Natural Language Processing - Module 5

Explain three types of referents that complicate the reference resolution problem. - Module 5 - IDG

Three types of referring expression that complicate the reference resolution	
Type	**Example**
Inferrables	I almost bought a Ford Escort, but a **door** had a dent.
Discontinuous Sets	John and Mary love their Escorts. **They** often drive them.
Generics	I saw 6 Ford Escorts today. **They** are the coolest cars.

1. **Inferrables**:
    - **Definition**: Inferables are referents that cannot be directly extracted from the text but require some level of reasoning or inference to determine their identity. In other words, they are implied or inferred based on contextual information.
    - **Complexity**: Inferables add complexity to reference resolution because they rely on the ability of the system to make inferences based on the available information. For example, in the sentence "After the game, he was exhausted," the referent "he" is an inferable, and the system needs to infer that it refers to a player in the game.
2. **Discontinuous Sets**:
    - **Definition**: Discontinuous sets refer to entities or referents that are scattered or distributed across the text rather than appearing in a single, continuous mention. These sets may have multiple related expressions that need to be connected.
    - **Complexity**: Discontinuous sets complicate reference resolution because the system needs to identify and link all the expressions within the set. This can

involve handling anaphoric references to different parts of the set. For example, in a text about a family, "The father works in finance. His son is an engineer. The daughter is a doctor," the referents "father," "son," and "daughter" form a discontinuous set.

3. **Generics**:
    - **Definition**: Generic expressions refer to a general category or type of entity rather than a specific, individual referent. They are often used to make statements about a class of objects or people.
    - **Complexity**: Generics can be challenging for reference resolution because they don't point to a particular instance, making it difficult to determine the specific referent. Resolving generics may require world knowledge or context about common attributes of the category. For example, in the sentence "Dogs are loyal animals," the reference resolution system needs to understand that "dogs" refers to the generic category of dogs, not a specific dog.

Define discourse & pragmatic analysis. Discuss reference resolution problem in detail. - Module 5
same as Discourse reference resolution

Pragmatic Analysis

To reinterpret what was said to what was actually meant
*
concerns how sentences are used in different situations and how it affects the interpretation of the sentence.

1. During this process, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.
2. It involves deriving those aspects of language which require real world knowledge.
3. Explains how extra meaning is read into texts without actually being encoded in them.
    a. This requires much world knowledge, including the understanding of intentions, plans, and goals.

4. Consider the following 2 sentences:
    - The college authorities refused the student union a permit because they feared violence.
    - The college authorities refused the student union a permit because they supported strike.
5. The meaning of "they" in the 2 sentences is different. In order to figure out the difference, world knowledge in knowledge bases and inferencing modules should be utilized.

Discourse Integration

Resolving references Between sentences
*
concerns how the immediately preceding sentences affect the interpretation of the next sentence.
*** processing of sequences of sentences**

1. The meaning of any sentence depends upon the meaning of the sentence just before it.
2. In addition, it also brings about the meaning of immediately succeeding sentence.
3. The meaning of any sentence depends upon the meaning of the sentence just before it.
4. In addition, it also brings about the meaning of immediately succeeding sentence.
5. Focuses on the properties of the text as a whole that convey meaning by making connections between component sentences
    a. I only like to solve theory questions in exam. Therefore I attempted only Q. 1 and Q.2.
    b. I only like to solve theory questions in exam. Nevertheless, I attempted all questions.
    c. In example (1), we can infer Q.1 and Q.2 are theory questions. (2) We can't infer same information that we inferred in first statement.

Reference Resolution

Five common types of referring expression	
Type	**Example**
Indefinite noun phrase	I saw **a** Ford Escort today.
Definite noun phrase	I saw a Ford Escort today. **The** Escort was white.
Pronoun	I saw a Ford Escort today. **It** was white.
Demonstratives	I like **this** better than **that**.
One-anaphora	I saw 6 Ford Escort today. Now I want **one**.
Three types of referring expression that complicate the reference resolution	
Type	**Example**
Inferrables	I almost bought a Ford Escort, but a **door** had a dent.
Discontinuous Sets	John and Mary love their Escorts. **They** often drive them.
Generics	I saw 6 Ford Escorts today. **They** are the coolest cars.

Reference resolution deals with understanding how words or phrases in a text refer to other words or entities. Reference phenomena encompass different referencing strategies.

- **Reference**: the process by which speakers use expressions to denote an entity.
- **Referring expression**: expression used to perform reference .
- **Referent**: the entity that is referred to.

