Inferring the Antecedent: Resolving Anaphora off the Deep End

Scott Martin
http://www.coffeeblack.org/

Natural Language Understanding and Artificial Intelligence Laboratory
Nuance Communications

AMPRA Panel on Computational and Experimental Approaches to Reference and Anaphoric Inference
UCLA, October 18, 2014

Background: Shallow vs. Deep Approaches to Anaphora Resolution

- Many approaches to resolving anaphora are more 'shallow,' avoiding parsing and semantic analysis
- Why? Because 'deep' approaches
  - Require large, sophisticated annotations (grammars, etc.)
  - Make categorical distinctions that lead to a built-in precision bias
  - Produce systems that are complicated to maintain
- But what if we didn't care about the first point because the necessary annotations were available?

This Talk

- I'll discuss a deep NLP technique for resolving anaphora in discourse I'm developing, in part jointly with Kathy Dahlgren
- The approach generalizes a well-known algorithm from the literature
- Most importantly, it leverages large-scale knowledge bases to get *bridging* inferences

Outline

What Makes It So Hard

- The problem is sometimes cast as the task of rewriting an anaphor using a previously-mentioned named entity, as in

  (1) Who is [the President of the U.S.]$_i$? How old is he$_i$? (Both slots rewritten with *Barack Obama*.)

What Makes It So Hard

- The problem is sometimes cast as the task of rewriting an anaphor using a previously-mentioned named entity, as in

  (1) Who is [the President of the U.S.]$_i$? How old is he$_i$? (Both slots rewritten with *Barack Obama*.)

- But this obviously doesn't generalize:

  (2) Find an Italian restaurant$_i$, and tell me how late it$_i$'s open. (Where's the named entity?)

'Bridging' the Gap between Shallow and Deep

- One of the apparently more difficult complicating factors, inference, may be a candidate for moving toward the deep end

(3) Find my reservation$_i$ at Bierhaus$_j$.

a. What time is it$_i$ for?

b. What's on the menu$_?$?

'Bridging' the Gap between Shallow and Deep

- One of the apparently more difficult complicating factors, inference, may be a candidate for moving toward the deep end

  (3) Find my reservation$_i$ at Bierhaus$_j$.

  a. What time is it$_i$ for?

  b. What's on the menu$_?$?

- Getting the anaphora in (3a) requires world knowledge—reservations are the kinds of things that have times, and menus aren't, so the reservation should be more salient than the menu

'Bridging' the Gap between Shallow and Deep

- One of the apparently more difficult complicating factors, inference, may be a candidate for moving toward the deep end

  (3) Find my reservation$_i$ at Bierhaus$_j$.

  a. What time is it$_i$ for?

  b. What's on the menu$_?$?

- Getting the anaphora in (3a) requires world knowledge—reservations are the kinds of things that have times, and menus aren't, so the reservation should be more salient than the menu
- Getting (3b), an instance of *bridging*, is even tougher: we have to know that
  1. Bierhaus is an instance of a restaurant,
  2. Restaurants, unlike reservations, are the kinds of things that have menus, and
  3. *the menu* refers to the menu at Bierhaus

Outline

Discourse Representation Theory

DRT can be thought of as an extension of first-order logic where

- Indefinites (*a, some*, etc.) introduce *discourse referents*, rather than existential quantification
- There is a notion of semantic scope, giving rise to a relation of *accessibility* for anaphoric antecedents
- An entire discourse's meaning can be captured, rather than just a single utterance

Benefits of DRT

- It can model everything that single-utterance first-order approaches to semantics can model
- There are well-known translations from DRT to first-order logic and back*
- There is a fairly good algorithm for resolving anaphora over DRT representations (more on that in a minute)
- Its notation is pretty intuitive, and a fairly large community of researchers exists around it

Simple DRT Example

(4) An engineer walked in. Then a scientist walked in, and he ordered a beer.

