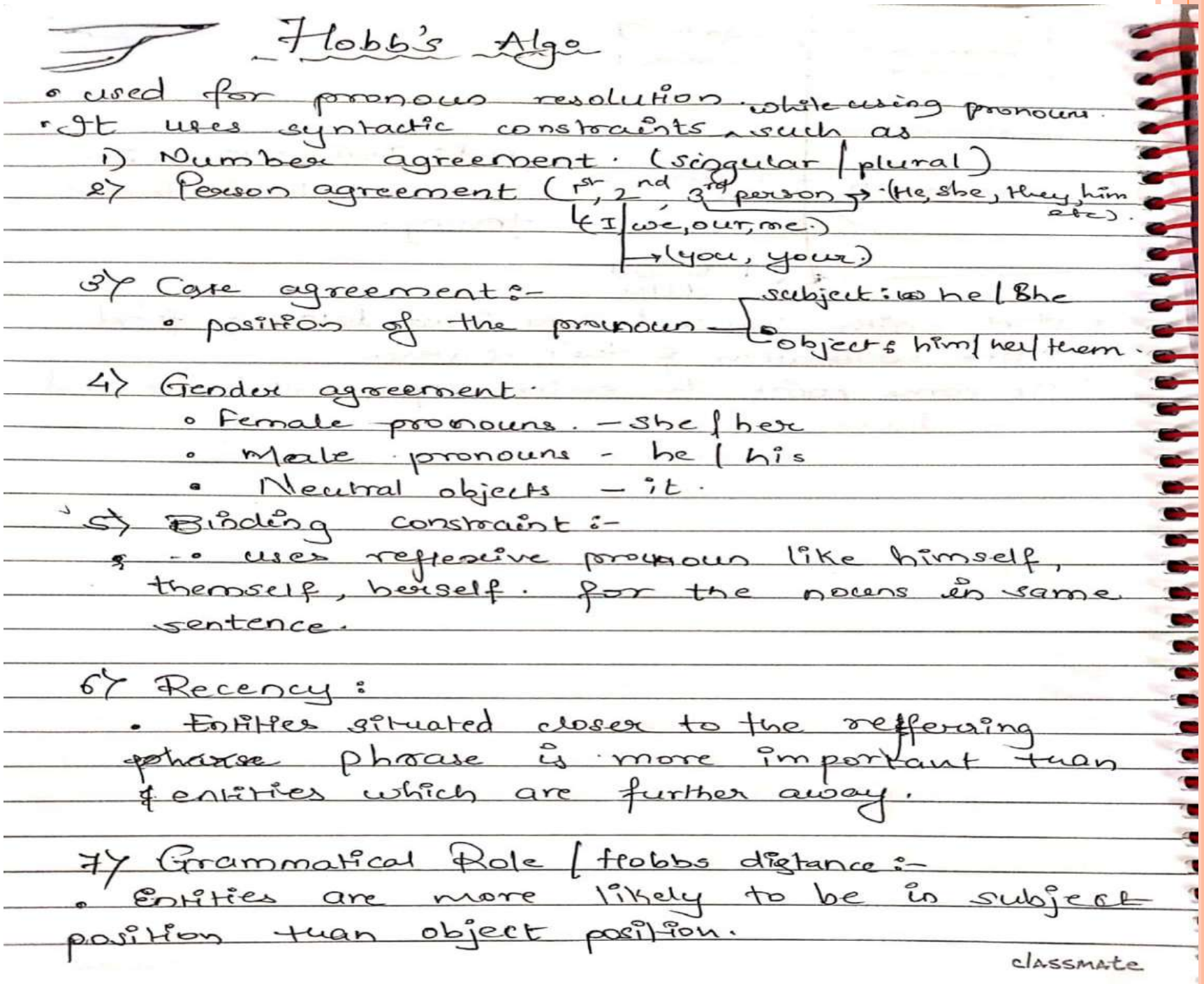
ANAPHORA RESOLUTION- HOBBS ALGORITHM

- Hobbs Algorithm is one of the technique used for **Pronoun Resolution**. But what is Pronoun Resolution?
- Let's understand this with an example.
- You all maybe familiar with this nursery rhyme. Read the text carefully.

