| Select correct constraint on coreference for given example "John and Mary have Hyundai cars. They love them". |
|---|
| Number agreement |
| Gender agreement |
| Person and Case agreement |
| Syntactic constraint. |

| What are the attachments for fragment of English? Explain with example. |
|---|

| Natural language processing is a sub-domain of, |
|---|
| Networking |
| Artificial Intelligence |
| Algorithms |
| Databases |

| What are the reference phenomenons? Explain types of referring expression. |
|---|

| This kind of ambiguity occurs when a sentence is parsed in different ways. |
|---|
| Lexical Ambiguity |
| Syntactic Ambiguity |
| Semantic Ambiguity |
| Pragmatic Ambiguity |

| "Appoint→Appointee" is an example of ---------- morphology. |
|---|
| Derivational |
| Inflectional |
| Compounding |
| Cliticization |

| Which of this is not a reference resolution algorithm? |
|---|
| Hobb's Algorithm |
| Lappin and Leass's Algorithm |
| Centering Algorithm |
| Lesk's Algorithm |

Shivaji→ शिवाजी

Is an example of:

| Translation |
|---|
| Transfer |
| Transliteration |
| Generation |

| |
|---|
| "John bought an Acura Integra today, but the engine seemed noisy." Which of the following is an Inferrable referent? |
| John |
| Acura |
| engine |
| noisy |

| |
|---|
| Pragmatic refers to |
| Literal meaning |
| Intended meaning |
| Structural meaning |
| Wordnet dictionary meaning |

| |
|---|
| "All boys love cricket ". How is this sentence represented in First Order Logic form? |
| ∃x boys(x)→love(x,cricket) |
| ∀x boys(x)→love(x,cricket) |
| ∃x,y love(x) ∧ cricket(y) |
| ∀x boys(x) ∧ love(x,cricket) |

| |
|---|
| The root form of a word in Wordnet dictionary is called |
| Stem |
| Sense |
| Gloss |
| Lemma |

| |
|---|
| Which of the following is an example of "hyponym-hypernym" semantic relationship? |
| Car-Vehicle |
| Car-Wheel |
| Wheel-Car |
| Car-Ford |

| |
|---|
| Which of the following is not a sequence labeling technique? |
| Maximum Entropy |
| Context Free Grammar |
| Conditional Random Fields |
| Hidden Markov Model |

| Which of the following is a Rule based POS tagger? |
| --- |
| HMM Tagger |
| Ngram Tagger |
| ENGTWOL Tagger |
| Brill Tagger |

| CFG captures ---------- |
| --- |
| Constituency and ordering |
| word meaning |
| relation between words |
| sentence meaning |

## rule based classifier: rule satisfication and coverage

rule pruning
rule extraction from dt. if then and or

ordered rule set= decision list (rule based ordering)
ranked wrt priority
default/ fallback rule for tuples not satisfying any rule,, fired when no other rule satisfied. evaluated at end

class based ordering . all the rules for
the most prevalent (or most frequent) class come first, the rules for the next prevalent class come next, and so on

size ordering (size of rule) scheme assigns the highest priority to the triggering rule that has the "toughest" requirements, where toughness is measured by the rule antecedent size. That is, the triggering rule with the most attribute tests is fired.

a record may trigger more than one rule. hence some ordering needed. not mutually exclusive

a record may not trigger any rule so default rule is needed. non exhaustive hua toh

for a rule to match, only lhs has to be same

Q.6 Problem on Ngram model

Consider following Training data:

<s> I am Sam </s>

<s> Sam I am </s>

<s> Sam I like </s>

<s> Sam I do like </s>

<s> do I like Sam </s>

Assume that we use a bigram language model based on the above training data.

What is the most probable next word predicted by the model for the following word sequences?

(1) <s> Sam ...

(2) <s> Sam I do ...

(3) <s> Sam I am Sam ...

(4) <s> do I like ...

