VP \to VBP\ VP$
$VP \to VBN\ PP$
$VP \to TO\ VP$
$SBAR \to IN\ S$
$ADJP \to JJ\ PP$
$PP \to IN\ NP$

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

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

Types of Parsing

- **Phrase structure** / **Constituency Parsing**: find phrases and their recursive structure.
Constituency - groups of words behaving as single units, or constituents.
  - **Shallow Parsing/ Chunking**: identify the flat, non-overlapping segments of a sentence: noun phrases, verb phrases, adjective phrases, and prepositional phrases.
- **Dependency Parsing**: find relations in sentences
- **Probabilistic Parsing**: given a sentence X, predict the most probable parse tree Y

Chunking

- Chunking - process of identifying and classifying the flat, non-overlapping segments of a sentence that constitute the basic **non-recursive phrases** corresponding to the major content-word parts-of-speech:
  - noun phrases
  - verb phrases
  - adjective phrases, and
  - prepositional phrases

Example

[$_{NP}$ The morning flight] [$_{PP}$ from] [$_{NP}$ Denver] [$_{VP}$ has arrived.]

- Two operations in this type of parsing:
  - segmenting - finding the non-overlapping extents of the chunks and
  - labeling - assigning the correct tag to the discovered chunks
- Some words may not be part of any chunk

From Jurafsky & Martin

IOB notation

- Chunking - IOB tagging
  - B - beginning of each chunk type
  - I - inside of each chunk type
  - O - one for tokens outside (O) any chunk
- Total: (2N + 1) tags for N chunk types

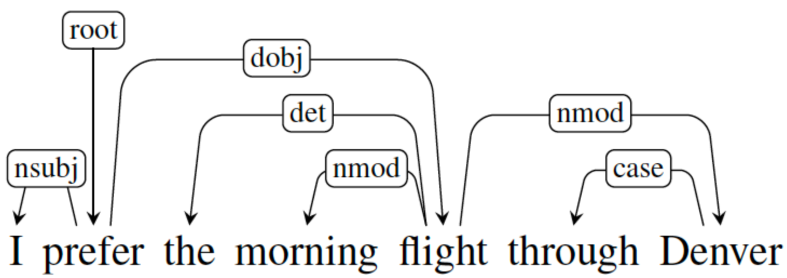
Example

The	*morning*	*flight*	*from*	*Denver*	*has*	*arrived.*
B_NP	I_NP	I_NP	O	B_NP	O	O

From Jurafsky & Martin

Dependency Parsing

- **Meaning** depends on
  - Words (lemmas) in a sentence
  - Their directed binary grammatical relations with other words
  - (and not on CFGs)
- **Notation**: Labeled arcs are from heads to dependents



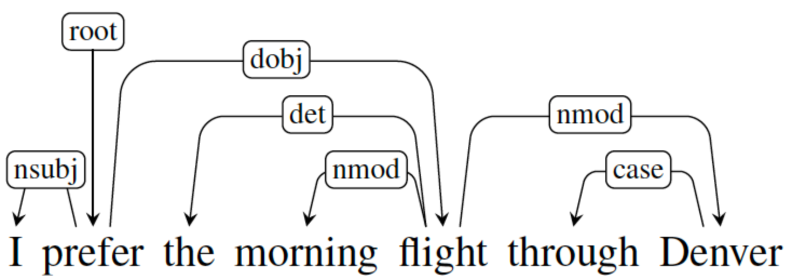
No node corresponding to phrasal constituents or lexical categories in the dependency parse

Dependency Conditions

1. There is a single designated root node that has no incoming arcs.
2. With the exception of the root node, each vertex has exactly one incoming arc.
3. There is a unique path from the root node to each vertex in V.

From Jurafsky & Martin

Dependency Parsing



Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	**Description**
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	**Description**
CONJ	Conjunct
CC	Coordinating conjunction

Edge: role that the dependent plays with respect to its head. Examples: subject, direct object and indirect object.

