

Semantics and Pragmatics of NLP

Segmented Discourse Representation Theory

Alex Lascarides

School of Informatics
University of Edinburgh

Outline

- 1 Present an extension of DRT with rhetorical relations
 - Logic for representing discourse semantics
 - Logic for constructing logical forms

- 2 Apply SDRT to some semantics tasks

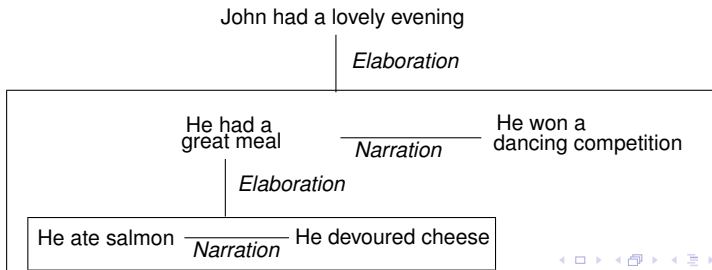
Claims

- ① Rhetorical relations are an essential component of discourse semantics
- ② *Constructing* logical form doesn't involve full access to the logic for *interpreting* logical form.
 - (1) a. There are unsolvable problems in number theory.
 - b. Any even number greater than two is equal to the sum of two primes, for instance.
- ③ In fact, constructing logical form has only partial access to:
 - Lexical semantics, domain knowledge, cognitive states etc.for similar reasons.

Need Rhetorical Relations: Some Motivating Data

Pronouns

- (2)
- John had a great evening last night.
 - He had a fantastic meal.
 - He ate salmon.
 - He devoured lots of cheese.
 - He won a dancing competition.
 - ??It was a beautiful pink.



More Motivation for Rhetorical Relations

Tense

- (3) Max fell. John helped him up.
- (4) Max fell. John pushed him.
- (5) John hit Max on the back of his neck. Max fell. John pushed him. Max rolled over the edge of the cliff.

Words

- (6) a. A: Did you buy the apartment?
b. B: Yes, but we rented it./ No, but we rented it.

Bridging

- (7) a. John took an engine from Avon to Dansville.
b. He picked up a boxcar./He also took a boxcar.

The Strategy

- 1 SDRSs: Extend DRT with rhetorical relations.
- 2 \mathcal{L}_{ulf} : Supply a separate logic for describing SDRSs (semantic underspecification).
- 3 *Glue logic*: Construct logical form for discourse via:
 - 1 default reasoning, over
 - 2 \mathcal{L}_{ulf} -formulae for clauses which are generated by the grammar and
 - 3 'shallow' representations of lexical semantics, domain knowledge, cognitive states. . .

Glue logic entails more consequences about content than the grammar does. These are *implicatures*.

Review of DRT

- 1 $f[\langle U, \emptyset \rangle]_M g$ iff $dom(g) = dom(f) \cup U$
- 2 $f[K \oplus \langle \emptyset, \gamma \rangle]_M g$ iff $f[K] \circ [\gamma]_M g$
- 3 $f[R(x_1, \dots, x_n)]_M g$ iff $f = g$ and $\langle f(x_1), \dots, f(x_n) \rangle \in I_M(R)$
- 4 $f[\neg K]_M g$ iff $f = g$ and there's no h such that $f[K]_M h$
- 5 $f[K \Rightarrow K']_M g$ iff $f = g$ and for all h such that $f[K]_M h$ there's an i such that $h[K']_M i$.

Logic of Information Content: Syntax

SDRS-formulae:

- DRSs
- $R(\pi, \pi')$, where R is a rhetorical relation and π and π' are labels.
- Boolean combinations of these

An SDRS is a structure $\langle A, \mathcal{F}, LAST \rangle$

- A is a set of labels
- \mathcal{F} maps labels to SDRS-formulae (i.e., labels tag content)
- $LAST$ is a label (of the last utterance)
- Where $Succ(\pi, \pi')$ means $R(\pi', \pi'')$ or $R(\pi'', \pi')$ is a literal in $\mathcal{F}(\pi)$: A forms a partial order under $Succ$ with a unique root.

SDRSs allow Plurality

Of Relations: *Contrast*(π_1, π_2), *Narration*(π_1, π_2)

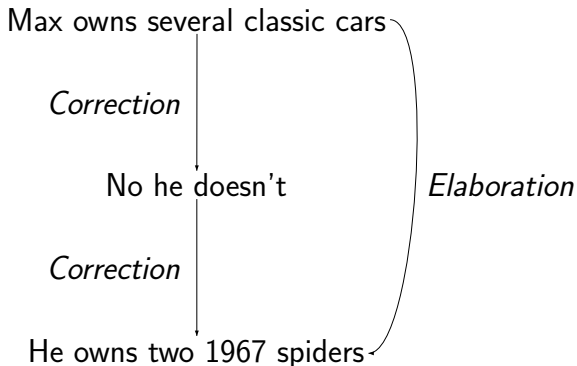
- (6) a. A: Did you buy the apartment?
b. B: Yes, but we rented it.