**(Solution: Bigram probabilities: P(Sam|<s>) = 3/5 P(I|<s>) = 1/5 P(I|Sam) = 3/5 P(</s>|Sam) = 2/5 P(Sam|am) = 1/2 P(</s>|am) = 1/2 P(am|I) = 2/5 P(like|I) = 2/5 P(do|I) = 1/5 P(Sam|like) = 1/3 P(</s>|like) = 2/3 P(like|do) = 1/2 P(I|do) = 1/2**

**(1) and (3): "I". (2): "I" and "like" are equally probable. (4): </s>)**

**25. Draw the shift-reduce parser in processing the sentence**

***The woman saw a puppy***

Explain CFG with suitable example. Discuss following potential problems in CFG such as : 1
1) Agreement 2) Sub categorization 3) Movement..

| Which of the following is not a type of referring expression? |
|---|
| Indefinite noun phrases |
| definite noun phrases |
| Inferrables |
| Demonstratives |

| Which of the following is not true about Question-Answer (QA) System? |
|---|
| Natural extension of Information Retrieval |
| A QA system receives a query expressed in Natural Language |
| Tries to provide a document containing the answer |
| Provides answer as a fact |

| Which of the following is a correct rule for "Yes-No-Question"? |
|---|
| $S \rightarrow NP\ VP$ |
| $S \rightarrow VP$ |
| $S \rightarrow Aux\ NP\ VP$ |
| $S \rightarrow WhWord\ NP\ VP$ |

| Which of the following is not an open class? |
| --- |
| Noun |
| Preposition |
| Verbs |
| Adverb |

| Which of the following is the example of surface segmentation? |
| --- |
| Achievability = achievabil + ity |
| Achievability = achiev + ability |
| Achievability = Achieve + able + ity |
| Achievability = achiev + abil + ity |

Q.1. The simplest augmentation of the context-free grammar is the Probabilistic Context PCFG Free Grammar (PCFG), also known as the Stochastic Context-Free Grammar SCFG (SCFG), first proposed by Booth (1969).

| The word "Putting" is handle and clean up by which stemming rule? "*Putting* → *Put*" |
| --- |
| {(X) –ing → ϵ} and {--CC→ C} |
| {(X) –eed → -ee} and {-at → -ate} |
| {(X) –ing → -ing} |
| {(X) –C1VC2 → C1VC2e} |

| Which of the following is not a Information Retrieval type? |
| --- |
| Text |
| Image |
| Speech |
| Structured Information |

| Syntax-driven semantic analysis is based on following, |
| --- |
| Principle of compositionality |
| Principle of integrity |
| Principle of comparability |
| Principle of tagging |

Q.2.

| | |
|---|---|
| S → NP VP | V → swam \| ran \| flew |
| VP → V NP | VP → swam \| ran \| flew |
| VP → VP PP | D → the \| a \| an |
| NP → D N | N → pilot \| plane |
| NP → NP PP | NP → Edinburgh \| Glasgow |
| PP → P NP | P → to |

CYK algorithm uses bottom-up parsing technique. Apply the CYK algorithm on the sentence "The pilot flew the plane to Glasgow" using the grammar G and state whether the sentence is recognized or not.

Q.4. Classify give test statement using Naïve Bayes algorithm.

| | docID | words in document | in c = China? |
|---|---|---|---|
| Training set | 1 | Chinese Beijing Chinese | yes |
| | 2 | Chinese Chinese Shanghai | yes |
| | 3 | Chinese Macao | yes |
| | 4 | Tokyo Japan Chinese | no |
| Test set | 5 | Chinese Chinese Chinese Tokyo Japan | ? |

Q.5 Find out pos tagging using Viterbi's algorithm.

Book a car.
Park the car.
The book is in the car.
The car is in a park.

| |
|---|
| The process of deciding what pronouns and other noun phrases refer to is known as |
| Inferable |
| Coreference Resolution |
| Reflexive |
| Verb Semantics |

Lesk algorithm
converts words to vectors
finds comparison between two words
measures overlap between sense definitions for all words in context
check for similarity between words in context

Natural Language Generation does not involve the following task
Producing meaningful phrases and sentences
Mapping the given input in natural language into useful representations.
Retrieving the relevant content from the knowledge base.
Mapping sentence plan into sentence structure

Explain lexicon, lexeme and the different types of relations that hold between lexemes

Why it is important to preprocess text data in natural language? Explain in detail the steps of preprocessing with examples.

Discuss dictionary based approach for WSD.

