

Coreference

Jordan Boyd-Graber

University of Maryland

Adapted from slides by Vincent Ng, Adrian Brasoveanu, and Kenton Lee

Outline

- What is coref?
- Rule-based vs. neural coref
- What makes coref hard
- How developments in coref shaped NLP research
- Why and how you should use coref in your own work

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 ...

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 ...

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 ...

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 ...

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: the dragon found the knight, but he flew away
- Some of these are easier than others to teach a computer

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: the dragon found the knight, but he flew away
- Some of these are easier than others to teach a computer

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: the dragon found the knight, but he flew away
- Some of these are easier than others to teach a computer

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: the dragon found the knight, but he flew away
- Some of these are easier than others to teach a computer

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...

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...

Understanding Evolution of Character Arcs

1 When **Jesus** had spoken these words, he went forth with his disciples over the brook Cedron, where was a garden, into the which he entered, and his disciples.

2 And **Judas** also, which betrayed him, knew the place: for **Jesus** oftentimes resorted thither with his disciples.

3 **Judas** then, having received a band of men and officers from the chief priests and Pharisees, cometh thither with lanterns and torches and weapons.

4 **Jesus** therefore, knowing all things that should come upon him, went forth, and said unto them, Whom seek ye?

5 They answered him, **Jesus of Nazareth**. **Jesus** saith unto them, I am he. And **Judas** also, which betrayed him, stood with them.

6 As soon then as he had said unto them, I am he, they went backward, and fell to the ground.

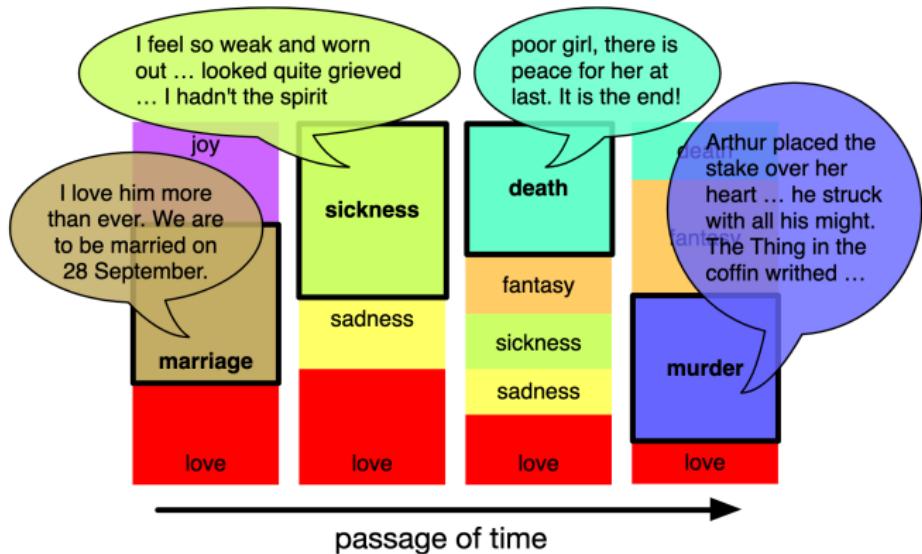
7 Then asked he them again, Whom seek ye? And they said, **Jesus of Nazareth**.

betrayed knew place disciples answered
Nazareth saith betrayed

Jesus
and
Judas

From Chang, Boyd-Graber, Blei 2009

Understanding Evolution of Character Arcs



From Iyyer, Guha, Chaturvedi, Boyd-Graber, and Daumé, 2016 (Best Paper, NAACL 2016)

Understanding Evolution of Character Arcs



From Iyyer, Guha, Chaturvedi, Boyd-Graber, and Daumé, 2016 (Best Paper, NAACL 2016)

Early Days: Hobbs' Algorithm (1978)

