Anaphora and Discourse

Miriam Butt

January 2014

(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) Prediate-Argument Structure: like(vincent,maria)

2) Quantifier Scope (J&M 18.3)

3) Reasoning/Deduction

Vincent knows every boxer.
Butch is a boxer.
Vincent knows Butch.

4) Lexical Semantics:

Semantic Nets, Ontologies (J&M 19)

Computational Semantics

Central Topics (cont.):

5) Pronoun Resolution (J&M 21.6)

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). Well 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. $\Rightarrow$

  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 754-756 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 sucessful 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 Toyota Prius at the dealership. (U_1)

<u>He</u> showed <u>it</u> to Bob. (U_2)

<u>He</u> bought <u>it</u>. (U_3)

Think of each sentence as an Utterance (U_n).

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

Definitions:

Grammatical Role Hierarchy:

Order antecedents according to the following hierarchy.

subject > object > indirect object/oblique > PP

Centering Theory

Definitions:

$C_b(U_{n+1})$ = most highly ranked element of $C_f(U_n)$ that is mentioned in U_{n+1}

$C_f(U_n)$ = all entities in U_n

$C_p(U_n)$ = highest on $C_f(U_n)$

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 Toyota Prius at the dealership. (U_1)

Mary showed a watch to Bob. (U_2)

<u>He</u> bought <u>it</u>. (U_3)

Discourse Transitions

	$C_b(U_{n+1})=C_b(U_n)$ 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

(from J&M:740)

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 Toyota Prius at the dealership. (U_1)

He showed it to Bob. (U_2)

He bought it. (U_3)

$C_f(U_1)$: {John, Toyota, dealership}

$C_p(U_1)$: {John}

$C_b(U_1)$: {undefined}

Applying the Algorithm

John saw a beautiful Toyota Prius at the dealership. (U_1)

He showed it to Bob. (U_2)

He bought it. (U_3)

$C_f(U_1)$: {John, Toyota, dealership}

$C_p(U_1)$: {John}

$C_b(U_1)$: {undefined}

Applying the Algorithm

John saw a beautiful Toyota Prius at the dealership. (U_1)

He showed it to Bob. (U_2)

He bought it. (U_3)

$C_f(U_2)$: {John(=he), it, Bob}

$C_p(U_2)$: {John(=he)}

$C_b(U_2)$: {John(=he)}

He must be resolved to *John* on purely morphosyntactic grounds as it is the only male gendered antecedent.

Applying the Algorithm

Possibility 1 for U_2: $C_f(U_2)$: {John, it=Toyota, 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, it=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.

$C_f(U_1)$: {John, Toyota, dealership}

Possibility 1: $C_f(U_2)$: {John, it=Toyota, Bob}

Applying the Algorithm

John saw a beautiful Toyota Prius at the dealership. (U_1)

He showed it to Bob. (U_2)

He bought it. (U_3)

$C_f(U_3)$: {he, Toyota(=it)}

$C_p(U_2)$: {he}

$C_b(U_2)$: {he}

It must be resolved to *Toyota Prius* on morphosyntactic grounds as it is the only neuter gendered antecedent.

Applying the Algorithm

Possibility 1 for U_3: $C_f(U_3)$: {John(=he), Toyota(=it)}

$C_p(U_3)$: {John(=he)}

Preferred

$C_b(U_3)$: {John(=he)}

Transition: Continue ($C_p(U_3)=C_b(U_3)=C_b(U_2)$)

Possibility 2 for U_3: $C_f(U_3)$: {Bob(=he), Toyota(=it)}

$C_p(U_3)$: {Bob(=he)}

$C_b(U_3)$: {Bob(=he)}

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.

original cover
1(first NP)+1(term preference)+1(referential distance)=3

original
1(first NP)+1(lexical iteration)+1(term preference)
+2(referential distance)=5 **Preferred**

original glass
1(term preference)-1(PP)+2(referential distance)=2

Evaluation

Manual	# of pronouns	% success rate
Minolta Copier	48	95.8
Portable Style Writer	54	83.8
Alba Twin Recorder	13	100.0
Seagate Hard Drive	18	77.8
Haynes Car Manual	50	80.0
Sony Video Recorder	40	90.6
Total	223	89.7

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 a Toyota Prius. 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 assertations of U_1 and U_2.