Jack and Jill went up the hill
to fetch a pail of water.
Jack fell down and broke his crown
and Jill came tumbling after.

- Now, the question is: To whom the pronoun **'his'** refers to ??
- Well to answer this, we as a human can easily relate that the word 'his' refers to Jack and not to the Jill, hill or the crown.
- But do you think is this task easy for computers as well?
- The answer to this is 'NO'. Guess why 😐 ?
- **Because computers lack Common sense.**

- **Hobbs algorithm** is one of the several approaches for pronoun resolution.
- The algorithm is mainly based on the syntactic parse tree of the sentences.



8. Repeated mentions :-

- Entities that are focused previously appear to be focused later too.

9. Parallelism :-

- He follows structual parallel concept.

10. Verb semantics :-

- Here the verb meaning / phrase helps to identify the entity which will be referred by the pronoun.

Example :-

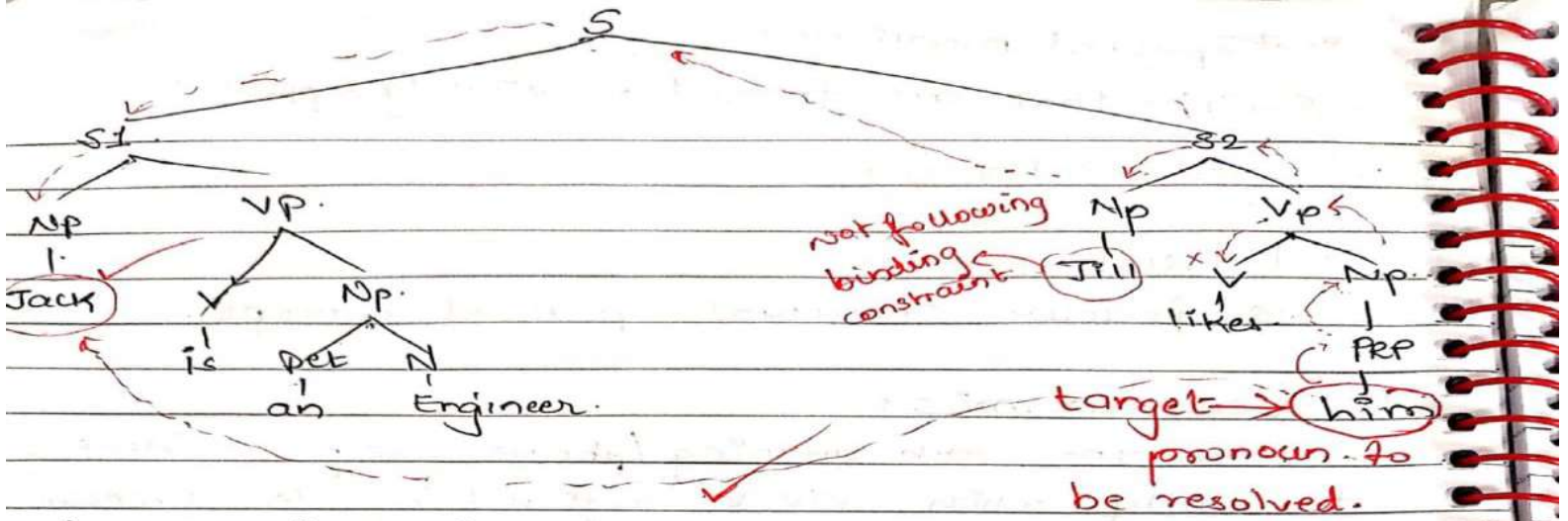
S1 : Jack is an engineer.
S2 : Jill likes him.

Now, the task is to identify what is referred by him? → Jack, Engineer, Jill ?

Steps :-

1) Construct the parse tree for both the sentences & identify the target pronoun which has to be resolved.

↳ In this example : target pronoun = him.



2) Basic idea is to start from the target pronoun & traverse the parse tree ↑ upwards ie towards the root node S & then go to the left of the root.

3) It follows breadth first left to right search.

4) While traversing, when we get a 'NP' look into its left path.

