Tutorial 6

1. Discuss the **THREE** (**3**) major obstacles involved in Natural Language Understanding (NLU).

Answer:

1. A large amount of human knowledge is assumed. Language acts describe relationships is an often complex world. Knowledge of these relationships must be part of any understanding system. Too much of assumptions make the understanding not "real".
2. Language is pattern based: phonemes are components of words and worls make phrases and sentences. Phoneme, word, and sentence orders are not random. Communication is impossible without a rather constrained use of these components. However, it is rather difficult to set the constraint of the relationships most of the time.
3. Limitations of programming tools make the coding becomes so difficult.
4. Ambiguity and Polysemy:Natural language is inherently ambiguous due to factors like polysemy (words having multiple meanings) and context-dependent interpretations. NLU systems must contend with identifying the correct meaning based on the surrounding context. Resolving ambiguity requires sophisticated techniques to disambiguate words and phrases to ensure accurate comprehension.
5. Communication with natural language heavily depends on our knowledge within the domain of disclosure, which involves transmission of words, knowledge, assumption and more.
6. Sentences formed with natural language cannot be understand through a simplistic, literal treatment of meaning as we need to consider the ambiguity or sarcasm meaning of a word or sentence..
7. Contextual Dependency and Inferences: Language meaning heavily relies on context, making contextual analysis a crucial aspect of NLU. Words or phrases may have different meanings depending on the surrounding text. Successful NLU involves not only understanding the immediate context but also making inferences about implied information and connections between different parts of the conversation.
8. Misspelled or misused words can create problems for text analysis. Autocorrect and grammar correction applications can handle common mistakes, but don't always understand the writer's intention.

Any other acceptable answers

2. What are the three typical stages involved in a Natural Language Processing (NLP) application development? Elaborate these three stages with the aid of appropriate diagrams.

Answer:

Parsing:

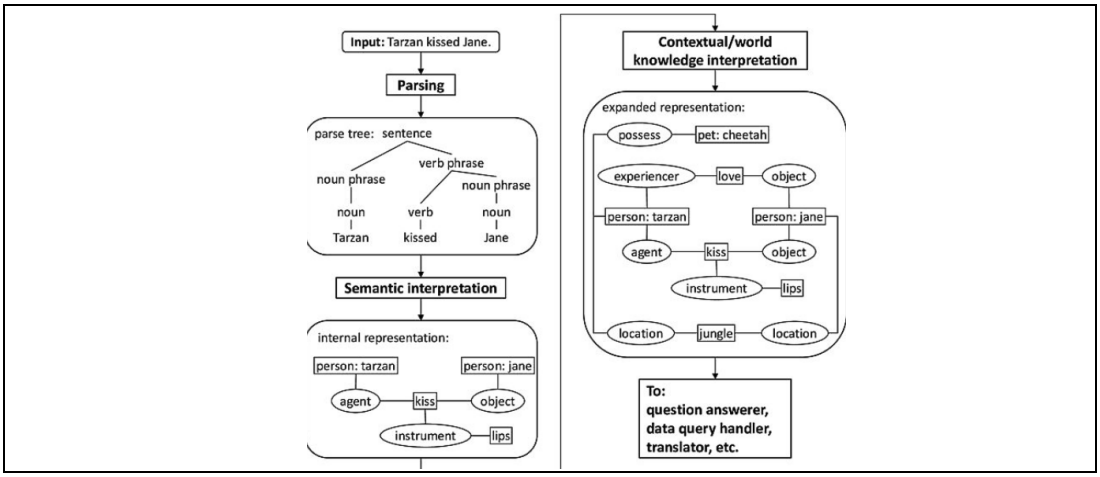
- Analyze the syntactic structure of sentences, by identifying the major relations such as subject-verb, verb-object, noun-modifier, e.g. Tarzan and Jane are nouns, and kiss is a verb.
- Often represented as parse tree
- Employs knowledge of language syntax, morphology, and some semantics.

Semantic interpretation:

- Produce a representation of the meaning of the text, such as conceptual graph, conceptual dependencies, frames, etc.
- Uses knowledge about the meaning of words and linguistic, e.g. the meaning of kiss is added "lips" as the instrument.
- Perform consistency checks, such as verb "kiss" may include constraints that only be performed by person to another person, but not tree.

World knowledge interpretation:

- Produce expanded representation of the sentence's meaning and add to knowledge base.
- Necessary for complete understanding such as Tarzan loves Jane, they live in jungle, Tarzan has pet Cheetah.
- The resulting representation is used for further processing such as handling database query, answering questions, translating the meaning, etc.