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

Hobbs' Algorithm Example

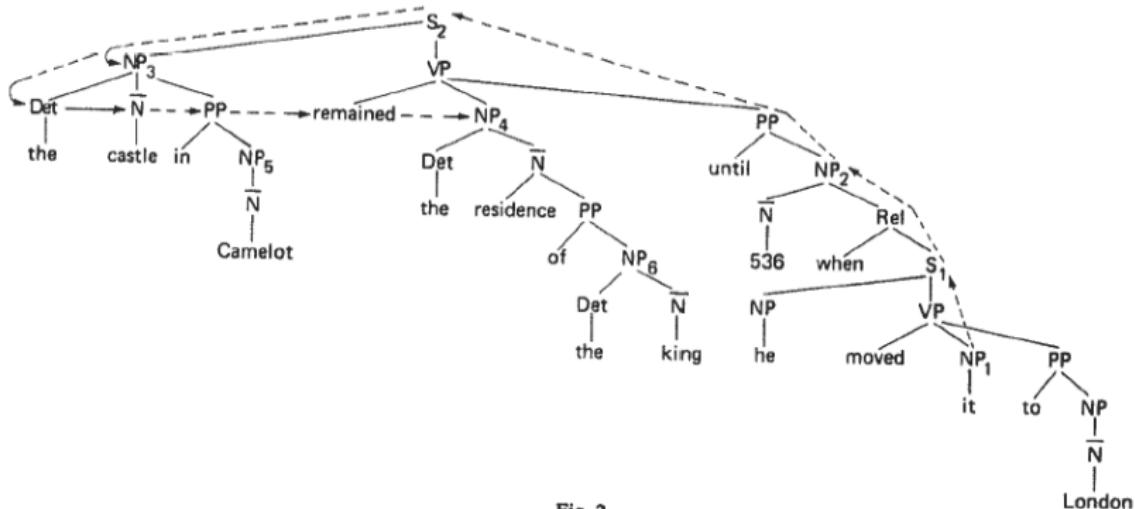
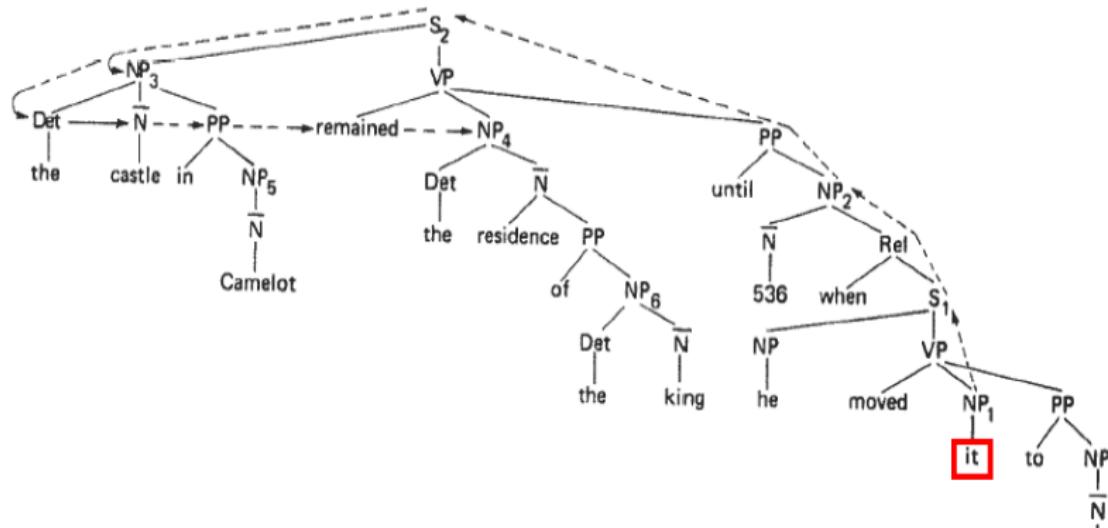


Fig. 2.

Setup

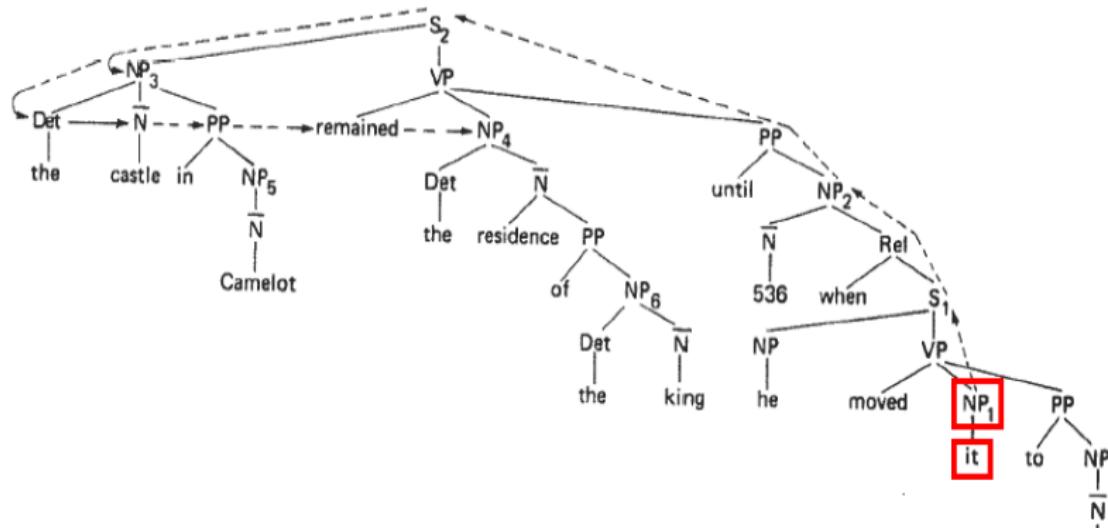
In this sentence, “it” refers to “the residence of the king”

