# NLP

# Introduction to NLP

*Discourse Analysis*

# Issues with Discourse

- Anaphora
  - I went to see my grandfather at the hospital. *The old man* has been there for weeks. *He* had surgery a few days ago.
- Referring expressions and antecedents
- Issues with both single sentences and multi-sentential text
- Needed
  - Models of discourse

# Coreference

- Sample use of anaphora
  - John saw Mary in the park. As every morning, she was walking her dog.
- What does "she" refer to?
- Candidate referents
  - John
  - Mary
  - The park
  - Every morning
  - Her dog

# MUC-7 Coreference Task

<COREF ID="6" TYPE="IDENT" REF="5" MIN="Aeroflot">The Russian airline Aeroflot</COREF> has been hit with <COREF ID="19">a writ</COREF> for loss and damages, filed in <COREF ID="15">Hong Kong</COREF> by <COREF ID="7" TYPE="IDENT" REF="4" MIN="families">the families of <COREF ID="22" MIN="passengers">seven passengers killed in <COREF ID="8" TYPE="IDENT" REF="9" MIN="crash">an air crash</COREF></COREF></COREF>.

All 75 people on board <COREF ID="12" MIN="Airbus">the <COREF ID="10" TYPE="IDENT" REF="6">Aeroflot</COREF> <COREF ID="25">Airbus</COREF> </COREF> died when <COREF ID="11" TYPE="IDENT" REF="12">it</COREF> ploughed into a Siberian mountain in March 1994.

# MUC-7 Coreference Task

<COREF ID="6" TYPE="IDENT" REF="5" MIN="Aeroflot">The Russian airline Aeroflot</COREF> has been hit with <COREF ID="19">a writ</COREF> for loss and damages, filed in <COREF ID="15">Hong Kong</COREF> by <COREF ID="7" TYPE="IDENT" REF="4" MIN="families">the families of <COREF ID="22" MIN="passengers">seven passengers killed in <COREF ID="8" TYPE="IDENT" REF="9" MIN="crash">an air crash</COREF></COREF></COREF>.

All 75 people on board **<COREF ID="12" MIN="Airbus">the <COREF ID="10" TYPE="IDENT" REF="6">Aeroflot</COREF> <COREF ID="25">Airbus</COREF> </COREF>** died when **<COREF ID="11" TYPE="IDENT" REF="12">it</COREF>** ploughed into a Siberian mountain in March 1994.

# Screwdriver on Wikipedia

- A screwdriver is a tool, manual or powered, for turning (driving or removing) screws. A typical simple screwdriver has a handle and a shaft, and a tip that the user inserts into the screw head to turn it. The shaft is usually made of tough steel to resist bending or twisting. The tip may be hardened to resist wear, treated with a dark tip coating for improved visual contrast between tip and screw—or ridged or treated for additional 'grip'. Handles are typically wood, metal, or plastic and usually hexagonal, square, or oval in cross-section to improve grip and prevent the tool from rolling when set down. Some manual screwdrivers have interchangeable tips that fit into a socket on the end of the shaft and are held in mechanically or magnetically. These often have a hollow handle that contains various types and sizes of tips, and a reversible ratchet action that allows multiple full turns without repositioning the tip or the user's hand.

# Screwdriver on Wikipedia

- A screwdriver is a tool, manual or powered, for turning (driving or removing) screws. A typical simple screwdriver has a handle and a shaft, and a tip that the user inserts into the screw head to turn it. **The shaft** is usually made of tough steel to resist bending or twisting. **The tip** may be hardened to resist wear, treated with a dark tip coating for improved visual contrast between tip and screw—or ridged or treated for additional 'grip'. **Handles** are typically wood, metal, or plastic and usually hexagonal, square, or oval in cross-section to improve grip and prevent the tool from rolling when set down. Some manual screwdrivers have interchangeable tips that fit into a socket on the end of the shaft and are held in mechanically or magnetically. **These** often have a hollow handle that contains various types and sizes of tips, and a reversible ratchet action that allows multiple full turns without repositioning the tip or the user's hand.

# Coreference Resolution

- Agreement constraints
  - gender, number, animacy
- Syntactic constraints
  - e.g., parallelism
- Sentence ordering
  - recency

# Salience Weights

| | |
|---|---|
| Sentence recency | 100 |
| Subject emphasis | 80 |
| Existential emphasis ("there is …") | 70 |
| Accusative (direct object) emphasis | 50 |
| Indirect object emphasis | 40 |
| Non-adverbial emphasis | 50 |
| Head noun emphasis | 80 |

[Lappin and Leass 1994]

# Lappin and Leass (cont'd)

- Recency handling
  - weights are cut in half after each sentence is processed.
- Examples:
  - An Acura Integra is parked in the lot. (subject)
  - There is an Acura Integra parked in the lot. (existential predicate nominal)
  - John parked an Acura Integra in the lot. (object)
  - John gave Susan an Acura Integra. (indirect object)
  - In his Acura Integra, John showed Susan his new CD player. (demarcated adverbial PP)

# Resolution of Anaphora Procedure (RAP)

- Collect the potential referents (up to four sentences back).
- Remove potential referents that do not agree in number or gender with the pronoun.
- Remove potential referents that do not pass intrasentential syntactic coreference constraints.
- Compute the total salience value of the referent by adding any applicable values for role parallelism (+35) or cataphora (−175).
- Select the referent with the highest salience value. In case of a tie, select the closest referent in terms of string position.
- When moving to a new sentence, halve all scores for the existing entities on the list.

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

# Example

- John saw a beautiful Acura Integra at the dealership last week. He showed it to Bill. He bought it.

|  | Rec | Subj | Exist | Obj | Ind Obj | Non Adv | Head N | Total |
|---|---|---|---|---|---|---|---|---|
| John | 100 | 80 |  |  |  | 50 | 80 | 310 |
| Integra | 100 |  |  | 50 |  | 50 | 80 | 280 |
| dealership | 100 |  |  |  |  | 50 | 80 | 230 |

Example from Jurafsky and Martin

# Example (cont'd)

| Referent | Phrases | Value |
|---|---|---|
| John | {John} | 155 |
| Integra | {a beautiful Acura Integra} | 140 |
| dealership | {the dealership} | 115 |

# Example (cont'd)

| Referent | Phrases | Value |
|----------|---------|-------|
| John | {John, he[1]} | 465 |
| Integra | {a beautiful Acura Integra} | 140 |
| dealership | {the dealership} | 115 |

# Example (cont'd)

| Referent | Phrases | Value |
|----------|---------|-------|
| John | {John, he[1]} | 465 |
| Integra | {a beautiful Acura Integra, it} | 420 |
| dealership | {the dealership} | 115 |

# Example (cont'd)

| Referent | Phrases | Value |
|----------|---------|-------|
| John | {John, he[1]} | 465 |
| Integra | {a beautiful Acura Integra, it} | 420 |
| Bill | {Bill} | 270 |
| dealership | {the dealership} | 115 |

# Example (cont'd)

| Referent | Phrases | Value |
|----------|---------|-------|
| John | {John, he[1]} | 232.5 |
| Integra | {a beautiful Acura Integra, it[1]} | 210 |
| Bill | {Bill} | 135 |
| dealership | {the dealership} | 57.5 |

# NLP