Dictionary-Based Approach for Word Sense Disambiguation (WSD):
- Definition: A WSD method that relies on pre-existing lexical resources such, as dictionaries or thesauri, to determine the correct sense of a word.
- Process: It involves matching the target word and its context to the definitions or senses provided in the dictionary. The sense with the closest match is selected.
- Advantages: Simplicity, reliance on established resources, and often fast processing.
- Challenges: Limited to the senses and information available in the dictionary, may not handle context-specific senses well.

Explain the different steps in text processing for Information Retrieval

Steps in Text Processing for Information Retrieval:

1. Document Collection: Gather a collection of documents that may contain the information the user is seeking. This collection can be a set of web pages, articles, books, or any textual content.

2. Text Preprocessing:
   - Tokenization: Split the text into words or tokens.
   - Lowercasing: Convert all words to lowercase to ensure consistency.
   - Stop Word Removal: Eliminate common words (e.g., "the," "and") that add little meaning.
   - Stemming or Lemmatization: Reduce words to their base or root form to group similar words (e.g., "running" and "ran" to "run").

3. Text Indexing:
   - Inverted Index Creation: Create an inverted index that maps terms (words) to the documents in which they appear. This speeds up retrieval by allowing quick access to relevant documents.
   - Term Frequency (TF) and Inverse Document Frequency (IDF): Calculate TF to measure the importance of a term within a document and IDF to measure its importance in the entire collection.

4. Query Processing:
   - Query Parsing: Break down the user's query into individual terms or concepts.
   - Query Expansion: Expand the query by adding synonyms or related terms to improve recall.

5. Ranking and Scoring:
  - Retrieval Model: Apply a retrieval model (e.g., TF-IDF, BM25) to calculate the relevance score of documents to the query.
  - Ranking: Sort documents based on their relevance scores.

6. User Interaction: Present the ranked documents to the user. The user may provide feedback or refine the query.

7. Results Presentation: Display the search results, often with snippets or summaries, to help users quickly assess document relevance.

8. Evaluation: Assess the retrieval system's performance using metrics like precision, recall, and F1-score to measure its effectiveness.

9. Feedback Loop: Incorporate user feedback to improve the system's performance and relevance ranking.

10. Relevance Feedback (Optional): Allow users to mark documents as relevant or non-relevant, and use this feedback to re-rank results for better accuracy.

Text processing in information retrieval aims to extract, index, and rank documents to provide users with relevant information from a large collection of textual data.
hmm.. limitations
My apologies for the oversight. Here are some limitations of Hidden Markov Models (HMMs):

1. Independence Assumption: HMMs assume that the current state is only dependent on the previous state. This might not accurately model systems with long-range dependencies.

2. Fixed State Space: HMMs require specifying a fixed set of states in advance. Adapting to changing or unknown states can be challenging.

3. Initial State Problem: HMMs can be sensitive to the choice of initial state probabilities, which might not always be known or reliable.

4. Model Complexity: Determining the appropriate number of states and transition probabilities in an HMM can be challenging, and more complex models require more data.

5. No Memory of Past Emissions: HMMs don't explicitly remember past emissions, which can limit their ability to capture certain patterns in the data.

6. Non-Stationary Processes: HMMs assume stationary state transition probabilities, which may not hold in non-stationary processes.

7. Overfitting: Like any model, HMMs can overfit the training data if they are too complex relative to the amount of data available.

8. Hidden State Identification: Determining the actual meanings or interpretations of the hidden states can be challenging, especially in complex models.

9. Learning Challenges: Training HMMs can be computationally intensive, and in some cases, it may be challenging to find the best model parameters.

10. Multimodal Data: HMMs are primarily designed for unimodal data, and extending them to handle multiple modalities (e.g., audio and visual data) can be complex.

It's important to note that HMMs have been highly successful in various applications, and many of these limitations can be addressed or mitigated through appropriate model design, training, and post-processing.
Yarowsky bootstrapping approach of semi supervised learning
Differentiate between Syntactic ambiguity and Lexical Ambiguity, pragmatic, semantic

**1. Both Lexical and Semantic Ambiguity:**
- Example: "I saw a man with a telescope." This sentence is both lexically and semantically ambiguous. "Man" could refer to a person or a group of humans, and "telescope" could mean an optical instrument or a place where people gather to stargaze. The meaning of the sentence depends on the context and the intended interpretations of these words.

