Empirical approaches to discourse

Day 3:
Solutions

ESSLLI 2012

Jennifer Spenader

Outline

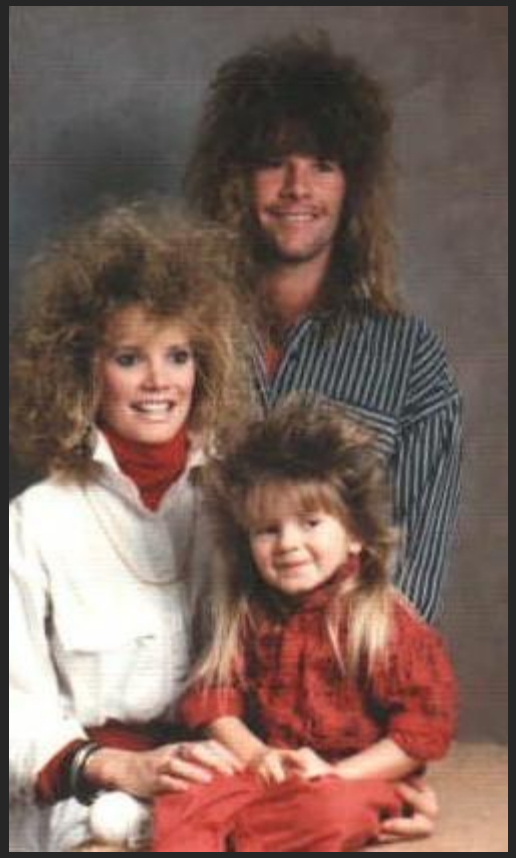
1. Problems with some theories of coherence relations
   - What should be a relation
   - Semantics/informational vs. Pragmatic/presentational relations

2. Possible solutions
   - SDRT (today)
   - The PDTB (Penn Discourse Treebank) (tomorrow)

SDRT:
Asher & Lascarides

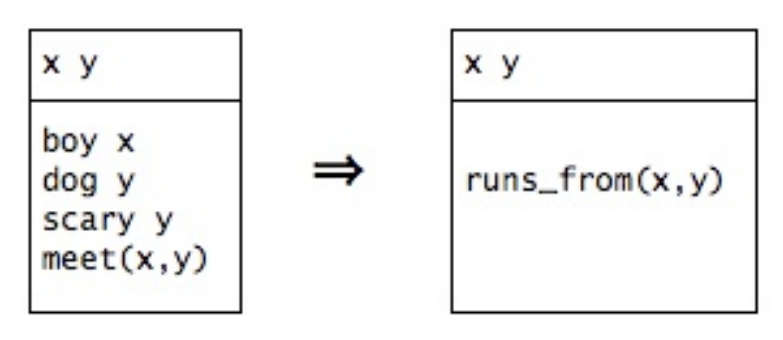
- Segmented Discourse Representation Theory (SDRT)
- Asher and Lascarides
- Paper starts by telling us how the 80's gave us two great things:
  - Dynamic Semantics
  - Theories of Rhetorical Structure

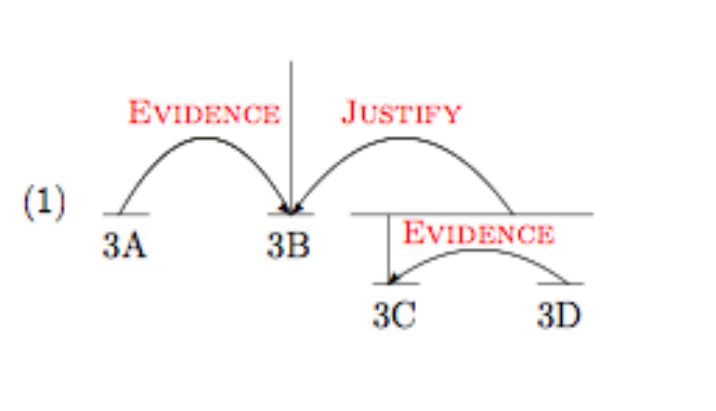
Many amazing things the 80's gave us haven't stood the test of time.



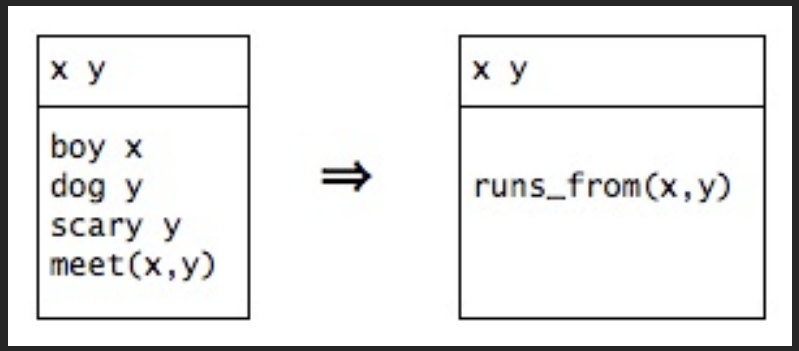
But some things will
never go out of style.

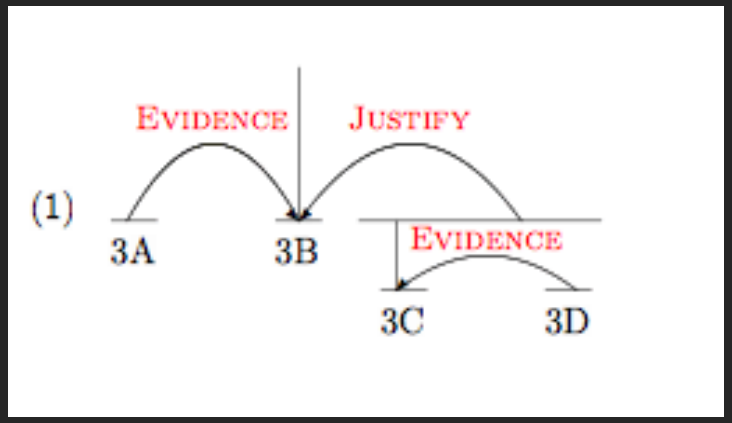
But some things will never go out of style.





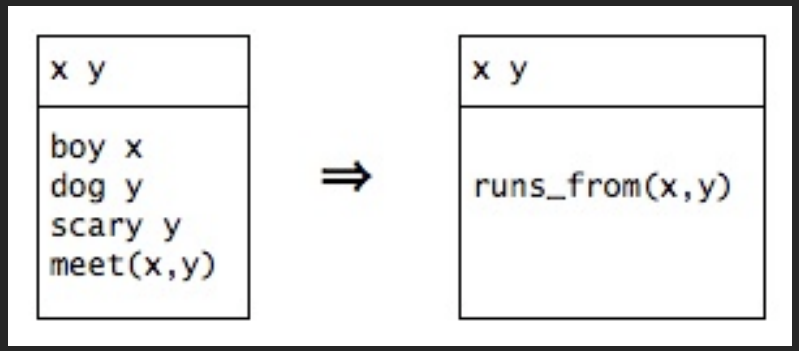
But some things will never go out of style.

