CSE556: NLP - Quiz 1 - Solutions

Answers are in red color.

1. Assign semantic-roles to the following sentence: “The old man the boat”. [2 Marks]

Ans. old - Agent, man - predicate (verb), boat - patient

2. Give an example to highlight the importance of punctuations to disambiguate the meaning of a sentence. [1 Mark]

Ans. No fixed answer.

If the example can be interpreted in more than one way without punctuation.

As per instructions, the following sentences are restricted as responses.

1. Let’s eat, Grandma! / Let’s eat Grandma!
2. A woman, without her man, is nothing / A woman: without her, man is nothing.
3. No parking! Violators will be towed / No parking Violators will be towed.

3. Define phonemes, graphemes, and morphemes. [3 Marks]

Ans.

Phonemes: Smallest distinctive unit sound of a language that distinguishes one word from another.

Graphemes: The smallest unit of writing system that represents a phoneme.

Morphemes: Morpheme is the smallest linguistic unit that has semantic meaning.

4. Which linguistic concept in the NLP hierarchy can you associate to the following sentence? “The use of shin bone is to locate furniture in a dark room.” [1 Mark]

Ans. Pragmatics.

For this quiz, if someone answered Pragmatics and Discourse, we will assign full marks. However, please note that Pragmatics and Discourse are different.

5. Select all correct options. [Note: Marks will be assigned only if all correct options are selected. No marks, if you fail to identify even one correct option] [2 Marks]

A. (aa*)* = a*
B. (aa*)* = a+
C. (a+)* = a+
D. (a+)* = a+ | epsilon
E. (a+)* = a* | epsilon

6. Define false-positives and false-negatives.

Ans.

False-positives: Incorrect identification/classification of something as “True” instead of “False”

False-negatives: Incorrect identification/classification of something as “False” instead of “True”

7. For the given merge rules [r$, er$, ew, new, lo, low, newer$, low$], tokenize the word "newest" using BPE. Show steps. [1 Marks]

Ans. n-e-w-e-s-t-$

n-ew-e-s-t-$

new-e-s-t-$

new e s t { tokenized - Four subword tokens}

8. Using Thompson's Algorithm, design a minimal DFA for the regular expression a*. Show all intermediate steps on paper. [Note: Marks will be awarded based on correct steps. Without step, the correct minimal DFA will fetch only 0.5 marks.] [3 Marks]

Ans.