Machine Learning for Entity Coreference Resolution: A Retrospective Look at Two Decades of Research

Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas

AAAI 2017

Entity Coreference

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

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

Entity Coreference

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

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

Entity Coreference

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

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

Entity Coreference

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

> Queen Elizabeth set about transforming her husband, King George VI, into <u>a viable monarch</u>. <u>A renowned speech therapist</u> was summoned to help the King...

Entity Coreference

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

> Queen Elizabeth set about transforming her husband, King George VI, into <u>a viable monarch</u>. <u>A renowned speech therapist</u> was summoned to help the King...

- Two coreferent mentions form a coreference relation
  - (Queen Elizabeth, her)

antecedent anaphor

Entity Coreference

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

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

- Some coreference relations are easier to identify than others
  - (King George VI, the King)

antecedent anaphor

Entity Coreference

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

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

- Some coreference relations are easier to identify than others
  - (husband, the King)

antecedent anaphor

Entity Coreference

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

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

- Inherently a clustering task
  - the coreference relation is transitive
    - Coref(A,B) $\wedge$ Coref(B,C) → Coref(A,C)

How hard is coreference? (Winograd, 1972)

The city council refused to give the women a permit because they feared violence.

The city council refused to give the women a permit because they advocated violence.

How hard is coreference? (Winograd, 1972)

The city council refused to give the women a permit because they feared violence.

The city council refused to give the women a permit because they advocated violence.

- This pronoun resolution task is known as the **Winograd Schema Challenge**
  - Lots of interest in the commonsense reasoning community
  - Easy for humans but challenging for machines
  - An appealing alternative to the Turing Test (Levesque, 2011)

Entity Coreference Resolution

- One of the most difficult tasks in NLP
  - reliance on sophisticated knowledge and inference mechanisms
  - Best English coreference resolver: ~0.65 F-measure

- Core task in information extraction from text
  - Consolidate textual information about an entity
  - Crucial for high-level NLP applications
    - E.g., question answering, machine translation, summarization

- State-of-the-art models employ supervised machine learning

Plan for the talk

- Models
- Features
- Challenges

Plan for the talk

- Models
- Features
- Challenges

The Mention-Pair Model

- a classifier that, given a description of two mentions, determines whether they are coreferent or not
  - coreference as a pairwise classification task

The Mention-Pair Model

- a classifier that, given a description of two mentions, determines whether they are coreferent or not
  - coreference as a pairwise classification task
  - But making pairwise classifications doesn't guarantee transitivity!

positive *negative*

[Jing] likes [him] but [she] …

positive

The Mention-Pair Model

- a classifier that, given a description of two mentions, determines whether they are coreferent or not
  - coreference as a pairwise classification task
  - But making pairwise classifications doesn't guarantee transitivity!

positive *negative*

[Jing] likes [him] but [she] …

positive

  - **Solution**: postprocess conflicting decisions using clustering
    - Closest-first: resolve anaphor to closest antecedent
    - Best-first: resolve anaphor to most probable antecedent

Weaknesses of the Mention-Pair Model

- **Can't determine which candidate antecedent is the best**

John is angry about Jim because he…

  - only determine how good a candidate antecedent is relative to "he", not how good it is relative to the other candidates

Weaknesses of the Mention-Pair Model

- **Can’t determine which candidate antecedent is the best**

John is angry about him because he…

  - only determine how good a candidate antecedent is relative to “he”, not how good it is relative to the other candidates

- **Solution**: formulate coreference as ranking, not classification
  - train a ranker that ranks candidate antecedents so that it assigns the highest rank to the correct antecedent

    → **mention-ranking model**

Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision

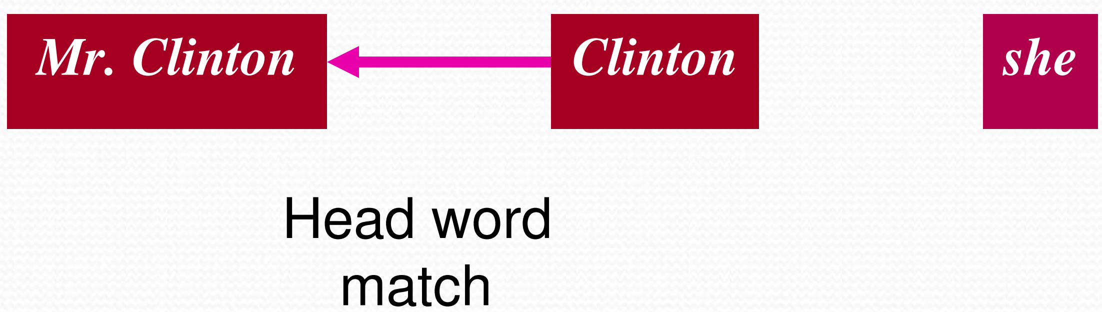
Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision

Mr. Clinton ***Clinton*** ***she***

Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



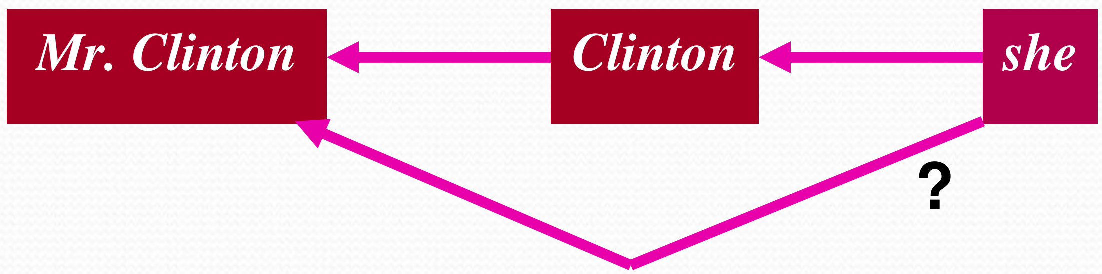
Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision

Proximity and lack of grammatical incompatibility

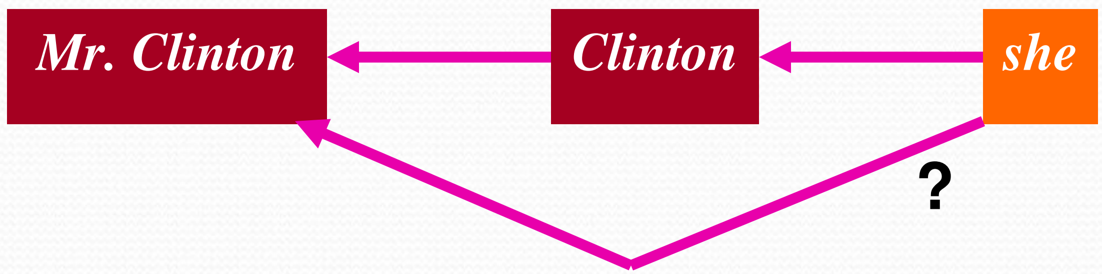
Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



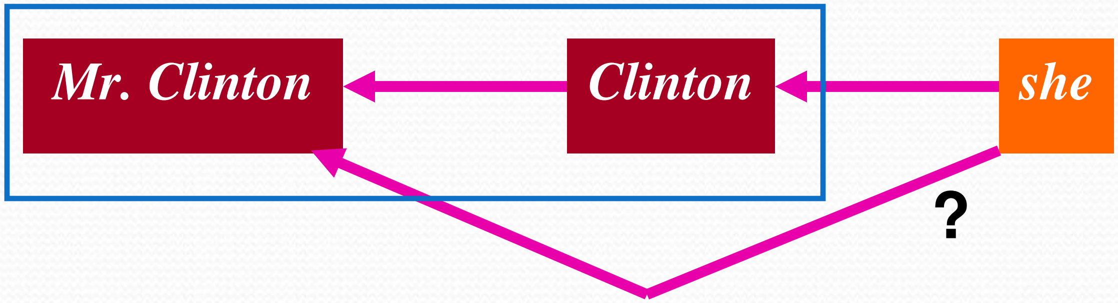
Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