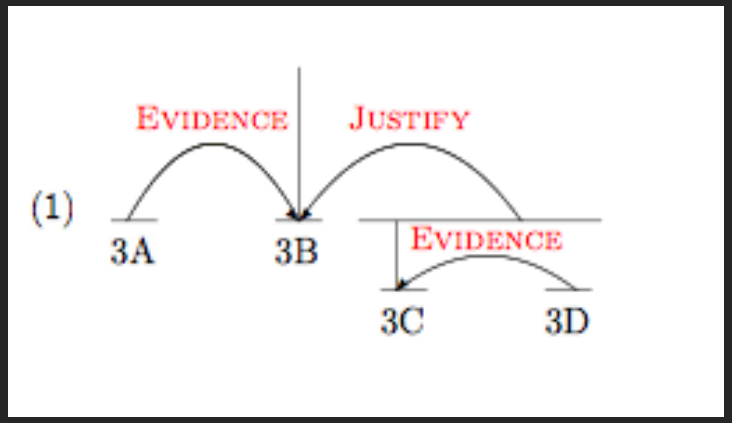




Dynamic Semantics and Rhetorical Structure.

But some things will never go out of style.





Dynamic Semantics and Rhetorical Structure.

Always in fashion.

SDRT

- SDRT combines Dynamic Semantics with a constrained theory of rhetorical structure
- It is well-defined and modular
  - different types of information are kept seperated by ‘porous fences’ to keep the theory manageable
  - Builds on ideas from Dynamic Semantics, but adds pragmatic reasoning/information to the representation

What is dynamic semantics?

- Dynamic semantics defines meaning in terms of context change potentials (CCPs)
  - the meaning contributed by a sentences is how it changes the context in which subsequent sentences will be interpreted
- Initially developed as logically based theories that could deal with pronouns.
  - DRT (Discourse Representation Theory)
    - Kamp & Reyle (1981)
  - Context change semantics
    - Heim (1983)

DRT vs. traditional logic

A man walked in. He ordered a beer.

$$\mathrm{a}'. \quad \exists x(man(x) \wedge walk\text{-}in(x)) \wedge \exists y(beer(y) \wedge order(z, y))$$

$$\mathrm{a}''. \quad \exists x(man(x) \wedge walk\text{-}in(x) \wedge \exists y(beer(y) \wedge order(x, y)))$$

DRT vs. traditional logic

A man walked in. He ordered a beer.

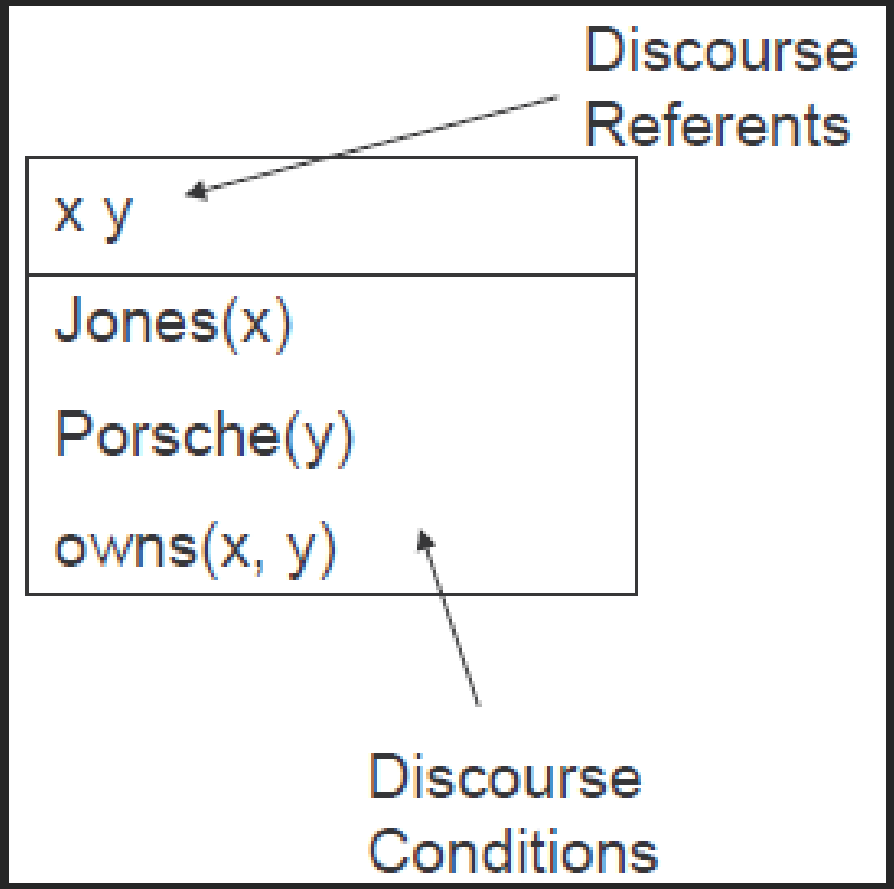
$$\mathrm{a}'. \quad \exists x(man(x) \wedge walk\text{-}in(x)) \wedge \exists y(beer(y) \wedge order(z, y))$$

$$\mathrm{a}''. \quad \exists x(man(x) \wedge walk\text{-}in(x) \wedge \exists y(beer(y) \wedge order(x, y)))$$

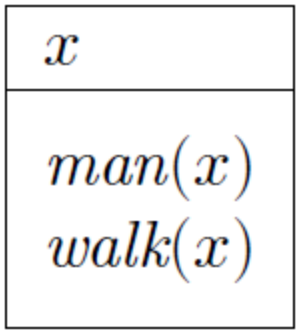
Simply moving the parentheses is awkward.

Do we really want connectives having scope over the entire text like this?

Jones owns a Porsche.

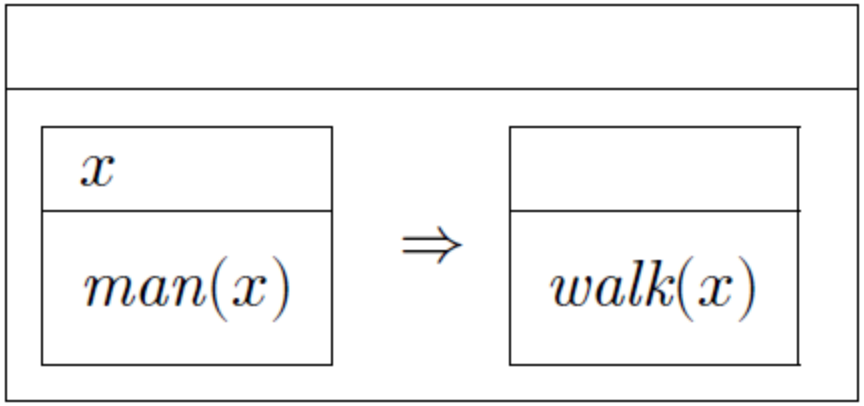


a.