- **Anaphora Resolution**:
  - **Definite Noun Phrases**: These refer back to a specific entity previously mentioned.
    - "The cat was hungry. **The cat** finished its food quickly."
  - **Relative Clauses**: These introduce additional information about a previously mentioned noun.
    - "I bought a new car. The car, **which is blue**, is very fuel-efficient."
- **Pronouns**:
  - Pronouns are words that replace or refer back to a previously mentioned noun or noun phrase. They are a common form of referring expression and are used to avoid repetition.
  - Examples:
    - **Personal Pronouns**: "He," "She," "They," "It"
      - "John is a talented musician. **He** plays the piano beautifully."
    - **Possessive Pronouns**: "His," "Her," "Their," "Its"

        - "Sara forgot **her** umbrella, so **she** got wet in the rain."

- **Demonstratives**:
    - Demonstratives are words that indicate or point to a specific object or group of objects in the immediate context. They provide a clear reference to something.
    - Examples:
        - **This** and **These** (for objects near the speaker):
            - "I need **this** book for my research."
        - **That** and **Those** (for objects farther from the speaker):
            - "Can you pass me **that** remote control?"

Example:

Text: “John met a cat on his way home. The cat was very friendly “

- **Reference Resolution**: In this text, "his" in the second sentence refers to "John." We understand this through reference resolution.
- **Reference Phenomena**: This example demonstrates the **anaphora** phenomenon. "The cat" in the second sentence refers back to "a cat" in the first sentence. This is an example of **cataphora**, where the reference comes after the antecedent.
- **Components of Reference Resolution - ACC goes BRrrrrrr**
    - **Anaphora**
        - Anaphora is a type of reference in which a word or phrase refers back to something mentioned earlier in the text.
        - It's essential to identify the antecedent, which is the specific word or phrase that the anaphor refers to. This often requires syntactic and semantic analysis.
        - For example, in the sentence: "Mary saw her friend, and she was happy," the anaphor "she" refers back to "Mary," so "Mary" is the antecedent in this case. The antecedent provides the necessary context to make sense of the anaphor's reference.
        - Anaphora is a broader category that includes expressions that refer back to something previously mentioned in a text or conversation. Anaphoric

expressions can include pronouns, but they are not limited to pronouns.

- **Cataphora**
  - Cataphora is less common than anaphora and involves a word or phrase that refers to something mentioned later in the text.
  - Resolving cataphoric references may require understanding the context and the information that follows.
  - For example, in the sentence: "If you find it, bring it to me," "it" is a cataphoric reference to something that is introduced later.
- **Coreference**
  - Coreference resolution involves identifying when two or more expressions in the text refer to the same entity or concept.
  - It is essential for understanding the relationships between different parts of a text.
  - For example, in the sentence: "John met Jane. He likes her," the resolution of "He" and "her" as referring to John and Jane, respectively, is an example of coreference resolution.
- **Bridging Reference**
  - Bridging reference occurs when an expression refers to an entity or concept introduced in a previous sentence or text segment, even if there is no direct mention.
  - It often relies on recognizing the connection between different parts of a discourse.
  - For example, in "The Eiffel Tower is a famous landmark. It stands in Paris," "It" is a bridging reference connecting to "The Eiffel Tower" from the previous sentence.

Explain Hobbs algorithm for pronoun resolution. - Module 5

> Anaphora resolution is also known as pronoun resolution

Anaphora resolution is the process of determining the referent of an anaphor, which is a word or phrase that refers back to something that has been mentioned earlier in the text. For example, in the sentence “John gave Mary a present. She loved it,” the pronoun “she” is an anaphor that refers back to the noun phrase “Mary.”

There are many different factors that can be used to resolve anaphora, including the following:

- The grammatical role of the anaphor and its antecedent. For example, in the sentence “John gave Mary a present. She loved it,” the pronoun “she” is in the same grammatical role (direct object) as its antecedent (Mary). This makes it more likely that “she” refers to Mary.
- The gender and number of the anaphor and its antecedent. For example, in the sentence “The cat chased the mouse. It ran away,” the pronoun “it” is singular and neuter, which matches the antecedent “mouse.”
- The proximity of the anaphor and its antecedent. In general, the closer the anaphor is to its antecedent, the more likely it is that they refer to the same thing.
- The context of the discourse. The meaning of the surrounding text can also be used to resolve anaphora. For example, in the sentence “John gave Mary a present. She loved it,” the context of the discourse tells us that John is the one who gave Mary the present.

Here are some other examples of anaphora resolution:

- “The man saw the dog. It barked.” (The pronoun “it” refers to the dog.)
- “The woman gave the book to the boy. He read it.” (The pronoun “he” refers to the boy.)
- “The cat chased the mouse. It ran away.” (The pronouns “it” and “the mouse” refer to the same thing.)