3. Besides the three main phases involved in NLP as discussed in Question 2, describe three other significant analyses that can be done to improve NLP applications.

Answer:

Refer lecture notes

1. Prosody – rhythm and intonation of language (mood, emotion, get out from this room)
2. Phonology – combination of sounds to form language (doubt)…focused on how speech sounds change and behave when in a syllable, word, or sentence
3. Morphology – components that make up words, such as prefixes (un-, non-, anti-) and suffixes (-ing, -ly) that modifying the meaning of root words
4. Syntax – rules for combining words into legal phrases and sentences
5. Semantics – the meaning of words, phrases, and sentences
6. Pragmatics – the study of the ways in which language is used and its effects on the listener. In contrast to semantic is more emphasis to the context/meaning of words, pragmatic refers to how words are being used in practical sense, words have different meaning at different situation, we will apply our understanding of symbols as we read or listen to others.

4. NLP application requires the use of knowledge about human languages. Suggest an example of NLP application and describe the difficulties that a researcher would face during the development of the application that you have suggested.

Answer:

Chatbot:

Spelling Variations:

Example: "color" (American English) vs. "colour" (British English)

Difficulty: Recognizing regional spelling preferences and suggesting the correct version based on the user's context or preference.

New Words and Slang:

Example: "LOL, that's so lit! "

Difficulty: Recognizing slang terms like "lit" and ensuring they are not flagged as errors.

Typographical Errors:

Example: "I am goiing to the party."

Difficulty: Detecting repeated letters ("goiing") and suggesting the correct form, "going."

Proper Nouns:

Example: "John Smith lives in New York."

Difficulty: Avoiding corrections to proper nouns like "John Smith" and "New York."

User Intent:

Example: "Their going to the park."

Difficulty: Understanding that "Their" should be corrected to "They're" based on the user's intent, despite the grammatical error.

False Positives and Negatives:

Example (False Positive): Correcting "I saw their cat" to "I saw there cat" (unnecessary correction).

Example (False Negative): Not correcting "I wont go" to "I won't go" (missed correction).

Machine Learning Models:

Difficulty: Developing and training machine learning models to balance correction accuracy without over-correcting or introducing errors in the text.

Any other acceptable answers: Idioms etc

5. In NLP, representation is important as it can solve issues like canonical form of sentences and syntactic problem of a sentence.
   a) Explain the meaning of canonical form of sentences. Provide examples to elaborate your answer.
   b) By referring to the below statement, there is a syntactic problem. Identify the problem and then draw **TWO** (**2**) different representation (by selecting either the semantic network or the conceptual graph) to solve the syntactic problem.

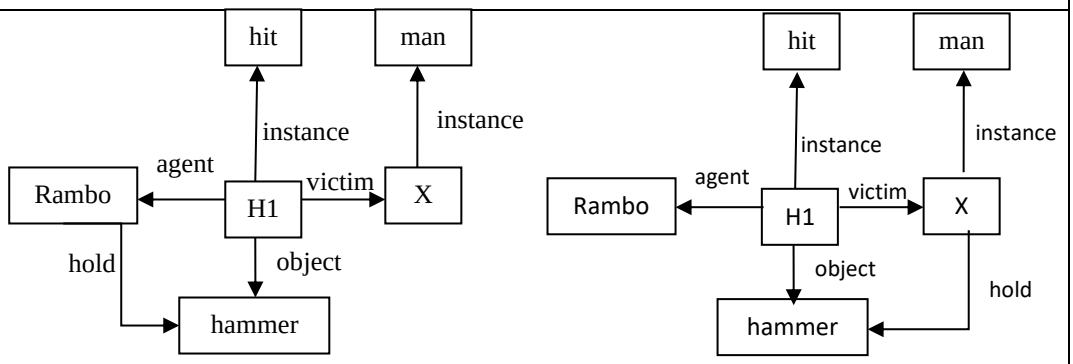
"Rambo hit the man with a hammer"

Answer:

a)

Canonical form of sentences means two different sentences have the same meaning. For example the sentences “Jen was given a gift by Sam” and “Sam gave Jen a gift” are having the same meaning, but written in different ways.

b)



1st diagram shows Rambo is the one who holds the hammer and hit the man

2nd diagram shows Rambo was hit by a man who holds a hammer

6. Parse tree is a popular tool used in one of the phases of NLP called parsing.
   a) Explain the importance of parsing in NLP.
   b) With the aid from the simple English grammar for simple transitive sentence as shown next page, draw the respective **parse trees** to verify the sentences "**the boy likes the girl**" and "**time flies like an arrow**"

