Anaphora and Discourse

Miriam Butt
January 2012
(First Version, October 2004)
Computational Semantics

Central Goals:

1) Automate the process of associating semantic representations with natural language expressions.

2) Use logical representations of natural language expressions to automate the process of drawing inferences.
Computational Semantics

Central Topics:

1) Predicate-Argument Structure: like(vincent, maria)

2) Quantifier Scope (Blackburn and Bos 1999:64)

3) Reasoning/Deduction

   Vincent knows every boxer.
   Butch is a boxer.

4) Lexical Semantics: Vincent knows Butch.

   Semantic Nets, Ontologies (B&B 1999:149)
Computational Semantics

Central Topics (cont.):

5) Pronoun Resolution (B&B 1999b:62)

There are several different (bad to better) solutions to this. We’ll look at **Centering Theory**.

6) Discourse Analysis (Presuppositions, Conversational Maxims, Discourse Coherence, Temporal Relations). We’ll look at a few examples of this kind of work.
Some Applications

• **Summarization**

  First Union is continuing to wrestle with severe problems. According to industry insiders, their president, John R. Georgius, is planning to announce his retirement tomorrow.  

  First Union president John R. Georgius is planning to announce his retirement tomorrow.

• **Information Retrieval**

• **Evaluating text coherence**

  student essays, papers, grant applications, other texts

  (cf. Oelke et al.)
Pronoun Resolution

Pronoun Resolution is not easy: it involves a good understanding of the interaction between the syntax, semantics and pragmatics of a language.

In theoretical linguistics, the treatment of anaphora (superset of pronoun resolution) remains a tricky (=unresolved) issue.

Dalrymple (1993) provides a nice overview of the theoretical problems and solutions within LFG.
Pronoun Resolution

Hobbs (1978, 1979) and works by Stanley Peters represent some complex semantic solutions to the problem.

The formulation of DRT (Discourse Representation Theory, Kamp and Reyle 1993) based on Heim’s (1982) file-change semantics provided a new method of resolving anaphora in discourse within computational linguistics (see Bos and Blackburn 1999 for some discussion).
Temporal Anaphora

One also speaks of temporal anaphora, whereby the interpretation of the reference time (R) of a sentence depends on the reference time of the previous sentence.

Fred arrived at 10. He had gotten up at 5, taken a long shower, ....

Max fell. John pushed him.

Again, information about the discourse context is needed.
Pronoun Resolution

One approach which has been quite successful is **Centering Theory**. This approach has been pioneered at UPenn (Grosz, Sidner, Webber: see J&M 740-742 for references).

Another approach: **Mitkov’s** robust, knowledge poor algorithm (Mitkov 2002)

Neither approach relies on in-depth syntactic and semantic knowledge, but rather on formulating successful heuristics for identifying pronouns and possible antecedent NPs, and then ranking them in terms of discourse importance.
Centering Theory

Sample Discourse:

John saw a beautiful Acura Integra at the dealership. (U₁)
He showed it to Bob. (U₂)
He bought it. (U₃)

Think of each sentence as an Utterance (Un).

Task: Build up a Discourse Model and resolve the pronouns.
Centering Theory

Assumptions:

Each Utterance has a *discourse center* (broadly equivalent to the idea of topic).

This center tends to be the *preferred antecedent* for a pronoun in a following utterance.

The first utterance in a discourse has an undefined discourse center (i.e., one needs to be established “on the fly”).
Centering Theory

Definitions:

*Backward Looking Center* $(C_b)$: current center of discourse.

*Forward Looking Centers* $(C_f)$: ordered list of entities mentioned in previous utterance $(U_n)$ which are candidates for the center of discourse in the current utterance $(U_{n+1})$.

