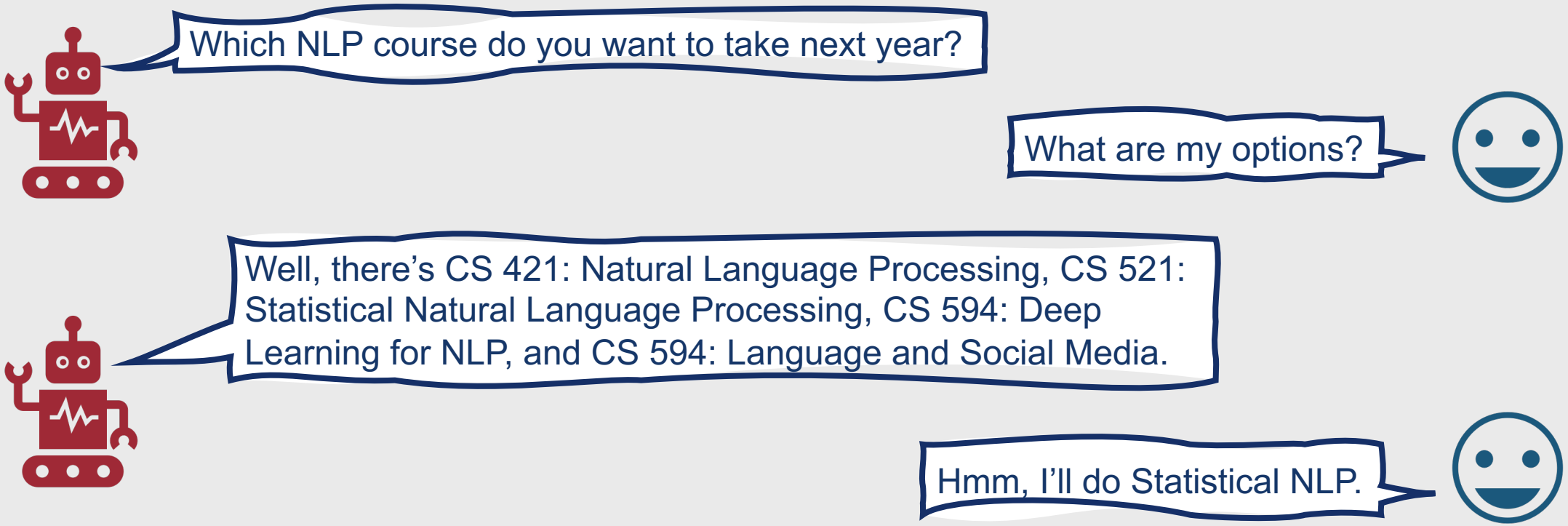
Overview of Coreference Resolution

Natalie Parde
UIC CS 421

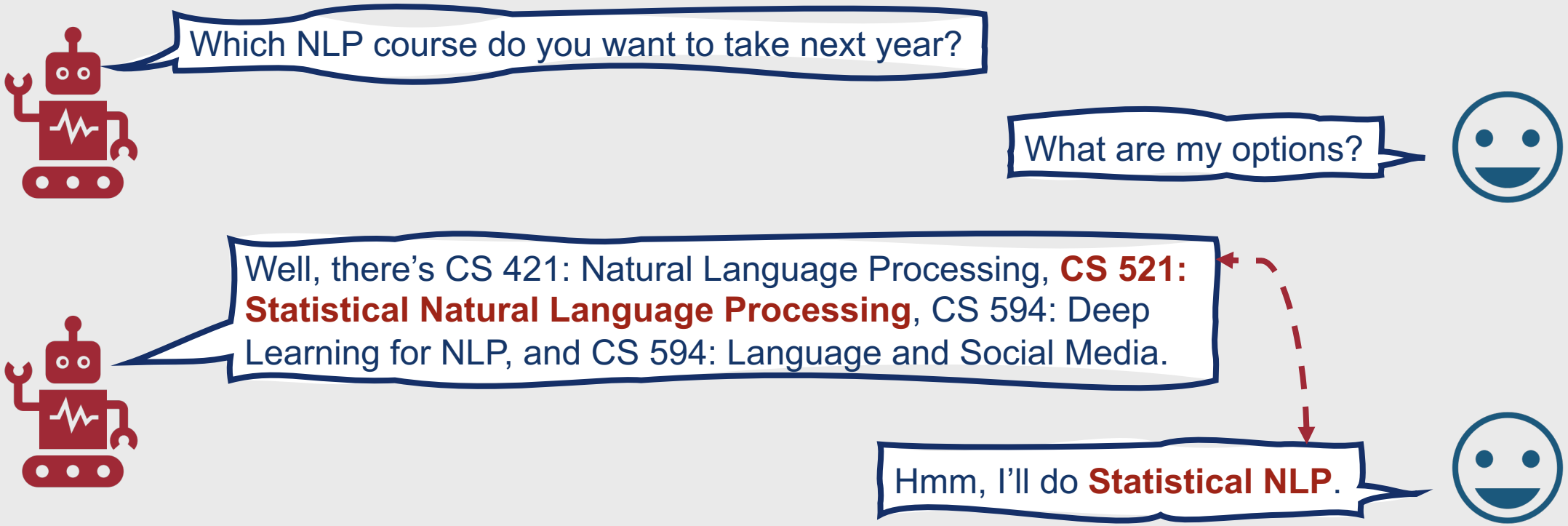
What is coreference resolution?

