Semantics and Pragmatics of NLP
Interpretation as Abduction

Alex Lascarides

School of Informatics
University of Edinburgh

Outline

1. Discourse interpretation yields more content than compositional semantics
2. Use abduction to model this
   - Logical metonymy and Compound nouns
   - Discourse structure

Interpretation amounts to *Explaining Adjacency*

Compounds: Prove relation between modifier and head.

- *tea cup* vs. *ceramic cup*.

Sentences: Prove predicate argument structure.

- John believes men work.

Don't explain adjacency of *believes* and *men*, but rather:

- *men* and *work*; *believes* and *men work*; *John* and *believes men work*

Discourse: Prove a coherence relation between the segments:

- *I collect classic cars. My favourite is an Alfa Spider.*

Lexical Choice and Interpretation

(1) A car hit a jogger last night.

- We infer a causal relation between hitting and jogging, which goes beyond what is given by compositional semantics.
- This is just the same sort of inference that will go on at the inter-sentential level.
- We'll look at inferences at the intra-sentential level first, and extrapolate up.

Solving Pragmatics by Abduction

- Abduction is inference to the best explanation.

$$\frac{\begin{array}{l} p \longrightarrow q \\ q \end{array}}{p}$$

- Abduction in NLP:
  - We must provide an explanation of why the sentence is true.

The Algorithm

To interpret a sentence:

- Prove the logical form of the sentence that's constructed in the grammar, together with the constraints that predicates impose on their arguments,
- allowing for coercions,
- Merging redundancies where possible,
- Making assumptions where necessary.

Proving: Prove logical form via FOL.

Redundancies: Merging redundancies $\approx$ the best explanation.

Abduction: Making assumptions is the abduction bit.

The Role of Abduction in Interpreting Utterances

S and *H* have

- their own beliefs
- mutual beliefs

The content of an utterance‘mixes’ mutual beliefs and *S*’s beliefs, and is an attempt to expand the set of mutual beliefs:

- The bits in mutual belief are *old information*
- The bits outside mutual belief are *new information*.
- The bits outside mutual belief will require abduction in order to prove them.

A Simple Example

(2) The Boston office called.

Three problems:

1. Determining the relation between *Boston* and *office*.
2. Determining the reference for *the Boston office*.
3. Resolving the metonymy to *Someone at the Boston office...*

Interpreting (2)

- We must prove the LF via abduction.

(2)′ $(\exists x, y, z, e)(call(e, x) \wedge person(x) \wedge rel(x, y) \wedge office(y) \wedge Boston(z) \wedge nn(z, y))$

- There's an event e of a person x calling.
- x may not be the explicit subject,
  but it must be related to it or coercible from it, represented by $rel(x, y)$.
- y is an office which bears some unspecified relation nn to Boston.
- Abduction must be used to find out why $nn(z, y)$ and $rel(x, y)$ are true.

Example Continued: The Mutual KB

- $Boston(B_1)$
  $office(O_1) \wedge in(O_1, B_1)$
  $person(J_1) \wedge work\text{-}for(J_1, O_1)$
- If y is in z, then y and z are in a possible compound relation:
  $\forall y \forall z (in(y, z) \rightarrow nn(y, z))$
- If x works for y, then y can be coerced from x:
  $\forall x \forall y (work\text{-}for(x, y) \rightarrow rel(x, y))$

Proving the Logical Form: Fix x to be J_1 and then...

- Everything in the LF can be proved from the KB except $call(e, x)$
- Abduction permits us to assume this, so we do and add it to the mutual belief set.
- $call(e, x)$ is the new information.
- We could have assumed $person(x)$, rather than proving it with $person(J_1)$.
  This would have given the less specific reading of (2) that someone called, rather than John called.
- Redundancy??

The Proof Graph

Logical Form:

$\boxed{call'(e,x)} \wedge person(x) \wedge rel(x,y) \wedge office(y) \wedge Boston(z) \wedge nn(z,y)$

Knowledge Base:

$person(J_1)$

$work\text{-}for(x,y) \supset rel(x,y)$

$work\text{-}for(J_1,O_1)$

$office(O_1)$

$Boston(B_1)$

$in(y,z) \supset nn(z,y)$

The Three Pragmatics Problems

They are all solved as a by-product:

- The implicit relation in
  the compound nominal *Boston Office* is *in*.
- *The Boston Office* is resolved to O_1.
- The metonymy has been expanded to:
  John, who works for the Boston office, called.

Problems with Logical Form

- You must be really careful to get the logical forms right.
  - You must have $call(e, x)$ and $person(x)$ rather than $call(e, y)$.
- Selectional restrictions aren't really a matter for grammar though!

More problems later...

Making Choices

- The problem of *which* inferences to make is *the* problem in pragmatics.
  - Eg., should we assume $person(x)$, or prove it with $person(J_1)$?
- Hobbs solves this by assigning **weights** to predicates, and guiding assumptions so that they have least cost:
  - cost = sum of weights on assumptions
- Weights are assigned *manually*:
  tweak weights using trial and error.
- Weights are 'context-free': they don't change as the KB changes.

