What is Coref?

Identify the noun phrases (or entity mentions) that refer to the same real-world entity.

**Example**

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. A renowned speech therapist was summoned to help the King overcome his speech impediment . . .

- Inherently a transitive clustering task
- Typical reframing: selecting antecedent for each mention $m_j$
What is Coref?

Identify the noun phrases (or entity mentions) that refer to the same real-world entity

Example

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. A renowned speech therapist was summoned to help the King overcome his speech impediment . . .

- Inherently a transitive clustering task
- Typical reframing: selecting antecedent for each mention $m_j$
What is Coref?

Identify the noun phrases (or entity mentions) that refer to the same real-world entity

Example

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. A renowned speech therapist was summoned to help the King overcome his speech impediment . . .

- Inherently a transitive clustering task
- Typical reframing: selecting antecedent for each mention $m_j$
What is Coref?

Identify the noun phrases (or entity mentions) that refer to the same real-world entity

Example

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. A renowned speech therapist was summoned to help the King overcome his speech impediment . . .

- Inherently a transitive clustering task
- Typical reframing: selecting antecedent for each mention $m_j$
Why it’s hard

- Many sources of information play a role
  - lexical / word: head noun matches President Clinton = Clinton =? Hillary Clinton
  - grammatical: number/gender agreement, …
  - syntactic: syntactic parallelism, binding constraints: John helped himself to... vs. John helped him to…
  - discourse: discourse focus, salience, recency, …
  - semantic: semantic class agreement, …
  - world knowledge

- Not all knowledge sources can be computed easily
Where was Mozart born?

Mozart was one of the first classical composers. He was born in Salzburg, Austria, in 27 January 1756. He wrote music of many different genres...

Haydn was a contemporary and friend of Mozart. He was born in Rohrau, Austria, in 31 March 1732. He wrote 104 symphonies...
Application: Question Answering

Where was Mozart born?

Mozart was one of the first classical composers. He was born in Salzburg, Austria, in 27 January 1756. He wrote music of many different genres...

Haydn was a contemporary and friend of Mozart. He was born in Rohrau, Austria, in 31 March 1732. He wrote 104 symphonies...
Hobb’s Algorithm

Intuition:

- Start with target pronoun
- Climb parse tree to S root
- For each NP or S
  - Do breadth-first, left-to-right search of children
  - Restricted to left of target
  - For each NP, check agreement with target
- Repeat on earlier sentences until matching NP found
Hobb’s Algorithm Example

S1
  NP1
    Det
    N
      mom
  NP2
    Lyn
  's

VP
  V
    is
  NP
    Det
    a
    N
      gardener

S2
  NP4
    Craig
  VP
    V
    likes
  NP5
    her
Machine Learning Approach

- Preprocessing
- Mention Detection
- Coreference
Machine Learning Approach

- Preprocessing
- Mention Detection
- Coreference

Not-so-trivial: extract the mentions (pronouns, names, nominals, nested NPs): Some researchers reported results on gold mentions, not system mentions
Machine Learning: Pairwise

Features:
- Exact string: are $m_i$ and $m_j$ same after determiners removed
- Grammatical: gender and number agreement
- Semantic: class agreement (country/company)
- Positional: distance between the two mentions
Problems

- Conflicts

- Constraints
More Advanced Coreference

- Anaphoric classifier
- Rank mentions
- Cluster assignment
- Pipeline approach
More Advanced Coreference

- Anaphoric classifier
- Rank mentions
- Cluster assignment
- Pipeline approach (Hand-crafted?)
More Advanced Coreference

- Anaphoric classifier
- Rank mentions
- Cluster assignment
- Pipeline approach

Harder to evaluate!
Possible Projects

- Improve QA (find mentions of candidate answers in Wikipedia)
- Use world knowledge to improve coref
- Better features / representations