From Jurafsky & Martin

Decisions with PCFG

Probability of parse tree T, given sentence S, is

$$P(T,S) = \prod_{i=1}^{n} P(RHS_i|LHS_i)$$

Definition:
Yield of a parse tree = String of words allowed by parse tree

Of all parse trees with a yield of S, the disambiguation algorithm for parsing picks the parse tree that is most probable given S:

$$\hat{T}(S) = \underset{T s.t. S = \text{yield}(T)}{\text{argmax}} P(T|S) \longrightarrow \hat{T}(S) = \underset{T s.t. S = \text{yield}(T)}{\text{argmax}} P(T)$$

choosing the parse with the highest probability

From Jurafsky & Martin

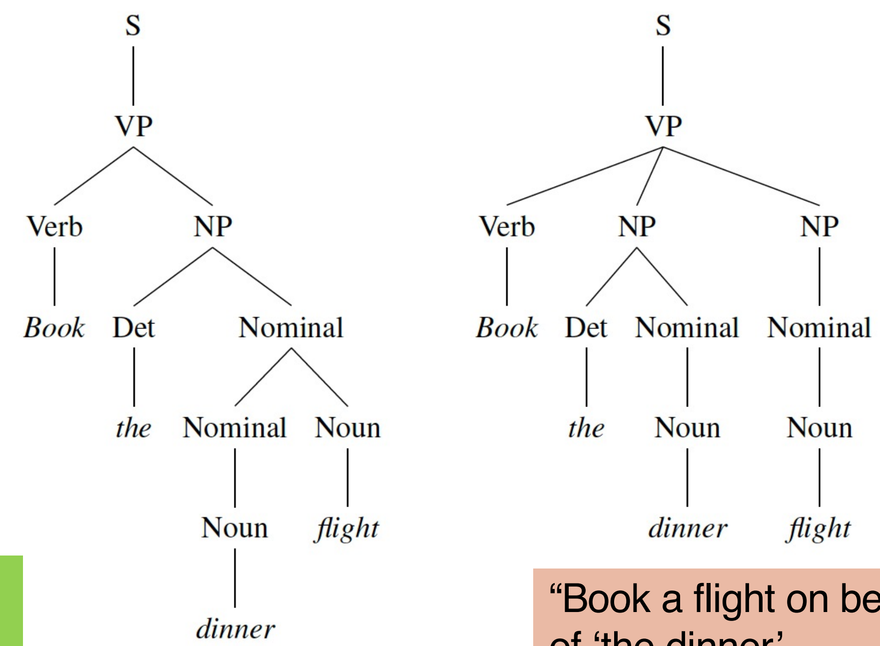
Interpretations of
"***Book the dinner flight***"

"Book a flight that serves dinner"

"Book a flight on behalf of 'the dinner'"

Example

Interpretations of
"***Book the dinner flight***"



"Book a flight that serves dinner"

"Book a flight on behalf of 'the dinner'"

$$P(T_{left}) = .05 * .20 * .20 * .20 * .75 * .30 * .60 * .10 * .40 = \mathbf{2.2 \times 10^{-6}}$$ ✓

$$P(T_{right}) = .05 * .10 * .20 * .15 * .75 * .75 * .30 * .60 * .10 * .40 = \mathbf{6.1 \times 10^{-7}}$$

Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flight	.40	Noun	→	flight	.40

From Jurafsky & Martin

Calculating Probability from Treebank

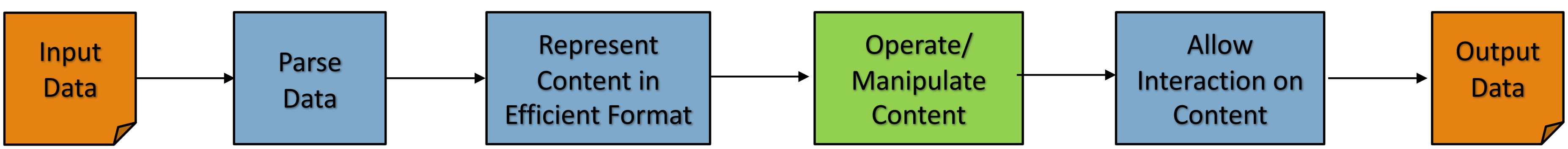
$$P(\alpha \to \beta|\alpha) = \frac{\text{Count}(\alpha \to \beta)}{\sum_{\gamma} \text{Count}(\alpha \to \gamma)} = \frac{\text{Count}(\alpha \to \beta)}{\text{Count}(\alpha)}$$