**2. Lexical Ambiguity but Not Semantic Ambiguity:**
- Example: "She's a bright student." This sentence is lexically ambiguous because "bright" can mean intelligent or shining with light. However, in this context, it's clear that "bright" refers to intelligence, so there is no semantic ambiguity.

**3. Semantic Ambiguity but Not Lexical Ambiguity:**
- Example: "They raced to the bank." In this sentence, there is semantic ambiguity because it's unclear whether "bank" refers to a financial institution or the side of a river. However, the individual words themselves (e.g., "raced," "bank") are not lexically ambiguous; it's the overall sentence meaning that is uncertain.

Syntactic Ambiguity:

Definition: Syntactic ambiguity occurs when a sentence or phrase has multiple valid syntactic parses or grammatical structures, leading to different possible interpretations.
Example: "Visiting relatives can be a nuisance." This sentence can be parsed in two ways: "Visiting relatives" is a nuisance, or "relatives who visit" are a nuisance.

Pragmatic Ambiguity:

Definition: Pragmatic ambiguity is related to how context, tone, and implied information affect the interpretation of a statement.
Example: "Can you pass the salt?" The intended meaning (a request to pass the salt) is influenced by the pragmatic context and the act of reaching for the salt shaker.

semantic ambiguity is when even a human is confused what the sentence means

imperative vs declarative sentences, yes-no and wh qns
close the door. please let me. join us : imperative
sun rises in the east, imma go to store : declarative

Compare Information Retrieval with Information Extraction system.

| Aspect | Information Retrieval (IR) | Information Extraction (IE) |
|---|---|---|
| **Goal** | Retrieve relevant documents or information in response to user queries. | Extract structured information or specific data from unstructured text. |
| **Input** | User queries (keywords, questions) and a large collection of unstructured documents. | Unstructured text corpus and, in some cases, domain-specific patterns/templates. |
| **Output** | A ranked list of relevant documents based on the query, with snippets or summaries. | Structured data, often in the form of triples or records, with specific pieces of information associated with entities and relationships. |
| **Examples** | Web search engines (e.g., Google), academic databases, document retrieval systems. | Named Entity Recognition (NER), event extraction, relation extraction, fact extraction systems. |
| **Challenges** | Relevance ranking, query expansion, handling ambiguous queries. Focus on retrieving entire documents. | Identifying patterns, entities, and relationships in text, handling language variations, and resolving ambiguities. Focus on extracting specific information. |

Explain five different types of dexis with a suitable example.

Deixis is a linguistic term that refers to words or expressions that rely on context to determine their meaning, particularly in terms of identifying specific people, objects, locations, or times.

1. Person Deixis:
   - Example: "I will call her after the meeting."
   - In this sentence, "I" and "her" are examples of person deixis. "I" refers to the speaker, and "her" refers to a specific person, relying on the context to identify who "her" is.

2. Spatial Deixis:
   - Example: "The keys are under the table."
   - Here, "under" is an example of spatial deixis, indicating the location of the keys in relation to the table.

3. Temporal Deixis:
   - Example: "She arrived yesterday."
   - In this sentence, "yesterday" is an example of temporal deixis, specifying a particular point in the past relative to the current time.

4. Discourse Deixis:
   - Example: "As I mentioned earlier, this is the solution."
   - "This" in the sentence refers to something previously introduced or discussed in the ongoing conversation, serving as discourse deixis.

5. Social Deixis:
   - Example: "Would you like some tea, ma'am?"
   - In this sentence, "ma'am" is a form of social deixis used to show respect or politeness when addressing someone.

These explicit examples demonstrate how deixis functions in language, relying on context to specify the intended referents, locations, times, and social dynamics in communication.

## Maximum Entropy model, Conditional random Field (CRF)

## Anaphora Resolution using Hobbs and Centering Algorithm

Anaphora resolution is the process of identifying and connecting anaphors (words or phrases that refer back to previously mentioned entities or concepts) to their antecedents (the entities or concepts they refer to) in a text

The Hobbs algorithm is a rule-based method for anaphora resolution. It follows a series of steps to identify potential antecedents and select the most likely one.

