Discourse Analysis

SLP3 Ch 21, 22; INLP Ch 15, 16

Definition of Discourse

- **Discourse** is the coherent structure of language above the level of sentences or clauses.
- A **discourse** is a coherent structured group of sentences.
- What makes a passage coherent?
  - It has meaningful connections between its utterances.

Example

- Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Lexical Chains

- Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Coreference Chains

- Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Discourse Markers

- Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Features for Discourse Cohesion

- There are three main classes of features for discourse cohesion
  - Lexical overlap/lexical chains
  - Coreference chains ← To be discussed in detail
  - Cue words/discourse markers

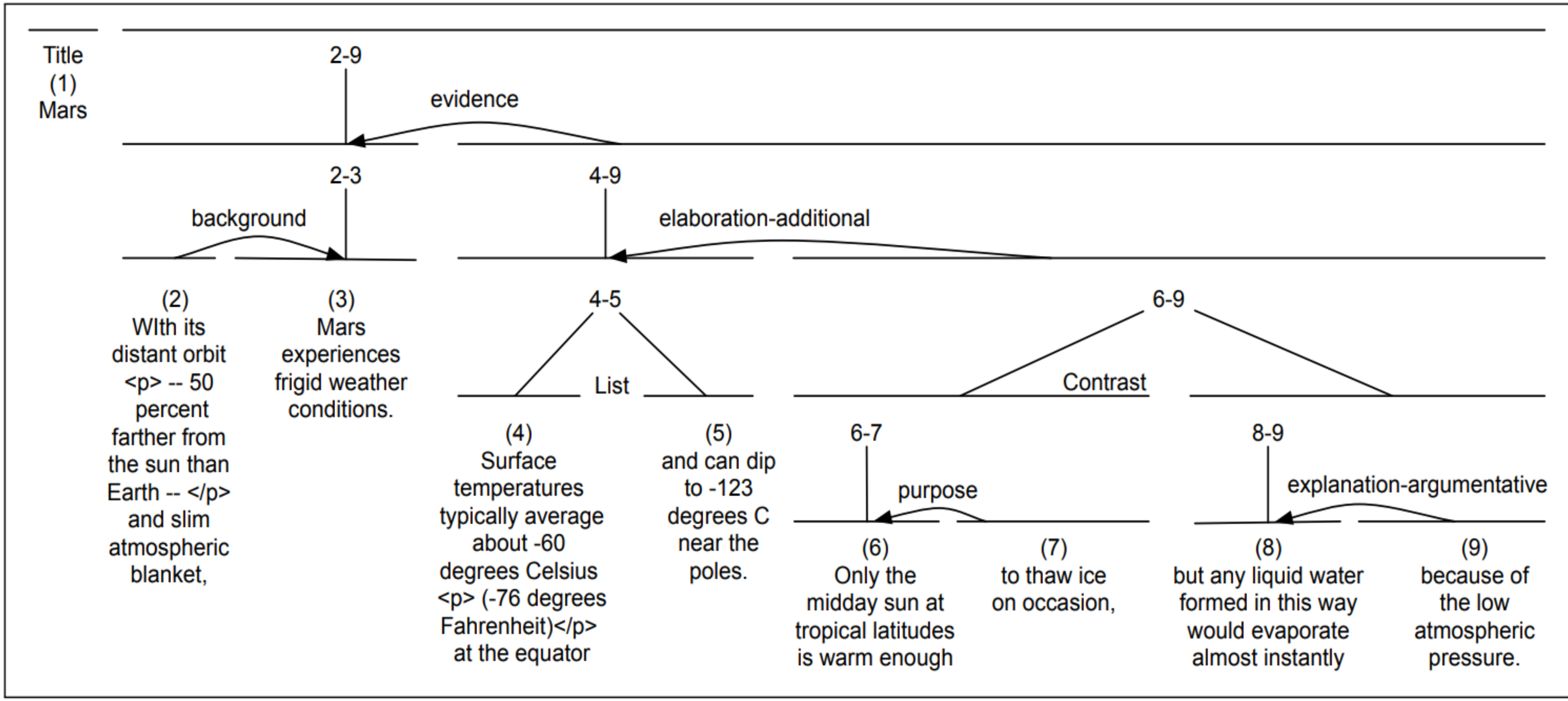
Coherence relations

- Connections between text spans can be specified as a set of coherence relations.
- Rhetorical Structure Theory (RST)
  - Most commonly used model of coherence relations
  - Most relations hold between a nucleus and a satellite.
    - Nucleus: central to the writer's purpose, interpretable independently
    - Satellite: less central and generally only interpretable with respect to the nucleus

