

Foundations of Language Science and Technology

Discourse: Co-Reference

Caroline Sporleder

Universität des Saarlandes

Wintersemester 2009/10

11.01.2010

Discourse

Why Discourse Processing?

Natural language rarely comes in isolated sentences. . .

- newspaper articles
- novels
- dialogues
- speeches by politicians
- . . .

Why Discourse Processing?

Natural language rarely comes in isolated sentences. . .

- newspaper articles
- novels
- dialogues
- speeches by politicians
- . . .

NLP applications need to be able to deal with discourse. . .

Why Discourse Processing?

Natural language rarely comes in isolated sentences. . .

- newspaper articles
- novels
- dialogues
- speeches by politicians
- . . .

NLP applications need to be able to deal with discourse. . .

- dialogue systems
- question answering
- text summarisation
- information extraction
- natural language generation
- natural language understanding
- . . .

Example: Co-reference

Campaigning has closed in Argentina ahead of Sunday's election to elect a successor to President Nestor Kirchner.

The front-runner in the opinion polls is the current first lady, Senator Cristina Fernandez de Kirchner.

She praised the economic record of her husband's government during a rally in Buenos Aires.

Example: Co-reference

She praised the economic record of **her husband**'s government during a rally in Buenos Aires.

Example: Co-reference

She praised the economic record of her husband's government during a rally in Buenos Aires.

Task: Co-reference Resolution

Example: Question Answering ("Why was Caesar killed?")

Caesar was proclaimed dictator for life, and he heavily centralised the bureaucracy of the Republic.

These events provoked a hitherto friend of Caesar, Marcus Junius Brutus, and a group of other senators, to assassinate the dictator on the Ides of March (March 15) in 44 BC.

Example: Question Answering (“Why was Caesar killed?”)

Caesar was proclaimed dictator for life, and he heavily centralised the bureaucracy of the Republic.

These events provoked a hitherto friend of Caesar, Marcus Junius Brutus, and a group of other senators, to assassinate the dictator on the Ides of March (March 15) in 44 BC.

Task: Inferring Discourse Relations (Discourse Parsing)

Example: Coherence/Text Generation

The Eurostar service between Britain, France and Belgium ran a limited service on Saturday, with a reduced service planned for Sunday and Monday.

Many countries across Europe have been hit by the bitter weather conditions.

Passengers have been urged to cancel or postpone their journeys if they do not have to travel.

Example: Coherence/Text Generation

Many countries across Europe have been hit by the bitter weather conditions.

The Eurostar service between Britain, France and Belgium ran a limited service on Saturday, with a reduced service planned for Sunday and Monday.

Passengers have been urged to cancel or postpone their journeys if they do not have to travel.

Example: Coherence/Text Generation

Many countries across Europe have been hit by the bitter weather conditions.

The Eurostar service between Britain, France and Belgium ran a limited service on Saturday, with a reduced service planned for Sunday and Monday.

Passengers have been urged to cancel or postpone their journeys if they do not have to travel.

Task: **Judging Text Coherence**

Example: Referring Expressions

He claims record

The 22-year-old computer science undergraduate from Bath is claiming a world record for the longest distance ridden on a unicycle in 24 hours.

A unicycling student covered exactly 282 miles at Aberystwyth University's athletics track.

Sam Wakeling was aiming to beat the existing record of 235.3 miles.

Example: Referring Expressions

Unicycling student claims record

A **student** is claiming a world record for the longest distance ridden on a unicycle in 24 hours.

Sam Wakeling covered exactly 282 miles at Aberystwyth University's athletics track.

The **22-year-old computer science undergraduate from Bath** was aiming to beat the existing record of 235.3 miles.

Example: Referring Expressions

Unicycling student claims record

A **student** is claiming a world record for the longest distance ridden on a unicycle in 24 hours.

Sam Wakeling covered exactly 282 miles at Aberystwyth University's athletics track.

The **22-year-old computer science undergraduate from Bath** was aiming to beat the existing record of 235.3 miles.

