

## Quiz1 (10.9.2020) with solutions

1. What is Natural Language Analysis? (multiple correct possible)
  - a) **Understanding of natural language input by the computer**
  - b) Conversion of structured data to natural language text
  - c) **Conversion of text to computer representation**
  - d) Generation of correct sentences
  
2. What does the Pragmatic level of NLP involve?
  - a) Deal with word order in sentences
  - b) Word understanding
  - c) Structure of Text
  - d) **Knowledge of outside world**
  
3. Coreference resolution is the task of finding all expressions that refer to the same entity in a text. Which level of language processing does it require? If multiple answers are possible indicate the most relevant one. (one answer)
  - a) Lexical Analysis
  - b) Morphological Analysis
  - c) Syntactic Analysis
  - d) Semantic Analysis
  - e) **Discourse Analysis**
  - f) Pragmatic Analysis
  
4. How many morphemes can you identify in the given words?
  - a) "Antidisestablishmentarianism"
    - a. **Correct answer 6**
  - b) "Demagnetizability"
    - a. **Answer: 5**
  
5. Write one Hindi word in Roman script having at least three morphemes and list the morphemes. You may use another language if you do not know Hindi

**Atma-nirbhar-sheel mita-byaye-ta etc**

6. These sentences have multiple interpretations. What type of ambiguity do they correspond to?

Possible ambiguities:

1. Part of speech Ambiguity
2. Lexical (word level)
3. Morphological (word decomposition)
4. Syntactic
5. Semantic

- a. Flying planes can be dangerous
  - i. POS
  - ii. Syntactic
- b. Local HS Dropouts Cut in Half
  - i. Semantic
- c. Stolen Paintings Found by Tree
  - i. Semantic

7. You are given the following data.

Vocabulary  $V = \{n_1, a_2, v_3\}$

Bigram probability distribution  $p$  on  $V \times V$  is given by:

$$p(n_1, n_1) = 0.25$$

$$p(a_2, a_2) = 0.0$$

$$p(v_3, v_3) = 0.25$$

$$p(a_2, n_1) = 0.125$$

$$p(n_1, v_3) = 0.25$$

$$p(n_1, *) = 0.5 \text{ (probability that } n_1 \text{ is the first of a pair)}$$

$$p(*, a_2) = 0.125 \text{ (probability that } a_2 \text{ is the last of a pair)}$$

Consider the following table for the bigram probabilities. The green cells show what is given.

	$n_1$	$a_2$	$v_3$	$(x, *)$
$n_1$	0.25		0.25	0.5
$a_2$	0.125	0		
$v_3$			0.25	
$(*, x)$		0.125		

The following values can be inferred (shown in blue and white)

	$n_1$	$a_2$	$v_3$	$(x, *)$
$n_1$	0.25	0	0.25	0.5
$a_2$	0.125	0	0	0.125
$v_3$	0	0.125	0.25	0.375
$(*, x)$	0.375	0.125	0.5	

$$p(n_1, *) = p(n_1, n_1) + p(n_1, a_2) + p(n_1, v_3)$$

Therefore  $p(n1, a2) = 0$

$$p(*, a2) = p(n1, a2) + p(a2, a2) + p(v3, a2)$$

Thus  $p(v3, a2) = 0.25$

Since all the 9 entries must sum to 1, the rest of the entries are 0.

7a. What is the probability of the sequence (n1, n1, v3) according to the bigram language model?

$$P(n1, n1, v3) = p(n1, *) * p(n1|n1) * p(v3|n1)$$

$$p(n1, *) = \frac{1}{2}$$

$$p(n1|n1) = \frac{p(n1, n1)}{p(n1, *)} = \frac{0.25}{0.5} = \frac{1}{2}$$

$$p(v3|n1) = \frac{p(n1, v3)}{p(n1, *)} = \frac{0.25}{0.5} = \frac{1}{2}$$

Therefore  $P(n1, n1, v3) = \frac{1}{8} = 0.125$

7b. Calculate the probability of  $p(a2|v3)$

$$p(a2|v3) = \frac{p(v3, a2)}{p(v3, *)} = \frac{0.125}{0.375} = \frac{1}{3}$$