Discourse: Reference

Ling571
Deep Processing Techniques for NLP
March 3, 2014

Roadmap

- Discourse
  - Motivation

  - Dimensions of Discourse

  - Coherence & Cohesion

  - Coreference

What is a Discourse?

- Discourse is:
  - Extended span of text

  - Spoken or Written

  - One or more participants

  - Language in Use

  - Goals of participants
    - Processes to produce and interpret

Why Discourse?

- Understanding depends on context
  - Referring expressions: it, that, the screen
  - Word sense: plant
  - Intention: Do you have the time?

- Applications: Discourse in NLP
  - Question-Answering
  - Information Retrieval
  - Summarization
  - Spoken Dialogue
  - Automatic Essay Grading

Reference Resolution

U: Where is A Bug's Life playing in Summit?
S: A Bug's Life is playing at the Summit theater.
U: When is it playing there?
S: It's playing at 2pm, 5pm, and 8pm.
U: I'd like 1 adult and 2 children for the first show.
    How much would that cost?

- Knowledge sources:
  - Domain knowledge
  - Discourse knowledge
  - World knowledge

Coherence

- *First Union Corp. is continuing to wrestle with severe problems. According to industry insiders at PW, their president, John R. Georgius, is planning to announce his retirement tomorrow.*
- Summary*:*
- *First Union President John R. Georgius is planning to announce his retirement tomorrow.*

- Inter-sentence coherence relations:
  - Second sentence: main concept (nucleus)
  - First sentence: subsidiary, background

Different Parameters of Discourse

- Number of participants
  - Multiple participants -> Dialogue

- Modality
  - Spoken vs Written

- Goals
  - Transactional (message passing) vs Interactional (relations,attitudes)
  - Cooperative task-oriented rational interaction

Spoken vs Written Discourse

- Speech

- Written text

Spoken vs Written Discourse

- Speech
  - Paralinguistic effects
    - Intonation, gaze, gesture
  - Transitory
  - Real-time, on-line

  - Less "structured"
    - Fragments
    - Simple, Active, Declarative
    - Topic-Comment
    - Non-verbal referents
    - Disfluencies
      - Self-repairs
      - False Starts
      - Pauses

- Written text
  - No paralinguistic effects

  - "Permanent"
  - Off-line. Edited, Crafted

  - More "structured"
    - Full sentences
    - Complex sentences
    - Subject-Predicate
    - Complex modification
    - More structural markers
    - No disfluencies

Spoken vs Written: Representation

- Spoken "text" "same" if:
  - Recorded (Audio/Video Tape)

  - Transcribed faithfully
    - Always some interpretation

  - Text (normalized) transcription
    - Map paralinguistic features
    - e.g. pause = ·,+,++
    - Notate accenting, pitch

- Written text "same" if:
  - Same words

  - Same order

  - Same punctuation (headings)

  - Same lineation

Coherence Relations

- John hid Bill's car keys. He was drunk.
- ?? John hid Bill's car keys. He likes spinach.

- Why odd?
  - No obvious relation between sentences
    - Readers often try to construct relations

- How are first two related?
  - Explanation/cause

- Utterances should have meaningful connection
  - Establish through **coherence relations**

Entity-based Coherence

- *John went to his favorite music store to buy a piano.*
- *He had frequented the store for many years.*
- *He was excited that he could finally buy a piano.*

- VS
  - *John went to his favorite music store to buy a piano.*
  - *It was a store John had frequented for many years.*
  - *He was excited that he could finally buy a piano.*
  - *It was closing just as John arrived.*

- Which is better? Why?
  - 'about' one entity vs two, focuses on it for coherence

Reference Resolution

- Match referring expressions to referents

- Syntactic & semantic constraints

- Syntactic & semantic preferences

- Reference resolution algorithms

Reference

- Queen Elizabeth set about transforming her husband, King George VI, into <u>a viable monarch</u>. Logue, a renowned speech therapist, was summoned to help the King overcome his <u>speech impediment</u>…

Referring expression: (refexp)
- Linguistic form that picks out entity in some model
- That entity is the "referent"
    - When introduces entity, "evokes" it
    - Set up later reference, "antecedent"
- 2 refexps with same referent "co-refer"

Reference (terminology)

- Queen Elizabeth set about transforming her husband, King George VI, into <u>a viable monarch</u>. Logue, a renowned speech therapist, was summoned to help the King overcome his <u>speech impediment</u>…

- Anaphor:
  - Abbreviated linguistic form interpreted in context
    - Her, his, the King
  - Refers to previously introduced item ("accesses")
    - Referring expression is then anaphoric

Referring Expressions

- Many alternatives:
  - Queen Elizabeth, she, her, the Queen, etc
  - Possible correct forms depend on discourse context
    - E.g. she, her presume prior mention, or presence in world

Referring Expressions

- Many alternatives:
  - Queen Elizabeth, she, her, the Queen, etc
  - Possible correct forms depend on discourse context
    - E.g. she, her presume prior mention, or presence in world

- Interpretation (and generation) requires:
  - Discourse Model with representations of:
    - Entities referred to in the discourse
    - Relationships of these entities
  - Need way to construct, update model
  - Need way to map refexp to hearer's beliefs