Task: Generating Referring Expressions

Example: Temporal Order

- 1 John arrived at an oasis. He saw the camels around the waterhole.

Example: Temporal Order

- 1 John arrived at an oasis. He saw the camels around the waterhole.

Example: Temporal Order

- 1 John arrived at an oasis. He saw the camels around the waterhole.
- 2 John arrived at an oasis. He left the camels around the waterhole.

Example: Temporal Order

- 1 John arrived at an oasis. He saw the camels around the waterhole.
- 2 John arrived at an oasis. He left the camels around the waterhole.

Task: Temporal Ordering

Co-reference Resolution

Example

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for **Europe**, not getting that much, but paying for **it**. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the **European Union itself**."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked **Georg Bernreuter** about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask **him** whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked **Georg Bernreuter** about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether **he** still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked **Georg Bernreuter** about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," **he** cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. **Some** have complaints, of course, but ask **them** how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (relatively straightforward)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between **Europe** and **its** biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (trickier)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (trickier)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (trickier)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (trickier)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later **we** head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: pronoun resolution (trickier)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... **it's** all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: NP co-reference resolution (also tricky)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: NP co-reference resolution (also tricky)

I asked **Georg Bernreuter** about the EU. **The Bavarian brewer** likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask **him** whether **he** still has faith in Europe. "Absolutely," **he** cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of **Georg's** beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: NP co-reference resolution (also tricky)

I asked Georg Bernreuter about **the EU**. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for **Europe**, not getting that much, but paying for **it**. Bureaucracy is growing faster than the **European Union itself**."

So I ask him whether he still has faith in **Europe**. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between **Europe** and **its** biggest member, and everyone is singing from the same hymn sheet. "**Europe** is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: NP co-reference resolution (also tricky)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and its biggest member, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Example: NP co-reference resolution (also tricky)

I asked Georg Bernreuter about the EU. The Bavarian brewer likes the family of nations - but not the bureaucracy

"We are paying for Europe, not getting that much, but paying for it. Bureaucracy is growing faster than the European Union itself."

So I ask him whether he still has faith in Europe. "Absolutely," he cuts across me, before I can finish the sentence. "The only way to go in Europe is this coming together of the nations."

Later we head off to a beer tent. People are sitting at long tables drinking enormous glasses of Georg's beer ... it's all quite mad.

Nearly everyone says they'll vote in the elections. Some have complaints, of course, but ask them how the relationship is between Europe and **its biggest member**, and everyone is singing from the same hymn sheet. "Europe is the future."

Adapted from <http://news.bbc.co.uk/2/hi/europe/8084685.stm>

Co-reference and Anaphora

- **Co-reference chain:** a set of co-referent referring expressions in a discourse
- **Anaphora:** co-reference of one referring expression with its **antecedent**
- **Anaphor:** a referring expression (often a pronoun) which **refers back to something** mentioned previously (e.g. *she, this day, the cat* ... but not *Peter* etc.)
- analogous: **cataphor** for expressions referring **forward** (e.g., *While he was in office, Bill Clinton* ...)
- **co-reference vs. anaphora**
 - cross-document co-reference (=not anaphoric)
 - some anaphora are not strictly co-referent (*Everybody has his own destiny.*)

Co-reference and Anaphora

- **Co-reference chain:** a set of co-referent referring expressions in a discourse
- **Anaphora:** co-reference of one referring expression with its **antecedent**
- **Anaphor:** a referring expression (often a pronoun) which **refers back to something** mentioned previously (e.g. *she, this day, the cat . . .* but not *Peter* etc.)
- analogous: **cataphor** for expressions referring **forward** (e.g., *While he was in office, Bill Clinton . . .*)
- **co-reference vs. anaphora**
 - cross-document co-reference (=not anaphoric)
 - some anaphora are not strictly co-referent (*Everybody has his own destiny.*)

Co-reference Resolution vs. Anaphora Resolution

Co-reference Resolution: find the co-reference chains in a text.