Example relations

- **Reason**
  - [$_{NUC}$ Jane took a train from Paris to Istanbul.] [$_{SAT}$ She had to attend a conference.]
- **Elaboration**
  - [$_{NUC}$ Dorothy was from Kansas.] [$_{SAT}$ She lived in the midst of the great Kansas prairies.]
- **Evidence**
  - [$_{NUC}$ Kevin must be here.] [$_{SAT}$ His car is parked outside.]
- **Attribution**
  - [$_{SAT}$ Analysts estimated] [$_{NUC}$ that sales at U.S. stores declined in the quarter, too]
- **List**
  - [$_{NUC}$ Billy Bones was the mate;] [$_{NUC}$ Long John, he was quartermaster]

Hierarchical discourse structure



Asymmetric relations are represented with a curved arrow from the satellite to the nucleus.

Discourse parsing

- Given a sequence of sentences, determine coherence relations between them
- Two stages:
  - EDU segmentation: segment the sentences into a sequence of elementary discourse units
    - Can be cast as sequence labeling
  - RST parsing: build a discourse structure over EDUs
    - Can be cast as constituency or dependency parsing

Coreference Resolution

What is Coreference Resolution

- Identify all mentions that refer to the same entity
- Example:
  - Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Coreference Resolution in Two Steps

- 1. Detect the mentions (easy)
  - "[I] voted for [Nader] because [he] was most aligned with [[my] values]," [she] said
  - mentions can be nested!
- 2. Cluster the mentions (hard)
  - "[I] voted for [Nader] because [he] was most aligned with [[my] values]," [she] said

Types of Mentions

- Indefinite noun phrases
  - *I saw a beautiful Ford Falcon today.*
- Definite noun phrases
  - *I read about it in the new magazine*.
- Pronouns
  - *Emma smiled as cheerfully as she could.*
- Demonstrative Pronouns
  - *Put it back. This one is in better condition.*
- Names
  - *Miss Woodhouse certainly had not done him justice.*
- Zero Anaphora
  - 我刚才精神太紧张。【0】现在比较平静了。

Mention Detection

- Use a POS tagger + constituency parser + named entity tagger to detect different types of NPs and pronouns
- But further filtering is needed
  - Many NPs are not mentions
    - Janet doesn't have <u>a car</u>.
    - Her pay jumped to <u>$2.3 million</u>.
  - Even pronouns may not be mentions
    - It is sunny.
    - It is known that …
    - It is necessary for them to …
- Methods
  - Rules
  - Binary classification
  - Joint inference with mention clustering

Mention Clustering

*"**I** voted for **Nader** because **he** was most aligned with **my** values," **she** said.*

I	Nader	he	my	she

Mention Clustering

*“**I** voted for **Nader** because **he** was most aligned with **my** values,” **she** said.*