John bought a Toyota Prius this weekend. He purchased a beautiful new Prius for $ 20 000 at Bill's dealership.

Conversational Implicatures

Grice pointed out that conversations follow certain *maxims* (J&M:856).

1) **Maxim of Quantity:** Be exactly as informative as required.
2) **Maxim of Quality**: Try to make a contribution be a true one.
3) **Maxim of Relevance**: Be relevant.
4) **Maxim of Manner:** Avoid being obscure, ambiguous, long-winded, disorganized.

Utterance: *I have 2 siblings*.

Inferences due to the Maxims: I have exactly 2 siblings, not 3 or more (though this could be truth-conditionally possible).

Computational Applications

Some Examples:

Lascarides and Asher (2003): Explain a number of discourse coherence phenomena by figuring out algorithms to reason about them (in implementations).

Glasbey (1993): Uses discourse relations to computationally disambiguate sentence-final *then* in English.

Lascarides and Asher

Discourse Relations: Explanation, Elaboration, Narration, Background, Result.

Defeasible Axioms: e.g., Penguin Principle, Nixon Diamond.

Examples: *Max fell. John pushed him.*

We know that Max fell because John pushed him because of the Penguin Principle.

? Max won the race. He was home with the cup.

We know this is odd because he couldn't be winning a race and being at home at the same time (Nixon Diamond).

Lascarides and Asher

Penguin Principle: If there are conflicting default rules that apply, and their antecedents are in logical entailment relations, then the consequent of the rule with the most specific antecedent is preferred.

Nixon Diamond: If there are conflicting default rules that apply but no logical relations between the antecedents, then no conclusions are inferred.

Lascarides and Asher

Discourse Structure: can assign a structure to a given discourse and see whether it is well-formed.

a. *Max had a great evening last night*.
b. *He had a fantastic meal*.
c. *He ate salmon*.
d. *He devoured lots of cheese*.
e. *He won a dancing competition*.

A good discourse structure can be built up according to the discourse relations and the axioms, however e is odd and can only be attached to the discourse if one assumes the axioms are **defeasible**.

Discourse Structure Representation

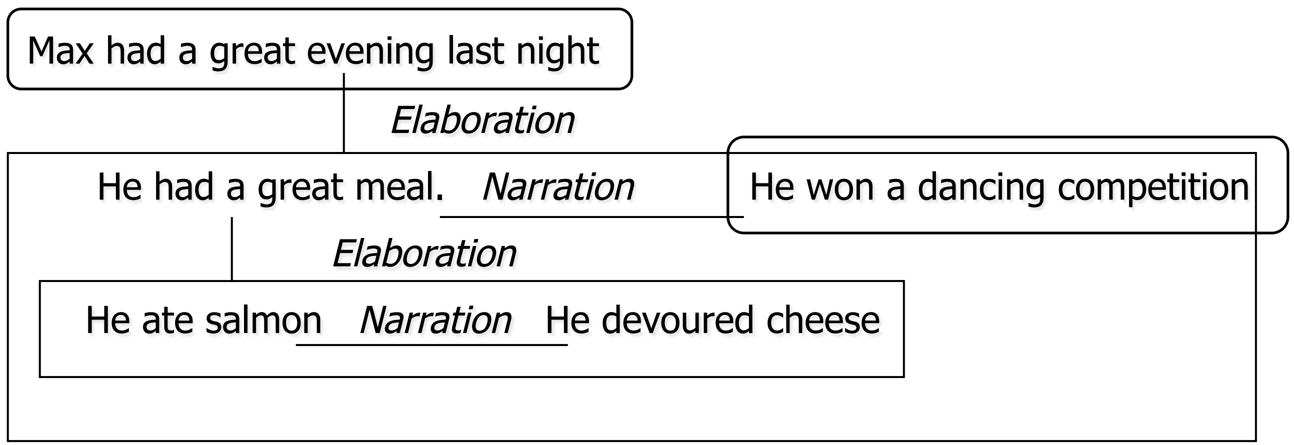
a. Max had a great evening last night.

b. He had a great meal.

c. He ate salmon.