Hobbs' Algorithm Example



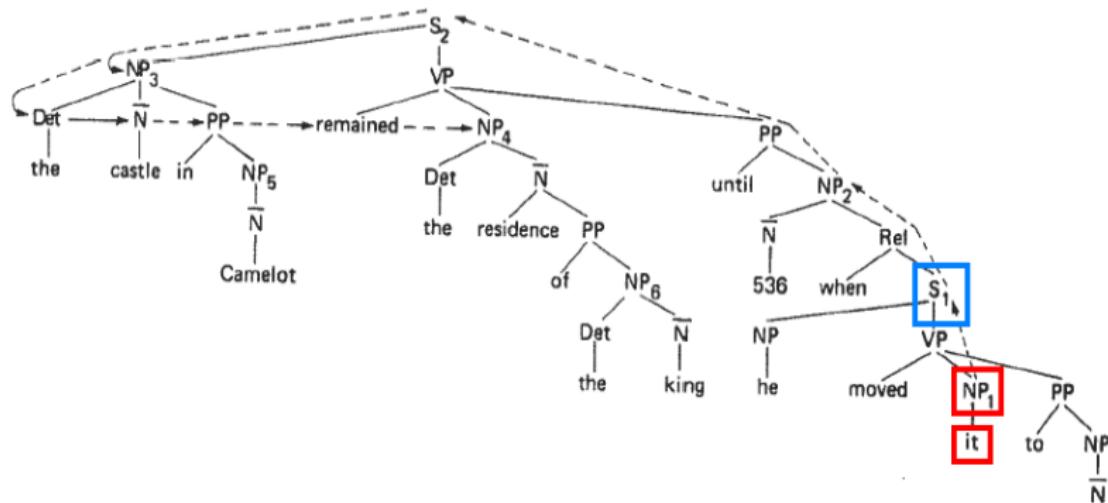
Let's try to find the referent for "it".

Hobbs' Algorithm Example



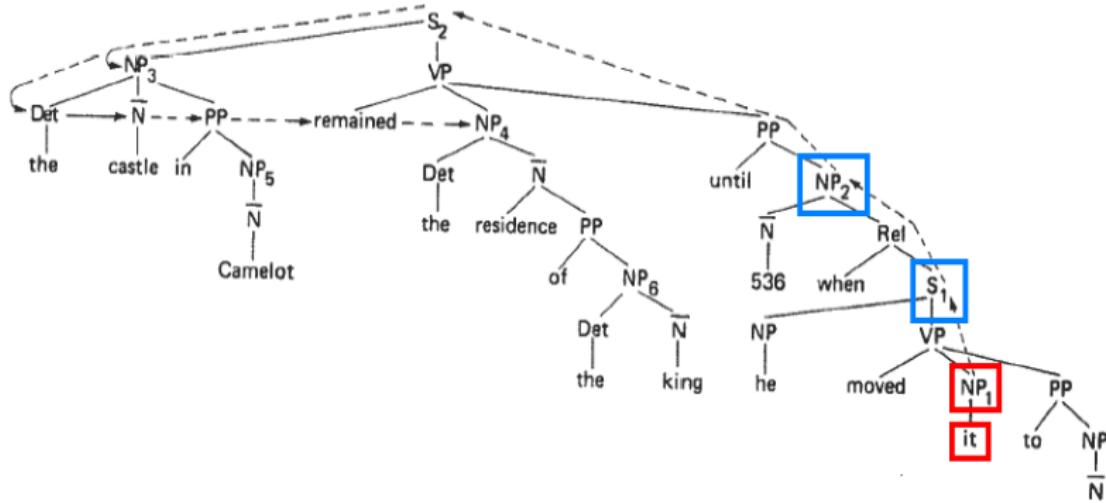
Begin at the NP immediately dominating the pronoun.

Hobbs' Algorithm Example



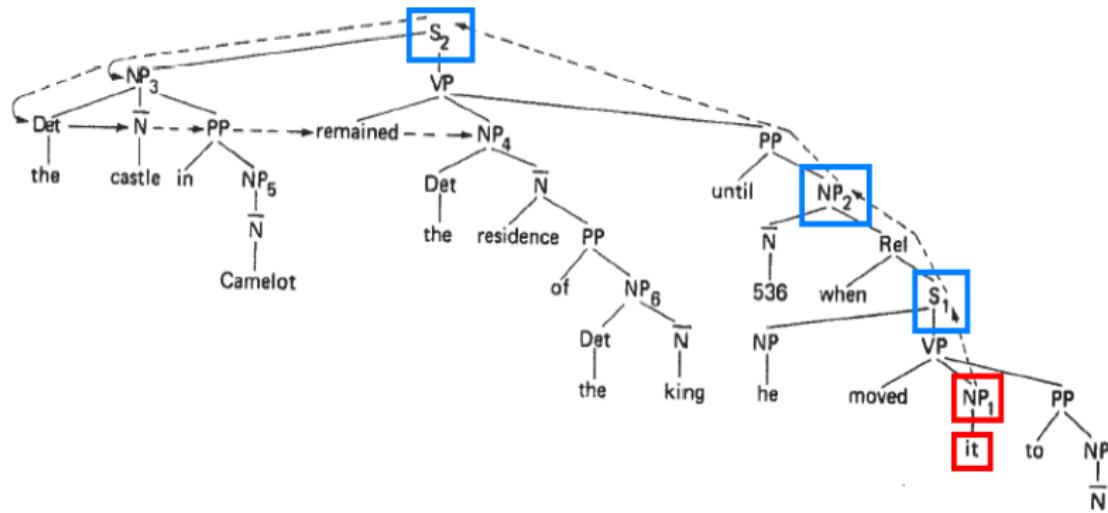
Go up tree to first NP or S encountered. Call node X, and path to it, p.
Search left-to-right below X and to left of p, proposing any NP node
which has an NP or S between it and X.
S1 has no candidates.