Anaphora Resolution: find the antecedent of an anaphor.

How would you model anaphora / co-reference resolution?

Which linguistic factors provide clues?

Ambiguity and Disambiguating Factors

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

Ambiguity and Disambiguating Factors

Jane told Peter he was in danger.

⇒ **Agreement** (gender, number etc.): *he = Peter*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

The cat did not come down from the tree. **It** was scared.

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

The cat did not come down from the tree. **It** was scared.

⇒ **selectional preferences**: *it = the cat*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

The cat did not come down from the tree. **It** was scared.

⇒ **selectional preferences**: *it = the cat*

*Jane told Mary **she** was in danger.*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

The cat did not come down from the tree. **It** was scared.

⇒ **selectional preferences**: *it = the cat*

*Jane told Mary **she** was in danger.*

⇒ **salience** (e.g., subject position): *she = Jane*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

The cat did not come down from the tree. **It** was scared.

⇒ **selectional preferences**: *it = the cat*

*Jane told Mary **she** was in danger.*

⇒ **salience** (e.g., subject position): *she = Jane*

*Jane told Mary **SHE** was in danger.*

Ambiguity and Disambiguating Factors

*Jane told Peter **he** was in danger.*

⇒ **Agreement** (gender, number etc.): *he = Peter*

*Peter said that John is running the business for **himself**.*

⇒ **syntactic constraints**: *himself = John*

*The cat did not come down from the tree. **It** was scared.*

⇒ **selectional preferences**: *it = the cat*

*Jane told Mary **she** was in danger.*

⇒ **salience** (e.g., subject position): *she = Jane*

*Jane told Mary **SHE** was in danger.*

⇒ **prosody**: *she = Mary*

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

⇒ **lexical semantics** (*warned*): *she* = Mary

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

⇒ **lexical semantics** (*warned*): *she* = Mary

Tony Blair met President Yeltsin. *The old man* had just recovered from a heart attack.

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

⇒ **lexical semantics** (*warned*): *she* = Mary

Tony Blair met President Yeltsin. *The old man* had just recovered from a heart attack.

⇒ **world knowledge**: *the old man* = Yeltsin

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

⇒ **lexical semantics** (*warned*): *she* = Mary

Tony Blair met President Yeltsin. *The old man* had just recovered from a heart attack.

⇒ **world knowledge**: *the old man* = Yeltsin

Georg Bernreuter ... Mr. Bernreuter

Ambiguity and Disambiguating Factors

Jane told Peter *he* was in danger.

⇒ **Agreement** (gender, number etc.): *he* = Peter

Peter said that John is running the business for *himself*.

⇒ **syntactic constraints**: *himself* = John

The cat did not come down from the tree. *It* was scared.

⇒ **selectional preferences**: *it* = the cat

Jane told Mary *she* was in danger.

⇒ **salience** (e.g., subject position): *she* = Jane

Jane told Mary *SHE* was in danger.

⇒ **prosody**: *she* = Mary

Jane warned Mary *she* was in danger.

⇒ **lexical semantics** (*warned*): *she* = Mary

Tony Blair met President Yeltsin. *The old man* had just recovered from a heart attack.

⇒ **world knowledge**: *the old man* = Yeltsin

Georg Bernreuter ... Mr. Bernreuter

⇒ **surface string similarity**

Difficulties:

Difficulties:

- different form \nrightarrow different referents
(*Georg Bernreuter vs. the Bavarian brewer vs. he*)

Difficulties:

- different form \nrightarrow different referents
(*Georg Bernreuter* vs. *the Bavarian brewer* vs. *he*)
- same form \nrightarrow same referents
(*the cat*, *Michael Jackson the singer* vs. *Michael Jackson the British general*)

Co-Reference Resolution Steps

- 1 identify anaphor / markable
difficulties: NPs which aren't referring expressions; pleonastic *it* (*It's raining.*) etc.
- 2 identify potential antecedents
- 3 find correct co-referent for each anaphor / markable

Before 1990 ...