Probability of each expansion of a non-terminal:

- counting the number of times an expansion occurs
- normalizing for all expansions

From Jurafsky & Martin

NLP Evaluation



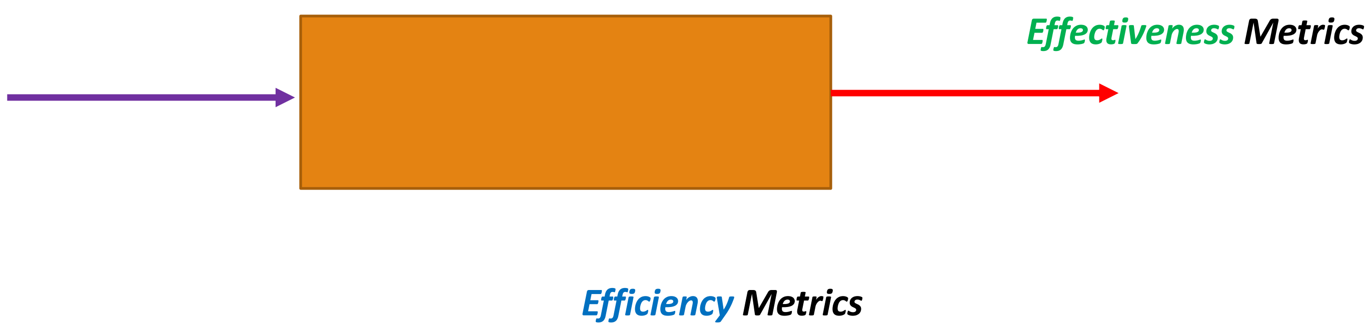
Metric Types

- **Effectiveness**: what the **user** of a system sees, primarily cares about

Extrinsic evaluation (esp. downstream applications)

- **Efficiency**: what the **executor** in a system sees, primarily cares about

Intrinsic evaluation



Example: Detecting Spam in Email

- **Effectiveness**: what the **user** of a system sees, primarily cares about
  - *How many spams identified?*
  - *How many spams missed?*
- **Efficiency**: what the **executor** in a system sees, primarily cares about
  - *How fast were spams detected?*
  - *How much memory was used per million emails processed ?*

Metrics: Accuracy, Precision, Recall

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Accuracy =
(TP+TN)/
(TP+FP+FN+TN)

Precision =
(TP)/
(TP+FP)

Recall =
(TP)/
(TP+FN)

F1 Score: *Harmonic Mean*

1/F1 = 1/Precision + 1/Recall

F1 = 2*(Recall * Precision) /
(Recall + Precision)

Evaluating Parsers - PARSEVAL

Degree to which the constituents in the hypothesis parse tree look like the constituents in a hand-labeled, gold-reference parse like PENN TreeBank

Overall measure is by F1 score

$$F_1 = \frac{2PR}{P+R}$$

$$\textbf{labeled recall:} = \frac{\text{\# of correct constituents in hypothesis parse of } s}{\text{\# of correct constituents in reference parse of } s}$$

$$\textbf{labeled precision:} = \frac{\text{\# of correct constituents in hypothesis parse of } s}{\text{\# of total constituents in hypothesis parse of } s}$$

From Jurafsky & Martin

Average Performance With Multiple Classes

- Setting
  - Class A: 1 TP and 1 FP
  - Class B: 10 TP and 90 FP
  - Class C: 1 TP and 1 FP
  - Class D: 1 TP and 1 FP

Precision =
(TP)/
(TP+FP)

- Average precision = ?

- Macro and micro average
  - A macro-average will compute the metric independently for each class and then take the average (hence treating all classes equally)
  - A micro-average will aggregate the contributions of all classes to compute the average metric.