Of Attachment sites: *Correction*(π_2, π_3), *Elaboration*(π_1, π_3)

- (8) π_1 A: Max owns several classic cars.
 π_2 B: No he doesn't.
 π_3 A: He owns two 1967 Alfa spiders.

- A single utterance can make more than one *illocutionary contribution* to the discourse.

A Diagram



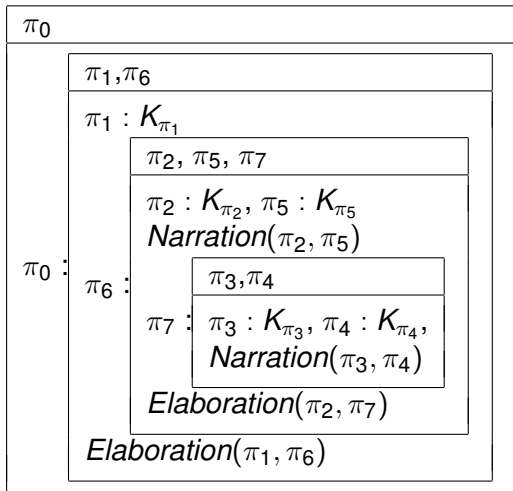
Example

- (2)
- 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 then won a dancing competition.

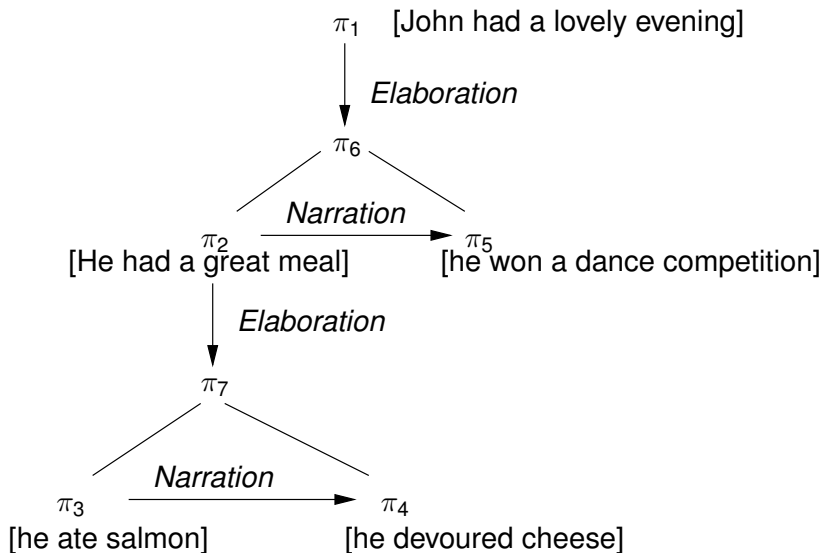
(2)' $\langle A, \mathcal{F}, LAST \rangle$, where:

- $A = \{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7\}$
- $\mathcal{F}(\pi_1) = K_{\pi_1}, \mathcal{F}(\pi_2) = K_{\pi_2}, \mathcal{F}(\pi_3) = K_{\pi_3},$
 $\mathcal{F}(\pi_4) = K_{\pi_4}, \mathcal{F}(\pi_5) = K_{\pi_5},$
 $\mathcal{F}(\pi_0) = \textit{Elaboration}(\pi_1, \pi_6)$
 $\mathcal{F}(\pi_6) = \textit{Narration}(\pi_2, \pi_5) \wedge \textit{Elaboration}(\pi_2, \pi_7)$
 $\mathcal{F}(\pi_7) = \textit{Narration}(\pi_3, \pi_4)$
- $LAST = \pi_5$

Other Ways of Showing This



Other Ways of Showing This



Availability: You can attach things to the right frontier

New information β can attach to:

- 1 The label $\alpha = LAST$;
- 2 Any label γ such that:
 - 1 $Succ(\gamma, \alpha)$; or
 - 2 $\mathcal{F}(I) = R(\gamma, \alpha)$ for some label I , where R is a subordinating discourse relation
(*Elaboration*, *Explanation* or \Downarrow)

We gloss this as $\alpha < \gamma$

- 3 Transitive Closure:
Any label γ that dominates α through a sequence of labels $\gamma_1, \dots, \gamma_n$ such that $\alpha < \gamma_1, \gamma_1 < \gamma_2, \dots, \gamma_n < \gamma$.

