# DLO8012: Natural Language Processing

# Module-5

# Pragmatics

## CO-5 [8-hrs]

CO-5: Understand the mathematical and linguistic foundations underlying approaches to the various areas in NLP.

# Discourse

- Earlier stages of natural language processing focused primarily on language phenomena that operate at the word or sentence level.
- Of course, language does not normally consist of isolated, unrelated sentences, but instead of collocated, related groups of sentences.
- Such a group of sentences refer as a **discourse. (Three types)**
- Here we will study discourse of particular sort: **a monologue (1)**
- Monologues are characterized by a *speaker* (a term which will be used to include writers, as it is here), and a *hearer* (which, analogously, includes readers).
- The communication flows in only one direction in a monologue, that is, from the speaker to the hearer.

# Discourse

- After reading this chapter, you may have a conversation with a friend about it, which would consist of a much freer interchange.
- Such a discourse is called a **dialogue (2)**.
- In this case, each participant periodically takes turns being a speaker and hearer.
- Unlike a typical monologue, dialogues generally consist of many different types of communicative acts: asking questions, giving answers, making corrections, and so forth.

# Discourse

- Finally, computer systems exist and continue to be developed that allow for *human-computer interaction*, or **HCI (3)**.
- **HCI** has properties HCI that distinguish it from normal human-human dialogue, in part due to the present-day limitations on the ability of computer systems to participate in free, unconstrained conversation.
- A system capable of HCI will often employ a strategy to constrain the conversation in ways that allow it to understand the user's utterances within a limited context of interpretation.

# Discourse

- Consider the discourse shown in example,

**“(18.1)John went to Bill’s car dealership to check out an Acura Integra. He looked at it for about an hour.”**

- In following section, What do pronouns such as *he* and *it* denote? No doubt that the reader had little trouble figuring out that *he* denotes John and not Bill, and that *it* denotes the Integra and not Bill’s car dealership.

# Discourse

**(18.2) I'd like to get from Boston to San Francisco, on either December 5th or December 6th. It's okay if it stops in another city along the way.**

- The system had to figure out that it denotes the flight that the user wants to book in order to perform the appropriate action.
- Similarly, information extraction systems must frequently extract information from utterances that contain pronouns. For instance,
- If an information extraction system is confronted with passage (18.3),

# Discourse

**(18.3) First Union Corp is continuing to wrestle with severe problems unleashed by a botched merger and a troubled business strategy.**

**According to industry insiders at Paine Webber, their president, John R. Georgius, is planning to retire by the end of the year.**

- It must correctly identify First Union Corp as the denotation of their (as opposed to Paine Webber, for instance) in order to extract the correct event.

# Module-5

## Contents:

- Discourse
- Reference resolution
- Reference phenomenon
- Syntactic & semantic constraints on co reference

# Reference Resolution

- The problem of **reference**, is the process by which speakers use expressions like *John* and *he* in passage (18.1) to denote a person named *John*.
- Our discussion requires 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 (18.1) are referring expressions, and John is their referent.

# Reference Resolution

- Two referring expressions that are used to refer to the same entity are said to **corefer**, thus *John* and *he* are corefer.
- There is also a term for a referring expression that licenses the use of another, in the way that the mention of John allows John to be subsequently referred to using he.
- We call *John* the **antecedent** of *he*.
- Reference to an entity that has been previously introduced into the discourse is called **anaphora**, and the referring expression used is said to be **anaphoric**.
- In passage (18.1), the pronouns *he* and *it* are therefore **anaphoric**.

