CSE842: Natural Language Processing

Lecture 17: Computational Discourse

Reminder

- Project proposal due on March 30 (less than 2 week)
- Survey and presentation starts on March 30.

Pragmatics and Discourse

- Pragmatics: context dependent meaning
- Discourse: anything longer than a single utterance or sentence, a group of sentences
  - Monologue
  - Dialogue:
    - May be multi-party
    - May be human-machine
- Topics
  - Discourse segmentation
  - **Text coherence**
  - **Reference resolution**

Text Coherence

(a) John hid Bill's car keys. He was drunk.
(b) *John hid Bill's car keys. He likes spinach.

Coherence: the meaning relations between two textual units.

Example of relations (Hobbs'79)

Result: The Tin Woodman was caught in the rain. His joints rusted.

Explanation: John hid Bill's car keys. He was drunk.

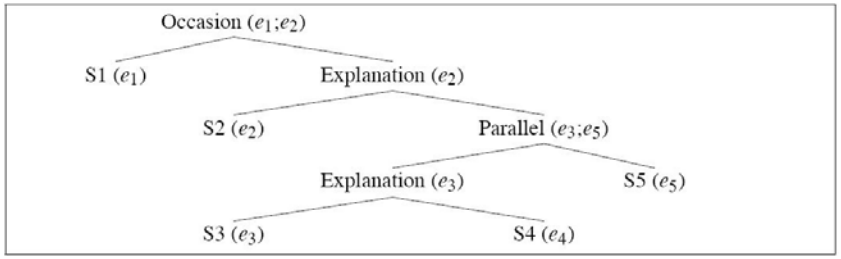
Parallel: The Scarecrow wanted some brains. The Tin woodman wanted a heart.

Elaboration: Dorothy was from Kansas. She lived in the midst of the great Kansas prairies.

Occasion: Dorothy picked up the oil-can. She oiled the Tin Woodman's joints

Text Coherence

S1: John went to the bank to deposit his paycheck
S2: he then took a train to Bill's car dealership
S3: He needed to buy a car.
S4: The company he works for now isn't near any public transportation.
S5: He also wanted to talk to Bill about their softball league.

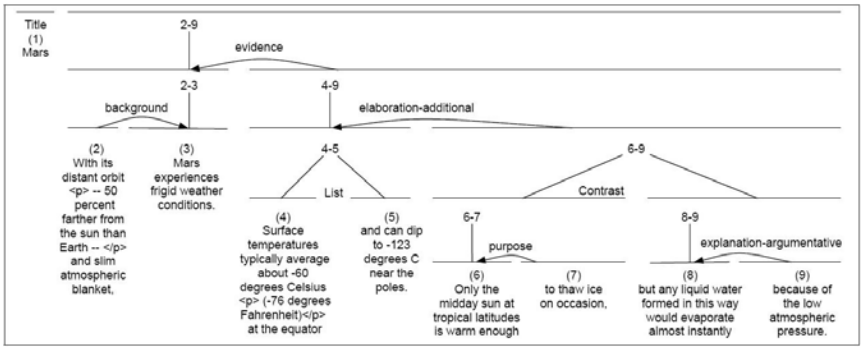


Other Discourse Theories

- Rhetorical Structure Theory (Mann and Thompson, 1988)
- Grosz and Sidner's attention, intention and structure of discourse (1986)

Rhetorical Structure Theory (RST)

- 23 rhetorical relations that can hold between spans of text (e.g., evidence, elaboration, contrast, etc.)
  - Nucleus: more central to the writer's purpose and interpretable independently;
  - Satellite: less central and only interpretable with respect to the nucleus.



Discourse Parsing

- Tasks:
  - Shallow processing: coherence relation assignment
  - Deep processing: extracting a tree or graph representing an entire discourse
- Rule-based or machine learning approaches
  - Take into consideration of *discourse markers*: although, but, because, for example, and, etc.
  - Other lexical features (e.g., I don't want a truck; I'd prefer a convertible.)
- Remain unsolved research problems.
- Resources:
  - RST Treebank
  - Penn Discourse Treebank

Reference Resolution

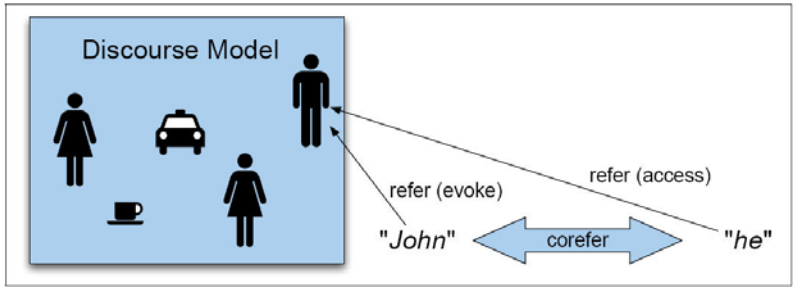
Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based financial-services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.