Here are some of the challenges in anaphora resolution:

- **Ambiguity**: The anaphor may have multiple possible antecedents. For example, in the sentence “The man saw the woman with the telescope. He bought it,” the

pronoun “he” could refer to either the man or the woman.

- **Incompleteness**: The antecedent of the anaphor may not be explicitly mentioned in the text. For example, in the sentence “The cat chased the mouse. It ran away,” the antecedent of the pronoun “it” is not explicitly mentioned.
- **Anaphora across sentences**: The anaphor may refer to something that was mentioned in a previous sentence. For example, in the sentence “The man saw the woman. She was wearing a red dress,” the pronoun “she” refers to the woman mentioned in the previous sentence.

Hobbs Algorithm

- The Hobbs algorithm is used to resolve anaphoric references, particularly pronouns. Here's an example:
  • Text: Mary saw a cat. She petted it.
  • Hobbs Algorithm in Action: The Hobbs algorithm can resolve "She" to "Mary" and "it" to "the cat" by applying a set of rules that consider the linguistic context and the proximity of possible antecedents.
- The Hobbs algorithm is a rule-based algorithm for anaphora resolution. It was first proposed by Jerry Hobbs in 1978.
- The algorithm works by first finding the **syntactic parse tree** of the sentence containing the anaphor.
- The parse tree shows the grammatical relationships between the words in the sentence.
- The Hobbs algorithm then uses the parse tree to identify the possible antecedents of the anaphor.
- The possible antecedents are the noun phrases that are in the same grammatical role as the anaphor and that are within the same scope of quantification as the anaphor.
- The Hobbs Algorithm is a rule-based approach used in natural language processing for pronoun resolution. It aims to determine the referent (antecedent) for a pronoun within a given text.

- The algorithm proceeds by analyzing the context and structure of a sentence to make the most suitable choice for the antecedent.
- Key Characteristics:
  - Rule-Based: Hobbs Algorithm is primarily rule-based, meaning it follows a set of predefined rules and conditions to identify the antecedent for a pronoun.
  - Parse Tree: It relies on the sentence's parse tree, which represents the grammatical structure of the sentence, to navigate through the text.
  - Conditions: The algorithm considers conditions such as proximity, grammatical role, number, gender agreement, definiteness, and the path in the parse tree to make antecedent selections.
  - Heuristic: It employs heuristics to prioritize and rank potential antecedents based on these conditions.

Creating a syntactic parse tree:

Example: The cat chased the mouse

Step 1: Tokenization

- Begin by tokenizing the sentence, breaking it down into individual words: "The," "cat," "chased," "the," and "mouse."

Step 2: Part-of-Speech Tagging

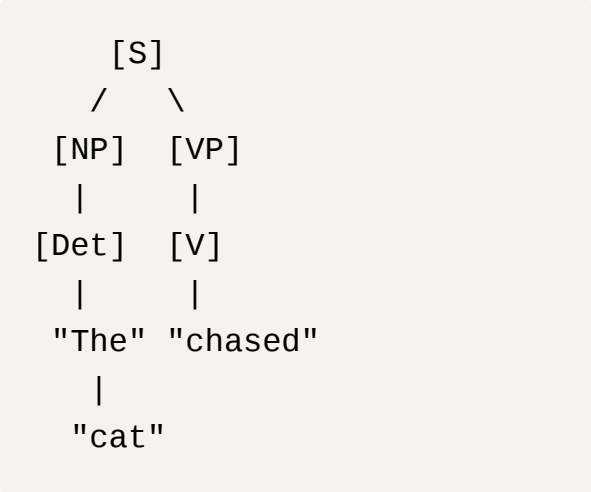
- Assign part-of-speech tags to each token. In this example, the tags might be: "The" (Determiner), "cat" (Noun), "chased" (Verb), "the" (Determiner), and "mouse" (Noun).

