

En comence le roman de la rose
ou tout fait d'amour est enclose



Autre genre dit
que en songes
Ne sont n' fables
et mensonges

Mais on veult telz songes sonner
Qui ne sont mie mensonges
Ains sont apres bien appareant
En cyne bien trouver garant

Ving acteur d'un homme macrobe
Qui ne tient mie sonce a sobe
Aucore escript la vision
Qui aduint au roy cyron
Quicunques cude ne qui die
Que ce soit vne misardie
De croire que songe aduente
Et n'ouldra v' fol men tement
Car en droit moy ay le fiance
Que songe soit sur fiance

Natural Language Processing

Info 159/259

Lecture 23: Coreference resolution (Nov. 14, 2017)

David Bamman, UC Berkeley

Discourse

- Discourse covers linguistic expression **beyond the boundary of the sentence.**
 - Dialogues: the structure of turns in conversation
 - Monologues: the structure of entire passages, documents



LUKE

I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!



LUKE

I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!



LUKE

I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true!
That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

Coreference resolution

- “Trump met Putin today; **he**’s the leader of the US.

Coreference resolution

Barack Hussein Obama II ( [/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/](#); born August 4, 1961) is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he served as president of the *Harvard Law Review*. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney and taught constitutional law at the University of Chicago Law School from 1992 to 2004. He served three terms representing the 13th District in the Illinois Senate from 1997 to 2004, running unsuccessfully for the United States House of Representatives in 2000.

Coreference resolution

attend graduate school at [Harvard University](#) on a scholarship. Obama's parents divorced in March 1964.^[11] Obama Sr. returned to Kenya in 1964 where he remarried; he visited Barack in Hawaii only once, in 1971.^[12] He died in an automobile accident in 1982 when his son was 21 years old.^[13]

Did Barack Obama die in an automobile accident in 1982?

Coreference resolution

“Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based financial services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.”

Coreference

“Referent”

The entities or individuals [in the real world](#) that text is pointing to.

- VICTORIA CHEN
- MEGABUCKS
- LOTSABUCKS

“Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based financial services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.”

Coreference

“Referring expression”

The text that points to entities.

“Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based financial services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.”

Coreference

“coreference”

The set of text strings that all refer to the same ENTITY.

“Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based financial services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.”

Event coreference

I stubbed my toe on the chair and **it** really hurt.

Frege

Mode of presentation (*Sinn*)
vs. reference

- The morning star/the evening star
- Mark Twain/Samuel Clemens



Worth solving?

English constraints

- Number
 - I have a car. They are blue [*they = car]
- Gender
 - My dad is shoveling snow. He's cold. [*he = snow]
- Person
 - We're watching a movie. He likes it [*he = you and I]

English exceptions

- Number

- *A: I have a new friend.*

- *B: What's their name?*

- *We are a grandmother* (Margaret Thatcher)

- Gender

- “The Nellie, a cruising yawl, swung to **her** anchor without a flutter of the sails, and was at rest.” (Heart of Darkness)

- *It puts the lotion in the basket* (Silence of the Lambs)

- Person

- ???

English preferences

- Recency: more recent NPs are preferred
- Grammatical role: subjects are preferred
 - Billy Bones went to the bar with Jim Hawkins. **He** called for a glass of rum.
- Repeated mention: more discourse-salient NPs are preferred
- Parallelism
 - Long John Silver went with Jim to the Old Parrot. Billy Bones went with **him** to the Old Anchor inn.
- Verb semantics
- Selectional restrictions

Verb semantics

- John telephoned Bill. He lost the laptop
- John criticized Bill. He lost the laptop.

Winograd challenge

- The trophy would not fit in the brown suitcase because **it** was too big. What was too big?
- The town councilors refused to give the demonstrators a permit because **they** feared violence. Who feared violence?
- The town councilors refused to give the demonstrators a permit because **they** advocated violence. Who advocated violence?

Selectional restrictions

- John parked his car in the garage after driving *it* around for hours.