- **Reference resolution**: process of determining what entities are referred to by which linguistic expressions.
- **Referring expressions**: *Victoria Chen, her, she, 37-year-old*
- **Referents**: the person named Victoria Chen
- **Corefer**: two or more referring expressions that are used to refer to the same entity. *Victoria Chen, she, 37-year-old, etc.*
- **Antecedent**: a referring expression (e.g., *Victoria Chen*) that licenses the use of another (e.g., *she*)
- **Anaphors**: Reference to an entity that has been previously introduced into the discourse.

Reference Resolution

- Can we use referring expressions freely?
  - Of course not. The type of referring expressions depend on the discourse context.
  - Each type of referring expressions encodes different signals about the place that the speaker believes the referent occupies within the hearer's mental model.
- A ***discourse model*** contains representations of the entities that have been referred to in the discourse and their relations as a part of hearer's mental model.
  - Methods that construct such models
  - Methods for mapping between the signals encoded by the referring expressions and the hearer's mental model
- Two fundamental operations
  - **Evoke**: when a referent is first mentioned, a representation for it is evoked into the model.
  - **Access**: Upon subsequent mention, this representation is accessed from the model.

Reference Resolution



- Two tasks:
  - Pronominal anaphora resolution
  - Coreference resolution

Types of Reference

- Entities, concepts, places, propositions, events, ...

  According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.
  - But **that** turned out to be a lie. (a speech act)
  - But **that** was false. (proposition)
  - **That** struck me as a funny way to describe the situation. (manner of description)
  - **That** caused Sue to become rather poor. (event)
  - **That** caused them both to become rather poor. (combination of several events)

Types of Referring Expressions

- **Indefinite noun phrases**: introduce entities that are new
  I saw an Acura Integra today
  I saw this awesome Acura Integra today.
- **Definite noun phrases**: refers to entities that are identifiable
  I saw an Acura Integra today. The Integra was white and needed to be washed.
  The Indianapolis 500 is the most popular car race in the US.
- **Pronoun**: refer to entities that have a high salience.
  - Usually referents were introduced no further than one or two sentences.
- **Demonstratives**: this and that
  I bought an Integra yesterday. It's similar to the one I bought five years ago. That one was really nice, but I like this one even better.
- **Names**: can refer to both new and old entities in the discourse

Inferrables

- A referring expression does not refer to an entity that has been explicitly evoked in the text, but instead one that is inferentially related to an evoked entity (bridging inference)
  - *I almost bought an Acura Integra today, but **a door** had a dent and **the engine** seemed noisy.*
  - *Mix the flour, butter, and water. Kneed **the dough** until smooth and shiny.*

Discontinuous Sets

- Entities **evoked** together but mentioned in different sentence or phrases
  - John and Mary love their Acuras. **They** drive **them** all the time.
  - John has an Acura, and Mary has a Mazda. **They** drive **them** all the time.

Generics and Non-referential

- Generic references
  - *I'm interested in buying a Mac laptop. **They** are very stylish.*
  - *In March in Boulder **you** have to wear a jacket.*
- Non-referential
  - ***It** was Frodo who carried the ring*
  - ***It** was good that Frodo carried the ring*

Pronominal Anaphora Resolution

- Task: given a pronoun (e.g., he, they, it) and the previous context, identify the antecedent of the pronoun in this context
- Constrains that filter potential referents
  - Number agreement
    John's parents like opera. John hates it/John hates them.
  - Person agreement
    differences between three forms of person
  - Gender agreement
    John has a Porsche. He/it is attractive.
  - Syntactic constraints: reflexives
    John bought himself a new Volvo. (himself = John)
    John bought him a new Volvo (him = not John)

