Speech and Language Processing

Discourse: Anaphora Resolution and Coherence

Dan Jurafsky

Thanks to Diane Litman, Andy Kehler, Jim Martin!!! This material is from J+M Chapter 18, written by Andy Kehler, slides inspired by Diane Litman + Jim Martin

LING 138/238 Autumn 2004
Outline

• Reference
  – Kinds of reference phenomena
  – Constraints on co-reference
  – Preferences for co-reference
  – The Lappin-Leass’ algorithm for coreference

• Coherence
  – Hobbs’ coherence relations
  – Rhetorical Structure Theory
Reference
Reference Resolution

- Two examples:
  - John went to Bill’s car dealership to check out an Acura Integra. He looked at it for half an hour
  - I’d like to get from Boston to San Francisco, on either December 5th or December 6th. It’s ok if it stops in another city along the way

- What is the target of “it”? 
  - First example: two possible targets
  - Second example: where is the target?
Why reference resolution?

- **Conversational Agents**: Airline reservation system needs to know what “it” refers to in order to book correct flight.
- **Information Extraction**: 
  
  *First Union Corp.* is continuing to wrestle with severe problems unleashed by a botched merger and a troubled business strategy. According to industry insiders at *Paine Webber*, their president, John R. Georgius, is planning to retire by the end of the year.
Some terminology

- John went to Bill’s car dealership to check out an Acura Integra. He looked at it for half an hour
- Reference: process by which speakers use words John and he to denote a particular person
  - Referring expression: John, he
  - Referent: the actual entity (but as a shorthand we might call “John” the referent).
  - John and he “corefer”
  - Antecedent: John
  - Anaphor: he
- Cataphora: pronoun before the referent
  - Before he bought it, John checked over the Integra very carefully
Many types of reference

• (after Webber, ‘91)

• 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)

• But we focus on references to entities
Reference Phenomena

- Indefinite noun phrases: new to hearer
  - *I saw an Acura Integra today*
  - *Some Acura Integras were being unloaded…*

- Definite noun phrases: identifiable to hearer because
  - Mentioned:
    *I saw an Acura Integra today*. *The Integra was white*
  - Identifiable from beliefs (common knowledge):
    *The Indianapolis 500*
  - Inherently unique:
    *The fastest car in Indianapolis 500…*
Reference Phenomena

• Pronouns

• Simple example:
  
  – *I saw an Acura Integra today. It was white* 

• Referent salience, in case of discourse:
  
  1. *John went to Bob’s party, and parked next to a beautiful Acura Integra*
  2. *He went inside and talked to Bob for more than an hour.*
  3. *Bob told him that he recently got engaged.*
  4. a) *He also said that he bought it yesterday.*
    b) *He also said that he bought the Acura yesterday.*
Reference phenomena

- Demostratives
  - 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

- A non-pronominal anaphora
  - I saw no less that 6 Acura Integra today. I want one
    - … one (of them)
Reference phenomena

• Inferrable
  – *I almost bought an Acura Integra today, but the engine seemed noisy.*

• Pronominal
  – *I saw no less than 6 Acura Integras today. They are the coolest cars.*
Pronominal Reference Resolution

- Given a pronoun, find the reference
- Constraints to leverage
  - Hard constraints on reference
  - Soft constraints on reference
- Algorithms which use/don’t use these constraints
Hard constraints: syntax

• Number agreement
  – *John has an Acura. They?? are red
  – John has an Acura. It is red

• Person and case agreement
  – *John and Mary have Acuras. We love them???
    (who/what???)
  – John and I have Acuras. We love them.

• Gender agreement
  – John has an Acura. He / it is attractive.

• Syntactic constraints
  – John bought himself a new Acura (himself == John)
  – John bought him a new Acura (him /= John)
Soft constraints

Pronoun Interpretation Preferences

• Selectional Restrictions
  – John parked his Acura in the garage. He had driven it around for hours.

• Recency
  – John has an Integra. Bill has a Legend. Mary likes to drive it.
Soft constraints

Pronoun Interpretation Preferences

• Grammatical Role: Subject preference
  – John went to the Acura dealership with Bill. He bought an Integra.
  – John and Bill went to the Acura dealership. He bought an Integra
Soft constraints

Repeated Mention preference

- John needed a car to get to his new job. He decided that he wanted something sporty. Bill went to the Acura dealership with him. He bought an Integra.
Soft constraints

Parallelism Preference

• Same structure
  – Mary went with Sue to the Acura dealership. Sally went with her to the Mazda dealership.