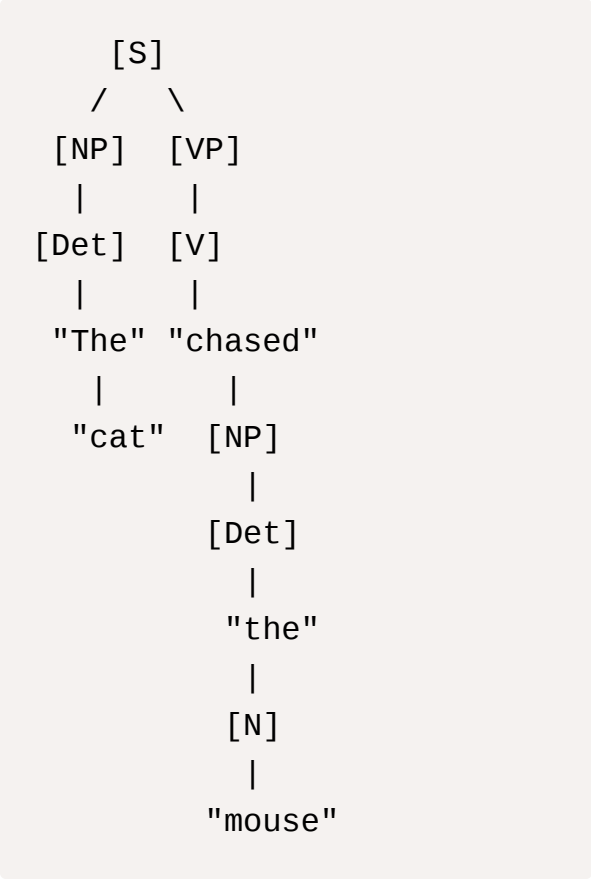
Step 3: Grammar Rules

- Define a set of basic grammar rules. In this simplified example, we will use the following rules:
  - S -> NP VP (A sentence consists of a noun phrase followed by a verb phrase)
  - NP -> Det N (A noun phrase consists of a determiner followed by a noun)
  - VP -> V NP (A verb phrase consists of a verb followed by a noun phrase)

Step 4: Tree Construction

- Use the defined grammar rules to build the parse tree. Start with the highest-level rule, which is "S."





- Advantages:
  • Transparency: The algorithm is based on clear and explicit rules, making it transparent and interpretable.
  • Linguistic Considerations: It takes into account various linguistic features, such as grammatical roles and agreement, which can lead to more accurate pronoun resolution in many cases.
  • Suitable for Coreference: It is effective for resolving coreference

- Disadvantages:
  • Limited Context: The algorithm often considers a limited local context, which may lead to incorrect antecedent selection in complex and ambiguous sentences.
  • Inefficiency: In practice, the Hobbs Algorithm can be computationally expensive and may not be suitable for large-scale applications with vast amounts of text.
  • Ambiguity: Pronoun resolution is inherently ambiguous in many cases, and the algorithm may

issues, where the same entity is referred to by multiple expressions in the text.

not handle all instances accurately, especially in cases requiring world knowledge or nuanced contextual understanding.
• Doesn't Learn: It doesn't learn from data like machine learning approaches, making it less adaptive to varying text types and domains.

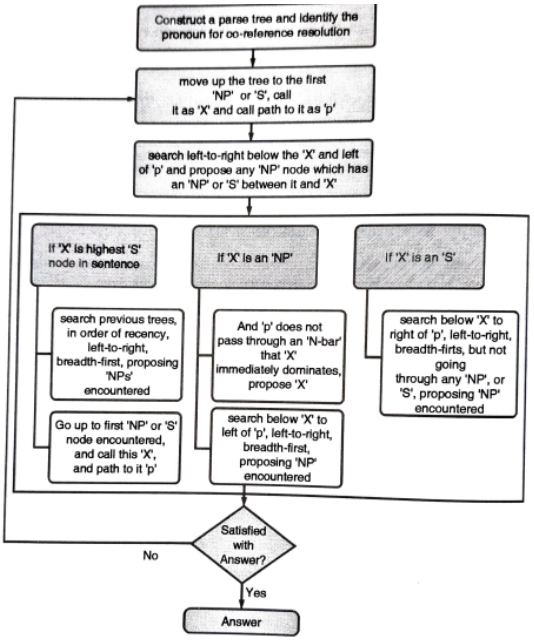


Fig 5.6.1 : Hobbs Algorithm

Hobbs's "Naïve" Algorithm

1. Begin at the NP immediately dominating the pronoun.
2. Go up tree to first NP or S encountered.
   • Call node X, and path to it, p.
   • Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.
3. If X is highest S node in sentence,
   • Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.
4. Otherwise, from X, go up to first NP or S node encountered,
   • Call this X, and path to it p.
5. If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.
6. Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.