Coreference Cluster 1

Nader

he

Coreference Cluster 2

she

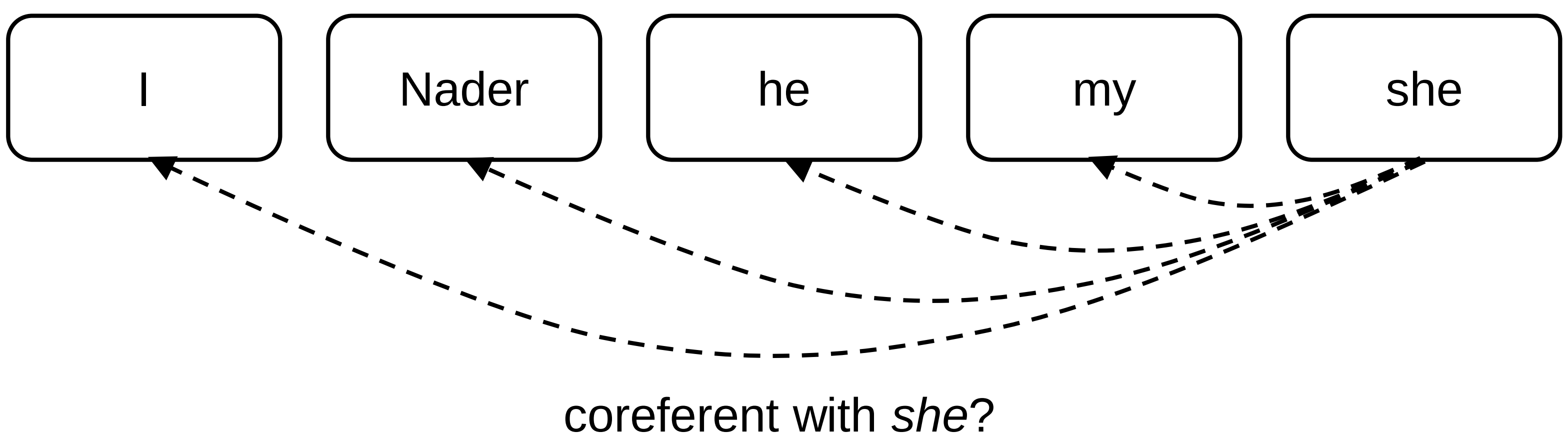
my

I

Mention Clustering

- Train a binary classifier that assigns every pair of mentions a probability of being coreferent
  - Ex: for “she”, look at all candidate antecedents (previously occurring mentions) and decide which are its coreferents

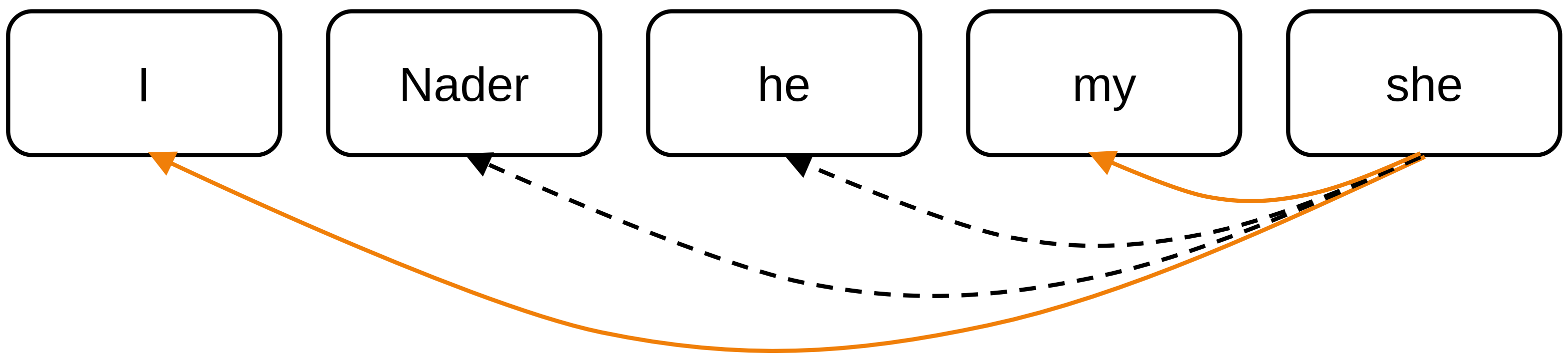
*“**I** voted for **Nader** because **he** was most aligned with **my** values,” **she** said.*



Mention Clustering

- Train a binary classifier that assigns every pair of mentions a probability of being coreferent
  - Ex: for "she", look at all candidate antecedents (previously occurring mentions) and decide which are its coreferents

*"**I** voted for **Nader** because **he** was most aligned with **my** values," **she** said.*