Preferences in Pronoun Interpretation

- Recency: more recent, more salient
  *The doctor found an old map in the captain's chest. Jim found an even older map hidden on the shelf. **It** described an island.*
- Grammatical Role: subject position is more salient
  *Billy Bones went to the bar with Jim Hawkins. **He** called for a glass of rum.*
- Repeated Mention:
  *Billy Bones had been thinking about a glass of run ever since the pirate ship docked. He hobbled over to the Old parrot bar. Jim Hawkins went with him. **He** called for a glass of rum.*
- Parallelism:
  *Long John Silver went with Jim to the Old parrot. Billy Bones went with **him** to the Old Anchor Inn.*
- Verb Semantics:
  *John telephoned Bill. **He** lost the laptop*
  *John criticized Bill. **He** lost the laptop*
- Selectional Restrictions:
  *John parked his car in the garage after driving **it** around for hours.*

More Challenges

- Metaphor
  John bought a new Acura. It drinks gasoline like you would not believe.
- World knowledge
  John parked his Acura in the garage. It is incredibly messy, with old bike and car parts lying around everywhere
  John parked his Acura in Beverly Hill. It is incredibly messy, with old bike and car parts lying around everywhere

Reference Resolution

- Given these types of constraints, can we construct an algorithm that will apply them such that we can identify the correct referents of anaphors and other referring expressions better than chance?
- Three approaches:
  - Hobbs '78: syntax tree-based referential search
  - **Centering (Grosz, Joshi, Weinstein, '95 and various): discourse-based search**
  - Machine learning approaches (e.g., log-linear Model, Naïve Bayes)

Centering Theory

- There is a single entity being "centered" on at any given point in the discourse which is to be distinguished from all other entities that have been evoked (Grosz et al., 1995)
- Given an utternace U_n,
  - **Forward-looking centers** $C_f(U_n)$: ordered list of potential focii referred to in U_n
    - E.g., C_f ordered: subj<exist. Pred. nom<obj<indobj-oblique<dem. advPP
  - **Preferred center** $C_p(U_n)$: the highest ranked forward-looking center
  - **Backward-looking center** $C_b(U_n)$: current focus after U_n interpreted. $C_b(U_{n+1})$ is the most highly ranked element of $C_f(U_n)$ mentioned in U_{n+1}. $C_b(U_0)$ is undefined.
- Entity-based coherence: "Centers" are the cohesive devices that link utterances together to contribute to a coherent discourse.

Centering-based Algorithm for Pronoun Resolution

- Transitions from U_n to U_{n+1}

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

- Rules:
  - Rule 1: If any element of $C_f(U_n)$ is realized by a pronoun in utterance U_{n+1}, then $C_b(U_{n+1})$ must be realized as a pronoun also
  - Rule 2: Preference: Continue > Retain > Smooth-Shift > Rough-Shift

Centering-based Algorithm for Pronoun Resolution

- Process
  - Generate C_b-C_f combinations for all possible reference assignments
  - Filter by constraints (syntactic contraints, selectional restrictions,…)
  - Rank by preference among transition orderings

Example

U_1:George saw a mouse on the floor.
U_2:He showed it to Harry.
U_3:He caught it.

- Grammatical role hierarchy
  - $C_f(U_1)$: {George,mouse,floor}
  - $C_p(U_1)$: George
  - $C_b(U_1)$: undefined
- Assume it=mouse
  - $C_f(U_2)$: {George, mouse,Harry}
  - $C_p(U_2)$: George
  - $C_b(U_2)$: George
  - Continue ($C_p(U_2)$=$C_b(U_2)$; $C_b(U_1)$ undefined

Assume it=floor

- $C_f(U_2)$: {George,floor,Harry}
- $C_p(U_2)$: George
- $C_b(U_2)$: George
- Continue ($C_p(U_2)=C_b(U_2)$; $C_b(U_1)$ undefined)

- What did George show to Harry?
  - How to break a tie? Ordering of previous C_f list

U_3: He caught it.

Assume he=George, it=mouse

- $C_f(U_3)$: {George,mouse}
- $C_p(U_3)$: George
- $C_b(U_3)$: George
- Continue ($C_p(U_3)=C_b(U_3)$; $C_b(U_3)=C_b(U_2)$)

Assume he=Harry, it=mouse

- C_f(U3): {Harry,mouse}
- C_p(U3): Harry
- $C_b(U_3)$: mouse
- Rough-shift ($C_p(U_3) \neq C_b(U_3)$; $C_b(U_3) \neq C_b(U_2)$)

- So, who caught the mouse?