- reference resolution = pronoun resolution
- rule-based (manually created rules)
- Examples:
 - SHRDLU (Winograd, 1972): complex heuristics (focus, obliqueness etc.)
 - Hobbs's (1976, 1978): heuristically directed search in parse trees
 - centering-based (Brennan et al. 1987)
 - Lappin & Leass (1994): agreement, syntax, salience

After 1990 ...

- corpus-based (co-occurrence statistics, machine learning)
⇒ Message Understanding Conference (MUC): annotated data
- reference resolution for non-pronominal expressions (definite NPs, bridging; z.B. Vieira & Poesio, 2000)

Rule-based Approaches

Resolution of Anaphora Procedure

Scope

- third person pronouns
- lexical anaphors (reflexives and reciprocals)

Software

numerous (re-)implementations, e.g.,

http:

`//wing.comp.nus.edu.sg/~qiu/NLPTools/JavaRAP.html`

Components

- procedure for identifying pleonastic/expletive pronouns
- morpho-syntactic filters
- salience weighting
- a resolution procedure

- pre-specified list of modal adjectives (*necessary, certain, good, possible . . .*)
- pre-specified list of cognitive verbs (*recommend, think, believe, expect . . .*)
- manually built rules, e.g.:

It is **modaladj** that **S**.

It is **cogv-ed** that **S**.

It is time to **VP**.

- expressions that don't agree in person, number and gender are not co-referent
- manually built syntactic filter rules (e.g., *John seems to want to see him.*, *His portrait of John is interesting.*)

Saliency Factors

- associated with one or more discourse referents (which are in its scope)
- each factor is weighted
- all weights decay as discourse goes on (at steps of -2 for each new sentence)
- factor is removed when weight reaches zero

Salience Factors

- **sentence recency**
- **subject emphasis**: *The postman delivered a parcel to Peter.*
- **existential emphasis**: *There are only a few restrictions on the courses one can choose.*
- **accusative emphasis**: *The postman delivered a parcel to Peter.*
- **indirect object and oblique complement emphasis**: *The postman delivered a parcel to Peter.*
- **head noun emphasis**: embedded NPs **don't** receive this factor (e.g., *Experts still discuss the impact of Opel's restructuring plans*)
- **non-adverbial emphasis**: any NP not contained in an adverbial PP demarcated by a separator (e.g., **not**: *In the first year, the company made a healthy profit.*)

Initial Weights

sentence recency	100
subj. emphasis	80
exist. emphasis	70
acc. emphasis	50
ind. obj and oblique compl. emphasis	40
head noun emphasis	80
non-adv. emphasis	50

Equivalence classes

- referring expressions are grouped into **equivalence classes** (note: no co-reference between definite NPs)
- each equivalence class has a saliency weight (= the sum of the weights of all saliency factors associated with the most recent expression in the class)

In a nutshell:

- 1 **classify referring NPs** in current sentence (definite NP, indefinite NP, pleonastic pronoun, other pronoun)
- 2 for all non-pleonastic pronouns **apply morpho-syntactic filters** and compute remaining potential antecedents
- 3 **modify salience scores** for possible anaphor antecedent pairs:
 - if antecedent follows anaphor, decrease weight by 175 (i.e., **cataphora are penalised**)
 - if grammatical roles between anaphor and antecedent are parallel increase weight by 35 (i.e., **parallelism is rewarded**)
- 4 **rank** possible antecedents by salience score
- 5 apply **salience threshold**
- 6 of antecedents above the threshold **choose highest scoring one**, in case of a tie select the antecedent closest to the anaphor

Example

John Smith talks about the EU.

Weights:

- **John Smith:** 100 (recency) + 80 (subj) + 80 (head noun) + 50 (non-adv) = 310
- **the EU:** 100 (recency) + 50 (acc) + 80 (head noun) + 50 (non-adv) = 280

Example

John Smith talks about the EU. He likes the family of nations.