Centering theory is a discourse-based approach to anaphora resolution. It focuses on the organization of information in discourse and the role of "centers" in establishing coherence. Centering theory includes the following components:

Centers: Think of "centers" as important things or people in a conversation. These centers are like main characters in a story or key ideas in a discussion. Each center has a potential antecedent (the word it refers to).

Transitions: Transitions are like the way you move from talking about one center to another. They can be smooth, where you keep talking about the same thing (like continuing to talk about a character in a story), or they can be a change, where you start talking about something new.

Rule of the Dominant: This rule helps decide which center is the most important at a given point in the conversation. It looks at how smoothly you can move from one center to another. If a smooth transition is possible, that center is often the "dominant" one.

conservations:
1. Alice: I just finished reading a fantastic book. It's a mystery novel.
2. Bob: Oh, really? What's the name of the book?
3. Alice: The book is called 'The Secret of the Old Mansion.' It's about a detective who solves a complex case.
4. Bob: That sounds interesting. Did you like the detective in the story?
5. Alice: Yes, he was a brilliant detective. I liked his character a lot.
Now, let's apply centering theory to analyze this conversation:

Centers:

Center 0: In sentence 1, "I" is the initial center, referring to Alice.
Center 1: In sentence 3, "The Secret of the Old Mansion" is introduced as a new center, referring to the book.
Center 2: In sentence 5, "the detective" becomes the new center, referring to the detective in the book.
Transitions:

From sentence 1 to sentence 3, we have a change of center (C→C), where the center shifts from Alice (C0) to the book (C1).
From sentence 3 to sentence 5, there is another change of center (C→C), moving from the book (C1) to the detective (C2).
Rule of the Dominant:

At each transition, we choose the most dominant center. In this case, we focus on the new elements introduced in the conversation.
In sentence 3, "The Secret of the Old Mansion" becomes dominant as it's a new topic.
In sentence 5, "the detective" becomes dominant as it's the new focus.

## Give any 3 different evaluation metrics available for text classification?

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

predicted sentinment analysis
slow and confusing in beginning, as it proceeded, story picked up with plot twists : fn

product was so good, gave it all away to my relatives: fp
dont listen to your haters. im one of em.
nice product, dont ever make again.

## Describe discourse segments.

Discourse segments, also known as discourse units or text segments, are fundamental elements in discourse analysis. They are distinct portions of a text or conversation that convey coherent and meaningful information. Discourse segments are typically defined by linguistic and structural features and play a crucial role in understanding the flow and organization of discourse. Here's a brief description:

1. Coherence and Cohesion: Discourse segments contribute to the coherence and cohesion of a text or conversation. They provide a logical structure that helps connect ideas and information in a meaningful way, making the discourse easier to follow.

2. Boundary Indicators: Discourse segments are often marked by boundary indicators, which can be explicit (e.g., transition words like "however" or "meanwhile") or implicit (e.g., shifts in the topic or tone). These indicators signal the beginning or end of a segment.

3. Unity of Meaning: Each discourse segment typically focuses on a particular theme, topic, or idea, and the content within a segment is expected to be thematically unified. This contributes to the clarity of communication.

4. Conversational Context: In spoken discourse, segments can correspond to turns in a conversation. Each person's contribution to the conversation represents a distinct segment, and the transition between speakers marks a boundary.

5. Variability: The length and complexity of discourse segments can vary depending on the type of discourse, language, and communication context. In written academic texts, segments may be longer and more structured, while in informal conversations, they can be shorter and more fluid.

6. Analytical Tool: In discourse analysis, researchers often break down a text or conversation into segments to analyze the structure, content, and relationships between these segments. This analytical approach helps uncover patterns and insights about how meaning is constructed in discourse.

Understanding discourse segments is important in fields such as linguistics, communication studies, and natural language processing, as it enables the examination and interpretation of how information is organized and communicated in various contexts.