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.

- Task: decide whether any pairs of noun phrases co-refer and create coreference chains

- *1. {Victoria Chen, Chief Financial Officer of Megabucks Banking Corp, the 37-year-old, the Denver-based financial-services company's president, she}*
- *2. {Megabucks Banking Corp, the Denver-based financial-services company, Megabucks }*

Coreference Resolution

- Why is it important?
  - Coreference is a pervasive phenomenon in natural language.
  - The problem lies at the intersection of syntax, semantics and discourse.
  - Coreference resolution is necessary for many NLP applications, such as information extraction, question answering, and summarization.

Coreference Resolution

- Most state-of-the-art approaches use machine learning.
  - A gold standard corpus labeled with coreference chains is used for training.
  - Supervised learning techniques are used to create a classifier that can predict whether two noun phrases co-refer.
  - A final clustering stage merges the pairwise coreference decisions made by the classifier into coreference chains.
- Active research field, good progress has been made.
  - Current systems achieve about 60-70% accuracy (for newswire articles)
  - Looking at semi-supervised and unsupervised approaches.

Typical features used in Classification

- **String Matching Features**
  A lot of resolutions can be made solely on the basis of string-based comparisons!
  - **Exact Matching:** exact match of NPs or head nouns.
  - **String Matching:** substring and partial matching of the NPs. E.g., "Apple" and "Apple Computer Inc."
  - **Word Overlap:** word overlap comparisons of the NPs. E.g., "John F. Kennedy" and "President John Kennedy"
  - **Edit Distance:** edit distance measures between the NPs. E.g., "Creutzfeldt-Jakob Disease" and "Creutzefeldt-Jacob Disease"
- **Proximity:** sentence and paragraph-based distance measures between the NPs.
- **Aliases:** titles, acronyms, etc.
- **Number:** number agreement (singular/plural)
- **Gender:** gender agreement (singular/plural)
- **Syntactic:** syntactic properties of the noun phrases
- **Appositive:** appositive recognition
- **Semantic:** animacy and semantic compatibility comparisons.
- **NER:** Named Entity Recognition

Evaluation of Coreference Resolution

- The system creates a set of hypothesis chains
  - i.e., A-B-C-D, E-F-G, H-I
    - Where A,B, etc. are mentions in the discourse
    - Mentions can be in at most 1 chain
- Manually annotated reference chains (i.e., gold standard)
- Compare hypothesis chains with reference chains
- Metrics
  - MUC Score
  - B-CUBED algorithm
  - CEAF (later in the presentation session)

MUC Score

- Proposed in 1995 for use in MUC-6
- Link-based evaluation
  - Compare the number of links from the hypothesis chains to the number of links from the reference chains.
  - Precision: the percentage of links out of the total number of links from the hypothesis chains that are correctly identified.
  - Recall: the percentage of links out of the total number of links from the reference chains that are correctly identified.
- Example:

  Hypothesis: A-B-C, D-E, F,G; Reference A-B, C-D-E-G, F

  Precision: 2/3, recall: ½
- Drawbacks:
  - Chains with a single mention are ignored.
  - Bias towards systems that return longer chains.

B-CUBED Algorithm

Mention-based evaluation (here entity means mention)

$$\text{Precision}_i = \frac{\textit{number of correct elements in the output chain containing entity}_i}{\textit{number of elements in the output chain containing entity}_i}$$

$$\text{Recall}_i = \frac{\textit{number of correct elements in the output chain containing entity}_i}{\textit{number of elements in the truth chain containing entity}_i}$$

$$\text{Precision} = \sum_{i=1}^{N} w_i \times \text{Precision}_i$$

$$\text{Recall} = \sum_{i=1}^{N} w_i \times \text{Recall}_i$$

Example:
Hypothesis: A-B-C, D-E, F,G; Reference A-B, C-D-E-G, F
Precision: 1/7(2/3+2/3+1/3+1+1+1+1) = 17/21
Recall: 1/7(1+1+1/4+2/4+2/4+1+1/4) = 9/14