7. If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.
8. Go to 2.

Describe in detail Centering Algorithm for reference resolution. - Module 5

- A discourse has a focus, or center. The center typically remains the same for a few sentences, then shifts to a new object.
- The center of a sentence is typically pronominalized.
- Once a center is established, there is a strong tendency for subsequent pronouns to continue to refer to it.
- The Centering theory and algorithm are used to analyze the organization of information in a text. It identifies the "center" of attention in discourse. Here's an example:
  • Dialogue:
  A: What did you think of the movie?
  B: The movie was great. The plot was a bit confusing.
- Centering Theory: In this dialogue, "the movie" is the initial center of attention. However, it shifts to "the plot" in B's response. The Centering algorithm would help identify this shift and understand how discourse organizes information and maintains coherence.
- Centering Theory introduces the concept of "centers" and "links" to represent the organization of entities and their relationships in discourse.
- **Key Components of Centering Theory**:
  - **Center**: A center is a prominent entity in discourse, typically the most recent entity mentioned in a sentence. Centers can be pronouns, noun phrases, or other expressions. In Centering Theory, the discourse context is represented by a list of centers, with the most recent one at the top.

- **Link**: A link is a connection between a pair of adjacent sentences that share a common center. Links are used to establish the relationships between centers as the discourse progresses. Links can be classified into "forward-looking" and "backward-looking" based on their directionality.

- **Centering Algorithms**:
  - Centering algorithms aim to determine how centers are related and how they change in a discourse. They provide a framework for resolving reference and determining which entity a pronoun or noun phrase refers to by examining the context and relationships between centers and links.
  - The primary steps involved in Centering Algorithms include:
    - **Identifying Centers**: The algorithm identifies potential centers within each sentence of a text. Centers are typically noun phrases, pronouns, or other referring expressions.
    - **Establishing Links**: Links are established between pairs of adjacent sentences in the discourse. These links connect sentences that share a common center.
    - **Computing the Score**: Centering algorithms compute a "center score" or "link score" for each pair of centers and their associated links. The scores are based on factors such as the type of entities involved, their salience in the context, and the directionality of the link.
    - **Selecting the Preferred Center**: The algorithm selects the most preferred center in the discourse based on the computed scores. The preferred center is the entity that is most prominent and salient in the context.
    - **Resolving Reference**: Once the preferred center is identified, the algorithm can resolve reference by determining which pronouns or referring expressions in the discourse refer to the preferred center.

- **Applications**:
  Centering algorithms have various applications in NLP, including:
  - Coreference resolution: Identifying when different expressions refer to the same entity.
  - Pronoun resolution: Determining the antecedent for a pronoun.

    - Improving discourse coherence: Understanding how information is organized in a text.
- In summary, centering algorithms are computational methods based on Centering Theory that help resolve reference and track entities in discourse. They play a crucial role in natural language understanding, coreference resolution, and discourse analysis in NLP applications.

Example:

- **Understanding Centers:**
    - **Backward-Looking Center (Cb)**: The most salient entity in the previous utterance that is being discussed or referred to in the current utterance.
    - **Forward-Looking Centers (Cf)**: Potential entities in the current utterance that could become the main topic of discussion in the next utterance.
    - **Preferred Center (Cp)**: Among the forward-looking centers, the one most likely to be the main focus or topic of the next utterance.
- **Example Scenario:**

Let's consider a short discourse to illustrate the centering algorithm:

1. **Utterance U1**: "Emily visited Paris last summer. She loved the museums."
2. **Utterance U2**: "The city fascinated her. She decided to learn French."

- **Applying Centering:**
    - **Utterance U1 Analysis**:
        - **Text**: "Emily visited Paris last summer. She loved the museums."
        - **Backward-Looking Center (Cb)**: None (as this is the first utterance).
        - **Forward-Looking Centers (Cf)**: [Emily, Paris]
        - Emily is the subject of the first sentence, making her a strong candidate for the preferred center in this utterance.
    - **Transition to Utterance U2**:
        - **Text**: "The city fascinated her. She decided to learn French."
        - **Backward-Looking Center (Cb)**: "She" from U1, which refers to "Emily."

        - **Forward-Looking Centers (Cf)**: [The city (Paris), her (Emily)]
        - The backward-looking center "She" (Emily) from U1 is maintained as the subject in the next sentence, indicating continuity and establishing Emily as the central entity. The city (Paris) is introduced as another entity, but Emily remains the main focus.

- **Understanding the Linking:**
    - **Centering Transition Types**:
        - **Maintain**: When the backward-looking center is the same as the preferred center of the current utterance. This usually happens when the discourse is smoothly focused on a single topic or entity.
        - **Shift**: When the backward-looking center changes to a new entity in the current utterance, indicating a shift in focus.
    - **In the Example**:
        - Between U1 and U2, we see a "maintain" transition type as the focus remains on "Emily" (She), even though the discussion includes "Paris" (the city). Emily's actions and experiences are central throughout the discourse, making her the "center" of this discussion.