## What is meant by knowledge representation?

Knowledge representation involves structuring information using subject-predicate structures

, ontologies,
a structured way of representing knowledge, emphasizing concepts, relationships, and hierarchies within a particular domain. They provide a standardized, organized framework for representing

information. Ontologies use subject-predicate-object triples to express relationships, similar to subject-predicate structures.

frames, or semantic networks to model knowledge, concepts, and relationships. It includes the use of inference and rule-based systems
These systems apply logical rules to the representations, allowing the system to make inferences or draw conclusions. For example, an inference rule might state that if "X is a parent of Y" and "Y is a child of Z," then "X is a grandparent of Z."

to draw conclusions from the represented information. Effective knowledge representation is essential for AI systems to process and utilize knowledge in various domains and tasks.
Explain in detail Lappin and Leass's Algorithm for Pronoun Resolution with suitable example.
Q.3 Explain three types of referents that complicate the reference resolution problem.
Ambiguous Referents: Words or phrases with multiple potential meanings, making it unclear which one is intended (e.g., "She saw the bird perched on the tree and was fascinated by it").

Deferred Referents: Expressions that refer to something not mentioned in the current context but in a previous or future part of the text, requiring information tracking (e.g., "the mysterious artifact" introduced earlier).
After weeks of intense training, she finally mastered it."

Non-Nominal Referents: References to abstract concepts or actions rather than concrete entities.
After weeks of intense training, she finally mastered it."
In this sentence, "it" is a non-nominal referent because it refers to a skill or task that's not a concrete noun but an abstract concept (mastery or proficiency).
Q.2 Discuss in detail attachments for fragments of English sentences.
Attachments in the context of sentence fragments or syntactic structures are crucial for understanding the grammatical relationships between words within a sentence. These attachments help determine how different components of a sentence fit together, creating a coherent and meaningful structure. Here, we'll discuss attachments in more detail.

Attachments in Sentence Fragments:

1. Subject-Verb Attachment: This attachment specifies the relationship between the subject and the verb in a sentence fragment. For instance, in the fragment "The cat that chased," the attachment shows how "The cat" relates to the verb "chased."

- Example: "The cat (subject) that chased (verb)..."

2. Modifier Attachment: Modifier attachments involve adjectives, adverbs, or other descriptive words that modify nouns or verbs. These attachments indicate how modifiers relate to the words they describe or enhance.

- Example: "The very (adverb) big (adjective) dog (noun)."
  - Here, the attachment determines that "very" modifies "big," and "big" modifies "dog."

3. Prepositional Phrase Attachment: Prepositional phrase attachments indicate how prepositional phrases relate to the rest of the sentence. They specify the nouns, verbs, or other components that the prepositional phrases modify.

- Example: "The book on the shelf (prepositional phrase)."
  - The attachment specifies that "on the shelf" modifies "The book."

4. Clause Attachment: In complex sentences or fragments, clause attachments specify how subordinate clauses relate to the main clause or other parts of the sentence. This is crucial for understanding the syntactic structure of complex sentences.

- Example: "While I was reading, (subordinate clause) I found a fascinating article (main clause)."
- The attachment shows how the subordinate clause "While I was reading" is related to the main clause "I found a fascinating article."

Attachments help disambiguate the meaning of sentence fragments and contribute to sentence parsing and understanding. Parsing is the process of assigning attachments to words and phrases within a sentence to create a structured representation of the sentence's grammatical and semantic structure.

In natural language processing and computational linguistics, attachment resolution is a critical step in syntactic and semantic analysis, enabling machines to understand and process human language effectively.

## Analyse how statistical methods can be used in machine translation

## Describe augmented grammar in syntactic analysis

In syntactic analysis, augmented grammar refers to an extended or modified context-free grammar (CFG) that includes additional rules, symbols, or features to capture specific linguistic phenomena or structural aspects not covered by a basic CFG. Augmented grammars are used in formal language theory and computational linguistics to handle various complexities and subtleties of natural language syntax. vocab extended, augmented grammars incorporate semantic information, such as lambda calculus expressions or logic forms, to represent the meaning of sentences or phrases. This allows for the integration of syntax and semantics in the analysis.,

## 17. Explain surface anaphora and the different methods for dealing with surface anaphora

Pronominal Anaphora Resolution: This method involves resolving pronouns and their antecedents.
Grammatical Agreement: Matching the pronoun with its antecedent in terms of gender, number, and person.
Proximity: Selecting the nearest noun phrase as the antecedent.
Semantic Compatibility: Ensuring that the pronoun and antecedent have compatible meanings.