• But… with similar structure
  – Mary went with Sue to the Acura dealership. Sally told her not to buy anything.
Soft constraints

Verb Semantics Preferences

• *John* telephoned *Bill*. *He* lost the pamphlet on Acuras.
• *John* criticized *Bill*. *He* lost the pamphlet on Acuras.
• Implicit causality
  – Implicit cause of criticizing is object.
  – Implicit cause of telephoning is subject.
Algorithms for pronoun anaphora resolution

• Knowledge-rich approach
  – Syntactic-based: Hobbs’ algorithm
  – Discourse-based: Centering Theory
  – Hybrid approaches: Lappin and Leas
  – Corpus-based: Charniak, Hale, and Ge

• Knowledge-poor approach
  – Machine Learning

• We’ll see the Lappin&Leas algorithms
Lappin and Leass

- Implements only recency and syntactic preferences
- Two steps
  - Discourse model update
    - When a new noun phrase is encountered, add a representation to discourse model with a salience value
    - Modify saliences.
  - Pronoun resolution
    - Choose the most salient antecedent
Salience Factors and Weights

- Salience given to NP

<table>
<thead>
<tr>
<th>Factor</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Recency</td>
<td>100</td>
</tr>
<tr>
<td>2 Subject emphasis</td>
<td>80</td>
</tr>
<tr>
<td>3 Existential emphasis</td>
<td>70</td>
</tr>
<tr>
<td>4 Accusative (direct object) emphasis</td>
<td>50</td>
</tr>
<tr>
<td>5 Ind. Obj and oblique emphasis</td>
<td>40</td>
</tr>
<tr>
<td>6 Non-adverbial emphasis</td>
<td>50</td>
</tr>
<tr>
<td>7 Head noun emphasis</td>
<td>80</td>
</tr>
</tbody>
</table>
Information leveraged

• NP *saliences*
  – according to recency of the NP (1)
  – according to syntactical role (2-7)
    • 6. A Demarcated adverbial PP: adverbial phrase introduced by coma or adverb ("his") → not a good candidate → give 50 to NP that are **not** a demarcated adverbial PPs: non-adverbial emphasis)
    • 7. Head noun emphasis: give 80 if NP is **not** part of a larger NP
  – cut in half after each sentence is processed
Syntactic roles

- Salience factors 2-6: Syntactic role preference
  - Subject > existential predicate nominal > object > indirect object > demarcated adverbial PP

- Examples for 2-5
  - *An Acura Integra is parked in the lot* (subject)
  - *There is an Acura Integra parked in the lot* (existential pred. nominal)
  - *John parked an Acura Integra in the lot* (object)
  - *John gave his Acura Integra a bath* (indirect obj)

- Add salience if 6 holds (*not* part of demarcated adverbial PP):
  - *Inside his Acura Integra, John showed Susan his new CD player*
    (here, it is part of demarcated adverbial PP → no salience)

- Add salience if 7 holds (*not* part of larger NP):
  - *The owner’s manual for an Acura Integra is on John’s desk*
Lappin and Leass Algorithm

• Collect the potential referents (up to 4 sentences back)
• Remove potential referents that do not agree in number or gender with the pronoun
• Remove potential references that do not pass syntactic coreference constraints
• Compute total salience value of referent from all factors, including, if applicable:
  – role parallelism (+35)
  – or cataphora (-175).
• Select referent with highest salience value. In case of tie, select closest.
Example

• John saw a beautiful Acura Integra at the dealership. He showed it to Bob. He bought it.