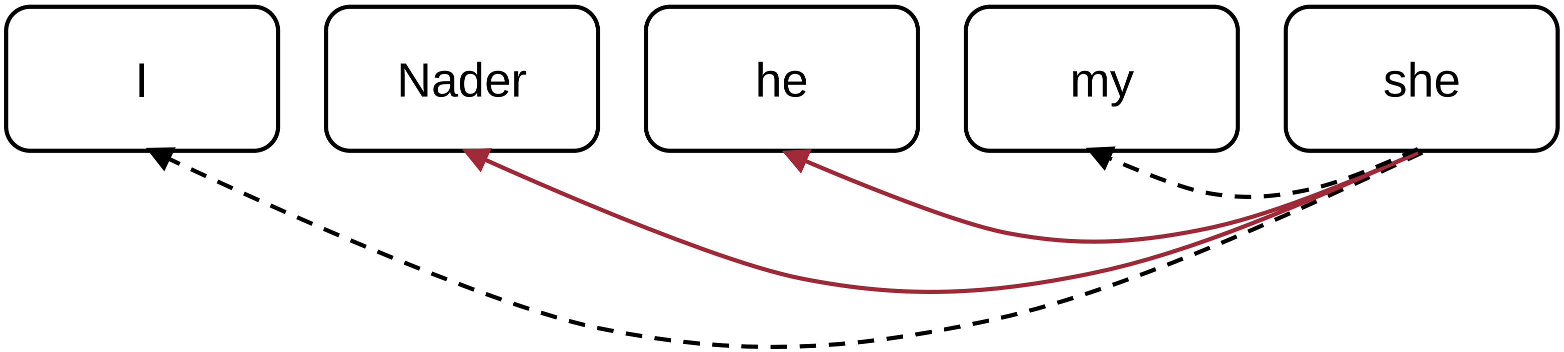


Positive examples: want $q(m_i, m_j)$ to be near 1

Mention Clustering

- Train a binary classifier that assigns every pair of mentions a probability of being coreferent
  - Ex: for “she”, look at all candidate antecedents (previously occurring mentions) and decide which are its coreferents

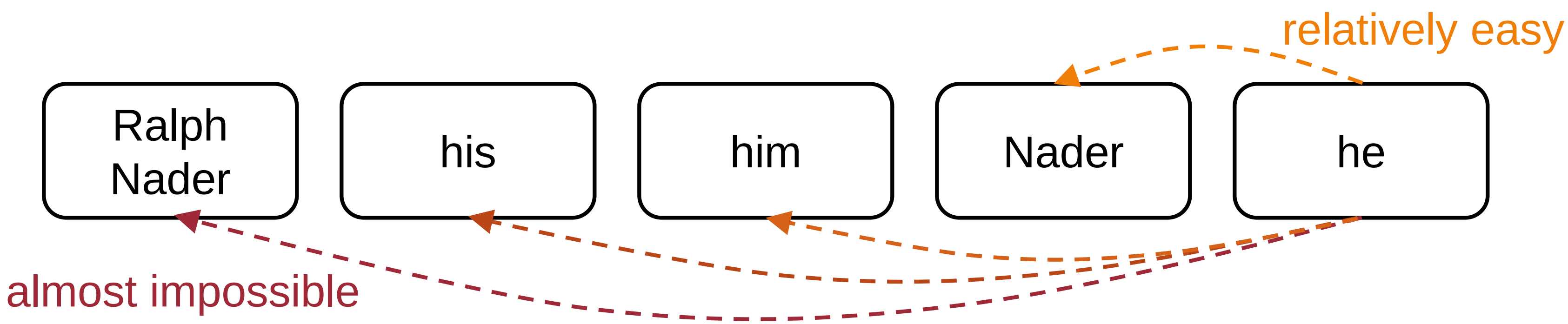
*“**I** voted for **Nader** because **he** was most aligned with **my** values,” **she** said.*



Negative examples: want $q(m_i, m_j)$ to be near 0

Mention Clustering: Disadvantage

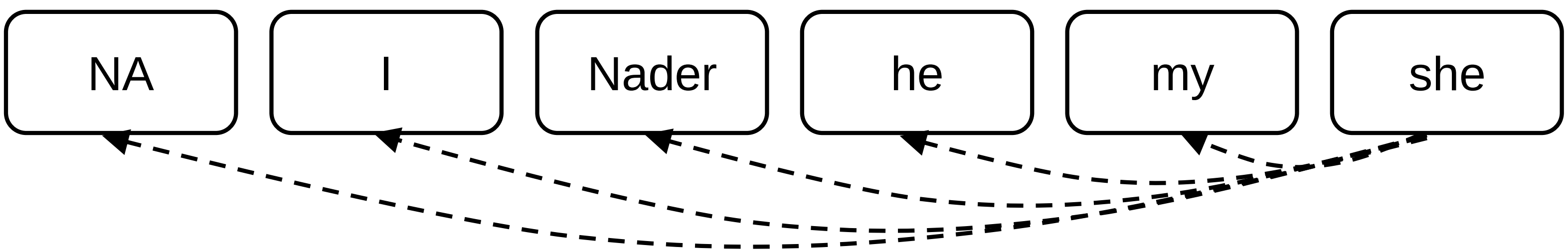
- Suppose we have a long document with the following mentions
  - **Ralph Nader** … **his** … **him** … <several paragraphs> … *voted for **Nader** because **he** …*