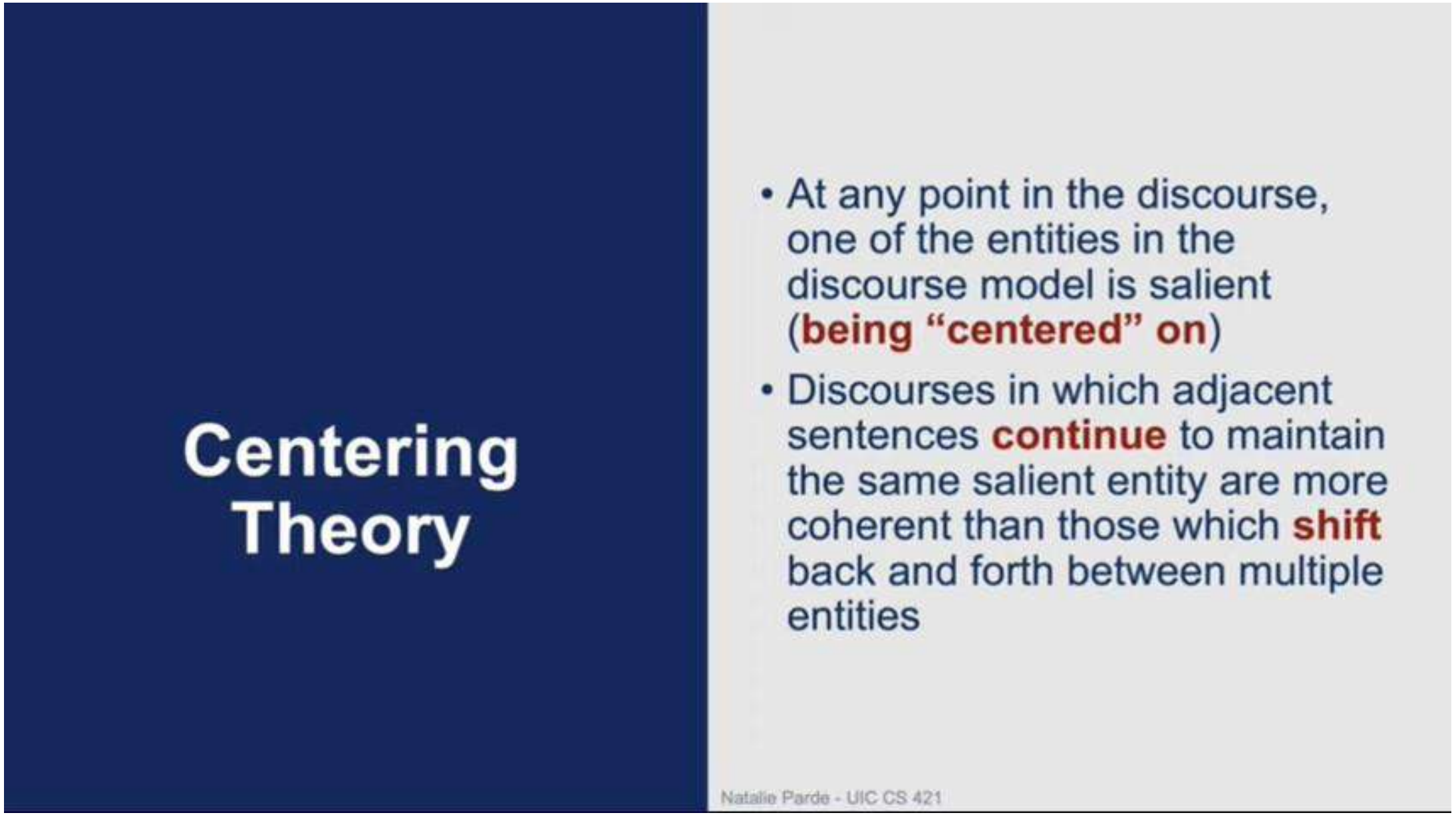
5) If the left path entity follows the syntactic constraints, then we have our answer ie the resolved pronoun.

6) If the left path does not have / or does not follow syntactic constraint, go ↑ upwards to the parent node and then towards root node.