Weights:

- **John Smith:** 98 (recency) + 78 (subj) + 78 (head noun) + 48 (non-adv) = 302
- **the EU:** 98 (recency) + 48 (acc) + 78 (head noun) + 48 (non-adv) = 272
- **the family of nations:** 100 (recency) + 50 (acc) + 80 (head noun) + 50 (non-adv) = 280
- **nations:** 100 (recency) + 50 (acc) + 50 (non-adv) = 200

Example

John Smith talks about the EU. He likes the family of nations.

Weights:

- **John Smith:** 98 (recency) + 78 (subj) + 78 (head noun) + 48 (non-adv) = 302
- **the EU:** 98 (recency) + 48 (acc) + 78 (head noun) + 48 (non-adv) = 272
- **the family of nations:** 100 (recency) + 50 (acc) + 80 (head noun) + 50 (non-adv) = 280
- **nations:** 100 (recency) + 50 (acc) + 50 (non-adv) = 200

Resolving “he”:

- “he” = “John Smith” by morpho-syntactic filter

Example

John Smith talks about the EU. He likes the family of nations.

Weights:

- **John Smith:** 100 (recency) + 80 (subj) + 80 (head noun) + 50 (non-adv) = 310
- **the EU:** 98 (recency) + 48 (acc) + 78 (head noun) + 48 (non-adv) = 272
- **the family of nations:** 100 (recency) + 50 (acc) + 80 (head noun) + 50 (non-adv) = 280
- **nations:** 100 (recency) + 50 (acc) + 50 (non-adv) = 200

Resolving “he”:

- “he” = “John Smith” by morpho-syntactic filter

Example

John Smith talks about the EU. He likes the family of nations. It is a good thing.

Weights:

- **John Smith:** 98 (recency) + 78 (subj) + 78 (head noun) + 48 (non-adv) = 302
- **the EU:** 96 (recency) + 46 (acc) + 76 (head noun) + 46 (non-adv) = 264
- **the family of nations:** 98 (recency) + 42 (acc) + 78 (head noun) + 42 (non-adv) = 272
- **nations:** 98 (recency) + 42 (acc) + 42 (non-adv) = 194
- **a good thing:** 100 (recency) + 50 (acc) + 80 (head) + 50 (non-adv) = 280

Example

John Smith talks about the EU. He likes the family of nations. It is a good thing.

Weights:

- **John Smith:** 98 (recency) + 78 (subj) + 78 (head noun) + 48 (non-adv) = 302
- **the EU:** 96 (recency) + 46 (acc) + 76 (head noun) + 46 (non-adv) = 264
- **the family of nations:** 98 (recency) + 42 (acc) + 78 (head noun) + 42 (non-adv) = 272
- **nations:** 98 (recency) + 42 (acc) + 42 (non-adv) = 194
- **a good thing:** 100 (recency) + 50 (acc) + 80 (head) + 50 (non-adv) = 280

Resolving “it”

- “the family of nations” (272) > “the EU” (264) > “nations” (194) > “a good thing” (105, cataphor)

Set-Up

unseen test set of 345 randomly selected sentence pairs (sentence with pronoun plus preceding sentence) subject to constraints:

- RAP generates a candidate list of at least two elements
- correct antecedent is on that list

Result

- 86% accuracy

Can you think of any cases that RAP would not do well on?

Machine Learning Approaches

Hybrid RAP

RAPSTAT (Dagan & Itai (1990, 1991)): RAP Hybrid with Statistics

Motivation

RAP disregards selectional preferences.

RAPSTAT (Dagan & Itai (1990, 1991)): RAP Hybrid with Statistics

Motivation

RAP disregards selectional preferences.

Example

We gave the bananas to the monkeys because **they** were hungry.

RAPSTAT (Dagan & Itai (1990, 1991)): RAP Hybrid with Statistics

Motivation

RAP disregards selectional preferences.

Example

We gave the bananas to the monkeys because **they** were hungry.