- Many mentions only have one clear antecedent
  - But we are asking the model to predict all of them
- Solution: train the model to predict only one antecedent for each mention

Coreference Models: Mention Ranking

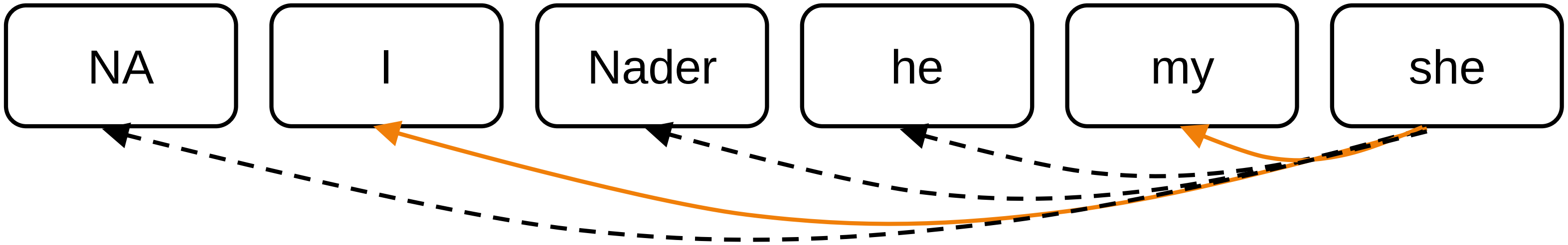
- Assign each mention its highest scoring candidate antecedent according to the model
  - Dummy NA mention allows model to decline linking the current mention to anything ("singleton" or "first" mention, or not a mention)



best antecedent for *she*?

Coreference Models: Mention Ranking

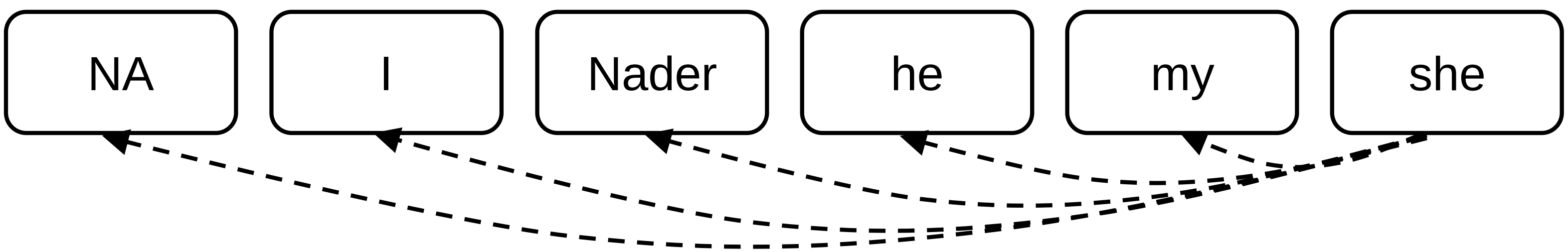
- Assign each mention its highest scoring candidate antecedent according to the model
  - Dummy NA mention allows model to decline linking the current mention to anything (“singleton” or “first” mention, or not a mention)



Positive examples: model has to assign a high probability to either one (but not necessarily both)

Coreference Models: Mention Ranking

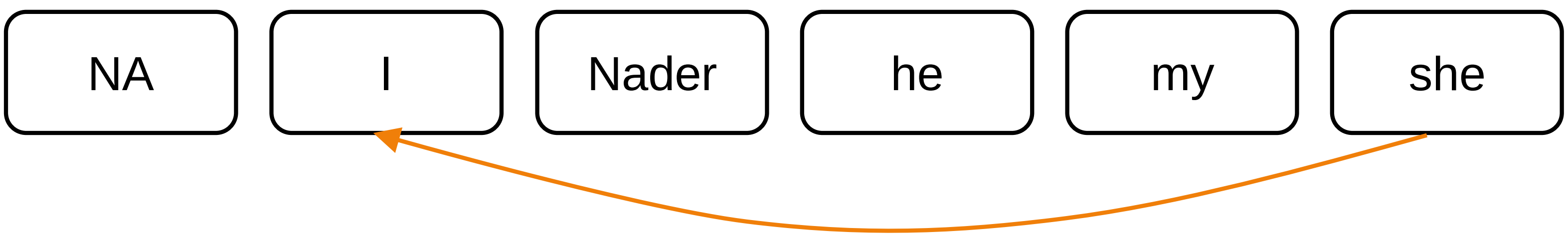
- Assign each mention its highest scoring candidate antecedent according to the model
  - Dummy NA mention allows model to decline linking the current mention to anything ("singleton" or "first" mention, or not a mention)