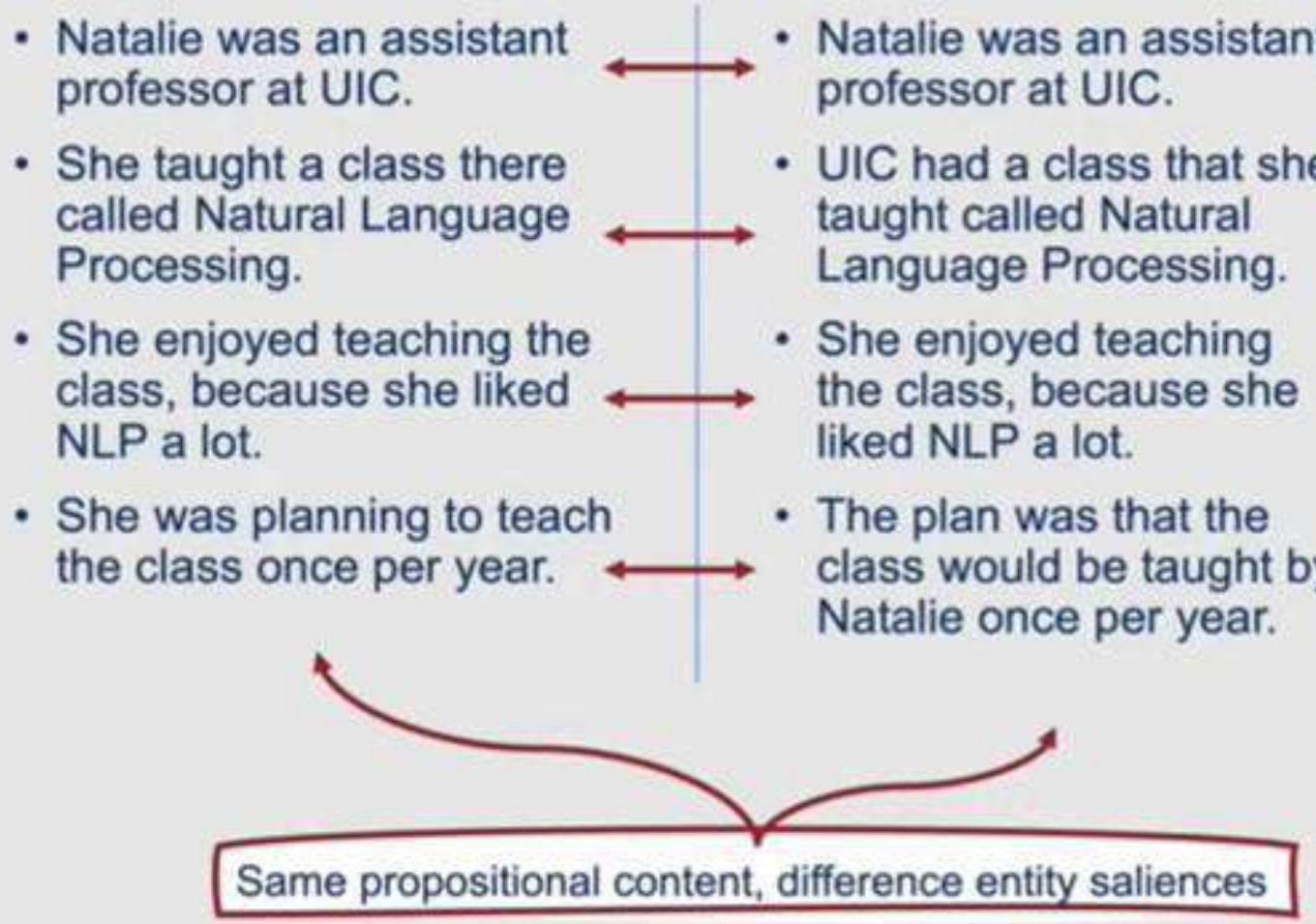
7) Once we reach node & we haven't found the correct resolution, traverse to left branch of the parse tree. & repeat step 5.

8) If correct resolution is not found, then move to parent node & then downwards towards the leaf nodes.

Centering Algorithm:



Centering Theory: Intuition



Centering Theory: Intuition

- Natalie was an assistant professor at UIC.
- She taught a class there called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- She was planning to teach the class once per year.

- Natalie was an assistant professor at UIC.
- UIC had a class that she taught called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- The plan was that the class would be taught by Natalie once per year.