Hobbs (1978) algorithm

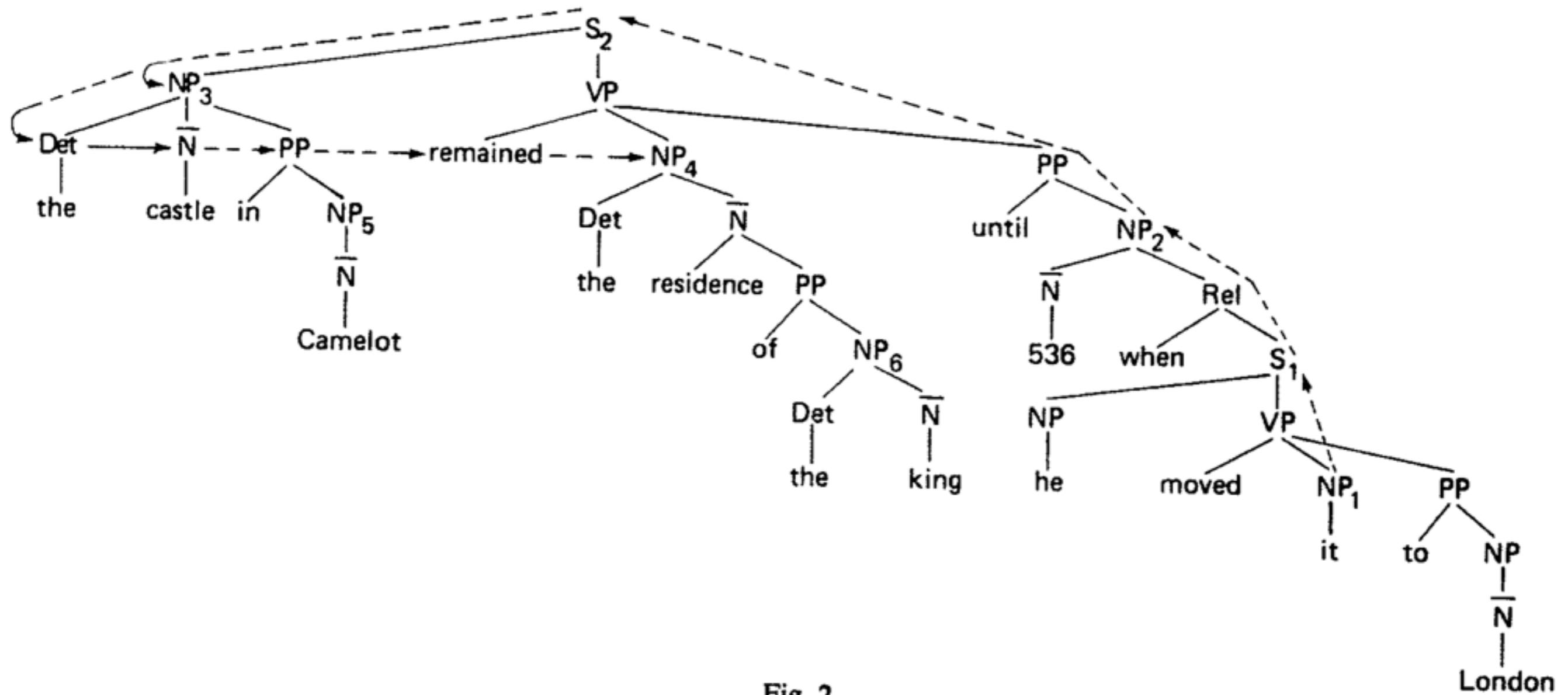