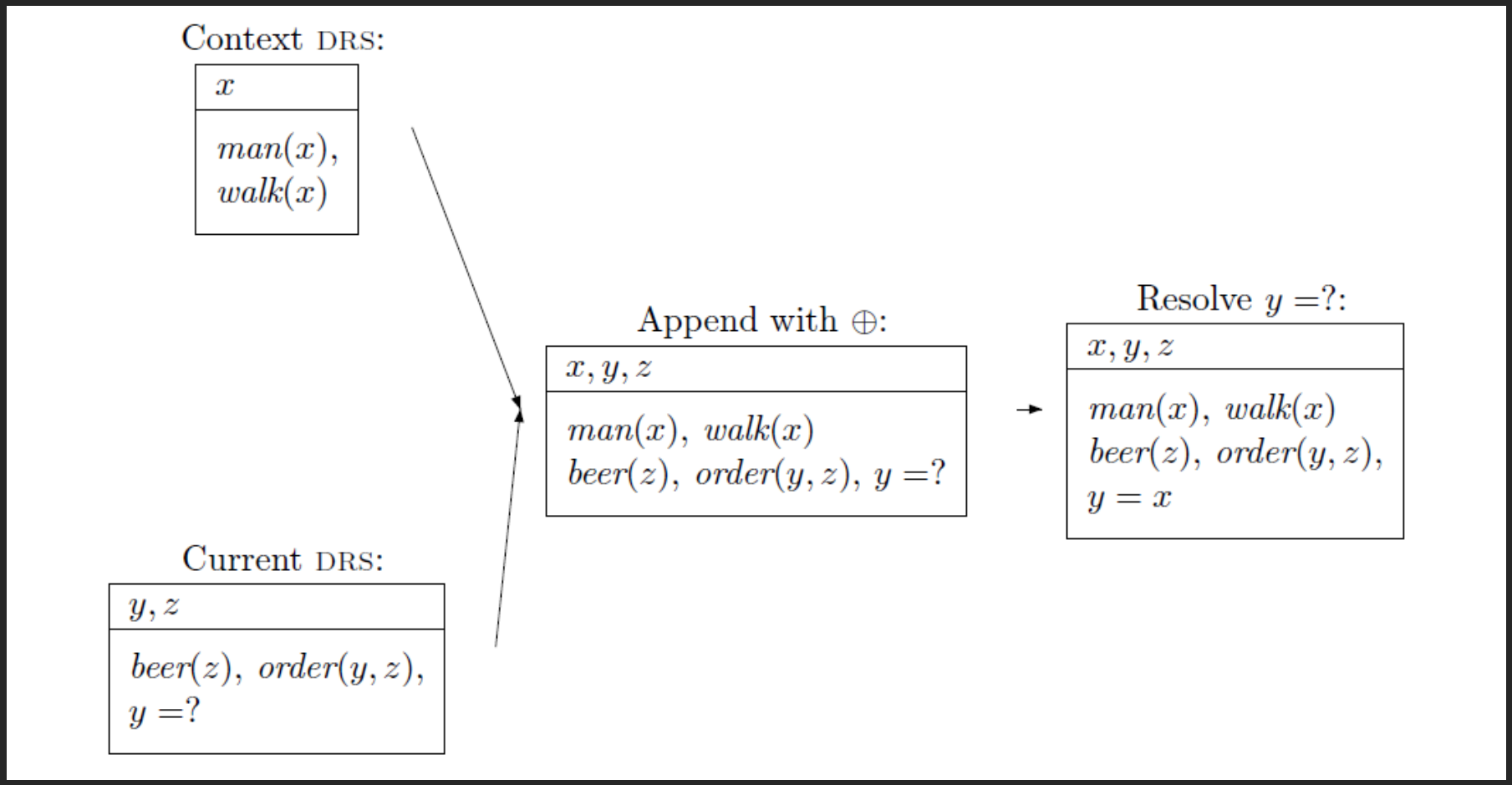
A man walks.

b.



Every man walks.

A man walked in. He ordered a beer.



- a sentence S is interpreted as a relation between an input context and an output one

- The introduction of new discourse referents into a drs K causes a transition from an input context (i.e. an information state) to an output one.
- Drs-conditions impose tests on the input context

- Accessibility constraints say what anaphoric links are possible and which are impossible

- As discourse referents are added, the assignment function gets extended, changing the model

- With more information, the set of possible worlds where the discourse could be true decreases

Accessibility governs pronoun interpretation possibilities

Discourse referents introduced in embedded DRSs are not accessible outside that DRS:

1. ? John doesn't have a car. It is red.

[x z : John(x), neg[y: car(y), owns(x,y)], red(z), z=???]

Why we need a representation of discourse structure.

1. Pronouns
2. Temporal Anaphora
3. Presuppositions

John

π_1. John had a great evening last night.

π_2. He had a great meal.

π_3. He ate salmon.

π_4. He devoured lots of cheese.

π_5. He won a dancing competition.

π_6. ??It was a beautiful pink.

π_1. John had a great evening last night.

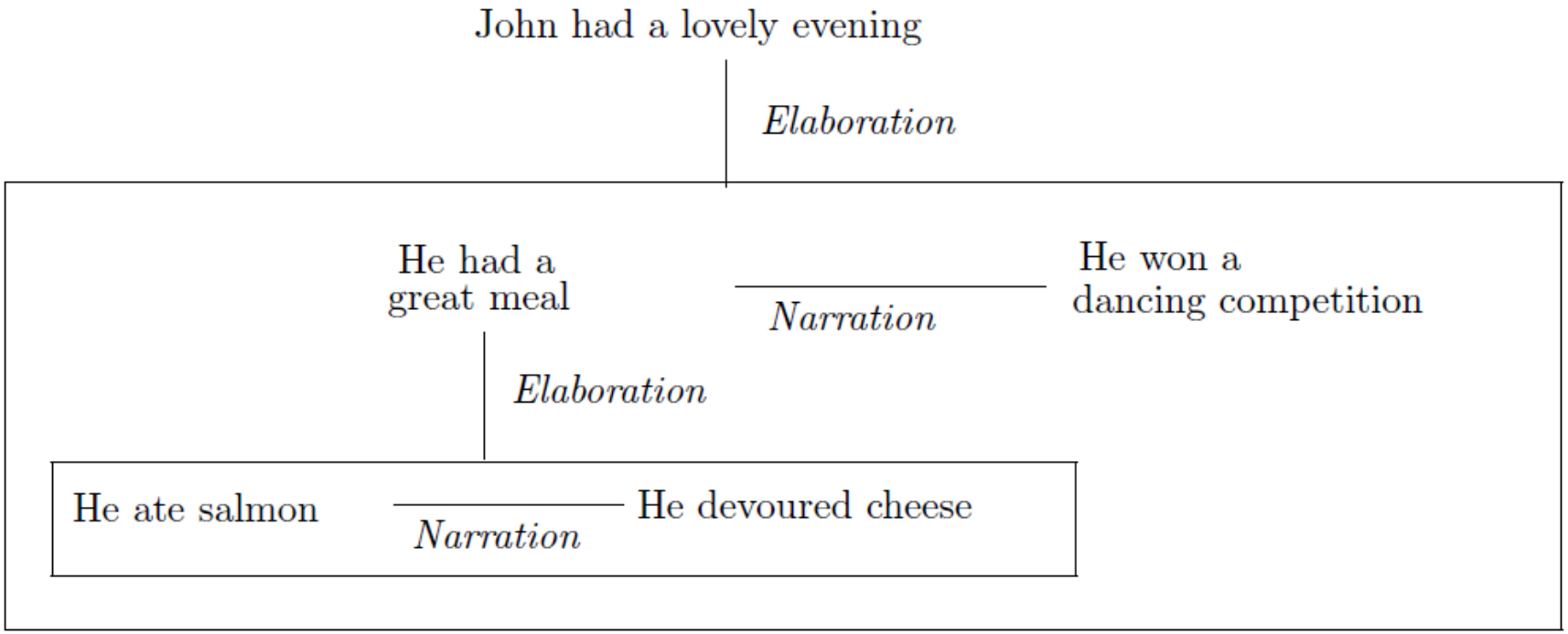
π_2. He had a great meal.

π_3. He ate salmon.

