

Introduction to Human Language Technologies

10. Coreference resolution



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



Exercise 1

- 1 Identify all the mentions for identity noun-phrase coreference resolution in the following text:

Mr. Smith was traveling when Lara came back home. He had never been far from his wife. Mrs. Smith closed the door and went to bed thinking of John.

- 2 Extract all positive and negative examples of coreferent mention pairs for closest-first and for best-first strategies.

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

Mr. Smith was traveling when Lara came back home. He had never been far from his wife. Mrs. Smith closed the door and went to bed thinking of John.

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back
3:[home]₃. 4:[He]₁ had never been far from 6:[5:[his]₁
wife]₂. 7:[Mrs. Smith]₂ closed 8:[the door]₄ and went to
9:[bed]₅ thinking of 10:[John]₁.

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back
3:[home]₃. 4:[He]₁ had never been far from 6:[5:[his]₁
wife]₂. 7:[Mrs. Smith]₂ closed 8:[the door]₄ and went to
9:[bed]₅ thinking of 10:[John]₁.

Extract all positive and negative examples of coreferent mention pairs for closest-first and for best-first strategies.

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,2)

e^-

Best-first strategy:

e^+ (1,2)

e^-

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,2)

e^-

Best-first strategy:

e^+ (1,2)

e^-

1 and 2 are not in the same chain

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (?,3)

e^-

Best-first strategy:

e^+ (?,3)

e^-

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (?,3)

e^-

Best-first strategy:

e^+ (?,3)

e^-

3 is a singleton

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4)

e^- (2,4) (3,4)

Best-first strategy:

e^+ (1,4)

e^- (2,4) (3,4)

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5)

e^- (2,4) (3,4)

Best-first strategy:

e^+ (1,4) + (4,5)

e^- (2,4) (3,4)

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5)

e^- (2,4) (3,4)

Best-first strategy:

e^+ (1,4) + (4,5)

e^- (2,4) (3,4)

5 is a pronoun. The antecedent can be a pronoun. No e^- found

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (5,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (5,6)$$

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (5,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (5,6)$$

mention 5 does not precede mention 6 as it starts in the same position. It is not taken as e^- for both strategies

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (4,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (4,6)$$

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

No mentions in between 6 and 7. So no e^-

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7) + (? ,8)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Best-first strategy:

$$e^+ (1,4) + (4,5) + (2,6) + (6,7) + (? ,8)$$

$$e^- (2,4) (3,4) + (4,6) (3,6)$$

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (?,8)

e^- (2,4) (3,4) + (4,6) (3,6)

Best-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (?,8)

e^- (2,4) (3,4) + (4,6) (3,6)

8 is a singleton as well as 9

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6)

Best-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6)

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6)

Best-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6)

10 is not a pronoun but 5 is, as well as 4. They are not e^+ for best-first strategy

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6) + (6,10) (7,10) (8,10) (9,10)

Best-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6)

Exercise 1

Identify all the mentions for identity noun-phrase coreference resolution in the following text:

1:[Mr. Smith]₁ was traveling when 2:[Lara]₂ came back 3:[home]₃.
4:[He]₁ had never been far from 6:[5:[his]₁ wife]₂. 7:[Mrs. Smith]₂
closed 8:[the door]₄ and went to 9:[bed]₅ thinking of 10:[John]₁.

Closest-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (5,10)

e^- (2,4) (3,4) + (4,6) (3,6) + (6,10) (7,10) (8,10) (9,10)

Best-first strategy:

e^+ (1,4) + (4,5) + (2,6) + (6,7) + (1,10)

e^- (2,4) (3,4) + (4,6) (3,6) + (2,10) (3,10) (6,10) (7,10) (8,10) (9,10)

Exercise 2

Assume we have already learned a mention-pair classifier. Consider m_1, \dots, m_9 as the ordered sequence of mentions in a text. Given the following probabilities for the mention pairs:

$$\begin{array}{ll} P(CO | \langle m_1, m_7 \rangle) = 0.8; & P(CO | \langle m_2, m_7 \rangle) = 0.6; \\ P(CO | \langle m_3, m_7 \rangle) = 0.4; & P(CO | \langle m_4, m_7 \rangle) = 0.5; \\ P(CO | \langle m_5, m_7 \rangle) = 0.7; & P(CO | \langle m_6, m_7 \rangle) = 0.6; \\ P(CO | \langle m_7, m_8 \rangle) = 0.9; & P(CO | \langle m_7, m_9 \rangle) = 0.5; \end{array}$$

provide the m_7 antecedent that results from applying:

- 1 Closest-first strategy
- 2 Best-first strategy

assuming a coreference threshold of > 0.6 .

Exercise 2

m7 antecedent using **closest-first strategy** assuming a coreference threshold of > 0.6 .

$$\begin{aligned}P(CO | < m1, m7 >) &= 0.8; & P(CO | < m2, m7 >) &= 0.6; \\P(CO | < m3, m7 >) &= 0.4; & P(CO | < m4, m7 >) &= 0.5; \\P(CO | < m5, m7 >) &= 0.7; & P(CO | < m6, m7 >) &= 0.6; \\P(CO | < m7, m8 >) &= 0.9; & P(CO | < m7, m9 >) &= 0.5;\end{aligned}$$

m1 m2 m3 m4 m5 m6 m7

Exercise 2

m7 antecedent using **closest-first strategy** assuming a coreference threshold of > 0.6 .

$$\begin{aligned}P(CO | < m1, m7 >) &= 0.8; & P(CO | < m2, m7 >) &= 0.6; \\P(CO | < m3, m7 >) &= 0.4; & P(CO | < m4, m7 >) &= 0.5; \\P(CO | < m5, m7 >) &= 0.7; & P(CO | < m6, m7 >) &= 0.6; \\P(CO | < m7, m8 >) &= 0.9; & P(CO | < m7, m9 >) &= 0.5;\end{aligned}$$

m1 m2 m3 m4 m5 m6 m7

Exercise 2

m7 antecedent using **closest-first strategy** assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>
0.8	0.6	0.4	0.5	0.7	0.6	

Exercise 2

m7 antecedent using **closest-first strategy** assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>
0.8	0.6	0.4	0.5	0.7	0.6	

Exercise 2

m7 antecedent using **closest-first strategy** assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>
0.8	0.6	0.4	0.5	0.7	0.6	

Exercise 2

m7 antecedent using **best-first strategy**, assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>
0.8	0.6	0.4	0.5	0.7	0.6	

Exercise 2

$m7$ antecedent using **best-first strategy**, assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

$m1$	$m2$	$m3$	$m4$	$m5$	$m6$	$m7$
0.8	0.6	0.4	0.5	0.7	0.6	

Exercise 2

m7 antecedent using **best-first strategy**, assuming a coreference threshold of 0.6.

$$\begin{aligned}P(CO | \langle m1, m7 \rangle) &= 0.8; & P(CO | \langle m2, m7 \rangle) &= 0.6; \\P(CO | \langle m3, m7 \rangle) &= 0.4; & P(CO | \langle m4, m7 \rangle) &= 0.5; \\P(CO | \langle m5, m7 \rangle) &= 0.7; & P(CO | \langle m6, m7 \rangle) &= 0.6; \\P(CO | \langle m7, m8 \rangle) &= 0.9; & P(CO | \langle m7, m9 \rangle) &= 0.5;\end{aligned}$$

<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>
0.8	0.6	0.4	0.5	0.7	0.6	