Much more coherent!

How does Centering Theory realize this intuition?

- Maintain two representations for each utterance U_n
  - $C_b(U_n)$: Backward-looking center of U_n
    - Salient entity being focused on in the discourse after U_n is interpreted
  - $C_f(U_n)$: Forward-looking centers of U_n
    - Set of potential future salient entities (potential $C_b(U_{n+1})$)
- Set of $C_f(U_n)$ are ranked based on a variety of factors (e.g., grammatical role)
- Highest-ranked $C_f(U_n)$ is the preferred center C_p

There are four possible intersentential relationships between U_n and U_{n+1}.

- These relationships depend on $C_b(U_{n+1})$, $C_b(U_n)$, and $C_p(U_{n+1})$

Speaker has continued talking about the same entity as in the previous utterance, and plans to keep doing that

	$C_b(U_{n+1}) = C_b(U_n)$ **or undefined** $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

There are four possible intersentential relationships between U_n and U_{n+1}.

- These relationships depend on $C_b(U_{n+1})$, $C_b(U_n)$, and $C_p(U_{n+1})$

Speaker is planning to shift to a new entity

	$C_b(U_{n+1}) = C_b(U_n)$ **or undefined** $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

There are four possible intersentential relationships between U_n and U_{n+1}.

- These relationships depend on $C_b(U_{n+1})$, $C_b(U_n)$, and $C_p(U_{n+1})$

Speaker shifted to a new entity

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

Based on these relationships, we can define two rules.

- Centered entities should be realized as pronouns when they are continued
- Transition states are ordered such that Continue > Retain > Smooth-Shift > Rough-Shift

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

With this in mind, we can revisit the sample texts from earlier….

- Natalie was an assistant professor at UIC.
- She taught a class there called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- She was planning to teach the class once per year.

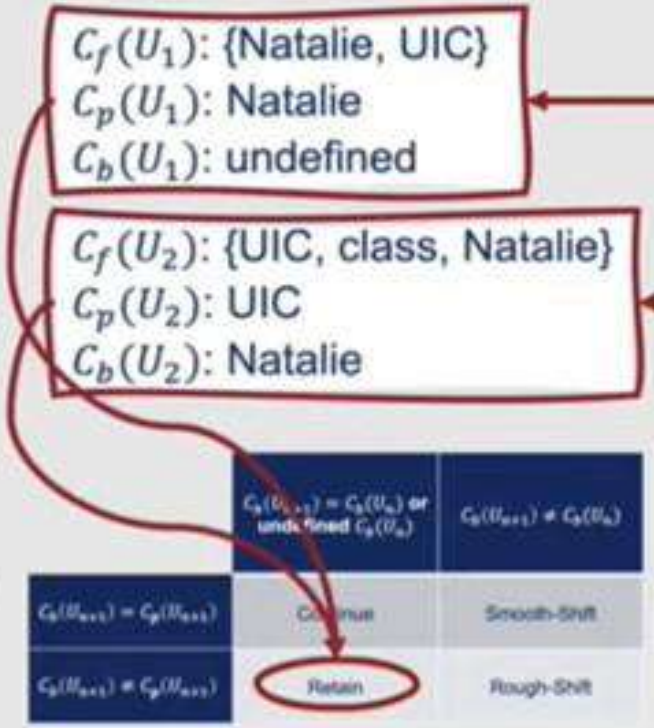
$C_f(U_1)$: {Natalie, UIC}
$C_p(U_1)$: Natalie
$C_b(U_1)$: undefined

$C_f(U_2)$: {Natalie, UIC, class}
$C_p(U_2)$: Natalie
$C_b(U_2)$: Natalie

- Natalie was an assistant professor at UIC.
- UIC had a class that she taught called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- The plan was that the class would be taught by Natalie once per year.

With this in mind, we can revisit the sample texts from earlier….

- Natalie was an assistant professor at UIC.
- She taught a class there called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- She was planning to teach the class once per year.



- Natalie was an assistant professor at UIC.
- UIC had a class that she taught called Natural Language Processing.
- She enjoyed teaching the class, because she liked NLP a lot.
- The plan was that the class would be taught by Natalie once per year.