Hobbs' Algorithm Example



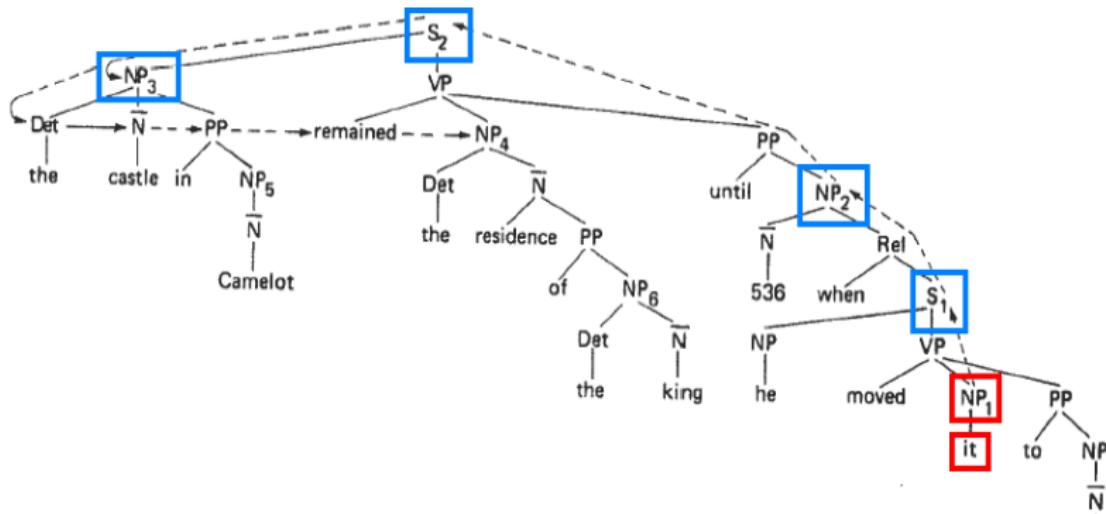
NP2 is proposed. Rejected by selectional restrictions (dates can't move).

Hobbs' Algorithm Example



From X, go up to first NP or S node encountered.

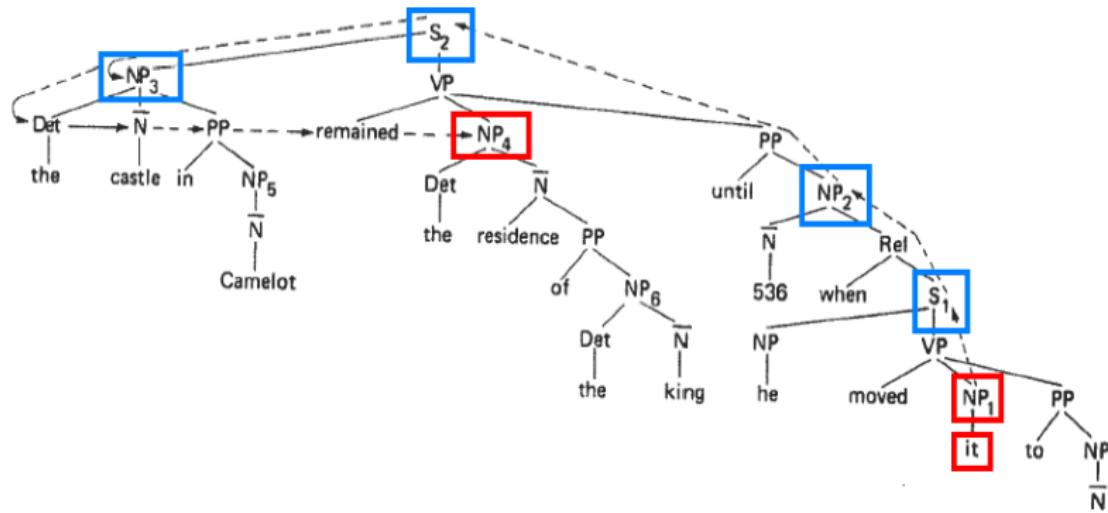
Hobbs' Algorithm Example



Search left to right.

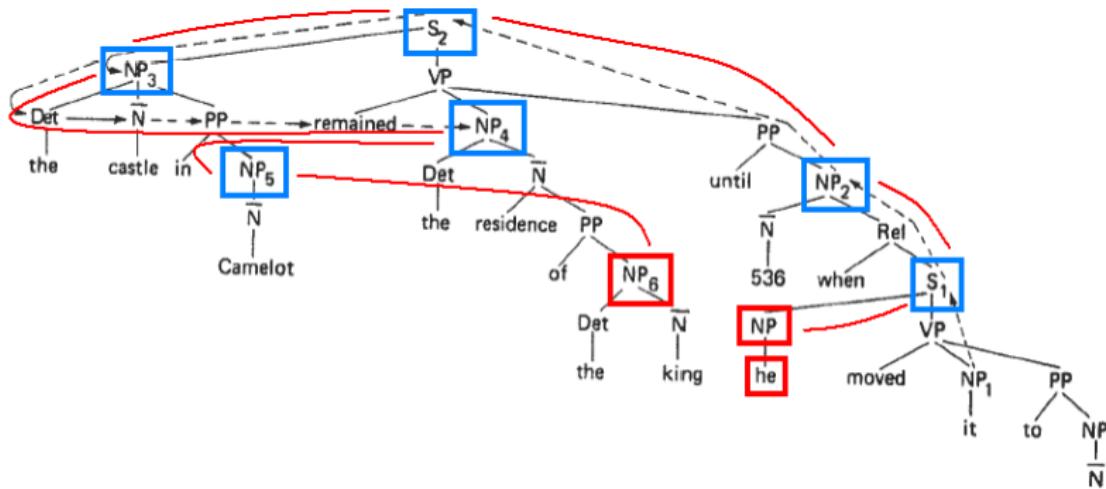
NP3: proposed. Rejected by rejected by selectional restrictions (can't move large fixed objects.)