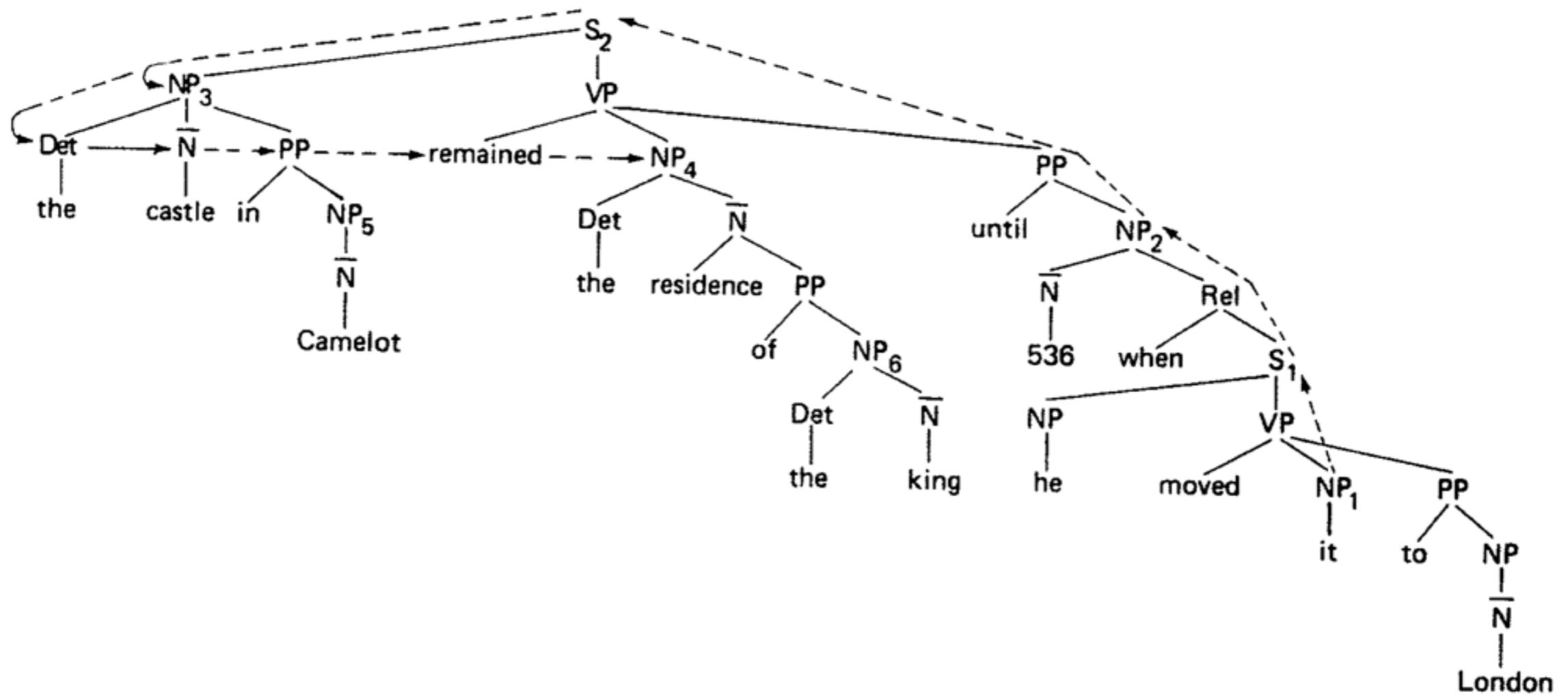
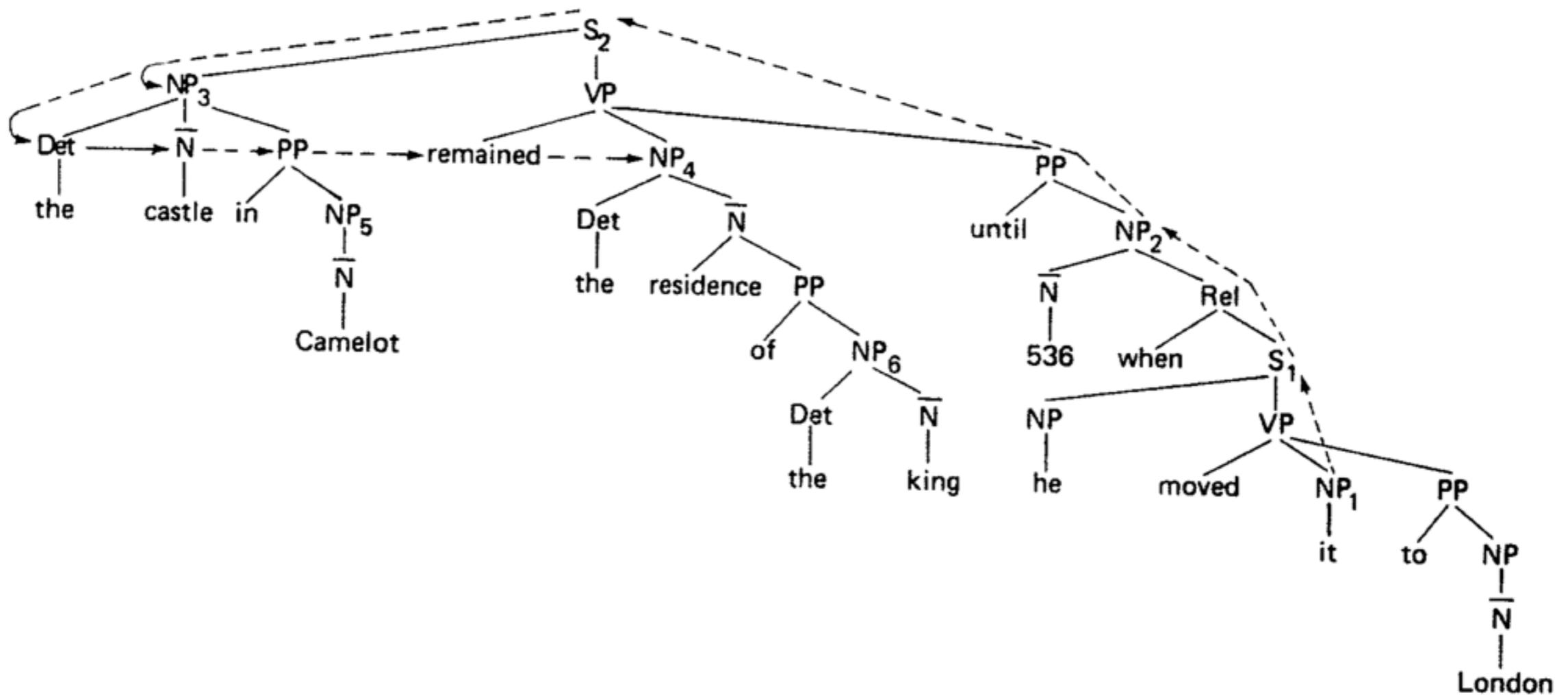