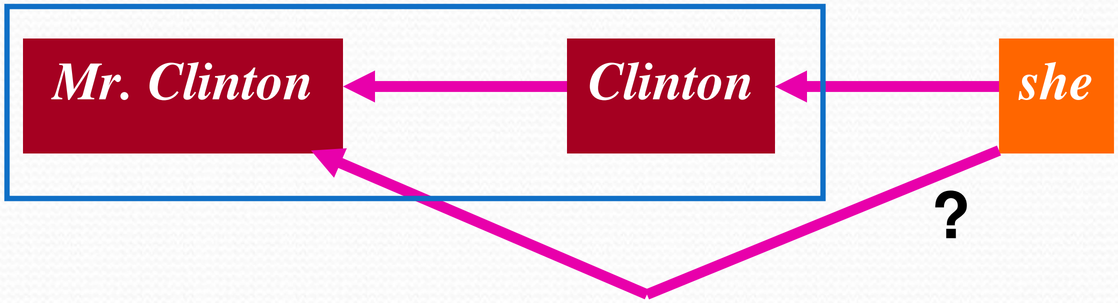
Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



Weaknesses of the Mention-Pair Model

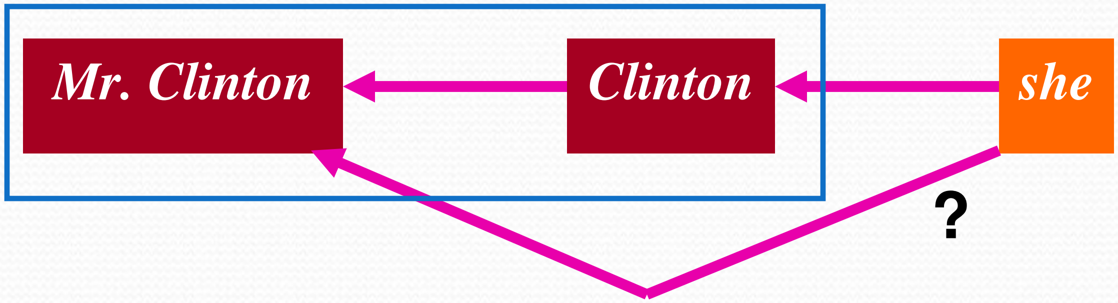
- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



  - **Idea**: train a classifier to determine whether she is coreferent with a preceding cluster

Weaknesses of the Mention-Pair Model

- **Limited expressiveness**
  - information extracted from two mentions may not be sufficient for making an informed coreference decision



  - **Idea**: train a classifier to determine whether she is coreferent with a preceding cluster

→ **Entity-mention model**

- More expressive than mention-pair model
  - can employ features over any subset of mentions in the cluster

Entity-Mention Model

- **Idea**:
  - Construct coreference clusters incrementally when processing the mentions in the text in a left-to-right manner
  - Later coreference decisions can exploit the partial clusters formed thus far

Entity-Mention Model

- **Idea**:
  - Construct coreference clusters incrementally when we process the mentions from left to right in the text
  - Later coreference decisions can exploit the partial clusters formed thus far
- **Strength**: improved expressiveness
- **Weakness**: error propagation
  - Partial clusters formed thus far can be wrong
    - We may be building on the wrong solution

Easy-First Approaches

- Rather than resolving the mentions in a left-to-right manner,
  - resolve the easy mentions first: more likely to be correct
  - Exploit partial clusters to make later coreference decisions
    - Exploit easy relations to discover hard relations

Easy-First Approaches

- Rather than resolving the mentions in a left-to-right manner,
  - resolve the easy ones first: more likely to be correct
  - Exploit partial clusters to make later coreference decisions
    - Exploit easy relations to discover hard relations
- Stanford resolver (winner of the CoNLL 2011 shared task)
  - **Rule-based** resolver organized as a pipeline of **sieves**

Sieve 1 → Sieve 2 → Sieve 3 → Sieve 4 → Sieve 5

Easy-First Approaches

- Rather than resolving the mentions in a left-to-right manner,
  - resolve the easy ones first: more likely to be correct
  - Exploit partial clusters to make later coreference decisions
    - Exploit easy relations to discover hard relation
- Stanford resolver (winner of the CoNLL 2011 shared task)
  - **Rule-based** resolver organized as a pipeline of **sieves**

Sieve 1 → Sieve 2 → Sieve 3 → Sieve 4 → Sieve 5

  - Each sieve uses rules to resolve a subset of the mentions
    - First sieve: resolves the easiest cases (e.g., string match, …)
    - Last sieve: resolves the hardest cases (pronouns)

Graph-Partitioning Approaches

- Not surprising
  - coreference is inherently a clustering task

- **Nodes**: mentions
- **Edges**: how likely the two mentions involved are coreferent

- Apply graph-partitioning algorithm to obtain coref clusters
  - Minimum cut, spectral clustering, correlation clustering, …

Recent Trend in Coreference Research

- Learn structured models for coreference resolution
  - **Input**: document
  - **Output**: a structure from which we can derive a partition

Partition-Based Models

- McCallum & Wellner (2004):
  - Since the goal is to output a coreference partition, why not learn to predict a partition directly?