Hobbs' Algorithm Example



NP4: proposed. Accepted.

Hobbs' Algorithm Example



The referent for “he”: we follow the same path, get to the same place, but reject NP4, then reject NP5. Finally, accept NP6.

Not like much of NLP!

- Normally when I show you a rule-based algorithm, it hasn't been used for decades
 - ▶ Lappin and Leass, 1994
 - ▶ Lee, Chang, Peirsman, Chambers, Surdeanu, Jurafsky, 2013
 - ▶ O'Connor and Heilman, 2013

Not like much of NLP!

- Normally when I show you a rule-based algorithm, it hasn't been used for decades
 - ▶ Lappin and Leass, 1994
 - ▶ Lee, Chang, Peirsman, Chambers, Surdeanu, Jurafsky, 2013
 - ▶ O'Connor and Heilman, 2013
- Rule-based algorithms still state of the art until 2016 (See Durrett and Klein, 2013)
- In some cases might still be (more on this in a second)

Machine Learning Approach

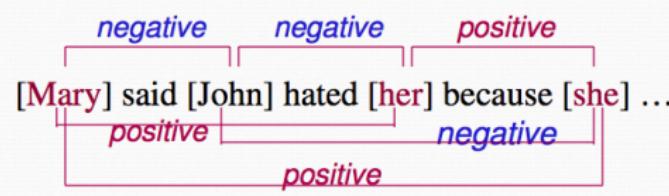
- Preprocessing
- Mention Detection
- Coreference

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

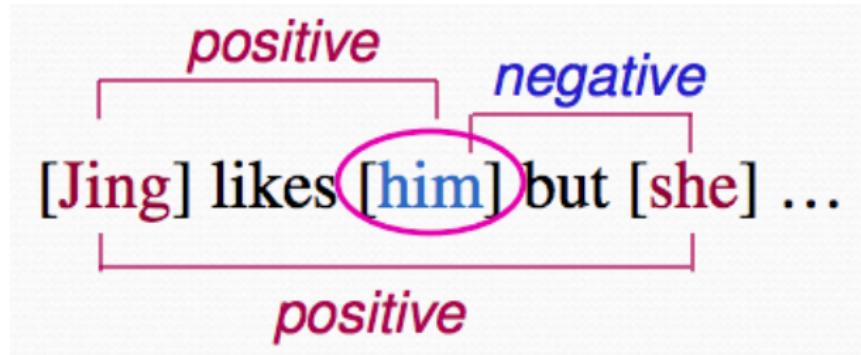


Features:

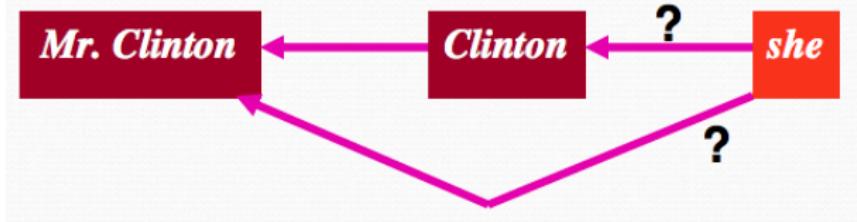
- Exact string: are m_i and m_j same after determiners removed
- Grammatical: gender and number agreement
- Semantic: class agreement (country/company)
- Positional: distance between the two mentions

Problems

- Conflicts



- Constraints



What broke the logjam?

End-to-end Neural Coreference Resolution

Kenton Lee[†], Luheng He[†], Mike Lewis[‡], and Luke Zettlemoyer^{†*}

[†] Paul G. Allen School of Computer Science & Engineering, Univ. of Washington, Seattle, WA

^{*} Allen Institute for Artificial Intelligence, Seattle WA

{kentonl, luheng, lsz}@cs.washington.edu

[‡] Facebook AI Research, Menlo Park, CA

mikelewis@fb.com

First model that didn't take a pipeline (often with hand-crafted rules) approach.