$x\ y\ z$
engineer(x)
walk-in(x)
scientist(y)
walk-in(y)
beer(z)
order(y, z)

Simple DRT Example

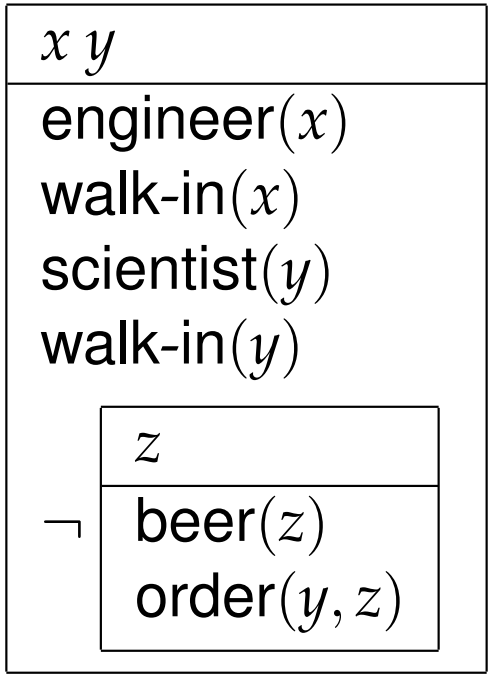
(4) An engineer walked in. Then a scientist walked in, and he ordered a beer.

$x\ y\ z$
engineer(x)
walk-in(x)
scientist(y)
walk-in(y)
beer(z)
order(y, z)

- The referents x, y, and z are left around so that more information can be added about them later

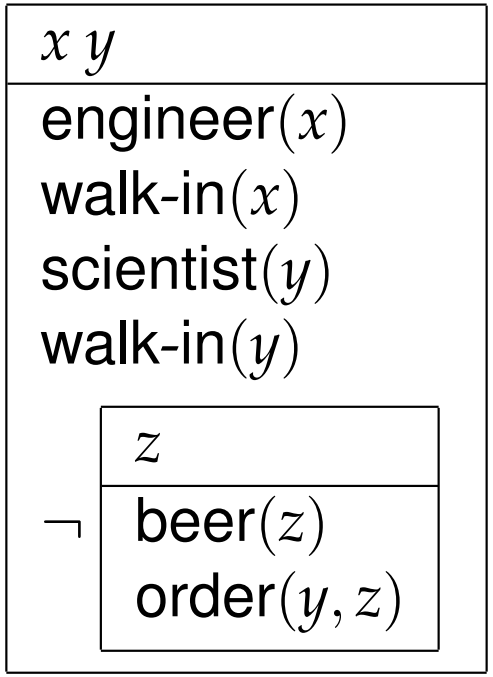
Slightly More Complicated DRT Example

(4) An engineer walked in. Then a scientist walked in, but he didn't order a beer.



Slightly More Complicated DRT Example

(4) An engineer walked in. Then a scientist walked in, but he didn't order a beer.

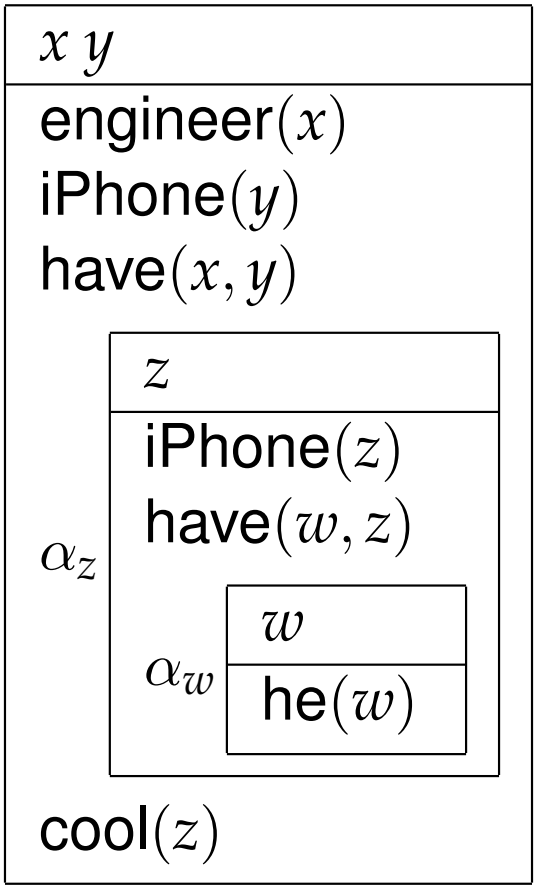


- In this version, the beer referent z is 'trapped' in the scope of negation and is not accessible as an antecedent for later anaphora

