ANISH ADNANI
D17B101

ASSIGNMENT 2

Q1)

1) Define discourse and pragmatic analysis. Discuss reference resolution problem in detail

→ * Structural definition: Discourse is a particular unit of language (above the sentence)
Eg: (from the color purple, Alice Whatson)

Jack is tall and kind and dont hardly say anything love children. Respect his wife, adessa and all odessa Amazon sisters (celie's diary)

* Functional Definition: Discourse is a particular focus of language use.

* Pragmatic Analysis: Pragmatics is defined as subfield of linguistics that studies the way in which context contributes to meaning. Pragmatics encompasses speech act theory, conversational implicature, talk in interaction and other approaches to language behaviour in philosophy, sociology conventional or "coded" in given language
Eg: You invited your friend over for dinner your child seen your friend reach for some cookies and says, "Better not take those, or you'll get even bigger!". You cant believe your child could be so rude.

* Reference Resolution Problem

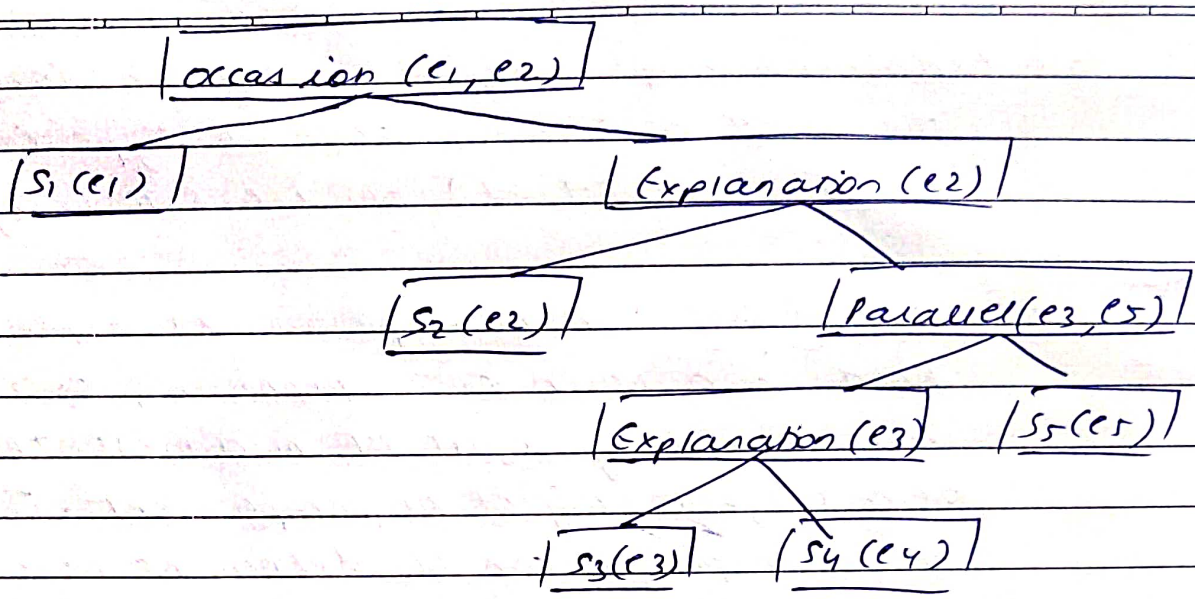
The process by which REFERENCE speakers use expressions (like John went to Bill's car dealership to check out an Acura integra. He looked at it for about an hour) to denote a person named John our discussion require that we first define some terminology.

A natural language expression used to perform reference is called a referring expression, and the entity that is referred to is called the referent. Thus, John and he in passage are referring expression, and John is their referent. (To distinguish between referring expressions and their referents, we italicize the former)

2) Explain hierarchical discourse structure & reference resolution.

→ The coherence of entire discourse can also be considered by hierarchical structure between coherence relations. For eg, the following passage can be represented as hierarchical structure -

S1 - Ram went to bank to deposit money
S2 - He then took a train to shayam's cloth shop
S3 - He wanted to buy some clothes
S4 - He do not have new clothes for party
S5 - He also wanted to talk to shyam regarding his health.