# Reference Resolution

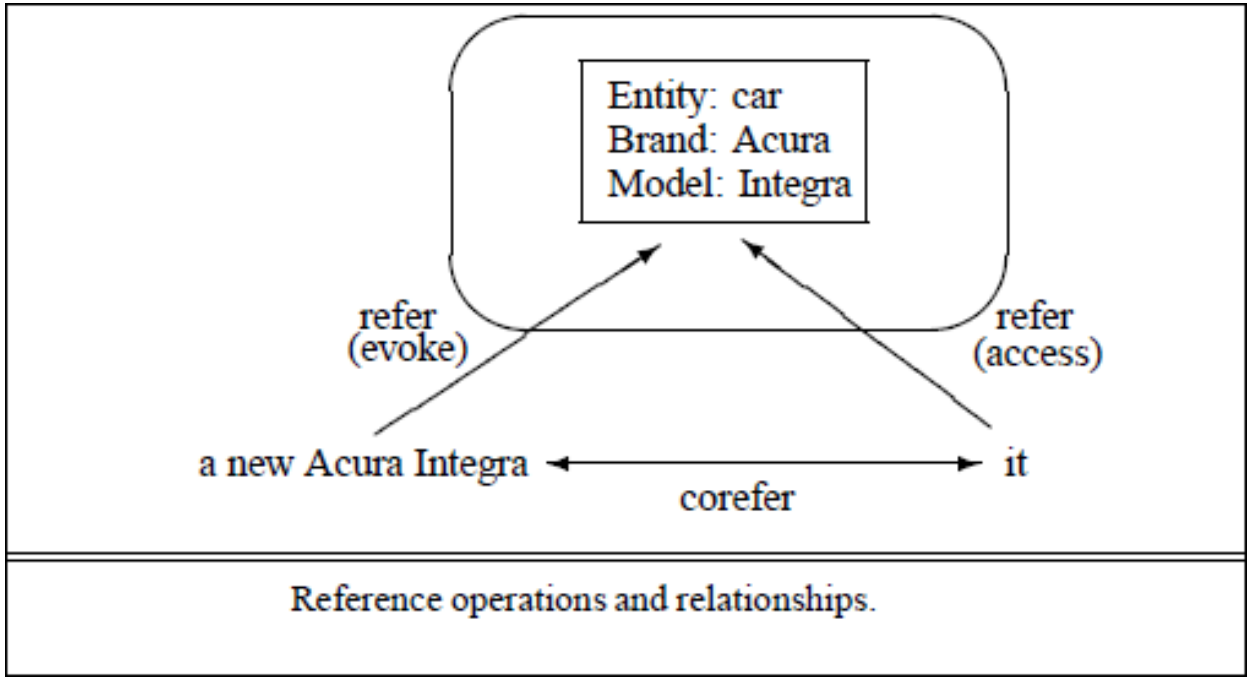
- Natural languages provide speakers with a variety of ways to refer to entities.
- Say that your friend has an Acura Integra automobile and you want to refer to it.
- Depending on the operative discourse context, you might *say it*, *this*, *that*, *this car, that car, the car, the Acura*, *the Integra*, *or my friend's car,* among many other possibilities.
- However, you are not free to choose between any of these alternatives in any context.
- For instance, you cannot simply say it or the Acura if the hearer has no prior knowledge of your friend's car, it has not been mentioned before, and it is not in the immediate surroundings of the discourse participants (i.e., the **situational context** of the discourse).

# Reference Resolution

- The discourse model contains representations of the entities that have been referred to in the discourse and the relationships in which they participate.
- Thus, there are two components required by a system to successfully produce and interpret referring expressions: a method for constructing a discourse model that evolves with the dynamically-changing discourse it represents, and a method for mapping between the signals that various referring expressions encode and the hearer's set of beliefs, the latter of which includes this discourse model.
- We will speak in terms of two fundamental operations to the discourse model.

# Reference Resolution

- When a referent is first mentioned in a discourse, we say that a representation for it is **evoked** into the model.
- Upon subsequent mention, this representation is **accessed** from the model.
- The operations and relationships are illustrated in Figure




Reference operations and relationships.

# Reference Resolution

- Consider the possibilities in example (18.4), adapted from Webber (1991).

(18.4) According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.

a. But that turned out to be a lie.

b. But that was false.

c. That struck me as a funny way to describe the situation.

d. That caused Sue to become rather poor.

e. That caused them both to become rather poor.

# Reference Resolution

- Consider the possibilities in example (18.4), adapted from Webber (1991).

(18.4) According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.

a. But that turned out to be a lie.

b. But that was false.

c. That struck me as a funny way to describe the situation.

d. That caused Sue to become rather poor.

e. That caused them both to become rather poor.

- The referent of that is a speech act in (18.4a), a proposition in (18.4b), a manner of description in (18.4c), an event in (18.4d), and a combination of several events in (18.4e).
- The field awaits the development of robust methods for interpreting these types of reference.