Sentence 1:

<table>
<thead>
<tr>
<th>Referent</th>
<th>1 Recency</th>
<th>2 Subject</th>
<th>3 Exist</th>
<th>4 Object</th>
<th>5 Ind-object</th>
<th>6 Non-adv</th>
<th>7 Head N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>100</td>
<td>80</td>
<td></td>
<td></td>
<td>50</td>
<td>80</td>
<td></td>
<td>310</td>
</tr>
<tr>
<td>Integra</td>
<td>100</td>
<td></td>
<td>50</td>
<td></td>
<td>50</td>
<td>80</td>
<td></td>
<td>280</td>
</tr>
<tr>
<td>dealership</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>80</td>
<td></td>
<td>230</td>
</tr>
</tbody>
</table>
After sentence 1

- Cut all values in half

<table>
<thead>
<tr>
<th>Referent</th>
<th>Phrases</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>{John}</td>
<td>155</td>
</tr>
<tr>
<td>Integra</td>
<td>{a beautiful Acura Integra}</td>
<td>140</td>
</tr>
<tr>
<td>dealership</td>
<td>{the dealership}</td>
<td>115</td>
</tr>
</tbody>
</table>
Sentence 2: *He showed it to Bob*

- **He** specifies male gender
- So Step 2 reduces set of referents to only **John**.
  - Referent for *He* found.
- Now update discourse model:
  - **He** in current sentence (recency=100), subject position (=80), not adverbial (=50) not embedded (=80), so add 310:

<table>
<thead>
<tr>
<th>Referent</th>
<th>Phrases</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>{John, he$_1$}</td>
<td>155+310</td>
</tr>
<tr>
<td>Integra</td>
<td>{a beautiful Acura Integra}</td>
<td>140</td>
</tr>
<tr>
<td>dealership</td>
<td>{the dealership}</td>
<td>115</td>
</tr>
</tbody>
</table>
Sentence 2: *He showed it to Bob*

- Need to add “it”, which can be “Integra” or “dealership”.
- Need to add weights:
  - Parallelism: “it” and “Integra” are objects (“dealership” is not), so +35 for “Integra”
  - Integra: 175, dealership: 115
    → pick Integra
      - Referent for *it* found.
- Update discourse model:
  - “it” is object, gets 100+50+50+80=280:
Sentence 2: *He showed it to Bob*

<table>
<thead>
<tr>
<th>Referent</th>
<th>Phrases</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>{John, he₁}</td>
<td>465</td>
</tr>
<tr>
<td>Integra</td>
<td>{a beautiful Acura Integra, it₁}</td>
<td>140+280</td>
</tr>
<tr>
<td>dealership</td>
<td>{the dealership}</td>
<td>115</td>
</tr>
</tbody>
</table>
Sentence 2:  
*He showed it to Bob*

- Bob is a new referent
- Update discourse model:
  - Bob is oblique argument, weight is 100+40+50+80=270

<table>
<thead>
<tr>
<th>Referent</th>
<th>Phrases</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>{John, he}_1</td>
<td>465</td>
</tr>
<tr>
<td>Integra</td>
<td>{a beautiful Acura Integra, it}_1</td>
<td>420</td>
</tr>
<tr>
<td>Bob</td>
<td>{Bob}</td>
<td>270</td>
</tr>
<tr>
<td>dealership</td>
<td>{the dealership}</td>
<td>115</td>
</tr>
</tbody>
</table>
Sentence 3:  
*He bought it*

- Drop weights in half:

<table>
<thead>
<tr>
<th>Referent</th>
<th>Phrases</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>{John, he$_1$}</td>
<td>232.5</td>
</tr>
<tr>
<td>Integra</td>
<td>{a beautiful Acura Integra, it$_1$}</td>
<td>210</td>
</tr>
<tr>
<td>Bob</td>
<td>{Bob}</td>
<td>135</td>
</tr>
<tr>
<td>dealership</td>
<td>{the dealership}</td>
<td>57.5</td>
</tr>
</tbody>
</table>

He$_2$ will be resolved to John, and it$_2$ to Integra
Evaluation

- Referential Rate (Byron, 2001)
- RR = C / (T+E)

  C: # pronouns correctly resolved
  T: all referential pronouns
  E: all excluded referential pronouns
Coherence
Text Coherence

• John hid Bill’s car keys. He was drunk
• ??John hid Bill’s car keys. He likes spinach

What makes a Discourse coherent?

• Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances
• Do you have a discourse?
  – Usually not
  – In general utterances, when juxtaposed, will not exhibit coherence
What makes a text coherent?

- Appropriate use of *coherence relations* between subparts of the discourse
  - rhetorical structure
- Appropriate *sequencing of subparts* of the discourse
  - discourse/topic structure
- Appropriate use of *referring expressions*
Hobbs 1979 Coherence Relations

Result

• Infer that the state or event asserted by S0 causes or could cause the state or event asserted by S1.