Representing Span Mentions

Span representation

the Postal Service



Word & character
embeddings



General



Electric



said



the



Postal



Service



contacted

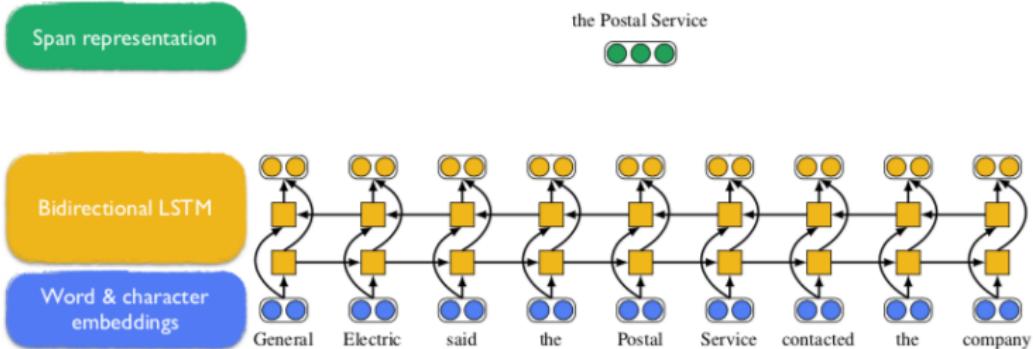


the

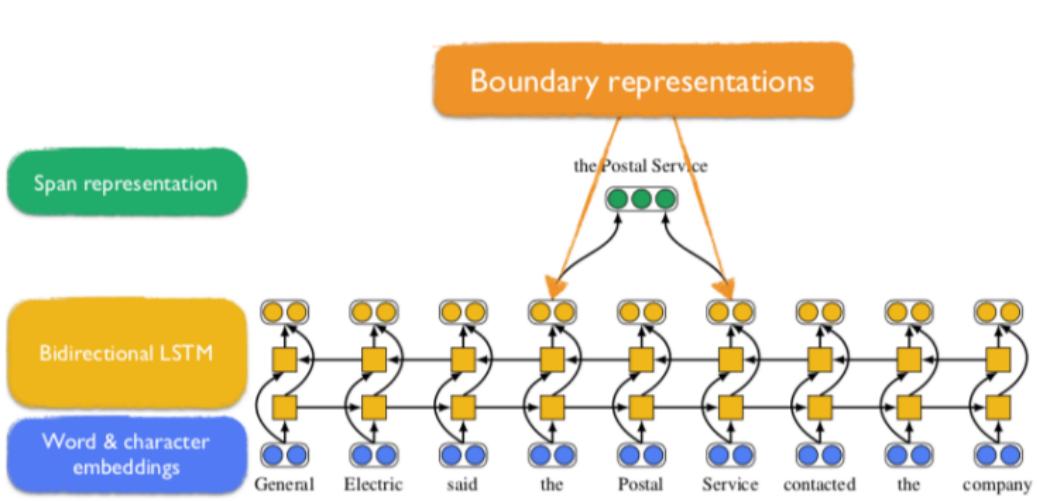


company

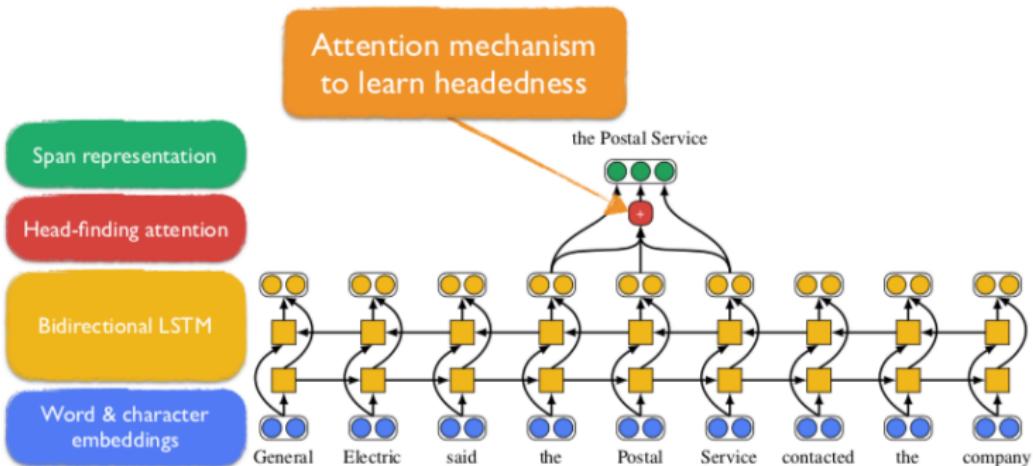
Representing Span Mentions



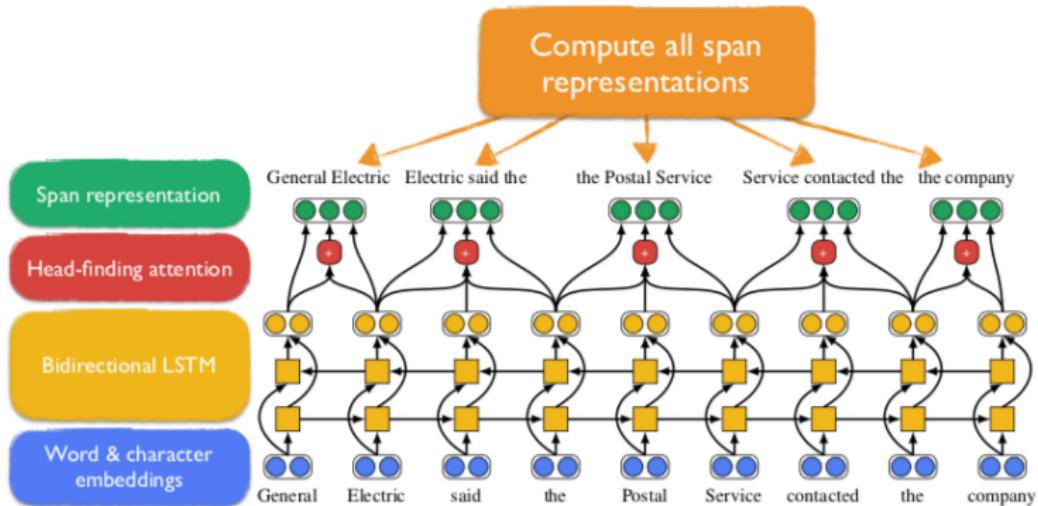
Representing Span Mentions



Representing Span Mentions



Representing Span Mentions



Given Span Representation, How to Decide which are Coreferent?

$$P(y_i \mid D)$$



Span representation



General Electric the Postal Service the company

Get all of the candidate spans (tricky to avoid $O(n^4)$) algorithm... Lee et al use beam.