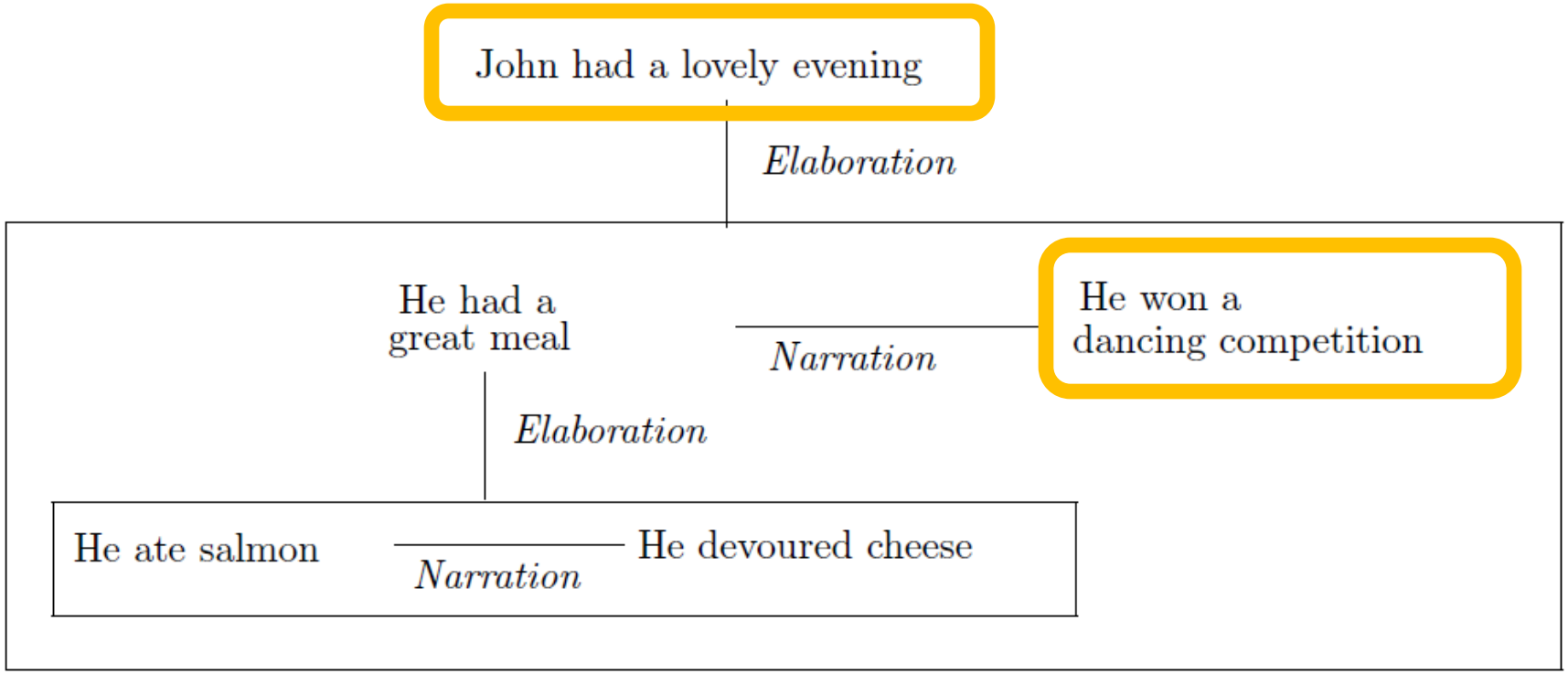
π_4. He devoured lots of cheese.

π_5. He won a dancing competition.

π_6. ??It was a beautiful pink.


John had a lovely evening
Elaboration
He had a great meal
Narration
He won a dancing competition
Elaboration
He ate salmon
Narration
He devoured cheese

The right-frontier constraint



Presuppositions

1. If baldness is heriditary, then **Jack's son** is bald.

   → Jack has a son. If baldness is heriditary, then he is bald.

1. If Jack has a son, then **Jack’s son** is bald.

→ Jack has a son. If Jack has a son, then Jack’s son is bald.

1. If Jack has a son, then **Jack's son** is bald.

→ Jack has a son. If Jack has a son, then Jack's son is bald.

→ If Jack has a son, then he is bald.

van der Sandt (1992): Accommodate presuppositions in the highest context, as long as the results is *informative* and *consistent.*

1. If John goes diving, he’ll take his regulator.

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

2. If David scuba dives, he'll take his dog.

   → David has a dog. If he scuba dives, he'll take the dog with him.

van der Sandt (1992): Accommodate presuppositions in the highest context, as long as the results is *informative* and *consistent.*

ONLY predicts 2 !!!

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

2. If David scuba dives, he'll take his dog.

   → David has a dog. If he scuba dives, he'll take the dog with him.

Beaver: (1996) Plausibility constraint on presuppositions

If domain knowledge predicts a dependency between the information in the antecedent of the conditional and the presupposition, prefer a conditonal presupposition.

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

2. If David scuba dives, he'll take his dog.

   → David has a dog. If he scuba dives, he'll take the dog with him.

Beaver: (1996) Plausibility constraint on presuppositions

Choose the more plausible interpretation.
If domain knowledge predicts a dependency between the information in the antecedent of the conditional and the presupposition, prefer a conditonal presupposition.

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

2. If David scuba dives, he'll take his dog.

   → David has a dog. If he scuba dives, he'll take the dog with him.

Beaver: (1996) Plausibility constraint on presuppositions

Choose the more plausible interpretation.
If domain knowledge predicts a dependency between the information in the antecedent of the conditional and the presupposition, prefer a conditonal presupposition.

Unfortunately, seems to predict conditional presupposition for 2. Wide scope reading says David owns a dog. That is less plausible than the more conservative narrow scope reading.

2. If David scuba dives, he'll take his dog.
   → David has a dog. If he scuba dives, he'll take the dog with him.

SOLUTION: Take coherence relations into account.

Asserted content is only coherent if it is rhetorically connected to the rest of the discourse.

Presupposed content is also only coherent if rhetorically connected.

Prefer discourse interpretations that maximize rhetorical links.

2. If David scuba dives, he'll take his dog.
   → David has a dog. If he scuba dives, he'll take the dog with him.

Maximize Discourse Coherence (or mdc)

Discourse is interpreted so as to maximize discourse coherence, where the ranking among interpretations are encapsulated in the following principles:

1. All else being equal, the more rhetorical connections there are between two items in a discourse, the more coherent the interpretation.

2. All else being equal, the more anaphoric expressions whose antecedents are resolved, the higher the quality of coherence of the interpretation.

3. Some rhetorical relations are inherently scalar… All else being equal, an interpretation which maximizes the quality of its rhetorical relations is more coherent than one that doesn't.

1. If David scuba dives, he'll take his regulator.

   → If David scuba dives, he has a regulator, and he'll take it with him.

- Relation is *Consequence,* triggered by ´if´
- Rhetorical relations of *Consequence* are better if John´s scuba diving connects to the content that he has a regulator
- This would be a better *Consequence.* There is **added value** in interpreting the presupposition as having narrow scope

1. If David scuba dives, he'll take his dog.

   → David has a dog. If David scuba dives, and he'll take it with him.

- Relation is still *Consequence,* triggered by ´if´

- But now, connecting David's dog to his scuba diving doesn't improve *Consequence,* there is no world knowledge that connects these two, so there is no reason to depart from the standard analysis of wide scope/global accommodation.

Adding rhetorical relations to DRT

- SDRSs : Segmented Discourse Representation Structures
- Both the coherence relation and the two segments it takes as arguments gets represented as speech act discourse referents

Max fell. John pushed him.