- The process of automatically identifying expressions that refer to the same entity

Coreference resolution is essential to creating high-performing NLP systems.



Coreference resolution is essential to creating high-performing NLP systems.



Both humans and NLP systems interpret language with respect to a discourse model.

- **Discourse model:** Mental model that is built incrementally, containing representations of entities, their properties, and the relations between them
- **Referent:** The discourse entity itself
  - (CS 521: Statistical Natural Language Processing)
- **Referring expression:** The linguistic expression referring to a referent
  - "CS 521"
  - "CS 521: Statistical Natural Language Processing"
  - "521"
  - "Statistical NLP"
- Two or more referring expressions that refer to the same discourse entity are said to **corefer**

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The University of Illinois at Chicago is an excellent place to study natural language processing. UIC has many faculty currently working in the area, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. The school is located in bustling downtown Chicago, and as a bonus it will be opening a snazzy new (non-brutalist) CS building in 2023.

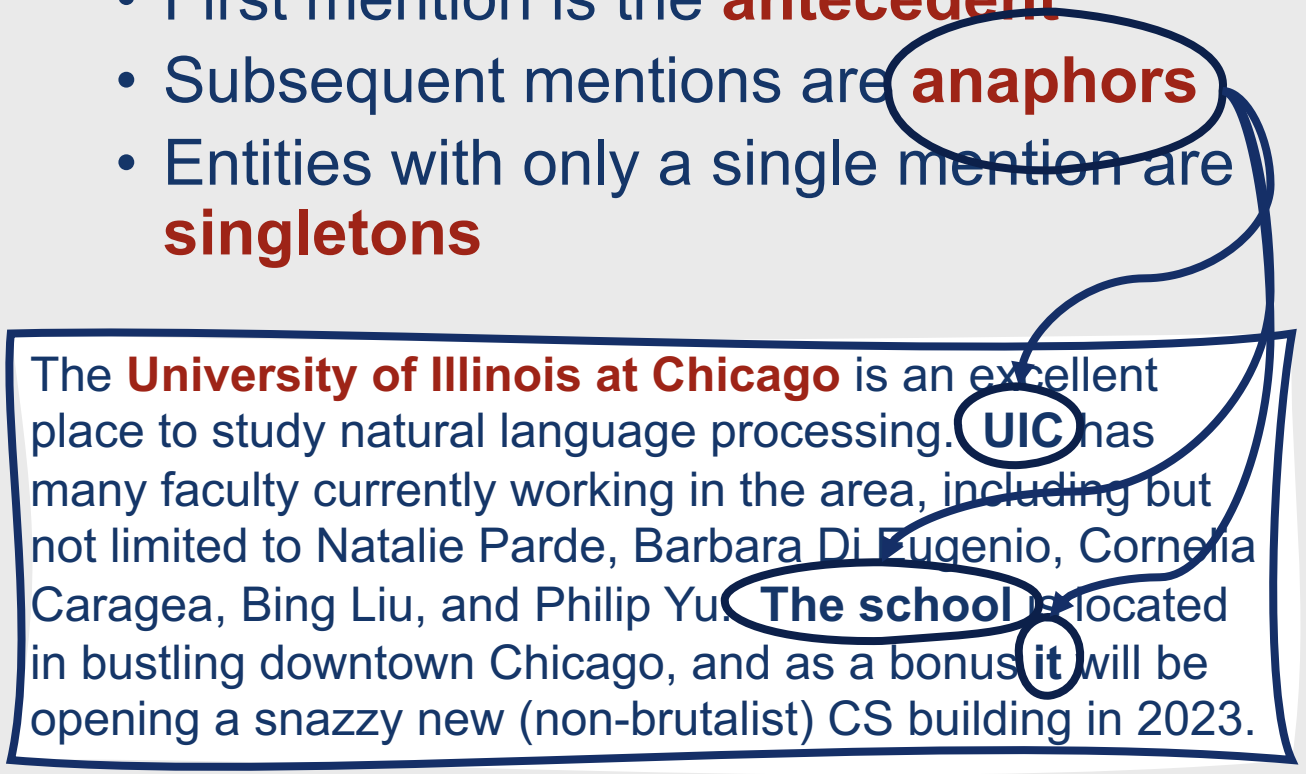
Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. UIC has many faculty currently working in the area, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. The school is located in bustling downtown Chicago, and as a bonus it will be opening a snazzy new (non-brutalist) CS building in 2023.

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**



Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. **UIC** has many faculty currently working in the area, including but not limited to **Natalie Parde**, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. **The school** is located in bustling downtown Chicago, and as a bonus **it** will be opening a snazzy new (non-brutalist) CS building in 2023.

Coreference Chains

- A set of coreferring expressions is often called a **coreference chain**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. **UIC** has many faculty currently working in the area, including but not limited to **Natalie Parde**, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. **The school** is located in bustling downtown Chicago, and as a bonus **it** will be opening a snazzy new (non-brutalist) CS building in 2023.

Two Key Tasks

- **Coreference resolution** thus generally comprises two key tasks:
  - Identify **referring expressions** (mentions of entities)
  - Cluster them into **coreference chains**
- We can also perform **entity linking** to map coreference chains to real-world entities
  - {“University of Illinois at Chicago”, “UIC”, “The school”, “it”} → https://en.wikipedia.org/wiki/University_of_Illinois_at_Chicago