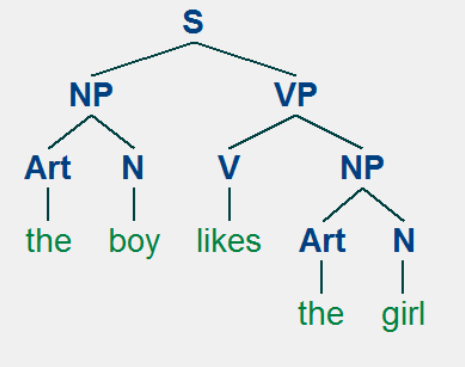
Sentence	-> noun_phrase verb_phrase
Noun phrase	-> noun
Noun-phrase	-> article noun
Verb phrase	-> verb noun_phrase
Verb phrase	-> verb preposition noun_phrase
preposition	[like]
article	[a, an, the]
noun	[flies, time, arrow, boy, girl]
verb	[like, flies]

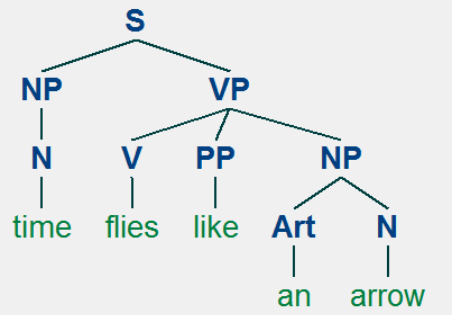
Answer:

a)

1. To analyze the syntactic structure of sentences (present, object can be used as noun or verb, diff meaning)
2. To identifying the major relations such as subject-verb, verb-object, noun-modifier

b)





7. Considering the sentence S = "**She beats George with one hand at the bank**".
   a) The sentence S consists of semantic ambiguity and syntactic ambiguity. Identify both of the ambiguities found from the sentence above.
   b) Given the grammar below, construct **ONE** (**1**) parse tree for the sentence S. (**Remark**: grammar in the parentheses **()** means it is optional.)

S -> NP VP (PP)
NP -> (DET) (ADJ) N (PP)
VP -> V NP (PP)
PP -> P NP
ADJ -> [one]
DET -> [a, the]
N -> [She, George, hand, bank]
V -> [beats]
P -> [with, at]

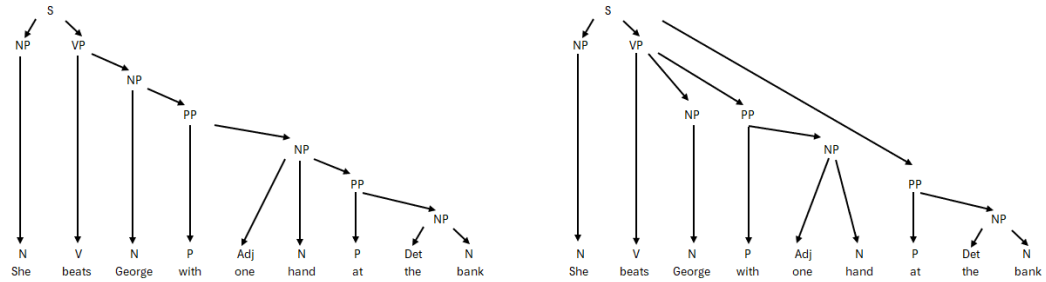
Answer:

a)

Semantic ambiguity: bank means different things, e.g. a river bank or a financial institution

Syntactic ambiguity: It is either she uses one hand to beat George, or George has only 1 hand

b)



8. Consider a sentence, *S = Ethan invited the person with microphone*.
   a) The sentence *S* consists of semantic ambiguity and syntactic ambiguity. Identify both of the ambiguities found from the sentence above.
   b) Given the grammar below, construct **ONE** (**1**) parse tree for the sentence *S*.

```
noun_phrase(NP)
verb_phrase(VP)
preposition_phrase(PP)
determiner(D)
noun(N)
verb(V)
S -> NP VP
NP -> N | N PP | D NP | D N
VP -> VP PP | V NP
PP -> P N
N -> [Ethan, person, microphone]
V -> [invited]
P -> [with]
D -> [the]
```

Answer:

a)

Syntactic ambiguity

Meaning 1: Ethan with microphone.

Meaning 2: The person with microphone

b)