$\langle A, \mathcal{F}, LAST \rangle$, where:

- $A = \{\pi_0, \pi_1, \pi_2\}$
- $\mathcal{F}(\pi_1) = \begin{array}{|l|}\hline x, e_{\pi_1} \\ \hline max(x), \\ fall(e_{\pi_1}, x),\ e_{\pi_1} \prec n \\ \hline \end{array}$

  $\mathcal{F}(\pi_0) = Explanation(\pi_1, \pi_2)$

  $\mathcal{F}(\pi_2) = \begin{array}{|l|}\hline y, e_{\pi_2} \\ \hline john(x),\ push(e_{\pi_2}, y, x) \\ e_{\pi_2} \prec n \\ \hline \end{array}$
- $LAST = \pi_5$

π_1. John had a great evening last night.

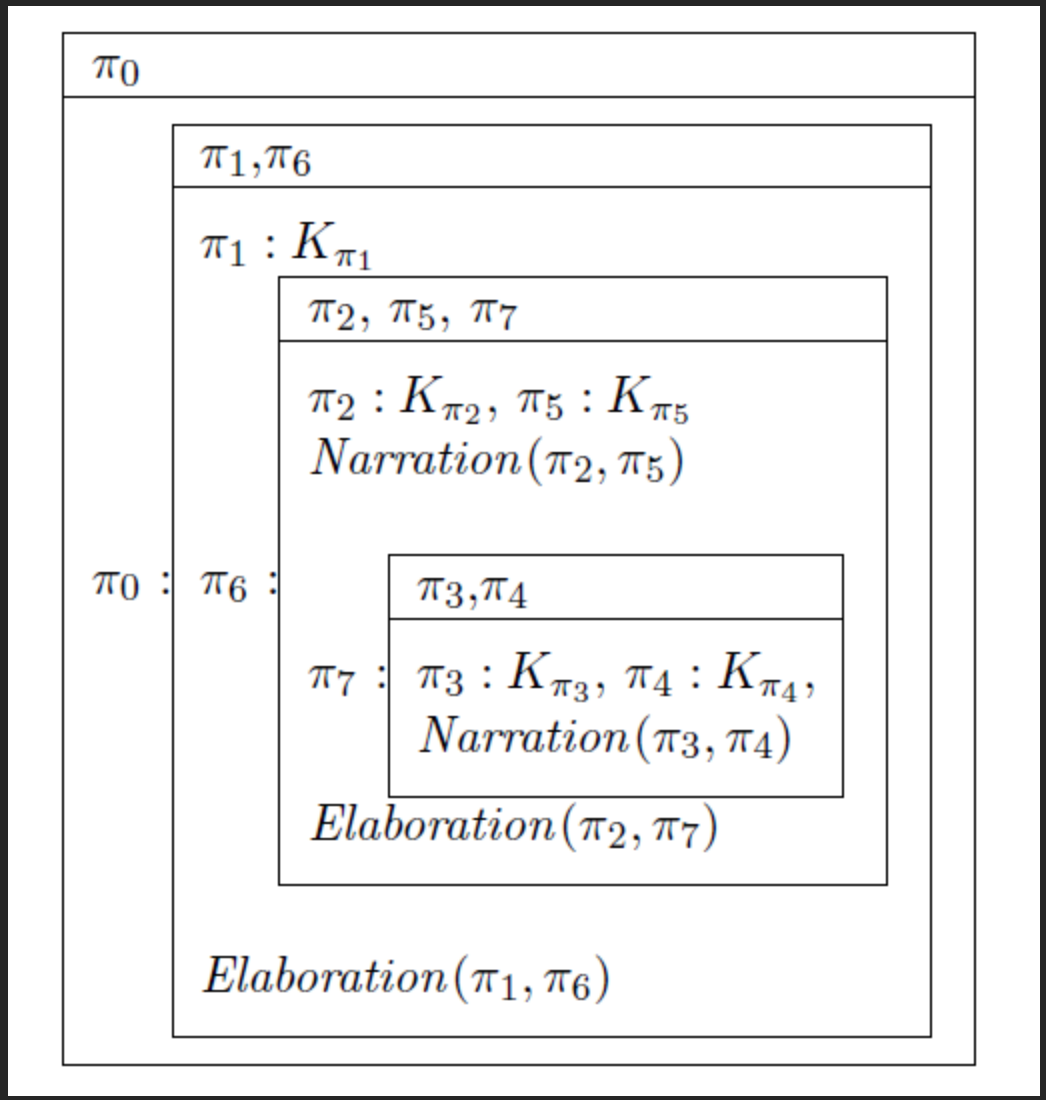
π_2. He had a great meal.

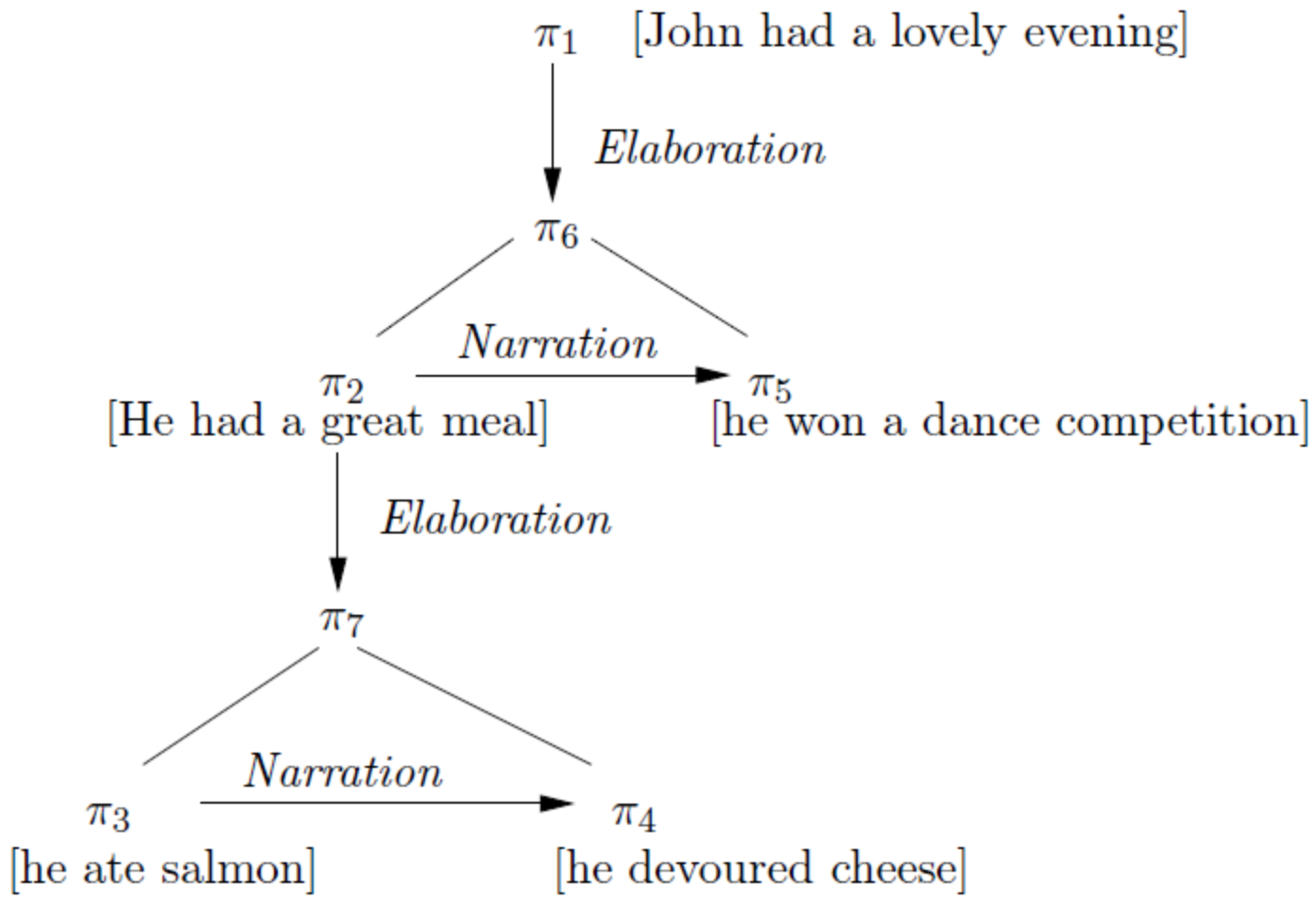
π_3. He ate salmon.