Van der Sandt-Bos Algorithm I

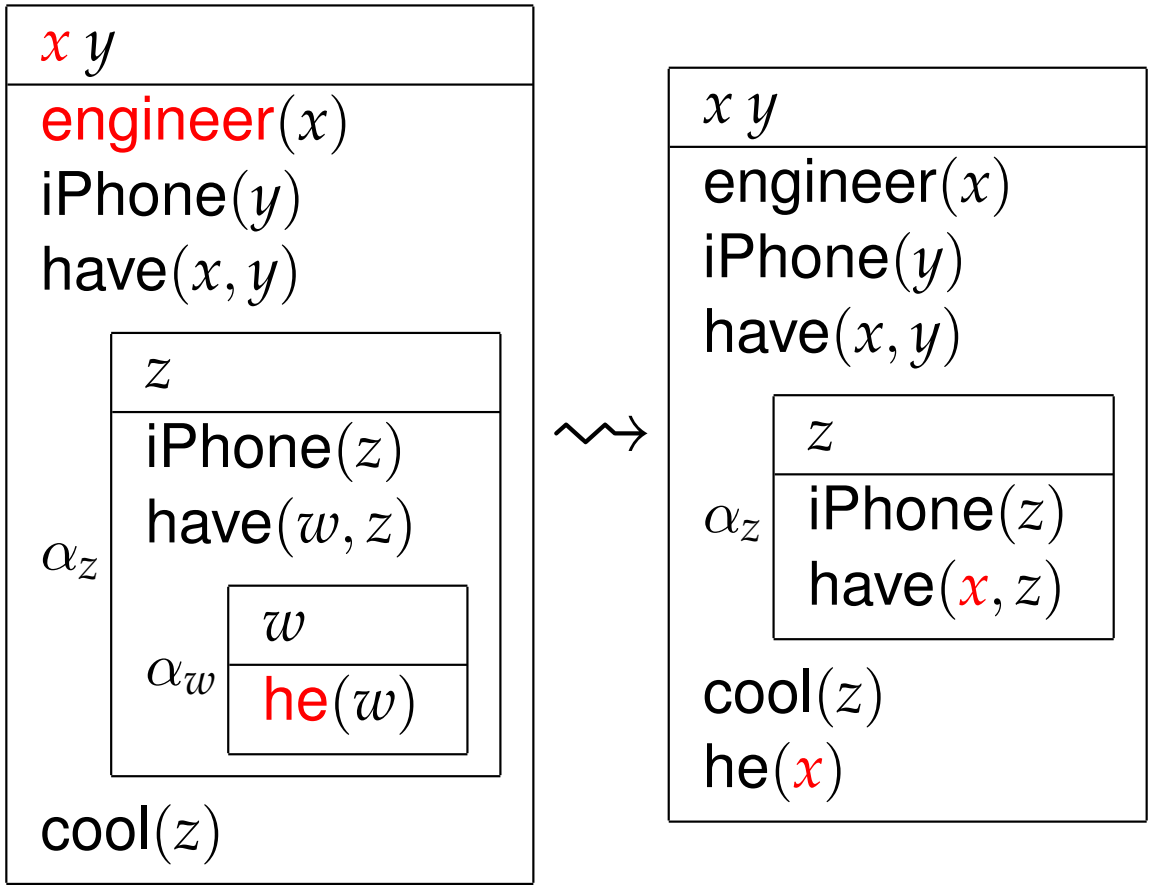
(5) An engineer$_i$ has an iPhone$_j$. His$_i$ iPhone$_j$ is cool.

- This example gets the initial representation below:



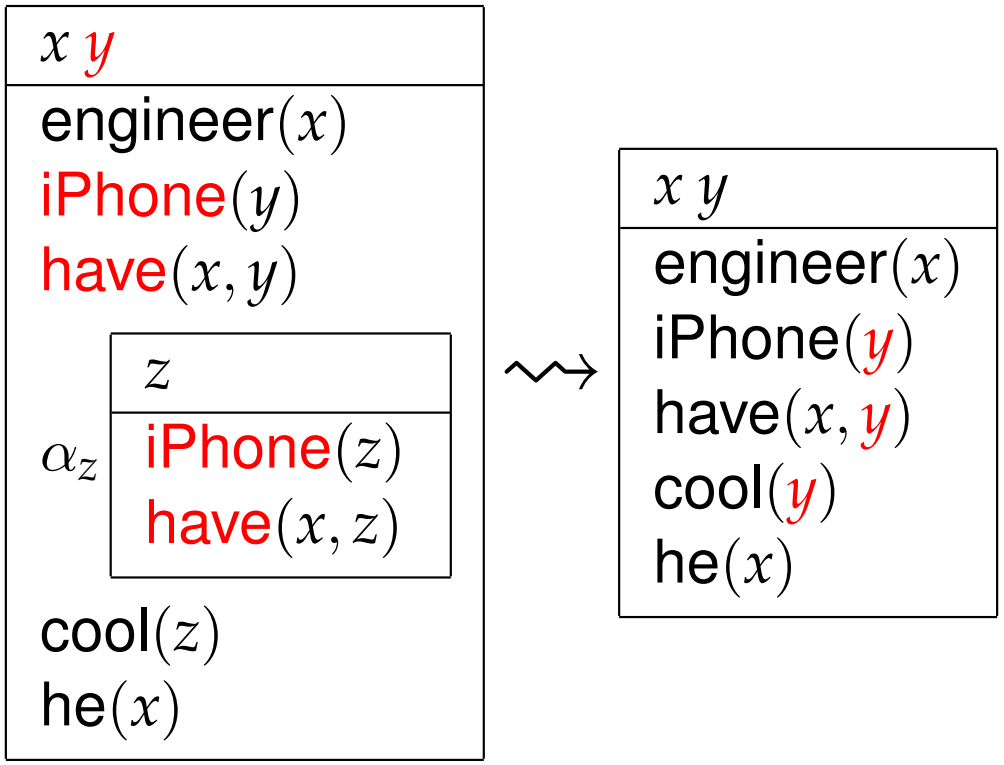
Van der Sandt-Bos Algorithm II

- Then the most deeply embedded anaphor is resolved to x, because *engineer* is consistent with *he*



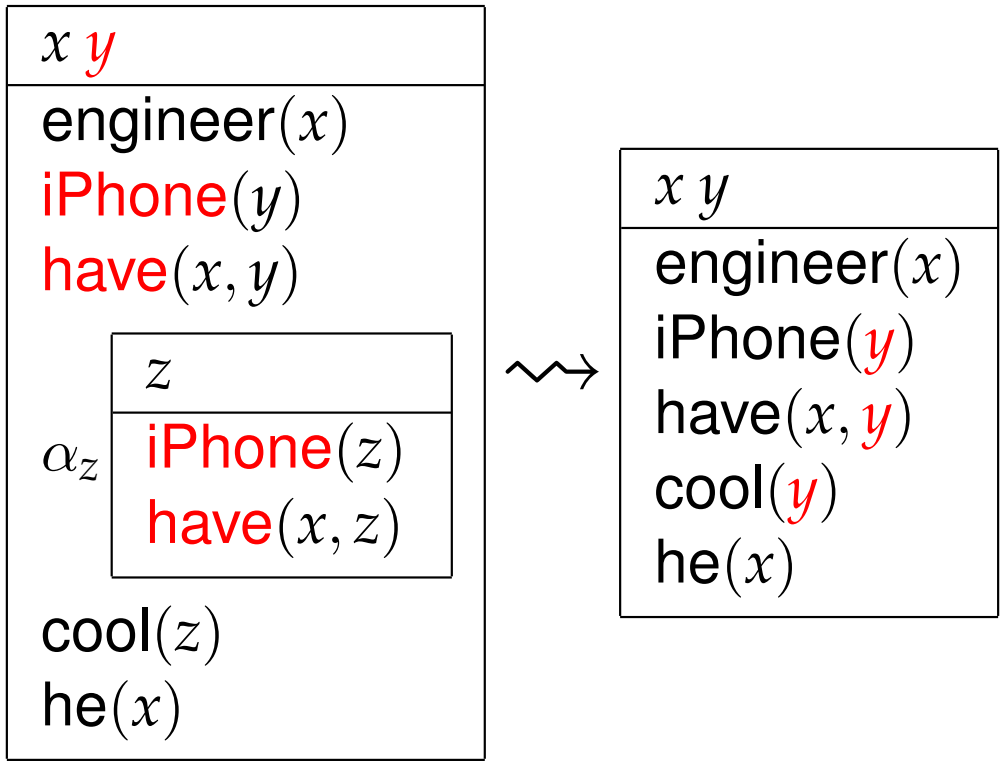
Van der Sandt-Bos Algorithm III

- Lastly, the remaining anaphor is resolved to the iPhone:



Van der Sandt-Bos Algorithm III

- Lastly, the remaining anaphor is resolved to the iPhone:



- This gives the desired interpretation: there is some male engineer x that has a cool iPhone y

Problems

- This process of traversing the accessibility hierarchy (‘going up the tree’) works well to rule out e.g. antecedents in the scope of negation

Problems

- This process of traversing the accessibility hierarchy ('going up the tree') works well to rule out e.g. antecedents in the scope of negation
- But a lot has to go on behind the scenes to decide whether an antecedent is suitable

  Pronouns Is the antecedent's descriptive content consistent with the pronoun's?