$$\left.\begin{aligned} &q(NA, she) = 0.1 \\ &q(I, she) = 0.5 \\ &q(Nader, she) = 0.1 \\ &q(he, she) = 0.1 \\ &q(my, she) = 0.2 \end{aligned}\right\} \quad \text{probabilities sum to 1}$$

Coreference Models: Mention Ranking

- Assign each mention its highest scoring candidate antecedent according to the model
  - Dummy NA mention allows model to decline linking the current mention to anything (“singleton” or “first” mention, or not a mention)



$$\left.\begin{array}{l} q(NA, she) = 0.1 \\ q(I, she) = 0.5 \\ q(Nader, she) = 0.1 \\ q(he, she) = 0.1 \\ q(my, she) = 0.2 \end{array}\right\} \text{probabilities sum to 1}$$

Coreference Models: Training

- We want the current mention m_i to be linked to *any one* of the candidate antecedents it's coreferent with.
- Mathematically, we want to maximize this probability:

$$\sum_{j=1}^{i-1} \mathbf{1}(y_{ij} = 1) \cdot q(m_j, m_i)$$

Iterate through candidate antecedents (previously occurring mentions)

For ones that are coreferent to m_i…

…we want the model to assign a high probability

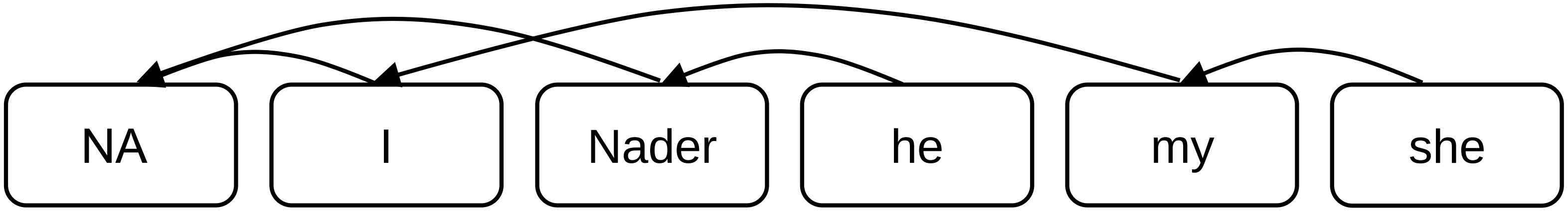
Computing the probabilities

- …from features
  - Person/Number/Gender agreement
    - Jack gave Mary a gift. She was excited.
  - Semantic compatibility
    - … the mining conglomerate … the company …
  - Certain syntactic constraints
    - John bought him a new car. [him can not be John]
  - More recently mentioned entities preferred for referenced
    - John went to a movie. Jack went as well. He was not busy.
  - Grammatical Role: Prefer entities in the subject position
    - John went to a movie with Jack. He was not busy.
  - Parallelism
    - John went with Jack to a movie. Joe went with him to a bar.
- Or just use neural networks

Inference

- Rank the scores of antecedents for each antecedent. The highest wins.