π_4. He devoured lots of cheese.

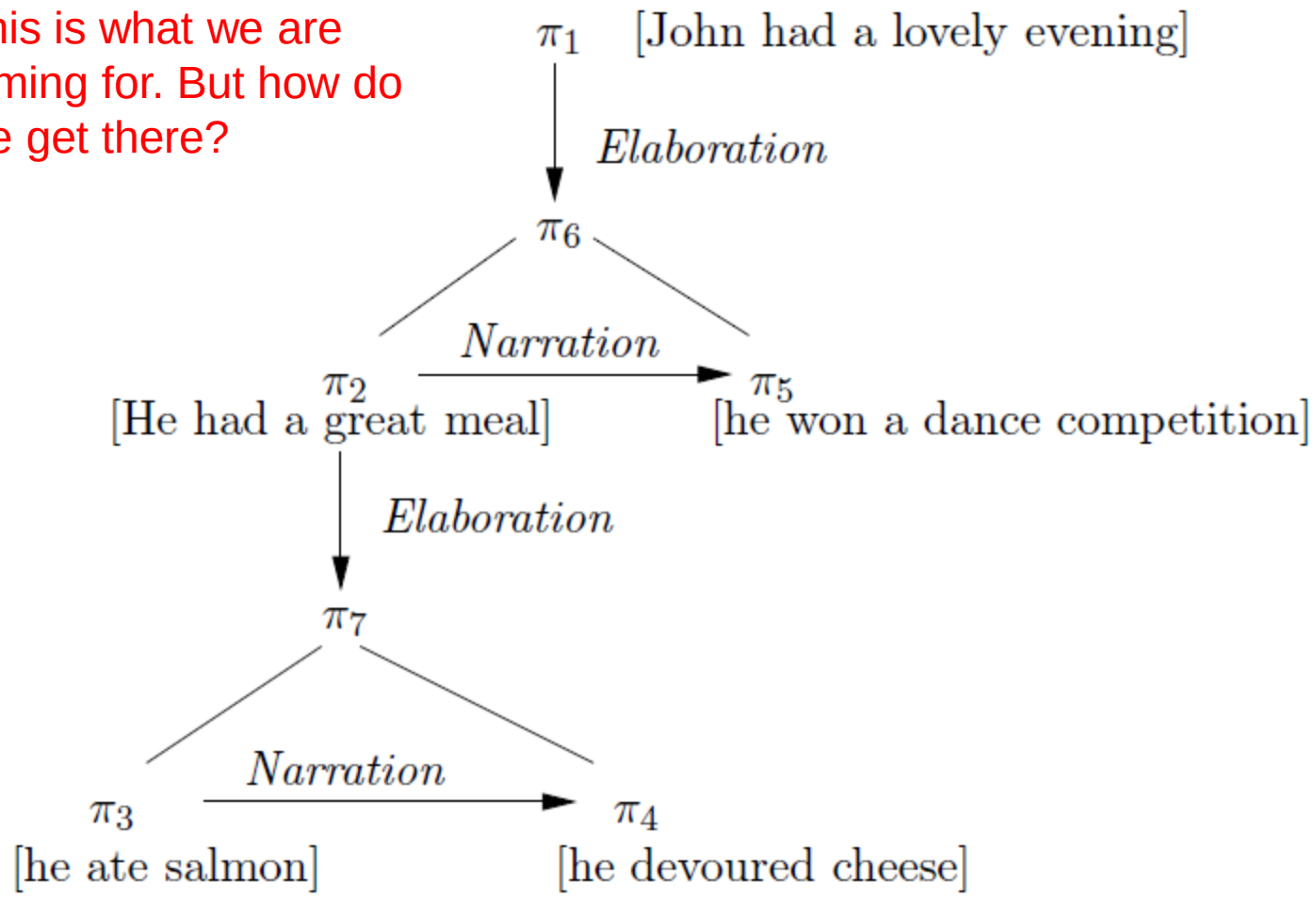
π_5. He won a dancing competition.

π_6. ??It was a beautiful pink.




π_1 [John had a lovely evening]
Elaboration
π_6
Narration
π_2
[He had a great meal]
π_5
[he won a dance competition]
Elaboration
π_7
Narration
π_3
[he ate salmon]
π_4
[he devoured cheese]

This is what we are aiming for. But how do we get there?



We have our text. How do we get from the text to the final representation in SDRT?

Unlike the previous theories (G&S, RST) we are going to be explicit about how both semantic and rhetorical interpretations combine to a final representation.

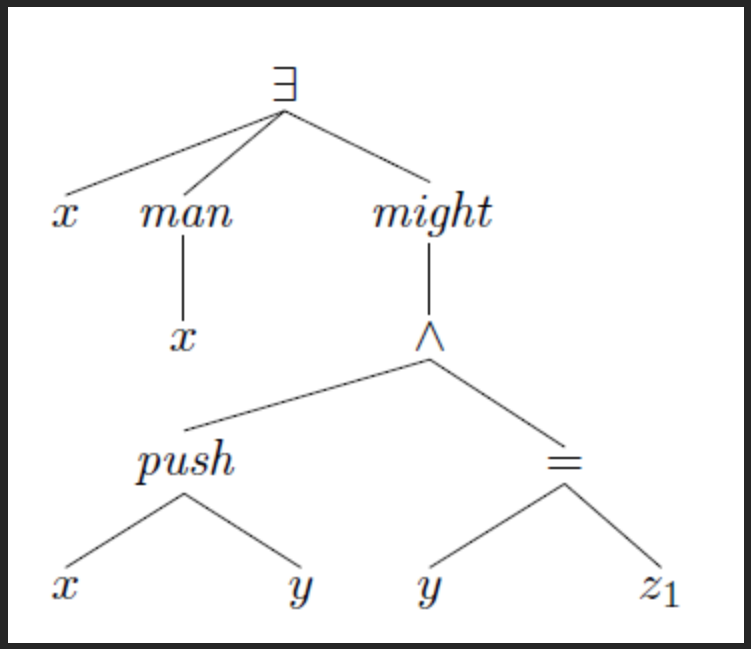
There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.