Fig. 2.



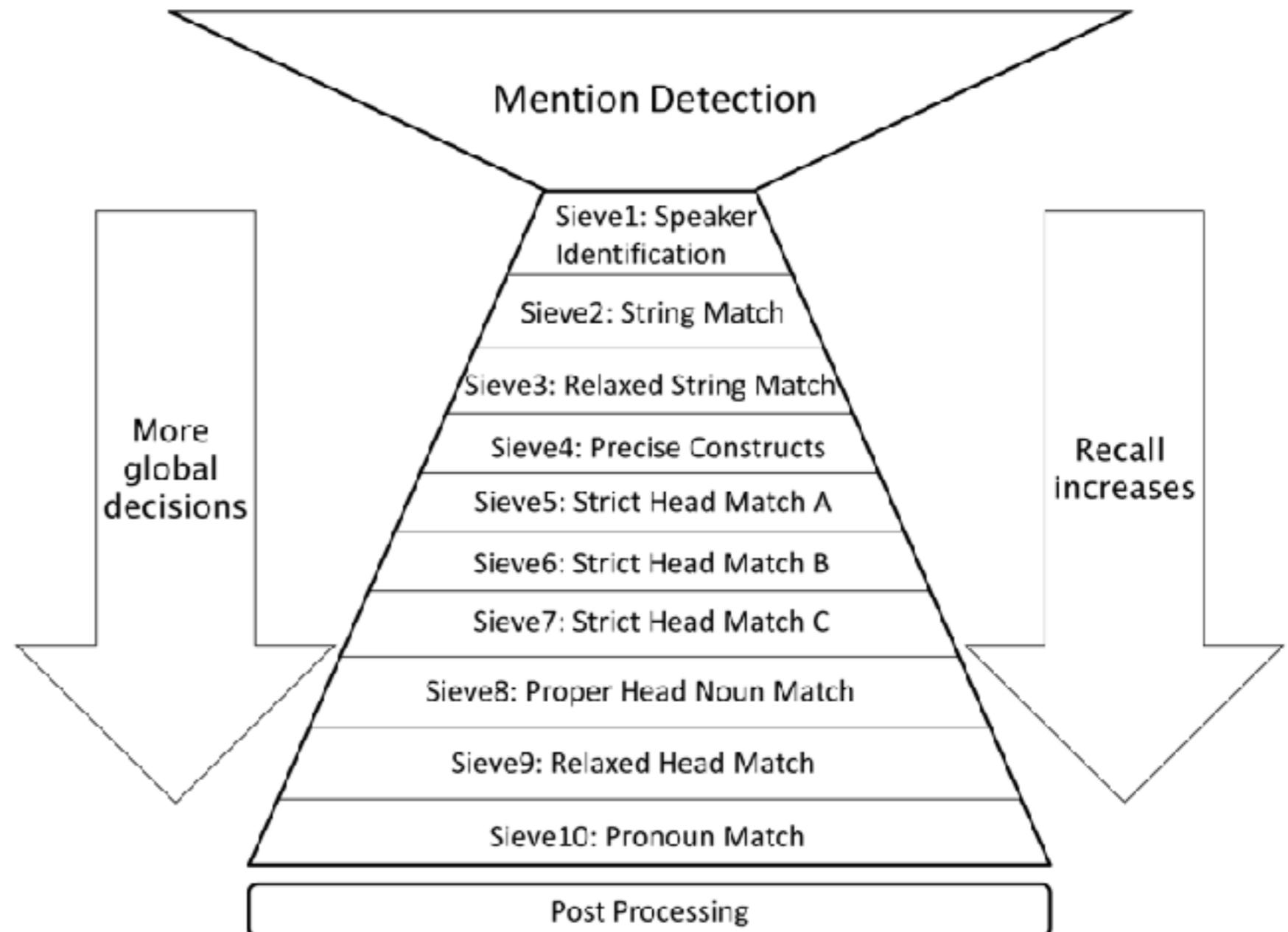
1. Begin at the noun phrase (NP) node immediately dominating the pronoun.
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node X, and call the path used to reach it p.
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any NP node that is encountered which has an NP or S node between it and X.
4. If node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent. If X is not the highest S node in the sentence, continue to step 5.



5. From node X, go up the tree to the first NP or S node encountered. Call this new node X, and call the path traversed to reach it p.
6. If X is an NP node and if the path p to X did not pass through the Nominal node that X immediately dominates, propose X as the antecedent.
7. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.
8. If X is an S node, traverse all branches of node X to the right of path p in a left-to-right, breadth-first manner, but do not go below any NP or S node encountered. Propose any NP node encountered as the antecedent.
9. Go to step 4.

Stanford “Sieve”

Sequence of pattern matching rules starting at high precision coreference links, progressing to higher recall.