# Reference Phenomenon

- The set of referential phenomena that natural languages provide is quite rich indeed.
- In this section, we provide a brief description of several basic reference phenomena.
- We first survey five types of referring expression: *indefinite noun phrases, definite noun phrases, pronouns, demonstratives, and one-anaphora.*
- We then describe three types of referents that complicate the reference resolution problem: *inferrables, discontinuous sets, and generics.*

# Reference Phenomenon

1 • **Indefinite Noun Phrases:** Indefinite reference introduces entities that are new to the hearer into the discourse context.

2 • The most common form of indefinite reference is marked with the determiner *a (or an)*, as in (18.5), but it can also be marked by a quantifier such as *some (18.6)* or even the determiner *this (18.7)*.

- *(18.5) I saw an Acura Integra today.*
- *(18.6) Some Acura Integras were being unloaded at the local dealership today.*
- *(18.7) I saw this awesome Acura Integra today.*

# Reference Phenomenon

**Indefinite Noun Phrases:**

- Such noun phrases evoke a representation for a new entity that satisfies the given description into the discourse model.
- The indefinite determiner a does not indicate whether the entity is identifiable to the speaker, which in some cases leads to a *specific/non-specific ambiguity*.
- Example (18.5) only has the specific reading, since the speaker has a particular Integra in mind, particularly the one she saw. In sentence (18.8), on the other hand, both readings are possible.
- *(18.8) I am going to the dealership to buy an Acura Integra today.*

# Reference Phenomenon

**Indefinite Noun Phrases:**

- That is, the speaker may already have the Integra picked out (specific) or may just be planning to pick one out that is to her liking (nonspecific).
- The readings may be disambiguated by a subsequent referring expression in some contexts; if this expression is definite then the reading is specific (*I hope they still have it*), and if it is indefinite then the reading is nonspecific (*I hope they have a car I like*).
- This rule has exceptions, however; for instance, definite expressions in certain modal contexts (*I will park it in my garage*) are compatible with the nonspecific reading.

# Reference Phenomenon

**Definite Noun Phrases:**

- Definite reference is used to refer to an entity that is identifiable to the hearer, either because it has already been mentioned in the discourse context (and thus is represented in the discourse model), it is contained in the hearer's set of beliefs about the world, or the uniqueness of the object is implied by the description itself.
- The case in which the referent is identifiable from discourse context is shown in (18.9).

*(18.9) I saw an Acura Integra today. The Integra was white and needed to be washed.*

# Reference Phenomenon

**Definite Noun Phrases:**

- Examples in which the referent is either identifiable from the hearer's set of beliefs or is inherently unique are shown in (18.10) and (18.11) respectively.

*(18.10) The Indianapolis 500 is the most popular car race in the US.*

*(18.11) The fastest car in the Indianapolis 500 was an Integra.*

- Definite noun phrase reference requires that an entity be accessed from either the discourse model or the hearer's set of beliefs about the world.
- In the latter case, it also evokes a representation of the referent into the discourse model.

# Reference Phenomenon

**Pronouns:**

1 • Another form of definite reference is pronominalization, illustrated in example (18.12).

*(18.12) I saw an Acura Integra today. It was white and needed to be washed.*

2 • The constraints on using pronominal reference are stronger than for full definite noun phrases, requiring that the referent have a high degree of activation or **salience** in the discourse model.

3 • Pronouns usually (but not always) refer to entities that were introduced no further than one or two sentences back in the ongoing discourse, whereas definite noun phrases can often refer further back.

# Reference Phenomenon

**Pronouns:**

- This is illustrated by the difference between sentences (18.13d) and (18.13d').

(18.13) a. John went to Bob's party, and parked next to a beautiful Acura Integra.

b. He went inside and talked to Bob for more than an hour.

c. Bob told him that he recently got engaged.

d. ?? He also said that he bought it yesterday.

d.' He also said that he bought the Acura yesterday.

- By the time the last sentence is reached, the Integra no longer has the degree of salience required to allow for pronominal reference to it.

# Reference Phenomenon

**Pronouns:**

- Pronouns can also participate in **cataphora**, in which they are mentioned before their referents are, as in example (18.14).

*(18.14) Before he bought it, John checked over the Integra very carefully.*

- Here, the pronouns *he* and *it* both occur before their referents are introduced.

- Pronouns also appear in quantified contexts in which they are considered to be **bound**, as in example (18.15).