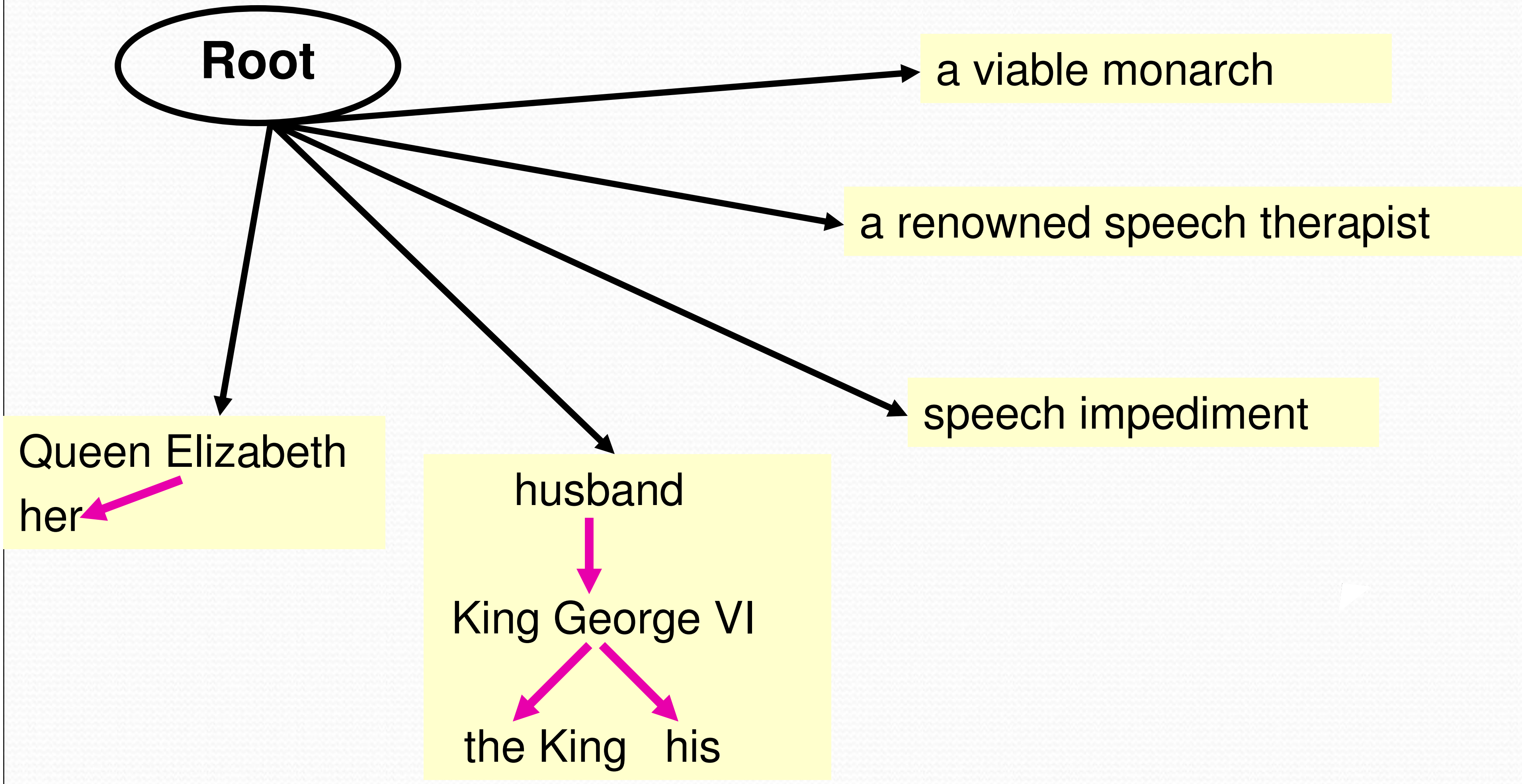
Partition-Based Models

- McCallum & Wellner (2004):
  - Since the goal is to output a coreference partition, why not learn to predict a partition directly?
- Learn a conditional random field to induce a distribution over coreference partitions
  - each training example corresponds to a document
  - two types of features
    - Features defined over each mention
    - Features defined over each pair of mentions
  - learn feature weights using a structured perceptron learner
  - decode using correlation clustering