Mention Detection

- All NPs, possessive pronouns, and named entity mentions are **candidate mentions**. Recall is more important than precision.
- Filters to remove candidates:
 - Remove mentions embedded within larger mentions with same headword
 - Remove numeric quantities (100 miles, 9%)
 - Remove existential there, it
 - Remove adjectival forms of nations
 - Remove 8 stop words (there, ltd., hmm)

John is a musician.
 He played a new
 song. A girl was
 listening to the song.
 “It is my favorite,”
 John said to her.

Mention Detection:	[John] ₁ ¹ is [a musician] ₂ ² . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁶ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁸ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Speaker Sieve:	[John] ₁ ¹ is [a musician] ₂ ² . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁶ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁸ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
String Match:	[John] ₁ ¹ is [a musician] ₂ ² . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁶ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁸ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Relaxed String Match:	[John] ₁ ¹ is [a musician] ₂ ² . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁶ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁸ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Precise Constructs:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁶ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁷ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Strict Head Match A:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁷ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Strict Head Match B,C:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁷ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Proper Head Noun Match:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁷ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Relaxed Head Match:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ³ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁷ is [[my] ₉ ⁹ favorite] ₈ ⁷ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ¹¹ .
Pronoun Match:	[John] ₁ ¹ is [a musician] ₂ ¹ . [He] ₃ ¹ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁴ is [[my] ₉ ⁹ favorite] ₈ ⁴ ,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ⁵ .
Post Processing:	[John] ₁ ¹ is a musician. [He] ₃ ¹ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁴ is [my] ₉ ¹ favorite,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ⁵ .
Final Output:	[John] ₁ ¹ is a musician. [He] ₃ ¹ played [a new song] ₄ ⁴ . [A girl] ₅ ⁵ was listening to [the song] ₆ ⁴ . “[It] ₇ ⁴ is [my] ₉ ¹ favorite,” [John] ₁₀ ¹⁰ said to [her] ₁₁ ⁵ .



Classification

A mapping h from input data x (drawn from instance space \mathcal{X}) to a label (or labels) y from some enumerable output space \mathcal{Y}

\mathcal{X} = set of all documents

\mathcal{Y} = {english, mandarin, greek, ...}

x = a single document

y = ancient greek



Classification

Positive examples =
pronouns paired with
closest antecedent (or
coreference chain)

Negative examples =
entities not in coreference
chain.

Classification

For every possible antecedent y for pronoun x , we frame a binary classification: is y coreferent with x ?
Every noun phrase is a candidate antecedent.

- I
- you
- you
- the power
- the power of the dark side
- the dark side
- Obi-Wan
- you
- your
- your father
- He
- me
- you

LUKE

I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed **him**.

VADER

No. I am your father.

LUKE

No. No. That's not true!
That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

Classifier

Let's brainstorm a supervised classifier.

Features

- John saw a beautiful 1961 Ford Falcon at the used car dealership
- He showed it to Bob.
- He bought it.

Features

- Unary features (valid of a single token)
 - token, lemma, part of speech
 - salience
- Binary features (valid of a pair of tokens)
 - number agreement (plural pronoun/plural NP)
 - compatible number (plural pronoun/??? NP)
 - gender agreement
 - compatible gender
 - sentence distance
 - Hobbs distance
 - syntax: grammatical role

Nominal coreference

- Pronominal coreference is a subset of the full coreference resolution problem because pronouns are nearly always **coreferent**.
- How would we extend the classification approach to general nominal referents?

Evaluation

- Evaluating general reference resolution (i.e., all noun phrase entities) is more complicated than straightforward accuracy/precision/recall

$$B_{precision}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|System_i|}$$

$$B_{recall}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|Gold_i|}$$

3 entities/coreference chains

LUKE

I ll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

7 elements
{I, you, you, your, me, your, your, You}

LUKE

I ll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

6 elements
{you, your father, you, him, I, your father}

LUKE
I'll never join **you!**

VADER
If you only knew the power of the dark side. Obi-Wan never told you what happened to **your father.**

LUKE
He told me enough! It was **you** who killed **him.**

VADER
No. **I** am **your father.**

LUKE
No. No. That's not true!
That's impossible!

VADER
Search your feelings. You know it to be true.

LUKE
No! No! No!

2 elements
{Obi-Wan, He}

LUKE

I'll never join you!

VADER

If you only knew the power of the dark side. **Obi-Wan** never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true!
That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

LUKE
I ll never join you!

VADER
If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE
He told me enough! It was you who killed him.

VADER
No. I am your father.

LUKE
No. No. That's not true!
That's impossible!

VADER
Search your feelings. You know it to be true.

LUKE
No! No! No!

Example system output: 4 entities

3 = {I, me, I}

8 = {you, you, you, your, you, your, your, you}

3 = {Obi-Wan, your father, your father}

2 = {He, him}

Evaluation

- More complicated than straightforward accuracy/precision/recall

$$B_{precision}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|System_i|}$$

$$B_{recall}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|Gold_i|}$$

n ranges over all entities in gold and system output



LUKE

I ll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!