* Reference Resolution

Interpretation of the sentences from any discourse is another important task and to achieve this we need to know who are what entity is being talked about. Here interpretation reference is the key element. Reference may be defined as the linguistic expression to denote an entity or an individual. For eg in the passage, Ram the manager of ABC bank, saw his friend shyam at shop. He went to meet him, the linguistic expression like Ram, His, He are reference.

on the same note, reference resolution may be defined as the task of determining what entities are referred to by which linguistic expression.

3) Explain noun phrase, verb phrase in language. Identify the head and morphological type (Noun, verb, adjective, adverbial) of the following sentence.

→ Noun phrase: A noun phrase is a phrase whose head is a noun or a pronoun, optionally accompanied by a set of modifiers. It can function as a subject, object or compliment. The modifiers of a noun phrase can be determine nouns or adjective phrase

NP → Pronoun

NP → Det. Noun

NP → Noun

We can combine all these rules in single phrase structure rules as follows:

NP ⇒ (Det) (Adj) Noun | Pronoun

Verb Phrase: Analogous to the noun phrase is the verb phrase, which is headed by verb. There is a fairly wide range of reasons that can modify of verb. This makes verb phrase a bit more complex. The verb phrase organises various elements of the sentence that depends syntactically on the verb of find the following are some example of phrases.

i) President of the company → Noun Phrase

ii) Look up the chimney → verb Phrase

iii) Angry as a Hippo → Adjective Phrase

iv) Rapidly like a bat → Adverbial Phrase.

Q2. Explain with suitable example Following relationships between word meaning: Homonymy, polysemy, antonymy, hyponomy, Hyponymy, meronymy.

→ * Homonymy: It may be defined as the words having same spelling or same form but having different and unrelated meaning.

For eg: The word "Bat" is a homonymy word because bat can be implement to hit a ball or bat is a flying mammal.

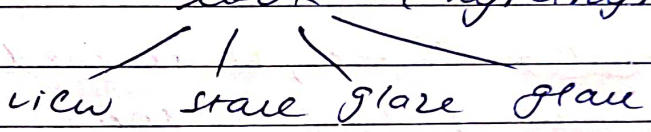
* Polysemy: This means "many signs" - It is a word or phrase with different but related sense. Polysemy has the same spelling but different meaning.
For eg: "Bank"
meanings i) Financial institution
ii) Synonym for "to rely on"

* Antonymy: It is the relation between 2 lexical items having symmetry between their semantic components relative to the axis.

For eg: rich/poor, life/death, etc.

* Hypernymy : Hypernymy is the condition of being inclusive of all other synonymy of a given synonyms

For eg:



* Hyponymy : It may be defined as the relationship between a generic term and instance of that generic term. Here generic term is called as hypernymy and its instance are called as hyponymys.

For eg: The word color is hyponomy and color blue, yellow, etc are hyponymy.

* Meronymy : It may be defined as a semantic relation specific to linguistics. A meronymy denotes a constituent part of a member of something.

For eg: finger is meronymy of hand

similarly, wheels are meronymy of automobile

Q3. Why is word sense disambiguation is challenging problem in natural language processing?

→ In natural language processing, word sense disambiguation is the problem of determining by a 'sense' (meaning) of a word is activated by the uses of the word in a particular context, a process which appears to be largely unconscious in people

WSD is a natural classification problem: Given a word and its possible senses, as defined by a dictionary, classify an occurrence of the word into context into one or more of its sense classes. The feature of the context provide the evidence for classification.

For eg: Little John was looking for mystery box. Finally he found it. The box was in the pen. John was very happy

1) pen - a writing implement with a point from which ink flows

2) pen - an enclosure for confying livestock

3) pen - a portable enclosure in which babies may be left to play.