Sentence: "John saw a dog, and he adopted it."
Resolution: "he" refers to "John," and "it" refers to "a dog."
Zero Anaphora:

Sentence: "食べた。"
Resolution: In Japanese, "食べた" means "ate." The subject (e.g., "I" or "John") is often omitted but implied based on context.
Ellipsis Resolution: Ellipsis is a form of surface anaphora in which one or more words are omitted because they are recoverable from the context.

Sentence: "Mary likes pizza, and Susan does too."
Resolution: The ellipsis in the second clause is "likes pizza," which is recoverable from the context.

write abt syntatic and semantic constraints, choeherecne and coreference also.

## 18. Explain the difference of discourse structure from other reference mechanisms

The key difference between discourse structure and reference mechanisms is their scope and function. Discourse structure pertains to the overall organization of a text, while reference mechanisms deal with specific linguistic elements used to refer to previously mentioned entities or concepts within that text.

Discourse structure focuses on the macro-level organization of a text, ensuring logical flow and coherence, whereas reference mechanisms operate at the micro-level for coreference anaphora etc, resolving ambiguity and connecting specific words or phrases to their referents.
Discourse structure is concerned with how information is organized within a larger piece of communication, while reference mechanisms are concerned with how specific words or expressions are linked to the entities or concepts they refer to.

Explain direct machine translation

21. Describe transfer model of Machine Translation. List out its three phases

Direct Machine Translation:
Direct Machine Translation, also known as Direct Translation, is an approach in machine translation where the source language text is translated directly into the target language without the need for an intermediate representation, such as an interlingua or pivot language. This method relies on mapping words, phrases, or sentences from the source language to their corresponding equivalents in the target language.

In Direct Machine Translation, there are typically two main components: the source language analysis and the target language generation. The source language analysis involves breaking down the input text into its linguistic components, such as words, phrases, and grammatical structures, and then identifying the corresponding elements in the target language. The target language generation step generates the translated text based on the identified elements.

While Direct Machine Translation can be effective for certain language pairs and contexts, it may face challenges in handling idiomatic expressions, cultural nuances, and language-specific variations. It may require extensive training data and language-specific models to achieve high translation quality.

Transfer Model of Machine Translation:
The Transfer Model is one of the traditional approaches to Machine Translation, particularly in rule-based and hybrid machine translation systems. It aims to bridge the gap between the source language and the target language, taking into account linguistic differences and structural variations.

The Transfer Model typically consists of three phases:

1. **Analysis Phase**: In this phase, the source language text is analyzed, and its linguistic structure is determined. This includes parsing the source language sentence to identify parts of speech, syntactic structures, and other linguistic features.

2. **Transfer Phase**: The transfer phase is where the linguistic information obtained from the source language is transformed into a form that is suitable for the target language. This may involve mapping source language structures to their corresponding structures in the target language. Rules and heuristics are often employed in this phase to handle linguistic and structural differences between the languages.

3. **Generation Phase**: In the generation phase, the transformed linguistic information is used to produce the target language output. This involves constructing sentences in the target language that convey the same meaning as the source language text. The generated output is expected to be fluent and grammatically correct.

The Transfer Model is particularly useful when translating between languages with significant linguistic differences, as it allows for more controlled and structured translation. However, it can be complex and may require a deep understanding of both the source and target languages, as well as extensive rule-based or linguistic knowledge. Modern machine translation approaches often incorporate elements of the Transfer Model within more data-driven and neural network-based models.

5 types of referring expressions

Definite Noun Phrases: These are noun phrases that specify a particular referent using the definite article "the." For example, "The book on the table is interesting."

Pronouns: Pronouns are words used to replace nouns, making language more concise. Common pronouns include "he," "she," "it," "they," and "them." For example, "She is going to the store."

Demonstratives: Demonstratives like "this," "that," "these," and "those" point to specific items in a context. For example, "This is my favorite song."

Proper Nouns: Proper nouns are specific names for individuals, places, or entities. For example, "John" refers to a specific person.

Indefinite Noun Phrases: These noun phrases refer to unspecified entities and are often introduced with indefinite articles like "a" or "an." For example, "I saw a cat in the garden."