*Preferred Center* $(C_p)$ for current utterance $(U_{n+1})$: highest forward looking center $(C_f)$ in this utterance $(U_{n+1})$.
Centering Theory

Ordering of $C_f$:

Use a grammatical role hierarchy (linguistically well motivated)

subject > object > indirect object or oblique (usually a PP in English) > adjunct PP
Centering Theory

**Discourse Transitions:** Based on these definitions, one can now define a number of relations which hold between sentences and which model how successful/acceptable transitions *between* utterances are.

This discourse is not smooth:

John saw a beautiful Acura Integra at the dealership. (U₁)
Mary showed a watch to Bob. (U₂)
**He bought it.** (U₃)
Discourse Transitions

\[
\begin{align*}
C_b(U_{n+1}) = C_b(U_n) & \quad C_b(U_{n+1}) \neq C_b(U_n) \\
\text{or undefined } C_b(U_n) & \\
C_b(U_{n+1}) = C_p(U_{n+1}) & \quad \text{CONTINUE} \quad \text{SMOOTH-SHIFT} \\
C_b(U_{n+1}) \neq C_p(U_{n+1}) & \quad \text{RETAIN} \quad \text{ROUGH-SHIFT}
\end{align*}
\]

(from J&M:692)

Utterances should be linked by these transitions and rough shifts should be dispreferred.
The Centering Algorithm

**Basic Rules:**

1) If an element was realized as a pronoun, keep referring to it as a pronoun.

2) The Transition states are ordered:
   Continue > Retain > Smooth-Shift > Rough-Shift
The Centering Algorithm

**Basic Steps:**

1) Generate possible $C_b$-$C_f$ combinations.

2) Filter the possible combinations by the basic rules, morphological/syntactic constraints and whatever else one may have defined.

3) Rank by Transition Orderings
Applying the Algorithm

John saw a beautiful Acura Integra at the dealership. (U₁)
He showed it to Bob. (U₂)
He bought it. (U₃)

\[ C_f(U₁) : \{John, Integra, dealership\} \]
\[ C_p(U₁) : \{John\} \]
\[ C_b(U₁) : \{undefined\} \]
Applying the Algorithm

Possibility 1 for $U_2$:
- $C_f(U_2)$: \{John, Integra, Bob\}
- $C_p(U_2)$: \{John\}
- $C_b(U_2)$: \{John\}

Transition: Continue ($C_p(U_2) = C_b(U_2)$; $C_b(U_1)$ undefined)

Possibility 2 for $U_2$:
- $C_f(U_2)$: \{John, dealership, Bob\}
- $C_p(U_2)$: \{John\}
- $C_b(U_2)$: \{John\}

Transition: Continue ($C_p(U_2) = C_b(U_2)$; $C_b(U_1)$ undefined)
Applying the Algorithm

Possibilities 1 and 2 are equally likely in terms of the discourse transitions. We could decide to slightly prefer Possibility 1 because of the initial ordering in $U_1$.

$$C_f(U_1): \{\text{John, Integra, dealership}\}$$
Applying the Algorithm

Possibility 1 for $U_3$: $C_f(U_3): \{\text{John, Acura}\}$
$C_p(U_3): \{\text{John}\}$
$C_b(U_3): \{\text{John}\}$
Transition: Continue ($C_p(U_3) = C_b(U_3) = C_b(U_2)$)

Preferred

Possibility 2 for $U_3$: $C_f(U_3): \{\text{Bob, Acura}\}$
$C_p(U_3): \{\text{Bob}\}$
$C_b(U_3): \{\text{Bob}\}$
Transition: Smooth-Shift ($C_p(U_3) = C_b(U_3); C_b(U_3) \neq C_b(U_2)$)
Mitkov’s Algorithm

1) Examine current sentence and 2 preceding ones (if available). Look for NPs to the left of the anaphor.
2) Select from set of NPs only those with gender/number compatibility.
3) Apply **antecedent indicators** to each candidate NP and assign scores. Propose candidate with highest score.
   - if equal score, compare immediate reference score
   - if still no resolution, compare collocational score
   - if still no resolution, compare indicating verbs score
   - if still no resolution, go for most recent NP
Mitkov’s Antecedent Indicators