*(18.15) Every woman bought her Acura at the local dealership.*

- Under the relevant reading, her does not refer to some woman in context, but instead behaves like a variable bound to the quantified expression every woman.

# Reference Phenomenon

**Demonstratives:**

1 • Demonstrative pronouns, like *this* and *that*, behave somewhat differently that simple definite pronouns like *it*.

2 • They can appear either alone or as determiners, for instance, *this Acura*, *that Acura*.

3 • The choice between two demonstratives is generally associated with some notion of spatial proximity: *this* indicating closeness and *that* signalling distance.

4 • Spatial distance might be measured with respect to the discourse participants' situational context, as in (18.16).

*(18.16) [John shows Bob an Acura Integra and a Mazda Miata]*

*Bob (pointing): I like this better than that.*

# Reference Phenomenon

**Demonstratives:**

- Alternatively, distance can be metaphorically interpreted in terms of conceptual relations in the discourse model.
- For instance, consider example (18.17).

*(18.17) I bought an Integra yesterday. It's similar to the one I bought five years ago.* ***That one*** *was really nice, but I like* ***this one*** *even better.*

- Here, that one refers to the Acura bought five years ago (greater temporal distance), whereas this one refers to the one bought yesterday (closer temporal distance).

# Reference Phenomenon

**One Anaphora:**

- *One*-anaphora, exemplified in (18.18), blends properties of definite and indefinite reference.

*(18.18) I saw no less than 6 Acura Integras today. Now I want one.*

- This use of *one* can be roughly paraphrased by *one of them*, in which *them* refers to a plural referent (or generic one, as in the case of (18.18), see below), and *one* selects a member from this set.
- Thus, *one* may evoke a new entity into the discourse model, but it is necessarily dependent on an existing referent for the description of this new entity.

# Reference Phenomenon

**One Anaphora:**

- This use of *one* should be distinguished from the formal, non-specific pronoun usage in (18.19), and its meaning as the number one in (18.20).

*(18.19) One shouldn't pay more than twenty thousand dollars for an Acura.*

*(18.20) John has two Acuras, but I only have one.*

# Reference Phenomenon

**Inferrables:**

- Now that we have described several types of referring expressions, we now turn our attention to a few interesting types of referents that complicate the reference resolution problem.
- For instance, in some cases a referring expression does not refer to an entity that has been explicitly evoked in the text, but instead one that is inferentially related to an evoked entity.
- Such referents are called *inferrables* (Haviland and Clark, 1974; Prince, 1981).
- Consider the expressions *a door* and *the engine* in sentence (18.21).

*(18.21) I almost bought an Acura Integra today, but **a door** had a dent and **the engine** seemed noisy.*

# Reference Phenomenon

**Inferrables:**

- The indefinite noun phrase *a door* would normally introduce a new door into the discourse context, but in this case the hearer is to infer something more: that it is not just any door, but one of the doors of the Integra.
- Similarly, the use of the definite noun phrase *the engine* normally presumes that an engine has been previously evoked or is otherwise uniquely identifiable.
- Here, no engine has been explicitly mentioned, but the hearer infers that the referent is the engine of the previously mentioned Integra.
- Inferrables can also specify the results of processes described by utterances in a discourse.

# Reference Phenomenon

**Inferrables:**

- Consider the possible follow-ons (a-c) to sentence (18.22) in the following recipe (from Webber and Baldwin (1992)):

(18.22) *Mix the flour, butter, and water.*

*a. Kneed **the dough** until smooth and shiny.*

*b. Spread **the paste** over the blueberries.*

*c. Stir **the batter** until all lumps are gone.*

- Any of the expressions the dough (a solid), the batter (a liquid), and the paste (somewhere in between) can be used to refer to the result of the actions described in the first sentence, but all imply different properties of this result.

# Reference Phenomenon

**Discontinuous Sets:**

- In some cases, references using plural referring expressions like *they* and *them* refer to sets of entities that are evoked together, for instance, using another plural expression (*their Acuras*) or a conjoined noun phrase (*John and Mary*):

*(18.23) John and Mary love their Acuras. They drive them all the time.*

- However, plural references may also refer to sets of entities that have been evoked by discontinuous phrases in the text:

*(18.24) John has an Acura, and Mary has a Mazda. They drive them all the time.*

# Reference Phenomenon

**Discontinuous Sets:**

3 • Here, *they* refers to John and Mary, and likewise *them* refers to the Acura and the Mazda.