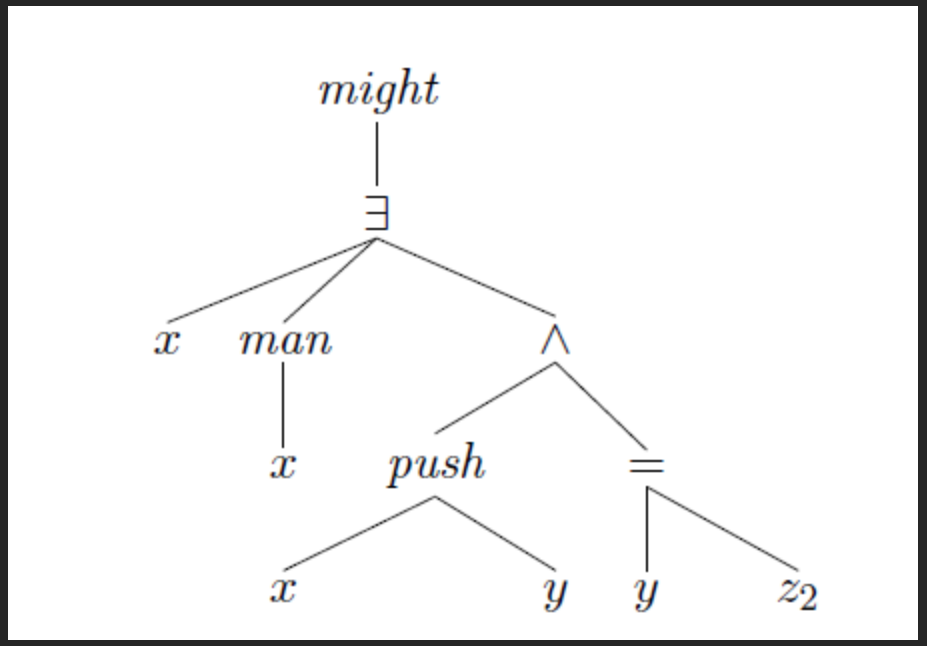
Donald Rumsfeld

Language of Underspecified Logical Forms (L_{ULF})

- In semantic interpretation we encounter many known unknowns.
- We can't always immediately resolve them
- That's why **everybody loves underspecification**
- And in fact, on main point of incorporating Rhetorical information in your theory is that it is supposed to help resolve some 'known unknowns'
-

A man might push him.





L_{ulf} can also express underspecified information about rhetorical connections.

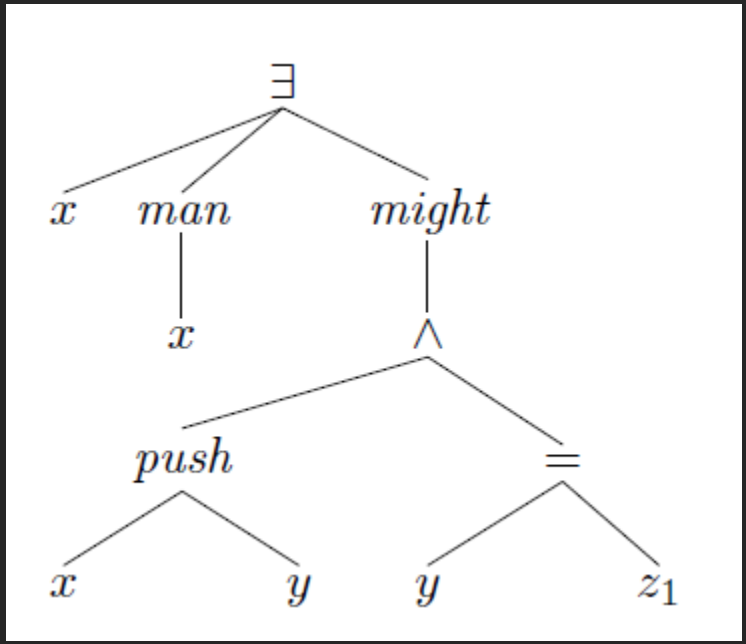
?(π_1; π_2; π_0)

π_1 and π_2 are rhetorically connected but we don't know the value of the rhetorical relation π_0 (`?' is a higher-order variable).

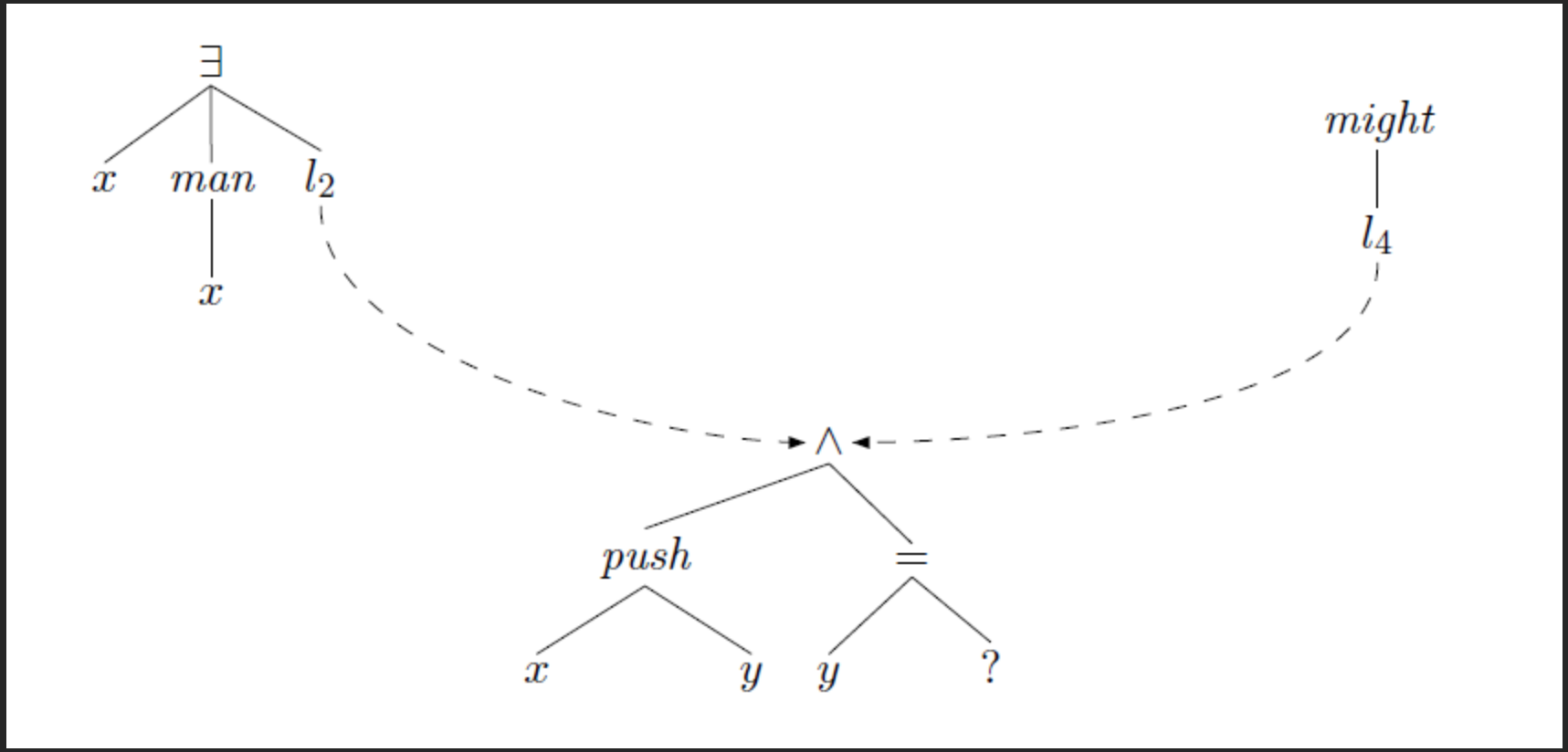
Underspecified Logical Forms

This ambiguity is represented as an ULF in L_{ULF}

l_1 and l_2 are nodes in the tree



$$l_1 : \exists(x, man(x), l_2) \wedge$$
$$l_3 : might(l_4) \wedge$$
$$l_5 : \wedge(l_6, l_7) \wedge l_6 : push(x, y) \wedge l_7 : x =? \wedge$$
$$outscopes(l_4, l_5) \wedge outscopes(l_2, l_5)$$


∃
x
man
l2
x
might
l4
∧
push
=
x
y
y
?