A macro-average will then compute: $Pr = \frac{0.5+0.1+0.5+0.5}{4} = 0.4$

A micro-average will compute: $Pr = \frac{1+10+1+1}{2+100+2+2} = 0.123$

Source and credit: https://datascience.stackexchange.com/questions/15989/micro-average-vs-macro-average-performance-in-a-multiclass-classification-settin

Code Sample – Metrics Calculation

Notebook:

https://github.com/biplav-s/course-nl-f22/blob/main/sample-code/l8-review-evalmetrics/Metric%20Calculations.ipynb

Review of Course Projects

Course Project – Deadlines and Penalty Rubric

What is next ?

- Create project plan and put in your G-drive; project sub-dir; File name: “Project plan”. Extension: .docx or .pdf
- File will contain
  * Project Title
  * Description: motivation and expected output
  * Illustrative Test cases: i.e., Example input / output
  * Data sources
  * Technique and tools to use
  * Metric for measuring output
  * How will you collect results
  * Format of report, presentation
  * Time schedule, by Week
- Penalty: **not** ready by Sep 15, 2022 **[-20%]**

- Other penalties
  - Project report **not** ready by Nov 10, 2022 **[-20%]**
  - Project presentations **not** ready by Nov 15, 2022 **[-10%]**

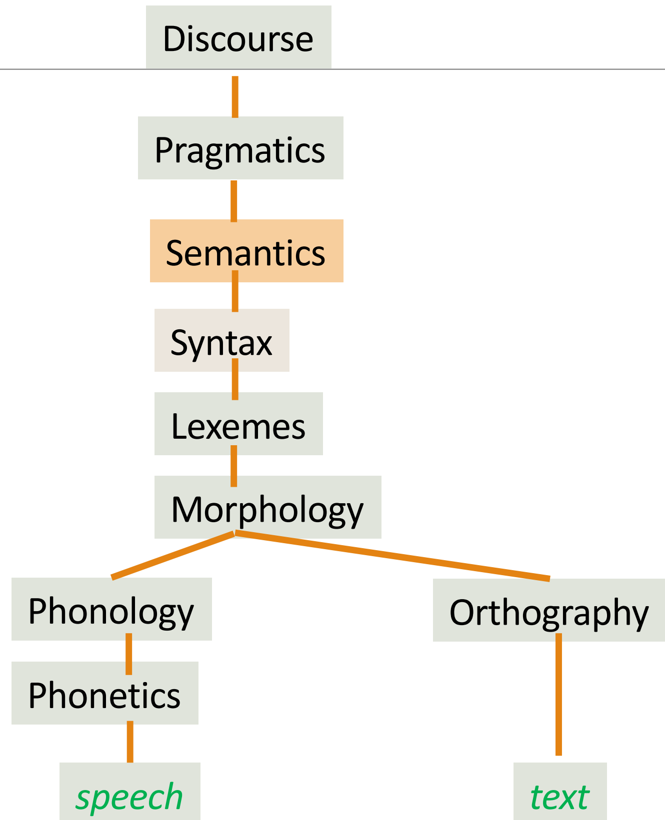
Lecture 8: Concluding Comments

- We reviewed
  - Quiz 1
  - Parsing
  - Projects
- We looked at evaluation measures
  - accuracy, precision, recall, F1
  - Macro and micro averages

Concluding Segment

About Next Lecture – Lecture 9

Lecture 9



- **Discourse:** study of group of sentences
- **Pragmatics:** how context contributes to meaning of sentences
- **Semantics: meaning of words and combinations of words**
- **Syntax:** rules for combining and using words/ phonemes.
- **Lexemes:** a set of words that are related through inflection (fly: verb, fly: noun)
- **Morphology**—rules that govern morphemes - the minimal meaningful units
  of language (lemmas and affixes)
- **Orthography**: convention for writing a language. E.g., spelling
- **Phonology:** organization of speech sound (i.e., phoneme)
- **Phonetics**: study of how sound is made and received