Tree-Based Models

- **Motivation**: learning a partition is hard: need to learn from all coreferent pairs, including both the easy and hard ones
- **Observation** (Fernandes et al., 2012):
  - We **don't** need all coreferent pairs to construct a partition
  - To construct a partition, we need to construct each cluster
  - To construct a cluster with *n* mentions, we need only *n*-1 links
  - What we can learn instead is a maximum spanning tree

Coreference Tree



Tree-Based Models

- Fernandes et al. (2012) claim that it is easier to learn a coreference tree than a coreference partition
  - may be able to avoid learning from the hard relations
  - winner of the CoNLL-2012 shared task

Neural Models (Wiseman et al., 2015, 2016)

- **Observation:** models developed by far are linear models

- Improve by learning non-linear models using neural nets
  - combine features in a non-linear fashion
  - learn useful task-specific **representations**

- Wiseman et al. (2015):
  - learn a **mention-ranking** model using a neural net

- Wiseman et al. (2016):
  - extend the neural model to exploit cluster information
  - achieve the best English result to date on OntoNotes
  - most promising approach to entity coreference

Plan for the talk

- Models
- Features
- Challenges

Features

- Early coreference models relied primarily on **lexical** and **syntactic** features

- Recently, the use of **semantics** and **world knowledge** for coreference resolution was made possible by
  - the development of large lexical knowledge bases
  - advances in corpus-based lexical semantics research

Semantic Class Agreement

- whether two mentions have the same semantic class
  - Cannot be coreferent if they don't
    - Barack Obama (PERSON) vs. country (LOCATION)

Soon et al. (2001)

Semantic Relations

- Two common nouns are more likely to be coreferent if they have certain semantic relations (e.g., synonymy, hypernymy)

Bengtson & Roth (2008)

Semantic Similarity

- Two words/phrases are more likely to be coreferent if they are semantically similar
  - e.g., if their WordNet distance is small

Ponzetto & Strube (2006)

Selectional Preferences

Companies set aside tax money because the government is going to collect it

Dagan & Itai (1990), Kehler et al. (2004), Yang et al. (2005)

Selectional Preferences

Companies set aside tax money because the government is going to collect it

- it cannot refer to "government" or "companies" because one cannot collect "government" or "companies"
- A verb has preferences/restrictions for certain arguments
  - Can exploit such preferences for selecting antecedents

Dagan & Itai (1990), Kehler et al. (2004), Yang et al. (2005)

World Knowledge

- Knowing that Donald Trump is U.S. president can help establish the coreference relation between two mentions "Donald Trump" and "president" in a document

- Knowledge attributes of a proper name can be extracted from knowledge bases such as Wikipedia and Freebase

Rahman & Ng (2011), Ratinov & Roth (2012), Hajishirzi et al. (2013)

Semantic Features

- Hard to draw general conclusions about the usefulness of different kinds of semantic features given that different researchers evaluated them under different conditions
- Performance gains beyond the current state of the art will likely come from the incorporation of sophisticated features

Plan for the talk

- Models
- Features
- Challenges

Challenges: New Models

- Can we jointly learn coreference resolution with other tasks?
  - Can exploit cross-task constraints to improve model learning
    - Jointly learn coreference with named entity recognition and entity linking with promising results (Durrett & Klein, 2014)

Challenges: New Features

- There is a limit on how far one can improve coreference resolution using machine learning methods
  - A good model can profitably exploit the available features, but if the knowledge needed is not present in the data, there isn't much that the model can do

Challenges: New Languages

- **Low-resource languages**
  - Lexical knowledge bases may not be available
    - How can we obtain world knowledge?
  - Coreference-annotated corpora may not be available
    - How can we learn a coreference model?