LUKE

I ll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

$$| \text{Gold}_i \cap \text{System}_i | = 2$$

$$| \text{Gold}_i | = 8$$

$$| \text{System}_i | = 3$$



I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!



I'll never join you!

VADER

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE

He told me enough! It was you who killed him.

VADER

No. I am your father.

LUKE

No. No. That's not true! That's impossible!

VADER

Search your feelings. You know it to be true.

LUKE

No! No! No!

$$|\text{Gold}_i \cap \text{System}_i| = 2$$

$$|\text{Gold}_i| = 6$$

$$|\text{System}_i| = 8$$

LUKE
I'll never join **you!**

VADER
If **you** only knew the power of the dark side. Obi-Wan never told **you** what happened to **your** father.

LUKE
He told me enough! It was **you** who killed him.

VADER
No. I am **your** father.

LUKE
No. No. That's not true!
That's impossible!

VADER
Search **your** feelings. **You** know it to be true.

LUKE
No! No! No!

LUKE
I ll never join you!

VADER
If **you** only knew the power of the dark side. Obi-Wan never told **you** what happened to **your** father.

LUKE
He told **me** enough! It was you who killed him.

VADER
No. I am **your** father.

LUKE
No. No. That's not true!
That's impossible!

VADER
Search **your** feelings. **You** know it to be true.

LUKE
No! No! No!

$$|\text{Gold}_i \cap \text{System}_i| = 6$$

$$|\text{Gold}_i| = 8$$

$$|\text{System}_i| = 8$$

LUKE
I'll never join you!

VADER
If you only knew the power of the dark side. **Obi-Wan** never told you what happened to **your father**.

LUKE
He told me enough! It was you who killed him.

VADER
No. I am **your father**.

LUKE
No. No. That's not true! That's impossible!

VADER
Search your feelings. You know it to be true.

LUKE
No! No! No!

LUKE
I'll never join you!

VADER
If you only knew the power of the dark side. **Obi-Wan** never told you what happened to your father.

LUKE
He told me enough! It was you who killed him.

VADER
No. I am your father.

LUKE
No. No. That's not true! That's impossible!

VADER
Search your feelings. You know it to be true.

LUKE
No! No! No!

| Gold_i ∩ System_i | = 1

| Gold_i | = 2

| System_i | = 3

LUKE
I'll never join **you!**

VADER
 **you** only knew the power of the dark side. Obi-Wan never told **you** what happened to **your** father.

LUKE
He told me enough! It was **you** who killed him.

VADER
No. I am **your** father.

LUKE
No. No. That's not true!
That's impossible!

VADER
Search **your** feelings. **You** know it to be true.

LUKE
No! No! No!

LUKE
I ll never join you!

VADER
 **you** only knew the power of the dark side. Obi-Wan never told **you** what happened to **your** father.

LUKE
He told **me** enough! It was you who killed him.

VADER
No. I am **your** father.

LUKE
No. No. That's not true!
That's impossible!

VADER
Search **your** feelings. **You** know it to be true.

LUKE
No! No! No!

$$|\text{Gold}_i \cap \text{System}_i| = 6$$

$$|\text{Gold}_i| = 8$$

$$|\text{System}_i| = 8$$

Evaluation

- More complicated than straightforward accuracy/precision/recall

$$B_{precision}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|System_i|}$$

$$B_{recall}^3 = \frac{1}{n} \sum_i^n \frac{|Gold_i \cap System_i|}{|Gold_i|}$$

n ranges over all entities in gold and system output

Hard coreference

“Between **him** and Darcy there was a very steady friendship, in spite of great opposition of character. Bingley was endeared to Darcy by the easiness, openness, and ductility of **his** temper, though no disposition could offer a greater contrast to **his** own, and though with **his** own **he** never appeared dissatisfied. On the strength of Darcy's regard, Bingley had the firmest reliance, and of **his** judgement the highest opinion. In understanding, Darcy was the superior. Bingley was by no means deficient, but Darcy was clever. **He** was at the same time haughty, reserved, and fastidious, and **his** manners, though well-bred, were not inviting. In that respect **his** friend had greatly the advantage. Bingley was sure of being liked wherever **he** appeared, Darcy was continually giving offense.”

- The Clinton campaign is circulating a fake photo of Barack Obama in Muslim clothes to damage his reputation. In the photo, Obama wears a long sari-like garment.
- The Clinton campaign is circulating a fake photo of Barack Obama in Muslim clothes to damage his reputation, but Obama never wore Muslim clothes.

- You cannot read **Cyril Connolly** for very long without wanting to acquire —and then developing— a relationship with the personality of **the man** himself. [. . .] With **Connolly** there is a marked difference and the difference is that **the artist** and **the man** are so conjoined and intermingled that you cannot savour **the one** without **the other** and vice versa.

Non-identity

- Non-Identity. The two NPs point to two different DEs. Even if they share any feature, they are not ‘the same thing.’
- “President Samaranch sent **a letter** to Sydney in which he asked for information. **A similar missive** has also been received by all the candidate cities to host the Olympic Games of 1996.”