Saliency Scores

- **the bananas:** 100 (recency) + 50 (acc) + 80 (head) + 50 (non-adv) = 280
- **the monkeys:** 100 (recency) + 40 (ind. obj) + 80 (head) + 50 (non-adv) = 270

RAPSTAT (Dagan & Itai (1990, 1991)): RAP Hybrid with Statistics

Motivation

RAP disregards selectional preferences.

Example

We gave the bananas to the monkeys because **they** were hungry.

Saliency Scores

- **the bananas:** 100 (recency) + 50 (acc) + 80 (head) + 50 (non-adv) = 280
- **the monkeys:** 100 (recency) + 40 (ind. obj) + 80 (head) + 50 (non-adv) = 270

Resolving “they”

- “they” = “the bananas”
- however:

$$p(\text{areHungry}(\text{bananas})) \ll p(\text{areHungry}(\text{monkeys}))$$

Any ideas how to do this?

Use statistics to improve anaphora resolution

- selectional preferences are automatically computed from corpus (co-occurrence statistics)
- if statistics point to another antecedent than RAP and the salience difference between the two potential antecedents is not too high, select statistically more plausible antecedent

Example

They held tax money aside on the basis that the government said it was going to collect it.

Example

They held **tax money** aside on the basis that the **government** said **it** was going to collect **it**.

Example

They held **tax money** aside on the basis that the **government** said **it** was going to collect **it**.

Subject(**it**, collect)

Object(**it**, collect)

Example

They held **tax money** aside on the basis that the **government** said **it** was going to collect **it**.

Subject(**it**, collect)

Object(**it**, collect)

co-occurrence statistics:

Subject(money,collect) = 5

Subject(government,collect) = 198

Object(money,collect) = 149

Object(government,collect) = 0

Example

They held **tax money** aside on the basis that the **government** said **it** was going to collect **it**.

Subject(**it**, collect)

Object(**it**, collect)

co-occurrence statistics:

Subject(money,collect) = 5

Subject(government,collect) = 198

Objekc(money,collect) = 149

Objekc(government,collect) = 0

⇒ **it** = **government**

⇒ **it** = **money**

Comparison RAP vs. RAPSTAT

- RAPSTAT has 89% accuracy (vs. 86% for RAP)
- overrules RAP's decision in 22% of the cases, 61% of these are correctly resolved by RAPSTAT

From Anaphora to Co-reference Resolution

Co-Reference Resolution

- **identity of reference** between two **markables** (definite NPs, proper names, demonstrative NPs, appositives, embedded NPs, pronouns etc.)
- annotated data from Message Understanding Conferences (MUC-6, MUC-7)

Example

Ms Washington's candidacy is being championed by several powerful lawmakers including her boss, Chairman John Dingell. She is currently a counsel to the committee.

Co-Reference Resolution

- **identity of reference** between two **markables** (definite NPs, proper names, demonstrative NPs, appositives, embedded NPs, pronouns etc.)
- annotated data from Message Understanding Conferences (MUC-6, MUC-7)

Example: markables

[[Ms Washington]'s candidacy] is being championed by [several powerful lawmakers] including [[her] boss], [Chairman John Dingell]. [She] is currently [a counsel] to [the committee].

Co-Reference Resolution

- **identity of reference** between two **markables** (definite NPs, proper names, demonstrative NPs, appositives, embedded NPs, pronouns etc.)
- annotated data from Message Understanding Conferences (MUC-6, MUC-7)

Example: co-reference resolution

[[**Ms Washington**]'s candidacy] is being championed by [several powerful lawmakers] including [[**her**] boss], [**Chairman John Dingell**]. [**She**] is currently [a counsel] to [the committee].

- supervised machine learning (C.5 - decision tree)
- on MUC-6 and MUC-7 data
- 12 shallow features

Generous

- all pairs in a co-reference chain are positive examples
- all other pairs are negative examples

More selective (Soon et al., 2001)

- adjacent pairs in co-reference chain are positive training data
- for all markables between the two co-referent expressions, pair the markable with either expression and label as 'negative'

Generous

- all pairs in a co-reference chain are positive examples
- all other pairs are negative examples

More selective (Soon et al., 2001)

- adjacent pairs in co-reference chain are positive training data
- for all markables between the two co-referent expressions, pair the markable with either expression and label as 'negative'

Note: in both cases (especially the first one) the training set will be imbalanced.