• *John bought an Acura. His father was not happy.*

• (S0) *John bought an Acura* as a direct consequence (S1) *His father was not happy.*
Explanation

• Infer that the state or event asserted by S1 causes or could cause the state or event asserted by S0

• *John hid Bill’s car keys. He was drunk*

• *(S0) John hid Bill’s car keys. because *(S1) He was drunk*
Parallel

- Infer proposition $P(a_1, a_2, \ldots)$ from the assertion of S0 and $P(b_1, b_2, \ldots)$ from the assertion of S1, where $a_i$ and $b_i$ are similar, for all i.

- John bought an Acura. Bill leased a BMW.

- \(S0\) John bought an Acura.
  \(\Rightarrow\) Possession(Person, Car)

- \(S1\) Bill leased a BMW.
  \(\Rightarrow\) Possession(Person, Car)
Elaboration

• Infer the same proposition $P$ from the assertions of $S_0$ and $S_1$:

• *John bought an Acura this weekend. He purchased a beautiful new Integra for 20 thousand dollars at Bill’s dealership on Saturday afternoon.*

• *(S0) John bought an Acura this weekend. *(S1) He purchased a beautiful new Integra ... S1 is just a more precise version of S0*
Occasion

• A change of state can be inferred from the assertion of $S_0$, whose **final state** can be inferred from $S_1$, or vice versa.

• *Dorothy picked up the oil-can. She oiled the Tin Woodman’s joints.*

• (S0) *Dorothy picked up the oil-can.* and because of this, at the end (S1) *She oiled the Tin Woodman’s joints.*
An example

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

He needed to buy a car. The company he works for now isn’t near any public transportation. He also wanted to talk to Bill about their softball league.
He needed to buy a car. The company he works for now isn't near any public transportation.

He also wanted to talk to Bill about their softball league.
The discourse structure

He needed to buy a car.

The company he works for now isn't near any public transportation.
Rhetorical Structure Theory

• One theory of discourse structure, based on identifying relations between segments of the text
  – Nucleus/satellite notion encodes asymmetry
  – Some rhetorical relations:
    • Elaboration (set/members, class/instances, whole/part…)
    • Contrast: multinuclear
    • Condition: Satellite presents precondition for N
    • Purpose: Satellite presents goal of the activity in N
    • Background: Satellite gives context for interpreting N
    • Attribution: multinuclear
    • List: multinuclear
    • Evidence: (see in the following)
In the original (Mann & Thompson 1987) formulation. An RST relation is formally defined by

- a set of constraints on the Nucleus (N) and satellite (S),
- having to do with the goals and the beliefs of the writer (W) and reader (R),
- and by the effect on the reader (R)
Relations

- **A sample definition**
  - Relation: *evidence*
  - Constraints on Nucleus: Reader might not believe Nucleus to a degree satisfactory to Writer
  - Constraints on Satellite: Reader will believ Satellite or will find it credible
  - Constraints on Nucleus+Satellite: Reader’s comprehending Satellite increases Reader’s belief of Nucleus
  - Effects: Reader’s belief of Nucleus is increased

N: Kevin must be here

S: His car is parked outside
An example

• An Example:

**Mars**
With its distant orbit –50 percent farther from the sun than Earth– and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 °C at the equator and can dip to -123 °C near the poles. Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, but any liquid water formed in this way would evaporate almost instantly because of the low atmospheric pressure.
A discourse tree (Marcu 2000)

Figure 21.4  A discourse tree for the *Scientific American* text in (21.23), from Marcu (2000a). Note that asymmetric relations are represented with a curved arrow from the satellite to the nucleus.
Automatic Rhetorical Structure Labeling

• Supervised machine learning
  – Get a group of annotators to assign a set of RST relations to a text
  – Extract a set of surface features from the text that might signal the presence of the rhetorical relations in that text
  – Train a supervised ML system based on the training set
Features

• Explicit markers: *because, however, therefore, then, etc.*

• Tendency of certain syntactic structures to signal certain relations: *Infinitives are often used to signal purpose relations: Use rm to delete files.*

• Ordering

• Tense/aspect

• Intonation
Some Problems with RST

• How many Rhetorical Relations are there?
• How can we use RST in dialogue as well as monologue?
• RST does not model overall structure of the discourse.
• Difficult to get annotators to agree on labeling the same texts