Identity

- Identity. The two NPs point to the same DE (i.e., they have the same set of attributes, as far as one can tell). They are (almost certainly) ‘the same thing.’
- “It began when a Hasidic Jewish family bought one of the town’s two meat-packing plants 13 years ago. First they brought in other Hasidic Jews, then Mexicans, Palestinians, Ukrainians.”

Identity

- Identity. The two NPs point to the same DE (i.e., they have the same set of attributes, as far as one can tell). They are (almost certainly) ‘the same thing.’
- “It began when **a Hasidic Jewish family** bought one of the town’s two meat-packing plants 13 years ago. First **they** brought in other Hasidic Jews, then Mexicans, Palestinians, Ukrainians.”

Near-identity

- A proper noun appears first, and a subsequent noun phrase refers to some aspect of the discourse entity
 - Role
 - Location
 - Organization
 - Information realization
 - Representation
 - Other

Role near-identity: A specific role or function performed by a human, animal or object, is distinguished from their other facets.

“Your father was the greatest” commented an anonymous old lady while she was shaking Alessandro’s hand —Gassman’s best known son. “I will miss *the actor*, but I will be lacking *my father* especially,” he said.

Location near-identity: The name of a location can be used to describe facets such as the physical place, the place associated with a (political) organization, the population living in that location, the ruling government, an affiliated organization, an event celebrated at that location, etc.

“The Jordan authorities arrested, on arriving in **Iraq**, an Italian pilot who violated the air embargo to **this country**.”

Organization near-identity: The name of a company or other social organization can be used to describe facets such as the legal organization itself, the facility that houses the organization or one of its branches, the company shares, a product manufactured by the company, etc.

“The strategy has been a popular one for **McDonalds** . . . It’s a very wise move on for them because if they would have **only just original McDonalds**, I don’t think they would have done so great.”

Information realization near-identity: A discourse entity corresponding to an informational object (e.g., story, law, review, etc.) can be split according to the format in which the information is presented or manifested (FRBR abstraction hierarchy)

She hasn't seen *Gone with the Wind*, but she's read *it*.

Representation near-identity: One noun phrase is a representation of the other--as in a picture or a starring of a person, or a toy replica of a real object.

We stand staring at two paintings of **Queen Elizabeth**. In the one on the left, **she** is dressed as Empress of India. In the one on the right, **she** is dressed in an elegant blue gown.

Other near-identity: (Any other case of metonymy not captured by the other classes)

Chevrolet is a brand of automobile produced by General Motors Company. **It** is feminine because of its sound.

Stuff-object near-identity: One noun phrase expresses the constituent material of the other noun phrase. Unlike components, the stuff of which a thing is made cannot be separated from the object.

Bangladesh Prime Minister Hasina and **President Clinton** expressed the hope that this trend will continue ...Both the **US government** and American businesses welcomed the willingness of Bangladesh.

Part-whole near-identity: One noun phrase mentions a part to refer to the whole expressed by the other noun phrase.

The City Council approved legislation prohibiting selling **alcoholic drinks** during night hours ...Bars not officially categorized as bars will not be allowed to sell **alcohol**.

Class near-identity: Two noun phrases share an is-a relationship, but they stand in a different position in the categorical hierarchy so that one can be viewed as more general or specific to the other.

Diego looked for information about **his character** in the novel forgetting that Saramago does not usually describe **them**.

Place near-identity: The same discourse entity is instantiated in different physical locations, each time resulting in a different discourse entity due to the change in the spatial feature. It is possible for them to coexist but not in the same place.

New York's New Year's Eve is one of the most widely attended parties in the world . . . Celebrating **it** in the Southern Hemisphere is always memorable, especially for those of us in the Northern Hemisphere.

Time near-identity: The same discourse entity is instantiated at different times

On homecoming night **Postville** feels like Hometown, USA, but a look around this town of 2,000 shows it's become a miniature Ellis Island . . . For those who prefer **the old Postville**, Mayor John Hyman has a simple answer.

Numerical function near-identity: The two noun phrases refer to the same function (e.g., price, age, rate, etc.) but have different numerical value due to a change in time or a change in space.

At 8, **the temperature** rose to 99°. This morning **it** was 85°.

Role near-identity: The two NPs refer to the same role (e.g., president, director, etc.) but is filled by a different person due to a change in time or space.

In France, **the president** is elected for a term of seven years, while in the United States **he** is elected for a term of four years.