Example

[[Ms Washington]'s candidacy] is being championed by [several powerful lawmakers] including [[her] boss], [Chairman John Dingell]. [She] is currently [a counsel] to [the committee].

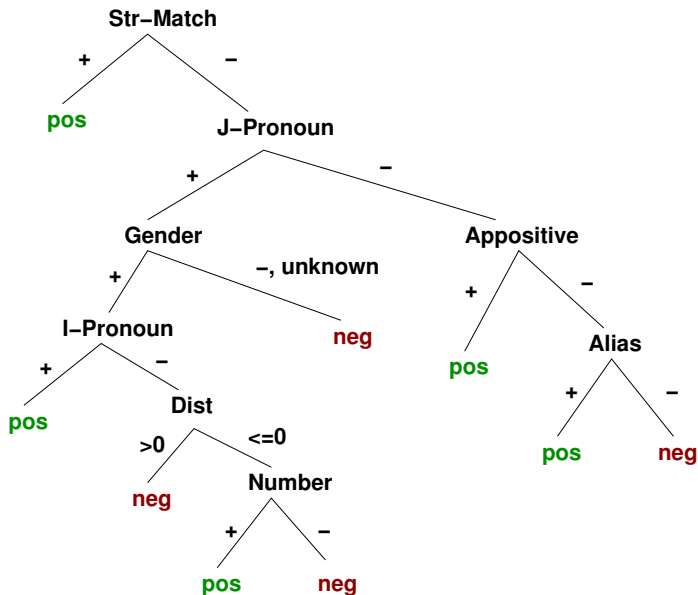
Training Data

- (Ms Washington, her): pos
- (her, she): pos
- (her boss, Chairman John Dingell): pos
- (Ms Washington, several powerful lawmakers): neg
- (her, several powerful lawmakers): neg

Twelve shallow features:

- distance (in terms of sentences): numeric
- pronoun features (i-pronoun, j-pronoun): boolean
- string match (excluding determiners): boolean
- j type features (def. NP, dem. NP): boolean
- number agreement: boolean
- semantic class agreement (WordNet, most frequent sense): true, false, unknown
- gender agreement: true, false, unknown
- both proper names (i and j): boolean
- alias feature (“Mr. Simpson” - “Bent Simpson”): boolean
- appositive feature: boolean

Soon et al. (2001): Decision Tree Learnt



Greedy chain building algorithm

- 1 compare each markable j with each preceding markable i , starting from the closest
- 2 apply decision tree to the pair (j, i)
- 3 stop as soon as decision tree returns 'true'

Scores for MUC-6 and MUC-7

- Recall: 56-59%
 - Precision: 66-67%
 - F-Score: 60-63%
- ⇒ (competitive with other systems)

How would you try to improve on
Soon et al. (2001)?

Ng and Cardie (2002):

- improve on Soon et al. through:
 - extra-linguistic changes to the learning framework
 - large-scale expansion of the feature set, incorporating “more sophisticated linguistic knowledge”
- MUC F-Scores: 70.4% and 63.4% (Soon et al: 62.6% and 60.4%)

Changes to the Learning Framework

Best-first instead of greedy clustering:

- Soon et al. search right-to-left for a possible antecedent and select the **first** (i.e., rightmost) expression which is classified as co-referent
- Ng and Cardie search right-to-left and select the **best** expression that is classified as coreferent (i.e., the one that scores highest)

Split string match feature:

- implement separate string match features for different types of expressions (pronouns, proper names, non-pronominal NPs)

Results (C4.5 and Ripper)

- statistically significant **gains in precision** over Soon et al. baseline
- **no drop in recall**

41 new features, e.g.:

- more complex **string matching**
- more **semantic features** (e.g., testing for ancestor-descendant relationships in WordNet, graph-distance in WordNet)
- 26 new **grammatical features**
- **hard-coded linguistic constraints**, indicator features (agreement, binding etc.)
- output of **rule-based pronoun resolution system**

Results

- significant increases in recall
- even bigger decreases in precision
⇒ **F-Score goes down**

Error Analysis

- drop in precision due to bad precision on common nouns
- counter intuitive rules were learnt

Example

(i,j) = coreferent iff

$properName(i) \wedge definiteNP(j) \wedge subject(j) \wedge semClass(i) =$
 $semClass(j) \wedge distance(i,j) \leq 1$

⇒ rule covers 38 examples with 18 exceptions

⇒ **this is a data sparseness problem!**

Solution: manual feature selection

- on data overall: increase in F-Score
- but large drop in precision for pronouns

Conclusion

pronoun and common noun resolution remain challenging

Don't treat co-reference resolution as a classification task!

Intuitively pairwise decisions are not what one wants

- ⇒ **ranking** instead of classification (e.g., Yang et al., 2003; Denis and Baldrige, 2007)
- ⇒ **graph partitioning** to convert pairwise scores into final coherent clustering (McCallum and Wellner, 2004)

Co-reference Resolution

- ... is a **heterogeneous task** (pronoun resolution, proper name matching, co-reference resolution for definite NPs)
⇒ one-size-fits-all may not be the best strategy
- ... is a **complex task**, many factors are involved (focus structure, similarity of surface strings, grammatical constraints, semantic constraints etc.)
- ... **maybe shouldn't be modelled as a classification task** (artificial pairwise decisions, class imbalance etc.)

-  Brennan, Susan E., Marilyn W. Friedman and Carl J. Pollard,
A centering approach to pronouns,
Proceedings of the 25th annual meeting on Association for Computational Linguistics, p.155-162, July 06-09, 1987, Stanford, California.
-  Dagan, Ido and Alon Itai,
Automatic Acquisition of Constraints for the Resolution of Anaphora References and Syntactic Ambiguities,
Proceedings of COLING, pp. 330-332, 1990.
-  Dagan, Ido and Alon Itai,
A Statistical Filter for Resolving Pronoun References,
In: Y. A. Feldman and A. Bruckstein (Eds.), *Artificial Intelligence and Computer Vision*, Elsevier Science Publishers B.V., 1991, pp. 125-135.
-  Denis, Pascal and Jason Baldridge,
A ranking approach to pronoun resolution,
In *Proceedings of IJCAI-2007*. Hyderabad, India, 2007.



Hobbs, Jerry R.

Pronoun Resolution,

Research Report 76-1, Department of Computer Sciences, City College, City University of New York, 1976.



Hobbs, Jerry R.,

Resolving Pronoun References,

Lingua, Vol. 44, pp. 311-338, 1978.



Lappin, Shalom and Herbert J. Leass,

An Algorithm for Pronominal Anaphora Resolution, *Computational Linguistics*, 20(4), pp. 535-561, 1994.



McCallum, Andrew and Ben Wellner,

Conditional Models of Identity Uncertainty with Application to Noun Coreference,

Neural Information Processing Systems (NIPS), 2004.

-  Ng, Vincent and Claire Cardie,
Improving Machine Learning Approaches to Coreference Resolution,
Proceedings of the ACL, 2002.
-  Soon, Wee Meng, Daniel Chung Yong Lim and Hwee Tou Ng,
A Machine Learning Approach to Coreference Resolution of Noun
Phrases,
Computational Linguistics 27(4), pp. 521-544, 2001.
-  Vieira, Renata and Massimo Poesio,
An Empirically-Based System for Processing Definite Descriptions,
Computational Linguistics, 26(4), p. 539-593, 2000.
-  Winograd, Terry
Understanding Natural Language,
Academic Press, 1972.
-  Yang, X., G. Zhou, J. Su, and C.L. Tan,
Coreference resolution using competitive learning approach.
In *Proceedings of ACL*, pages 176183, 2003.