4) pen - a correctional institution for those convicted of major crimes

5) pen - female swan

Words are typically assumed to have a finite and discrete set of senses, a gross simplification of the complexity of word meaning. There are different algorithms for different applications. In machine Translation, the problem takes the form of target word selection. Here the 'senses' are words in the target language which often correspond to significant meaning distinctions in the source language.

For eg: Bank could translate into Financial Bank or edge of river.

Q4. Draw and explain shift reduce parsing in natural language processing.

→ The shift reduce parsers by maintaining a state of the current parsed tree with the words of the sentence on a queue and partially completed trees on a stack and applying transitions to the state until the queue is empty and the current stack only contains a finished tree.

The initial state is to have all of the words in order on the queue with an empty stack. The transitions which can be applied are:

* shift: A word moves from the queue onto the stack

* unary reduce: A label of the first constituent on the stack changes. There is a different unary transition for every possible unary nodes in the tree bank used for training.

* Binary reduce: The first two nodes on the stack are combined into a new label. This are either right sided or left sided indicating which child is treated as the head.

* Finalize: A tree is not considered finished until the parser chooses the finalize transition

part of speech tags are not assigned by the parser. and are in fact used as features - This is accomplished by pretagging the text, meaning the POS annotator needs to be used.

For eg: "The dog saw a man in the park"