  Definites Does the antecedent's descriptive content entail the definite's?
- In case an anaphor can't be resolved, we have to decide whether to fail or *accommodate* (for which the van der Sandt-Bos algorithm has a special second phase)

Problems

- This process of traversing the accessibility hierarchy ('going up the tree') works well to rule out e.g. antecedents in the scope of negation
- But a lot has to go on behind the scenes to decide whether an antecedent is suitable

  Pronouns Is the antecedent's descriptive content consistent with the pronoun's?

  Definites Does the antecedent's descriptive content entail the definite's?
- In case an anaphor can't be resolved, we have to decide whether to fail or *accommodate* (for which the van der Sandt-Bos algorithm has a special second phase)
- For bridging inferences, we additionally have to ask about aspects related to the antecedent, i.e., if the antecedent is a restaurant, that it has a menu

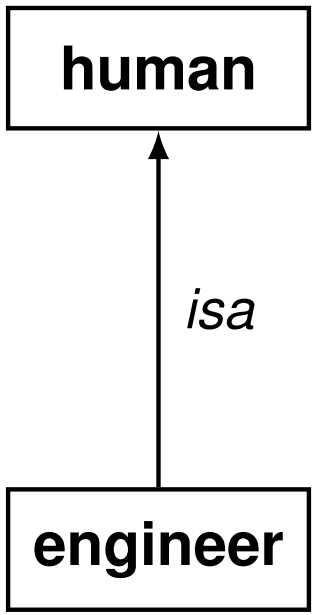
Outline

Making the Algorithm 'Smarter'

- Using DRT representations by itself only tells us what *can't* be possible antecedents (accessibility)
- Comparing descriptive contents for consistency and entailment, and doing inferences to find unmentioned antecedents, is beyond its scope
- Fortunately, there are many resources that are up to the task: ontologies like Cognition, Cyc, Freebase, SUMO, …

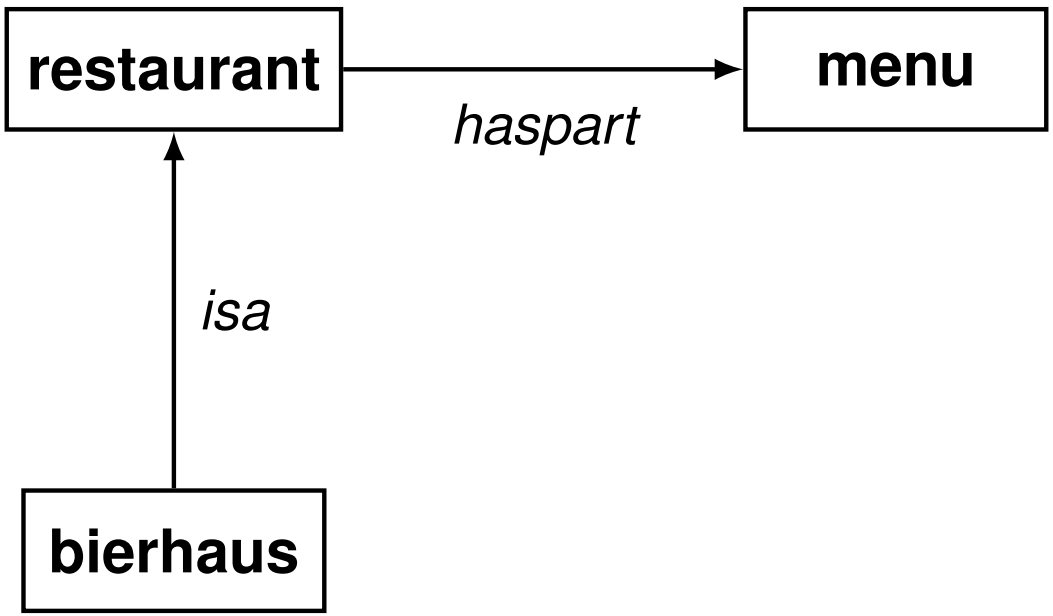
Simple Entailments

- In the simplest case, an ontology can help out with basic entailments needed for consistency
- For example, if the ontology has an *isa* hierarchy (or one can be deduced), we can get that an engineer is a human
- This is useful because humans can be referred to as *he/him* or *she/her*