Given Span Representation, How to Decide which are Coreferent?

$$P(y_i | D)$$



Span representation



General Electric the Postal Service the company

span i



Create mention representation for each candidate

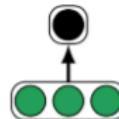
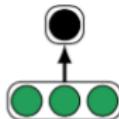
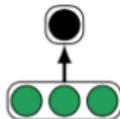
Given Span Representation, How to Decide which are Coreferent?

$$P(y_i | D)$$



$$s_m(i)$$

Span representation



General Electric the Postal Service the company

Is this actually a mention?

Given Span Representation, How to Decide which are Coreferent?

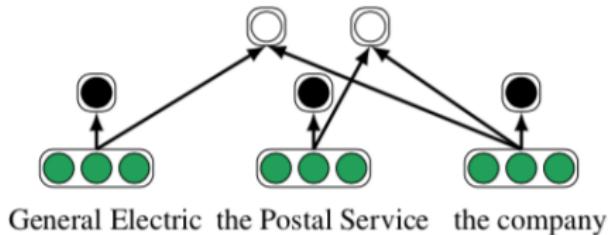
$$P(y_i | D)$$



$$s_a(i, j)$$

$$s_m(i)$$

Span representation



$s_a(i, j)$ are mention i and j coreferent

Given Span Representation, How to Decide which are Coreferent?

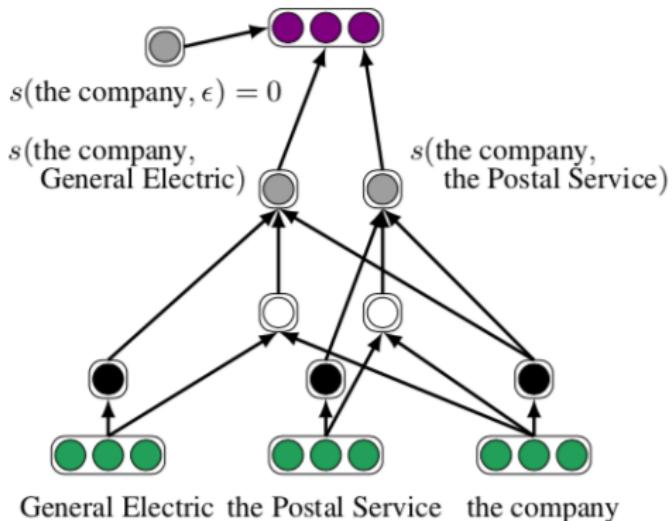
$P(y_i | D)$

$s(i, j)$

$s_a(i, j)$

$s_m(i)$

Span representation



Cluster assignment (ϵ denotes null cluster)

What Makes Coref Hard?

- Training data is from Newspapers
- Generic Mentions
- Multiple Universes
- Sub-token annotation
- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

What Makes Coref Hard?

- Training data is from Newspapers

Guha, Iyyer, Bouman, Boyd-Graber, 2015

Journalists are instructed to limit the number of entities mentioned in a sentence, and there are strict rules for referring to individuals (Boyd et al., 2008). Furthermore, writers cannot assume that their readers are familiar with all participants in the story, which requires that each entity is explicitly introduced in the text (Goldstein and Press, 2004).

- Generic Mentions
- Multiple Universes
- Sub-token annotation
- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

What Makes Coref Hard?

- Training data is from Newspapers
- Generic Mentions

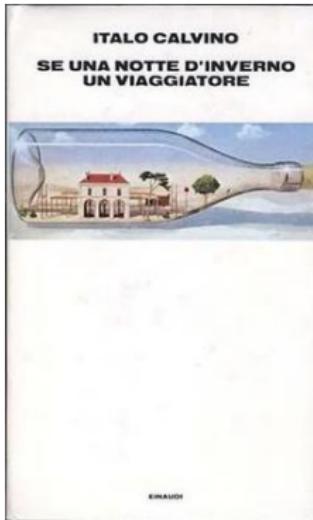
Miss Sara Sampson (Gotthold Ephraim Lessing), example from Rosiger, Schulz and Reiter, 2018

Der Wirt: Was liegt mir daran, ob ich es weiss, oder nicht, was Sie für eine Ursache hierher führt, und warum Sie bei mir im Verborgnen sein wollen? **Ein Wirt** nimmt sein Geld, und lässt seine Gäste machen, was ihnen gut dünkt.

- Multiple Universes
- Sub-token annotation
- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

What Makes Coref Hard?