4 • Note also that the second sentence in this case will generally receive what is called a *pairwise* or *respectively* reading, in which John drives the Acura and Mary drives the Mazda, as opposed to the reading in which they both drive both cars.

# Reference Phenomenon

**Generics:**

1 • Making the reference problem even more complicated is the existence of *generic* reference.

2 • Consider example (18.25).

*(18.25) I saw no less than 6 Acura Integras today.* ***They*** *are the coolest cars.*

3 • Here, the most natural reading is not the one in which *they* refers to the particular 6 Integras mentioned in the first sentence, but instead to the class of Integras in general.

# Module-5

## Contents:

- Discourse
- Reference resolution
- Reference phenomenon
- Syntactic & semantic constraints on co reference

# Syntactic & semantic constraints on co reference

syntactic and semantic constraints on coreference refer to rules that help filter possible referents of a referring expression.

- One step that needs to be taken in any successful reference resolution algorithm is to filter the set of possible referents on the basis of certain relatively hard-and-fast constraints.

1 **Number Agreement :**

- Referring expressions and their referents must agree in number; for English, this means distinguishing between *singular* and *plural* references.
- A categorization of pronouns with respect to number is shown in Figure

# Syntactic & semantic constraints on co reference

| Singular | Plural | Unspecified |
|---|---|---|
| she, her, he, him, his, it | we, us, they, them | you |

Number agreement in the English pronominal system.

- The following examples illustrate constraints on number agreement.

(18.26) John has a new Acura. It is red.

(18.27) John has three new Acuras. They are red.

(18.28) * John has a new Acura. They are red.

(18.29) * John has three new Acuras. It is red.

# Syntactic & semantic constraints on co reference

2 **Person and Case Agreement:**

- English distinguishes between three forms of person: first, second, and third.
- A categorization of pronoun types with respect to person is shown in Figure.

| | First | Second | Third |
|---|---|---|---|
| Nominative | I, we | you | |
| Accusative | me, us | | |
| Genitive | | | |

Person and case agreement in the English pronominal system

# Syntactic & semantic constraints on co reference

**Person and Case Agreement:**

- The following examples illustrate constraints on person agreement.

(18.30) You and I have Acuras. We love them.

(18.31) John and Mary have Acuras. They love them.

(18.32) * John and Mary have Acuras. We love them. (where *We=John* and Mary)

(18.33) * You and I have Acuras. They love them. (where *They=You* and I)

- In addition, English pronouns are constrained by case agreement; different forms of the pronoun may be required when placed in subject position (nominative case, e.g., *he*, *she*, *they*), object position (accusative case, e.g., *him*, *her*, *them*), and genitive position (genitive case, e.g., *his Acura*, *her Acura*, *their Acura*).

# Syntactic & semantic constraints on co reference

**Gender Agreement:**

- Referents also must agree with the gender specified by the referring expression.
- English third person pronouns distinguish between male, female, and nonpersonal genders, and unlike many languages, the first two only apply to animate entities.
- Some examples are shown in Figure

| masculine | feminine | nonpersonal |
|---|---|---|
| he, him, his | she, her | it |

Gender agreement in the English pronominal system

# Syntactic & semantic constraints on co reference

**Gender Agreement:**

- The following examples illustrate constraints on gender agreement.

(18.34) John has an Acura. He is attractive. (he=John, not the Acura)

(18.35) John has an Acura. It is attractive. (it=the Acura, not John)

**Syntactic Constraints**

- Reference relations may also be constrained by the syntactic relationships between a referential expression and a possible antecedent noun phrase when both occur in the same sentence.

# Syntactic & semantic constraints on co reference

**Syntactic Constraints**

- For instance, the pronouns in all of the following sentences are subject to the constraints indicated in brackets.

  John bought himself a new Acura. [himself = John]

  John bought him a new Acura. [him ≠ John]

  John said that Bill bought him a new Acura. [him ≠ Bill]

  John said that Bill bought himself a new Acura. [himself = Bill]

  He said that he bought John a new Acura. [He ≠ John; he ≠ John]

# Thank You…