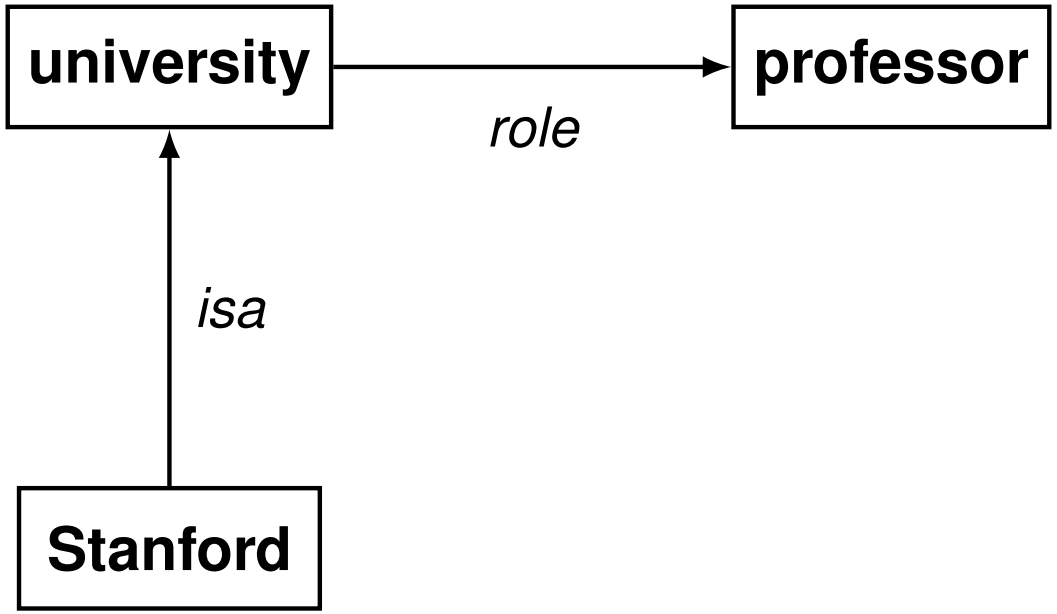
Broadening the Coverage: Bridging Inferences I

- An ontology with *haspart* relations between concepts or classes lets us get bridging inferences
- Only a single graph traversal is needed in addition to what was already needed for simple entailments



Broadening the Coverage: Bridging Inferences II

- The capabilities for bridging are limited only by the kinds of relations that the ontology can countenance
- For instance, if the ontology has a notion of *role* in addition to *isa* subsumption and *haspart*, we can get e.g. *Kim attended Stanford. She liked the professors.*



Generalizing the Algorithm

Assuming a suitable ontology is available, an extension of the van der Sandt-Bos algorithm for bridging goes like this:

1. *Assume*: R is an unresolved DRT representation
2. For each anaphoric condition a in R, starting with the most deeply embedded:
   - 2.1 Traverse the accessibility hierarchy (ordered by relative salience), and for each candidate antecedent c, if a is related to c:
     - 2.1.1 If the relation is subsumption, resolve a to c
     - 2.1.2 Otherwise, introduce a new, bridged discourse referent b in the DRS where c occurs that bears the relevant descriptive content
3. Fail unless R is fully resolved

Generalizing the Algorithm: Bridging Examples I

- This variant algorithm is best understood via some examples. As background, consider our earlier example

  (3) Find my reservation at Bierhaus.