Abduction over Default Rules

Default Rules:

- Gricean maxims; Domain knowledge; Reasoning about dialogue agents

Abduction on hard rules:

$$p \rightarrow q \text{ and } q \text{ permits us to assume } p.$$

We can represent default rules as hard rules plus a predicate *etc*:

- Birds fly: $\forall x((bird(x) \wedge etc_n(x)) \rightarrow fly(x))$
- From knowing Tweety flies, we can prove via abduction that Tweety is a normal bird.

Proving Discourse

(3) Max fell. John pushed him.

- You must prove that (3) is a discourse segment.
- You do this by proving a coherence relation between the sentences from rules like the following:

1. $\forall e_1, e_2, e(\mathit{CoherenceRel}(e_1, e_2, e) \rightarrow \mathit{Segment}(e))$
2. $\forall e_1, e_2, e((\mathit{Info}(e_1, e_2) \wedge \mathit{etc}_i) \rightarrow \mathit{CoherenceRel}(e_1, e_2, e))$

- *CoherenceRel* is coordinating: e must be computed from e_1 and e_2 together.
- *CoherenceRel* is subordinating: e is either e_1 or e_2.

Rules for (3)

- $\forall e_1, e_2, e(CoherenceRel(e_1, e_2, e) \rightarrow Segment(e))$
- $\forall e_2, e_1(cause(e_2, e_1) \rightarrow Explanation(e_1, e_2, e_1))$
- $\forall e_1, e_2, e(Explanation(e_1, e_2, e) \rightarrow CoherenceRel(e_1, e_2, e))$

- Abduce (i.e. assume) *cause*, and the appropriate conclusion follows.
- So abduce pushing caused the falling, and then you are assured that (3) is a coherent discourse segment.

Occasion

(4) a. At 5:00 the train arrived in Chicago.
    b. At 6:00 Bill Clinton held the press conference.

Instead of Explanation, we have *Occasion*, which is proved when:

- Both events describe a change in state, and the final state of the first is the initial state of the second.

Parallel

- *Parallel*(e_1, e_2, e) is proved if:
  - The first segment S_1 (plus assumptions) entails $p(x_1, \ldots, x_n)$
  - The second segment S_2 (plus assumptions) entails $p(y_1, \ldots, y_n)$
  - x_i is similar to y_i in that they share some property.
- It's a coordinating relation.

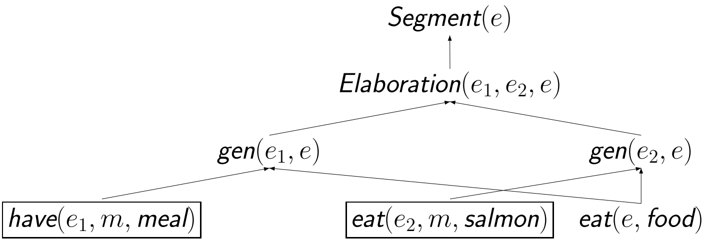
(5) John drank beer. Fred drank wine.

Elaboration: a limiting case of Parallel

- Entities are not merely similar, but identical.
- At some level, both segments say the same thing.
- Proving Elaboration:
  If there is an event e that is generated by both e_1 and by e_2,
  then they are connected by Elaboration,
  and e acts as the summary.
  - $\forall e_1, e_2, e(gen(e_1, e) \wedge gen(e_2, e) \rightarrow Elaboration(e_1, e_2, e))$
- Elaboration is a subordinating relation.

Proving an Elaboration

(6) Max had a great meal.
He ate lots of salmon.



Contrast (A Coordinating Relation)

(7) John has black hair. Jill has brown hair.

(8) John is graceful. Jill is an elephant.

To prove Contrast, prove:

1. Segment S_1 entails $p(x_1, \ldots, x_n)$
2. Segment S_2 entails $\neg p(y_1, \ldots, y_n)$, where x_i are similar to y_i.

- (7) can be interpreted as Contrast or Parallel.
- The sense extension of *elephant* in (8) is a by-product of trying to prove the Contrast relation:
  - You have to prove *elephant* implies $\neg$*graceful*.

Another Example

(9) a. The police prohibited the women from demonstrating.
b. They feared violence.

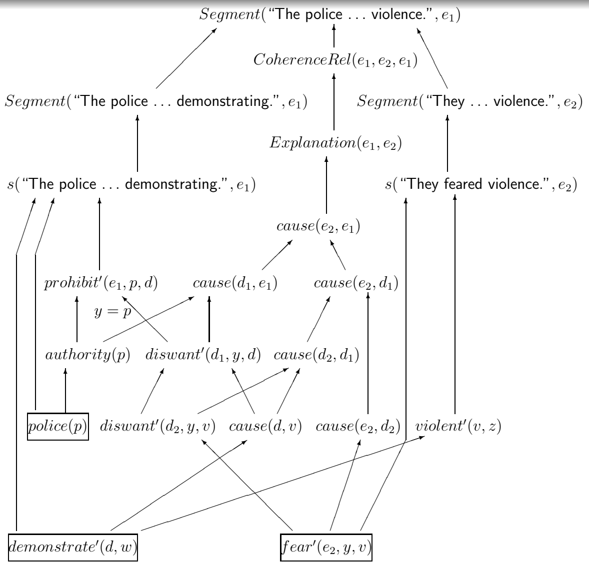
1. Prove that (9)a and (9)b are sentences.
2. Prove that together they form a segment.
   1. Aim for Explanation relation.
   2. So prove:
      - There is a prohibiting event e_1 of the police.
      - There is a fearing event e_2 of “them”
      - e_2 caused e_1.