- Training data is from Newspapers
- Generic Mentions
- Multiple Universes



- Sub-token annotation
- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

What Makes Coref Hard?

- Training data is from Newspapers
- Generic Mentions
- Multiple Universes
- Sub-token annotation

Miss Sara Sampson (Gotthold Ephraim Lessing), example from Rosiger, Schulz and Reiter, 2018

Eine schlechte Vorbereitung, eine **trosts**uchende Betrühte zu empfangen. Warum sucht sie **ihn** auch bei mir?

- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

What Makes Coref Hard?

- Training data is from Newspapers
- Generic Mentions
- Multiple Universes
- Sub-token annotation
- Poot and van Cranenburgh, 2020: Rule-based systems for Dutch are more robust to long documents
- Hard to fit long contexts into modern models (more on this later)

How coref has shaped NLP

- Mohit (2016) & Anupam (2015) needed better coref systems



- Mohit went to Luke's group, ELMO paper resulted
- Look familiar?
- Coref needed to go beyond word-level representations: needed a tailored representation of a word in context

How coref has shaped NLP

- Mohit (2016) & Anupam (2015) needed better coref systems
- Mohit went to Luke's group, ELMO paper resulted

Deep contextualized word representations

Matthew E. Peters[†], Mark Neumann[†], Mohit Iyyer[†], Matt Gardner[†],
`{matthewp, markn, mohiti, mattg}@allenai.org`

Christopher Clark*, Kenton Lee*, Luke Zettlemoyer^{†*}
`{csquared, kentonl, lsz}@cs.washington.edu`

[†]Allen Institute for Artificial Intelligence

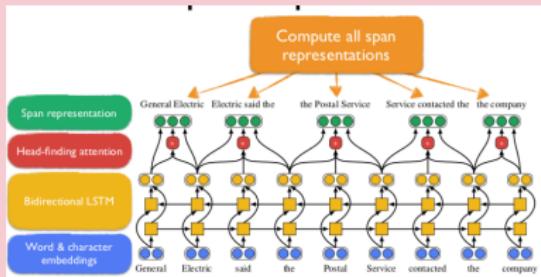
*Paul G. Allen School of Computer Science & Engineering, University of Washington

- Look familiar?
- Coref needed to go beyond word-level representations: needed a tailored representation of a word in context

How coref has shaped NLP

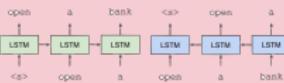
- Mohit (2016) & Anupam (2015) needed better coref systems
- Mohit went to Luke's group, ELMO paper resulted
- Look familiar?

End-to-end Neural Coreference Resolution, 2017

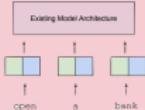


Deep Contextualized Word Representations

Train Separate Left-to-Right and Right-to-Left LMs



Apply as "Pre-trained Embeddings"



- Coref needed to go beyond word-level representations: needed a tailored representation of a word in context

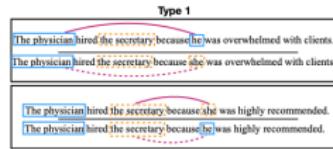
Maryland Connection

- Kermit (Jim, BS 1960),
1955
- ELMO (Mohit, PhD 2019)
- BERT (Jacob, MS 2009)



Wrapup

- Coref requires context-dependent representations
- Driving NLP research where it needs to go
- Important applications
 - ▶ Digital humanities
 - ▶ Bioinformatics
 - ▶ Question answering
- Difficult to solve
 - ▶ Curation: Getting training data
 - ▶ Algorithmic: Making sure unbiased
 - ▶ Technical: Using world knowledge and long context



Zhao, Wang, Yatskar,
Ordonez, Chang
(2018)

Link to Wikipedia

Britain would be “prosperous and dynamic and contented” after completing its exit from the EU, Johnson declared, free to strike trade deals around the world.

Link to Wikipedia

Britain would be “prosperous and dynamic and contented” after completing its exit from the EU, Johnson declared, free to strike trade deals around the world.

United_Kingdom

Link to Wikipedia

Britain would be “prosperous and dynamic and contented” after completing its exit from the EU, Johnson declared, free to strike trade deals around the world.

European_Union

Link to Wikipedia

Britain would be “prosperous and dynamic and contented” after completing its exit from the EU, [Johnson](#) declared, free to strike trade deals around the world.

[Boris_Johnson](#)

Link to Wikipedia

Britain would be “prosperous and dynamic and contented” after completing its exit from the EU, Johnson declared, free to strike **trade** deals around the world.

Trade_agreement

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing **The Hundred Secret Senses** , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of **his daughter by another woman** ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and **her newfound half-sister Kwan** . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that **her mother** 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between **Pearl** and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and **Winnie Louie** in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in **her novel The Kitchen God 's Wife** . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .

Find the Coref Groups

One of this author 's novels involves a promise made to a dying husband to take care of his daughter by another woman ; in that work , Olivia feels that her mother 's attentions are split unfairly between her and her newfound half-sister Kwan . In addition to writing The Hundred Secret Senses , she wrote about the mother-daughter relationship between Pearl and Winnie Louie in her novel The Kitchen God 's Wife . She is better known for a work that chronicles the story of four mother-daughter pairs of Chinese descent . For 10 points , name this author of The Joy Luck Club .