i) Initial state

Stack | Remaining Text

The dog saw a man in the park

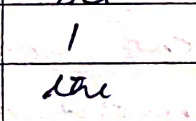
2) After the shift

Stack | remaining text

the | dog saw a man in the park

3) reduce shift reduce

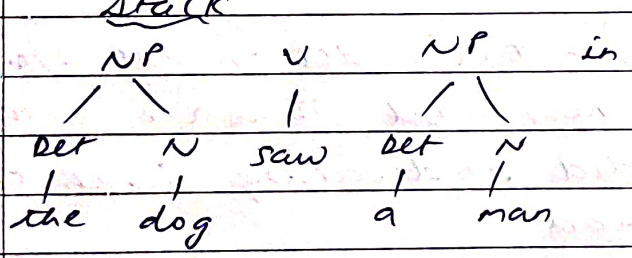
Stack | remaining text



saw a man in the park

4) After recognizing the second NP

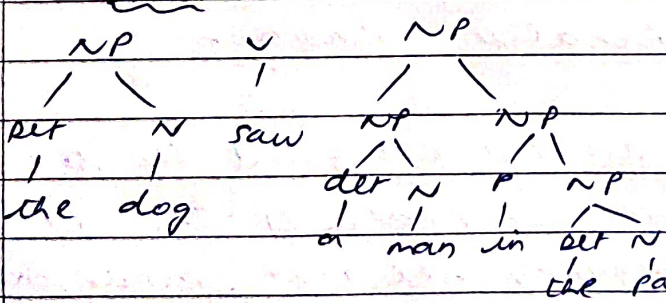
Stack | remaining text



the park

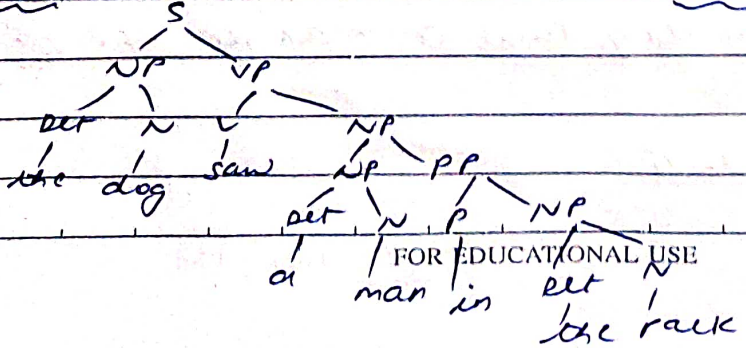
5) After building a complex NP

Stack | remaining text



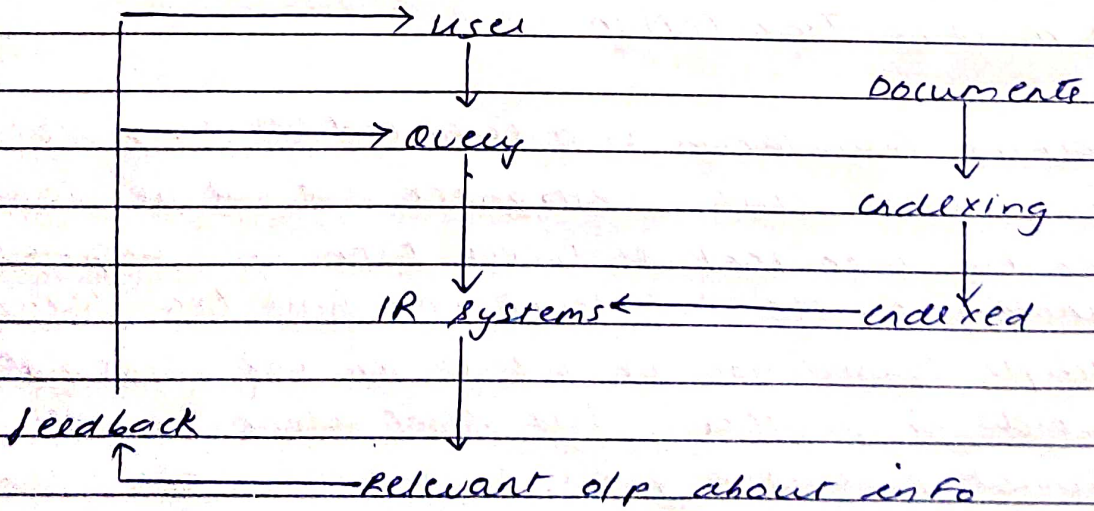
6) Build a complete parse Tree

Stack | remaining text



Q5. Write a short note on:
a) Information Retrieval

→ Information retrieval (IR) may be defined as a software program that deals with organization, storage, retrival and evaluation of information from document repositories particularly textual information. The system assists users in finding information they require but it does not explicitly return the answers of the questions. It informs the existence and location of documents that might consist of the required information. The documents that satisfy user's requirement are called relevant documents. A perfect IR system will retrieve only relevant documents.



From the above diagram, it is clearly shown that user who needs info will have to formulate a req in the form of a query in natural language. Then the IR systems will respond by retrieving the relevant o/p in the form of documents, about the read information.

Mathematically, a retrieval model consist of:

D – Representation of documents
R – Representation for queries
F – The modelling framework for D, C along with the relationship b/w them.
R(A,d.) – A similarity function which orders the doc wrt query. It is also called as ranking.

b) Machine Translation

Machine Translation is a subfield of computational linguistics that investigates the use of software to translate text or speech from one language to another. On the basic level, machine translation performs simple substitution of words in one language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counter parts in the target language is needed

Improved o/p quality can also be achieved by human investigation. For eg: some systems are able to translate more accurately if the user has unambiguously identified which word in the text are proper names.

With the assistance of these techniques, machine translation has proven useful as a tool to assist human translators and in very limited number of cases can even produce o/p that can be used as eg: weather reports

Major issues:

* Machine translation could produce some non-understandable phrases
* Disambiguity
* Non-standard speech
* Named Entities.

c) Word Net

→ Word Net is a lexical database of semantic relations between words in more than 200 languages. Word net links words, into semantic relations including synonyms, meronymys etc. The synonyms are grouped into synsets with short definitions and usage examples word NET can thus be seen as a combination of dictionary and the saurus. While it is accessible to human users via a web browser, it primary is in automatic text analysis and artificial intelligence applications.

WordNET was first created in English language and tools and WORDNET database have been released and are freely available for download from that WORDNET website

Limitations

- It is easy to create hyponyms relationship to capture that a 'conifer' is a type of 'tree' but it is difficult to classify emotions like "Fear" or "happiness" into equally deep and well defined hyponyms relationships.

- WordNet database include info about etymology and pronunciation of words and it contains only limited info about usage.

- WordNET does not include much domain specific terminology.