1) First NP gets +1 (generally topic)
2) NPs immediately following an *indicating verb* get +1
   • Examples: *assess, check, cover, define, describe*
   • Empirical evidence suggests that these NPs have high salience.
3) If an NP is repeated twice or more in paragraph, do +2. For single repetition, do +1.
4) Collocation Match: If NP has an identical collocation pattern to that of the pronoun, do +2 (weak preference).
   • Example: Press **the key** down and turn the volume up... Press **it** again.
Mitkov’s Antecedent Indicators

5) Immediate reference gets +2. Restricted to certain contexts: (You) V NP CONJ (you) V it.
   - Example: you can stand the printer up or lay it flat

6) Sequential instructions get +2
   - Example: To turn on the printer, ... To program it...

7) Term Preference: if NP is a term typical of the text genre, do +1.

8) Indefinite NPs get -1 (tend not to be antecedents).

9) NPs in PPs get -1 (tend not to be antecedents).

10) Referential distance: NPs in previous clause but same sentence +2, in previous sentence +1, etc.
An Example

Raise the original cover. Place the original face down on the original glass so that it is centrally aligned.

<table>
<thead>
<tr>
<th>Original Cover</th>
<th>Calculation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1(first NP) + 1(term preference) + 1(referential distance) = 3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Calculation</th>
<th>Score</th>
<th>Preferred</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1(first NP) + 1(lexical iteration) + 1(term preference) + 2(referential distance) = 5</td>
<td></td>
<td>Preferred</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original Glass</th>
<th>Calculation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1(term preference) - 1(PP) + 2(referential distance) = 2</td>
<td></td>
</tr>
</tbody>
</table>
## Evaluation

<table>
<thead>
<tr>
<th>Manual</th>
<th># of pronouns</th>
<th>% success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minolta Copier</td>
<td>48</td>
<td>95.8</td>
</tr>
<tr>
<td>Portable Style Writer</td>
<td>54</td>
<td>83.8</td>
</tr>
<tr>
<td>Alba Twin Recorder</td>
<td>13</td>
<td>100.0</td>
</tr>
<tr>
<td>Seagate Hard Drive</td>
<td>18</td>
<td>77.8</td>
</tr>
<tr>
<td>Haynes Car Manual</td>
<td>50</td>
<td>80.0</td>
</tr>
<tr>
<td>Sony Video Recorder</td>
<td>40</td>
<td>90.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>223</strong></td>
<td><strong>89.7</strong></td>
</tr>
</tbody>
</table>
More Discourse Factors

Text or Discourse Coherence is governed by a number of further factors:

1) Turn-Taking
2) Coherence Relations
3) Conversational Implicatures
Coherence Relations

That the flow of a discourse can seem more or less natural to us (i.e., we find some discourses “odd”) can be explained via the fact that discourses in general have structures and that these structures are governed by coherence relations (see J&M:723-729).
Coherence Relations

Some Coherence Relations proposed by Hobbs (1979):

**Result:** Infer that state or event asserted by U$_1$ could cause the state or event asserted by U$_2$.

*John bought an Acura. His father went ballistic.*

**Explanation:** Infer that state or event asserted by U$_2$ could explain/cause the state or event asserted by U$_1$.

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

**Elaboration:** Infer the same proposition P from the assertions of U$_1$ and U$_2$.

*John bought an Acura this weekend. He purchased a beautiful new Integra for $20 000 at Bill’s dealership.*
Further Concepts

• Coherence vs. Text Cohesion
  – Lexical cohesion (are words in the paragraph semantically related?)
  – Cohesion: grouping of units into a single unit
  – Coherence: meaning relation between the units

• Entity Based Coherence
  – Example from Grosz et al, p. 717 in J&M
References


References


Jurafsky, Daniel and James Martin. 2000. *Speech and Language Processing*. Prentice Hall.