- We start by assuming the following representation of (3):

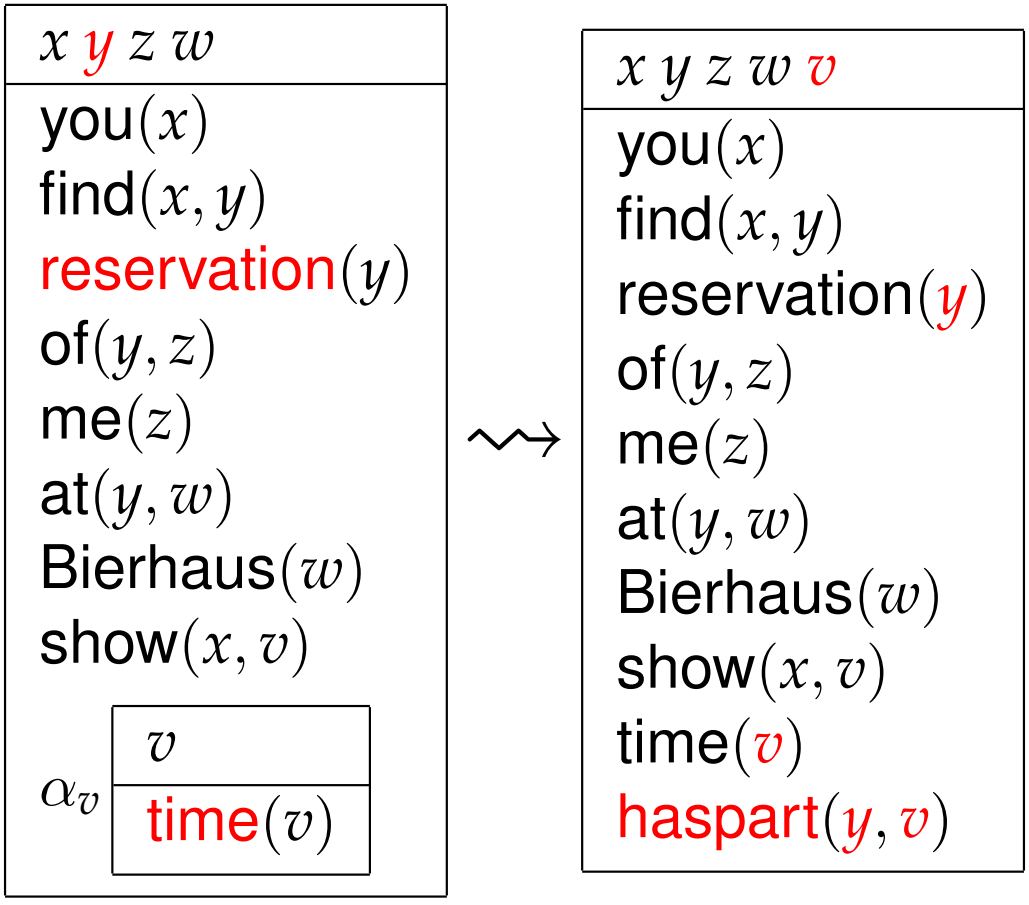
$x\ y\ z\ w$
you(x)
find(x, y)
reservation(y)
of(y, z)
me(z)
at(y, w)
Bierhaus(w)

Generalizing the Algorithm: Bridging Examples I

- Then, recalling that reservations have times, following up (3) with

  (3a′) Show the time.

  gives:

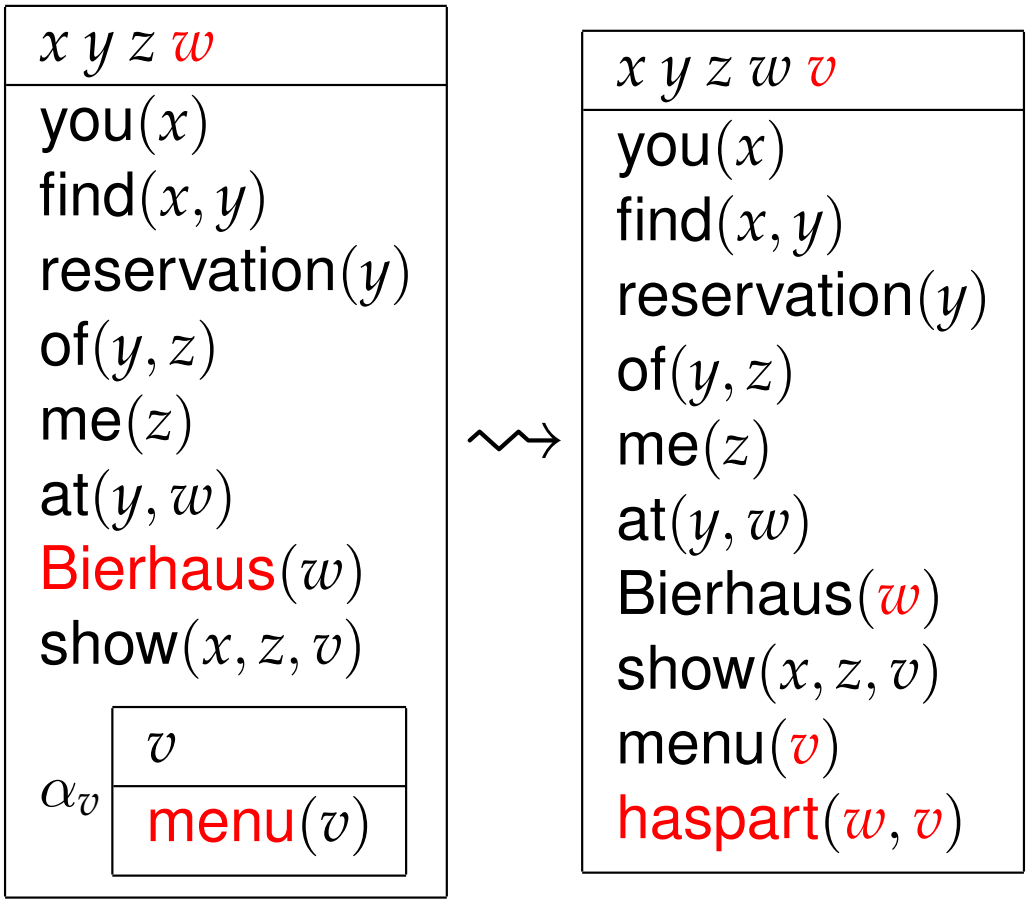


Generalizing the Algorithm: Bridging Examples II

- On the other hand, following up (3) with

  (3b′) Show the menu.

  gets a different representation, since restaurants have menus:



Taking Stock

- This extended algorithm allows us to improve on shallow methods, because shallow methods would be hard pressed to tell the difference between the *time* and *menu* variants
- But some caveats:
  - These are only simple cases—comparing descriptive content with an antecedent for suitability, in the general case, requires (the equivalent of) first-order theorem proving
  - Many ontologies are quite dense, so how do you know when to quit traversing?
  - When you're constantly asking about covert antecedents, does that slow down an implementation too much?
  - Already, without bridging, the cost of resolving a DRT structure is not exactly cheap

Outline

A Smarter Anaphora Resolution Algorithm

- Using simple ontology reasoning, a deep approach to anaphora resolution can improve on more shallow methods in terms of coverage

A Smarter Anaphora Resolution Algorithm

- Using simple ontology reasoning, a deep approach to anaphora resolution can improve on more shallow methods in terms of coverage
- With bridging inferences cast as a special (non-subsumptive) case of the kind of inferences needed to do anaphora resolution anyway, we get inferred antecedent selection without much extra work
  - Usually, just a single additional hop in the ontology traversal is needed

A Smarter Anaphora Resolution Algorithm

- Using simple ontology reasoning, a deep approach to anaphora resolution can improve on more shallow methods in terms of coverage
- With bridging inferences cast as a special (non-subsumptive) case of the kind of inferences needed to do anaphora resolution anyway, we get inferred antecedent selection without much extra work
  - Usually, just a single additional hop in the ontology traversal is needed
- The process of adding bridged antecedents can be seen as a highly constrained replacement for the *accommodation* phase of the van der Sandt-Bos algorithm

Tying it in with Shallow Methods

- I'm very optimistic that the AI winter will be followed by an NL summer, with shallow and deep NLP joining forces

Tying it in with Shallow Methods

- I'm very optimistic that the AI winter will be followed by an NL summer, with shallow and deep NLP joining forces
- For the case of the approach to anaphora resolution that I sketched, we could use:
  - A named entity recognizer trained from data
  - A parser trained from data
  - A sense disambiguation regime trained from data
  - A salience ranking algorithm trained from data

Tying it in with Shallow Methods

- I'm very optimistic that the AI winter will be followed by an NL summer, with shallow and deep NLP joining forces
- For the case of the approach to anaphora resolution that I sketched, we could use:
  - A named entity recognizer trained from data
  - A parser trained from data
  - A sense disambiguation regime trained from data
  - A salience ranking algorithm trained from data
- But also, for evaluation we can now use:
  - The Groningen Meaning Bank, a Penn Tree Bank sized corpus of texts with accompanying DRT representations
  - ISNotes, an extension to the OntoNotes corpus that has labels for anaphora resolution, including $\sim$650 instances of bridging