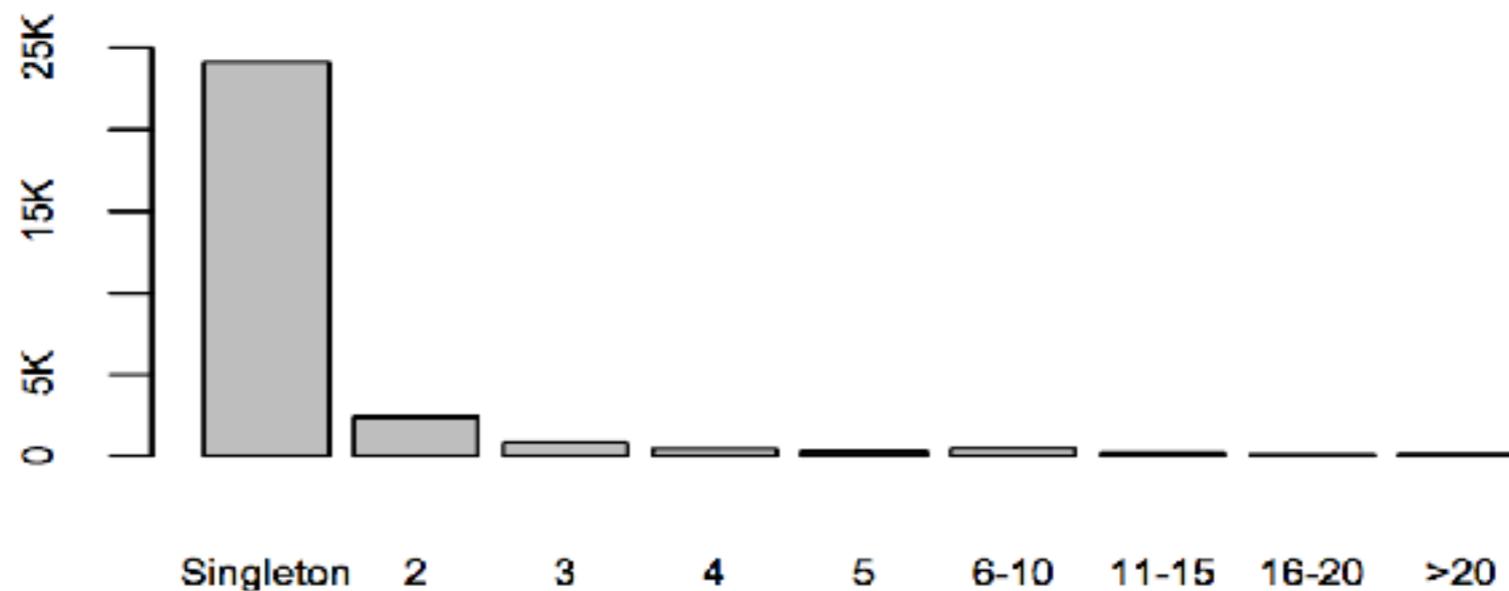
Singletons

- At test time we don't have access to true mentions

[John] saw [a beautiful 1961 Ford Falcon] at [the used car dealership]. [He] showed [it] to [Bob].
[He] bought [it].

Singletons

- Most noun phrases in a discourse are not coreferent. They are **singleton** mentions.



Singletons

- We can built a classifier to predict for any noun phrase, whether it will be a part of a coreference chain or a singleton (78% accurate).

	Estimate	P-value
Type = pronoun	1.21	< 0.001
Type = proper noun	1.88	< 0.001
Animacy = inanimate	-1.36	< 0.001
Animacy = unknown	-0.38	< 0.001
Person = 1	1.05	< 0.001
Person = 2	0.13	< 0.001
Person = 3	1.62	< 0.001
Number = singular	0.61	< 0.001
Number = unknown	0.17	< 0.001
Quantifier = indefinite	-1.49	< 0.001
Quantifier = quantified	-1.23	< 0.001
Number of modifiers	-0.39	< 0.001

	Estimate	P-value
Presence of negation	-0.18	< 0.001
Presence of modality	-0.22	< 0.001
Under an attitude verb	0.03	0.01
AttitudeVerb * (Type = pronoun)	0.29	< 0.001
AttitudeVerb * (Type = proper noun)	0.14	< 0.001
Modal * (Type = pronoun)	0.12	0.04
Modal * (Type = proper noun)	0.35	< 0.001
Negation * (Type = pronoun)	1.07	< 0.001
Negation * (Type = proper noun)	0.30	< 0.001
Negation * (Quantifier = indefinite)	-0.37	< 0.001
Negation * (Quantifier = quantified)	-0.36	0.23
Negation * (Number of modifiers)	0.11	< 0.001

Solve it

- Ontonotes
 - <http://catalog.ldc.upenn.edu/LDC2013T19>
- MUC 7
 - <http://catalog.ldc.upenn.edu/LDC2001T02>
- ACE 2003
 - <http://catalog.ldc.upenn.edu/LDC2001T02>

Thursday 11/16

- Read one of the following:
 - Voigt et al. 2017, Language from police body camera footage shows racial disparities in officer respect
 - Cheng et al. 2017, Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions
 - Underwood 2016, The Life Cycles of Genres