Proving the Causal Relation

(c) This can be proved if we have the following WK axioms:
 (i) If e_2 is a fearing by y of v, then this causes y not to want v
 (ii) If e_1 is a demonstration, then e_1 causes violence (v).
 (iii) If y doesn't want v, then this causes y to prevent v from happening.

(d) If we assume "they" is the police, then the proof of causation follows by the above WK axioms.

The Proof Graph



The Problem of Choice in Abduction

(3) Max fell. John pushed him.

(10) Max fell. John helped him.

A $(\langle e_1, e_2 \rangle \wedge cause(e_2, e_1)) \rightarrow Explanation(e_1, e_2, e_1)$

B $(\langle e_1, e_2 \rangle \wedge cause(e_1, e_2)) \rightarrow Narration(e_1, e_2, e)$

- Need (B) to prove (10) is a segment. *Be Orderly*.
- But you can abduce on (B) to get the wrong interpretation of (3).
- There's a choice of what to abduce. How do we choose?

Hobbs *et al*'s Solution

- Assign costs to predicates.
- Guide abduction so that you abduce things that give the smallest overall cost.
- This amounts to the least risk strategy.

Falling and Pushing:

$$(\langle e_1, e_2\rangle \wedge \text{FALL}(e_1, x) \wedge \text{PUSH}(e_2, y, x) \wedge \text{ETC}_n(e_1, e_2)) \rightarrow \text{CAUSE}(e_2, e_1)$$

ETC predicates generally assigned low weights.

Problems

- Ad hoc!
- Costs on predicates aren't context sensitive enough.

(11) John hit the back of Max's neck.
Max fell. John pushed him.
Max fell over the edge of the cliff.

- So the costs on predicates must be a function of the *whole KB*!
Definitely context-sensitive, then!
- But then assigning weights is intractable!!

More on Intractability

- Abduction (without weights) over first order logic is intractable anyway, because consistency checking over first order logic is beyond what's recursively enumerable.
- So computing these implicatures is uncomputable.
- It's thus inadequate as a theory of semantic competence:
  - It doesn't explain why by-and-large we agree on what was said.

Problems: Anaphora

Interpreting amounts to updating beliefs:

(2) The Boston Office called

- The interpreter abduces that John, who works for the Boston office, called.
- So John features in the representation of (2).
- But then John is available for future anaphoric reference:

(12) The Boston office called. ?He was very angry.

- The representation of linguistic content should be separate from the effects of content on beliefs.

Another Reason for Separating Content from Beliefs

(13) a. A: John went to jail.
b. He was caught embezzling funds.
c. B: No! He was caught embezzling funds, but he went to jail because he committed tax fraud.

- You *can't possibly* prove things that you believe to be false from:
  - your private beliefs, or
  - mutual beliefs (which you must believe)
- So *B* won't prove *A*'s segment as an *Explanation* unless he performs *all* the reasoning over (only) his model of *A*'s private beliefs (*not* mutual beliefs).
- But you don't need to do this: just use lexical semantics instead.

Confusing What is Said with Evaluating What is Said

The logic in which you construct logical form shouldn't have full access to the logic in which you interpret logical form:

(14) a. There are unsolvable problems in number theory.
b. Every even number greater than two is equal to the sum of two primes is undecidable, for instance.

- Abducing *Elaboration* involves checking it's consistent.
- That involves checking (14)b is satisfiable.
- But we don't know how to do that!!

Even mathematically inept people interpret (14) as an Elaboration.

Knowledge Interaction: Modularity is Crucial!

(15)
a. Jane saw Mary.
b. She asked a question.
c. She answered her *no*

- (15)b: low weight for resolution as in centering theory.
- But this conflicts with preferred interpretation of (15)c!

The rule for doing (15)c is then very *ad hoc*:

- A respondent to a question is different from the questioner, and this rule overrides preferences from Centering.

  (Stone and Thomason, 2002)

1. Misses generalisations about organisation of knowledge.
2. Can't be expressed in the logic anyway.

Summary

- People infer more content than just compositional semantics.
- The inferences they use involve *weighted abduction*.
- In proving a sentence you do a number of tasks as a byproduct:
  - Resolve logical metonymies and compute sense extensions
  - Resolve anaphora
  - Infer causal relations
  - and more. . .

Problems

- Interpretation as belief update:
  - Inferences are more complex than they need to be.
  - Should use linguistic knowledge sources rather than reasoning with other people's beliefs whenever possible.
- Modularity needed to:
  1. Make constructing logical form computable (and therefore the basis for explaining semantic competence)
  2. Separate computing what is said from evaluating whether what is said is true.
  3. Express generalisations about the relative priority of sources of information.