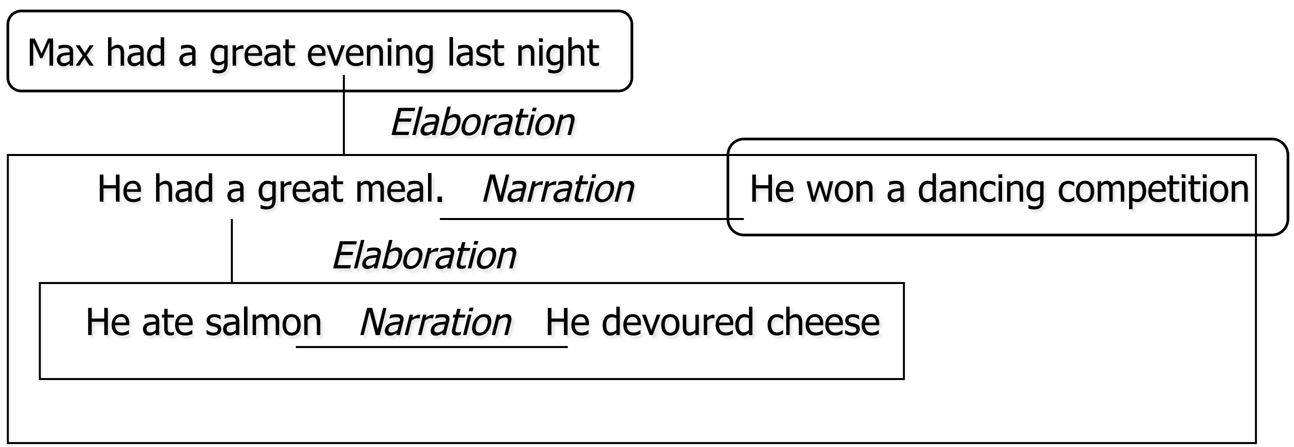
d. He devoured lots of cheese.

e. He then won a dancing competition.



Right Frontier Constraint

Right Frontier Constraint: discourse is important in anaphora resolution. So, f cannot be resolved properly because the discourse structure prohibits it.



f. ??It was beautiful pink

Sentence-Final Then

Emily climbed Ben Nevis in July.
Fiona climbed Snowden then. (Explicit Temporal Reference)

If there is no explicit time phrase in the preceding sentence, then one has to **infer** a different relation: **elaboration**.

Emily climbed Ben Nevis.
She achieved her ambition then. (Elaboration)

Glasbey defines an algorithm to disambiguate sentence-final then in computational applications based on discourse relations.

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

Blackburn, Patrick and Johan Bos. 1999. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. http://www.comsem.org

Blackburn, Patrick and Johan Bos. 1999. *Working with Discourse Representation Theory: An Advanced Course in Computational Semantics*. http://www.comsem.org

Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. Stanford, CA: CSLI Publications.

Glasbey, Sheila.1993. Temporal Connectives in a Discourse Context. *Proceedings of the Sixth Conference of the European Chpater of the Association for Computational Linguistics (EACL),* OTS, Utrecht.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts,Amhert.

References

Hobbs, Jerry. 1978. Resolving pronoun references. *Lingua* 44:311-338.

Hobbs, Jerry. 1979. Coherence and Coreference. *Cognitive Science* 3:67-90.

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

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.

Lascarides, Alex and Nicholas Asher. 1993. Temporal Interpretation, Discourse Relations and Commonsense Entailment. *Linguistics and Philosophy* 16:437-493.

Mitkov, Ruslan. 2002. *Anaphora Resolution*. Longman.