1. L_{ULF} gives us ULFs
   - these are like `preliminary SDRSs'
2. The glue logic then resolves ambiguities, fills in information, following MDC
3. Result is a fully specified interpretation

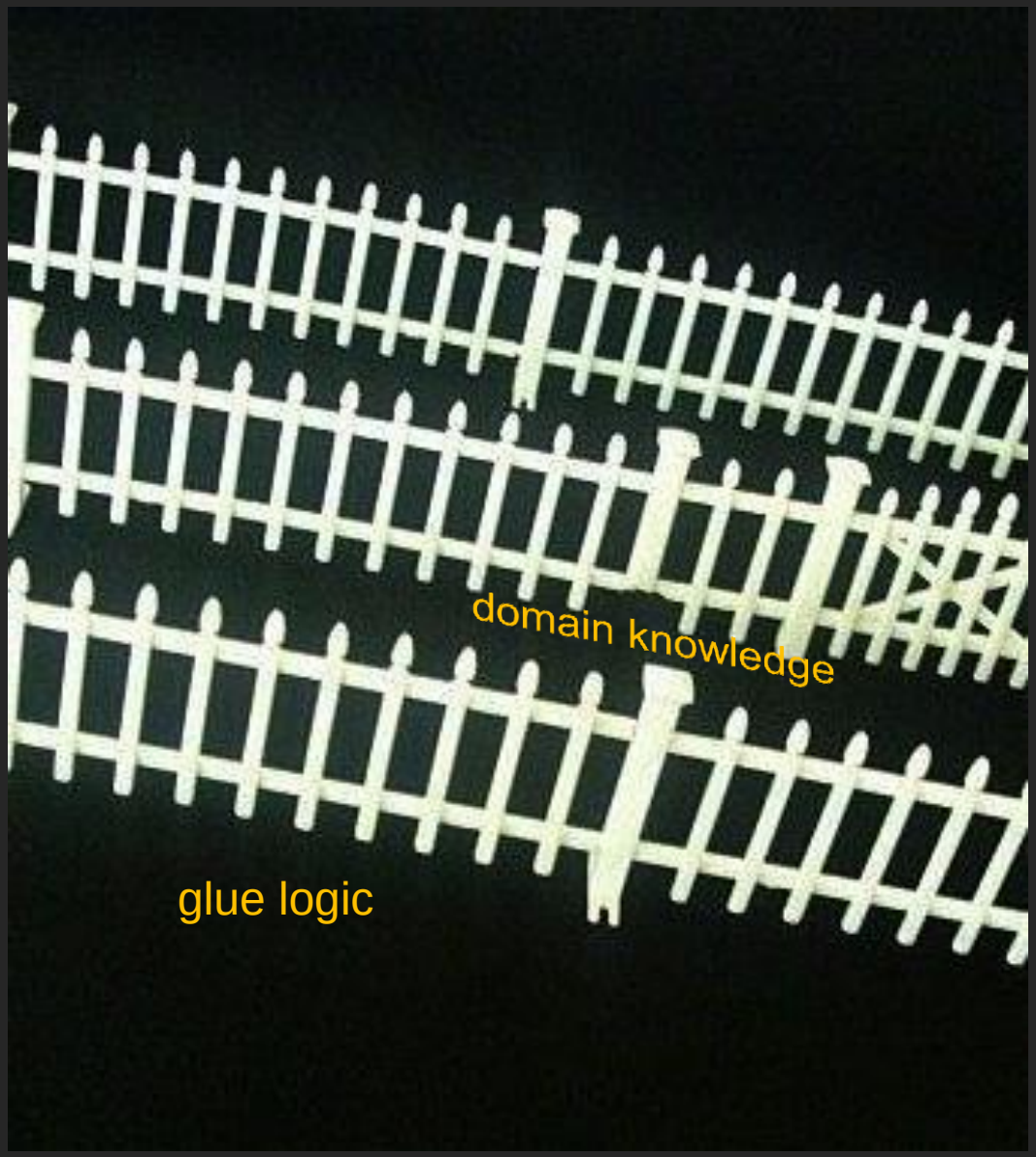
The Glue Logic

Determines 3 things:

1. the (pragmatically preferred) values of certain underspecied conditions that are generated by the grammar;

2. which labels are rhetorically connected to which other labels (this is equivalent to the task of text segmentation);

3. the values of the rhetorical relations.



‘’ Porous fences”

How the glue logic works

The ULF is enriched/resolved by using inferences over default axioms within the glue logic,

- “A > B “(which is read as If A then normally B).
- information about pragmatically preferred values of underspecifed conditions in a given ULF
- SDRT thus enriches dynamic semantics with contributions from pragmatics in a constrained way.
- In this way, very semantic information, such as pronoun interpretation, presupposition interpretation (word sense disambiguation and temporal anaphora) can be influenced in a prinicpled way by rhetorical information

What happens in an SDRT analysis:

1. The text gets interpreted into an Underspecified Logical Form
2. This creates a set of 'preliminary SDRSs'
3. From this set of preliminary SDRS's the Glue Logic determines the preferred interpretation:
   - the highest ranked SDRS's according to MDC are those with the minimum number of labels, the maximum number of rhetorical connections, the fewest unresolved semantic ambiguities (including anaphoric conditions) and no inconsistencies.
   - Glue logic axioms are used to determine which SDRS best fulfills the MDC.

What has SDRT done?

- It's very explicit, principled. Definately a theory.
  - it makes testable predications
- It outlines a full procedure of how we go from text to full interpretation
  - in implementation terms: Rule-based method
  - contrasts with how people have applied e.g. RST
    - humans do annotations
    - the knowledge and information that they use to do that annotation is implicit
    - unsupervised machine learning methods are applied to the annotations to try to see if there are useful patterns that can be used to make rules for automatic annotation
    - we guess at what information we should include in the annotation mark-up for input
    - definately dirty method, but so-far without so much success...

Does SDRT solve the problems talked about yesterday

Yes and no.

- Core semantics is separated from coherence structure
  - we can exploit semantics for those things that semantics does well
  - we can use coherence information to help when needed
- SDRT doesn't have intentional coherence relations
  - there is no "motivation", instead we would see it simply as its information relation,
- SDRT does have a distinction between coordinating and subordinating relations, similar to N and S in RST
- But its not clear if these are problems for SDRT
  - it's aims seem to be more modes than RST

Is SDRT easier to annotate?

"inspired by SDRT"

DISCOR and ANNODIS (French).

- Adam, Marianne Vergez-Couret: Exploiting naive vs expert discourse annotations: an experiment using lexical cohesion to predict Elaboration / Entity-Elaboration confusions

- Naive annotation: 86 texts; 3 annotators (other postgraduate students) Kappa: 0.4 (week to moderate inter-annotator agreement)