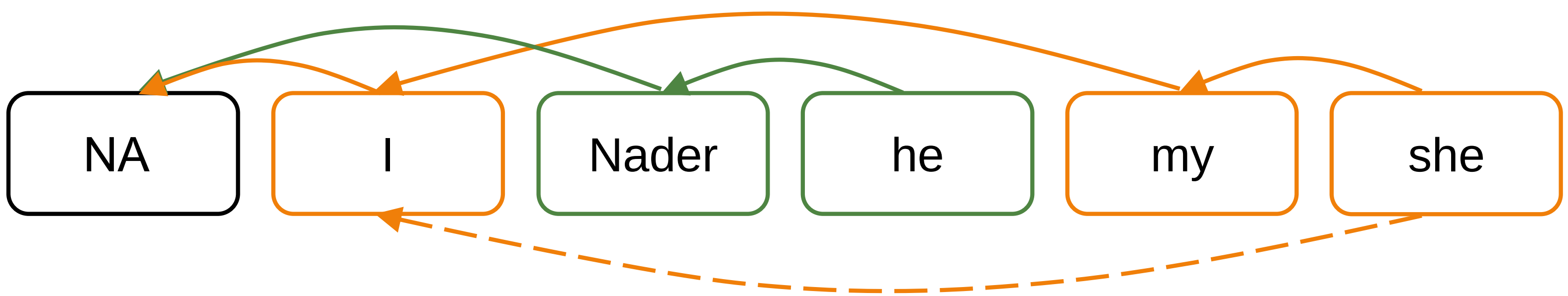
*"**I** voted for **Nader** because **he** was most aligned with **my** values," **she** said.*



Inference

- Rank the scores of antecedents for each antecedent. The highest wins.
- Take the transitive closure to get the clustering
  - Excluding NA

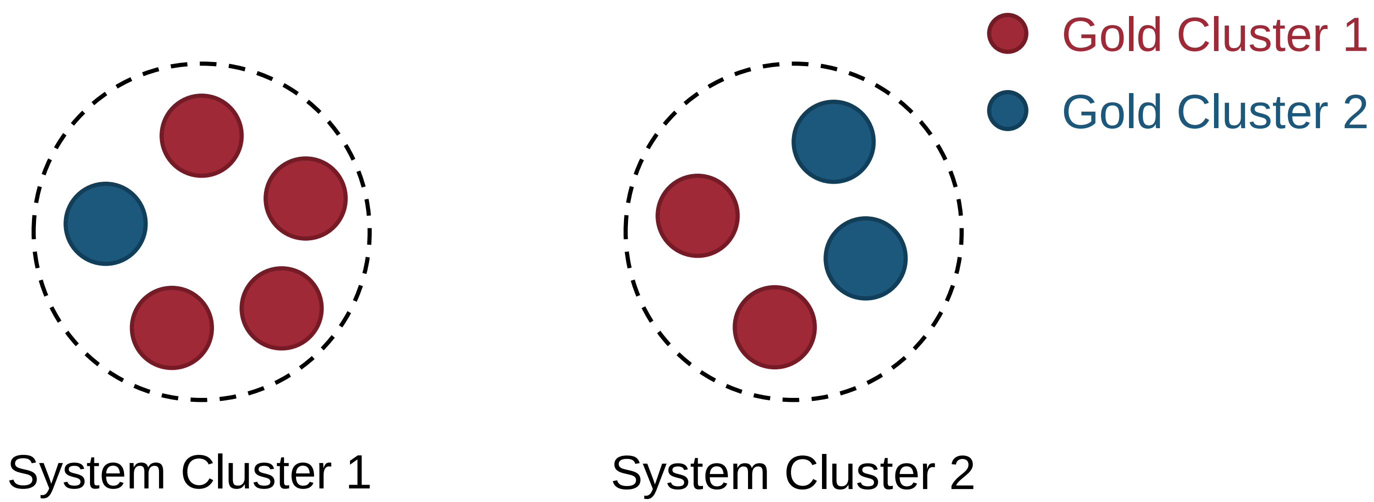
*“**I** voted for **Nader** because **he** was most aligned with **my** values,” **she** said.*



Even though the model did not predict this coreference link,
I and *she* are coreferent due to transitivity.

Evaluation

- Essentially regard coreference as a clustering task and evaluate the quality of the clustering
- Many different metrics
  - Refer to the text clustering chapter
- People often report the average over a few different metrics



The difficulty of coreference resolution

- She poured water from the pitcher into the cup until it was full.
- She poured water from the pitcher into the cup until it was empty.
- The city council refused the women a permit because they feared violence.
- The city council refused the women a permit because they advocated violence.
- These are called **Winograd Schema**
  - Requiring world knowledge and/or sophisticated reasoning to solve. This is beyond NLP!

Summary

Discourse Analysis

- A discourse is a coherent structured group of sentences.
  - Text spans are connected with coherence relations.
  - These relations form a hierarchical structure.
  - Discourse parsing: EDU segmentation + RST parsing

- Coreference Resolution
  - Mention Detection
  - Mention Clustering
    - Binary classification vs. ranking