Transfer learning

related tasks. cross utilize the knowledge learnt

activa non gear se gear se car seekhna

use in ml of tl:

use of pretrained models, instead of building from scratch since no enough data

idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones.

reuse the output for a first model as the starting pt for second model. Modify or add new layers on top of the pretrained model, which are specific to the target task. this is called fine tuning

eg. inception v3 model.

challenges:

domain mismatch needs domain adaptation (reduce the distribution shift between athe source and target domains by aligning them )

too much fine tuning can result in overfitting

complexity

---

Convolutional Neural Networks (CNN): A neural architecture commonly used for image processing but also applied to text data for feature extraction.

Recurrent Neural Networks (RNN): Neural networks designed to process sequences of data, suitable for text classification tasks. has an architecture for sequential processing of data.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU): Special types of RNNs that address the vanishing gradient problem by having improved memory capabilities.

Transfer Learning: The practice of using pre-trained models (e.g., BERT, GPT) for document classification tasks.

seq2seq models: Neural architectures for tasks like machine translation and text generation.

Transformers are a class of neural network architectures to capture contextual relationships in sequences by employing self-attention mechanisms, which allow each element in a sequence to focus on relevant parts of the input, enabling powerful and parallelized processing of data.

transformers:
global context. long range dependenccies
parallel archi: use entire sequence simultaneously
word order decided by positional encodings since no inherent sequence order.

Discourse and Pragmatic Processing:
- Discourse: Think of a conversation, context imp
bank: financial vs river

- Pragmatics: beyond the literal interpretation of words
Pragmatics includes various aspects, such as implicature (what is implied but not explicitly stated), speech acts (how language can be used to perform actions), and the influence of context, tone

friend 1: "Can you pass the salt?"  ans type: yes or no
you: "yep, here it is"
vs.
friend 1: "Please pass me the salt."
you: "here u go"
.....

Reference Resolution:

"Alice found a cat. The cat was fluffy," "The cat" refers to the cat Alice found.

It's like when you have two toys, one is a red ball, and the other is a blue ball. When you say, "I like the blue one," "the blue one" refers to the blue ball.

Sentence: "Anmol met her at the park for the book that she had to return."
Revised Sentence with resolved issues by gender agreement: "Anmol met sanjana at the park for the book that anmol had to return"

(issue was, anmol is a gender neutral name)
---
Coreference Resolution:
Coreference is a subset of anaphora and specifically refers to a situation where two or more words or phrases in a text refer to the same entity or concept.  Now u can have here multiple people, or multiple forms of same person

her daughter's safety pin, her husband's mirror
she was their home, never an error
even if mistakes, even if hard times
lets empathize, its daisy's first attempt at life
.....
For Definite Noun Phrase Resolution: If you read, "The red apple fell," you use clues in the text to identify which apple is being talked about.
-

Taxonomy Hierarchy of Reference Resolution:
- Anaphora Resolution: In a text, if you read, "John lost his wallet. He was upset," "He" refers back to "John."
-

sarah bought a car. dean, her brother seems to loves her car more than she does.

.....
Syntactic and Semantic Constraints on Coherence:
structure/syntactic
- coherent syntactic sentence: "I want to play soccer because I like sports."
incoherent syntactic sentence: Running the circus with a banana.

- Semantic Coherence: "I saw a dog. It tasted like chocolate." The second part doesn't logically connect to the first.
.....
Number agreement, in the context of language, pertains to words matching in terms of singular or plural form. Here's an example of number agreement:

- Singular Agreement: "The cat is on the roof."

- Plural Agreement: "The cats are on the roof." n and v are in plural form
Certainly, here are examples of person agreement and case agreement in sentences:
.....
Person Agreement:

1. First Person: "I am."

2. Second Person: "You are".

3. Third Person: "He sings".
.....
Case Agreement:

1. Nominative Case: "SHE is a doctor." SUBJECT

2. Accusative Case: "I saw HER at the park."  OBJ

3. Genitive Case: "That is JOHN'S car." possesion

Person and case agreement ensure that words within a sentence align in terms of both the person of the subject/object and the grammatical case to convey meaning accurately and coherently.