Available Anaphora (Not *Parallel* or *Contrast*)

Situation:

- $\beta : K_\beta$;
- K_β contains anaphoric condition φ .

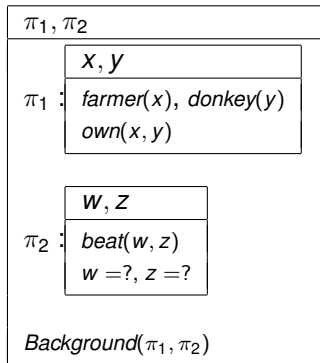
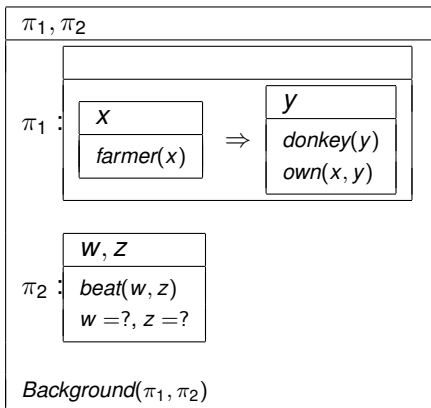
Available antecedents are:

- 1 in K_β and DRS-accessible to φ
- 2 in K_α , DRS-accessible to any condition in K_α , and there is a condition $R(\alpha, \gamma)$ in the SDRS such that $\gamma = \beta$ or $Succ * (\gamma, \beta)$ (where R isn't structural).

Antecedent must be DRS-accessible on the right frontier

Example: Uses Accessibility from DRT

- (9) Every farmer owns a donkey. ??He beats it.
 (10) A farmer owns a donkey. He beats it.



- *Contrast* and *Parallel* work a bit differently:

Improvement on DRT: The Dansville Example

- (7) π_1 John took an engine to Dansville. (π_1)
 π_2 He picked up a boxcar (π_2)
 π_3 It had a broken fuel pump (π_3)

DRT:

- Flat structure:
An engine is accessible to it

SDRT:

- *Narration*(π_1, π_2);
- So π_1 isn't available to π_3 : $R(\pi_1, \pi_3)$ can't hold for any R
- So the engine is not an available antecedent to *it*

Semantics: Veridical Relations

Speech Acts!!

- Satisfaction Schema for Veridical Relations:

$$f[R(\pi_1, \pi_2)]_M g \text{ iff } f[K_{\pi_1}]_M \circ [K_{\pi_2}]_M \circ [\phi_{R(\pi_1, \pi_2)}]_M g$$

Veridical:	<i>Explanation, Elaboration, Background, Contrast, Parallel, Narration, Result, Evidence...</i>
Non-veridical:	<i>Alternation, Consequence</i>
Divergent:	<i>Correction, Counterevidence</i>

Some Meaning Postulates: Defining $\phi_{R(\alpha,\beta)}$ for various R

- Axiom on Explanation:

(a) $\phi_{Explanation(\alpha,\beta)} \Rightarrow (\neg e_\alpha \prec e_\beta)$

(b) $\phi_{Explanation(\alpha,\beta)} \Rightarrow (event(e_\beta) \Rightarrow e_\beta \prec e_\alpha)$

Max went to bed. He was sick. Max fell. John pushed him.

- Axiom on Elaboration:

$$\phi_{Elaboration(\alpha,\beta)} \Rightarrow Part-of(e_\beta, e_\alpha)$$

Max ate a big dinner. He had salmon.

More Meaning Postulates

- Axiom on Background:

$$\phi_{\text{Background}(\alpha,\beta)} \Rightarrow \text{overlap}(e_\beta, e_\alpha)$$

Max entered. The room was dark.

- Axiom on Narration:

$$\phi_{\text{Narration}(\alpha,\beta)} \Rightarrow \text{(a) } e_\alpha \prec e_\beta \text{ and}$$

- (b) things don't move location between
the end of e_α and start of e_β
(unless adverbials indicate otherwise).

Max went to Paris. He visited a friend.

A Simple Example

- (7) π_1 John took an engine from Avon to Dansville.
 π_2 He picked up a boxcar.

Grammar produces (slightly simplified):

π_1	
	j, x, e_1, a, d
$\pi_1 :$	$john(j), engine(x),$ $avon(a), dansville(d)$ $take(e_1, j, x), e_1 \prec n$ $from(e_1, a), to(e_1, d)$

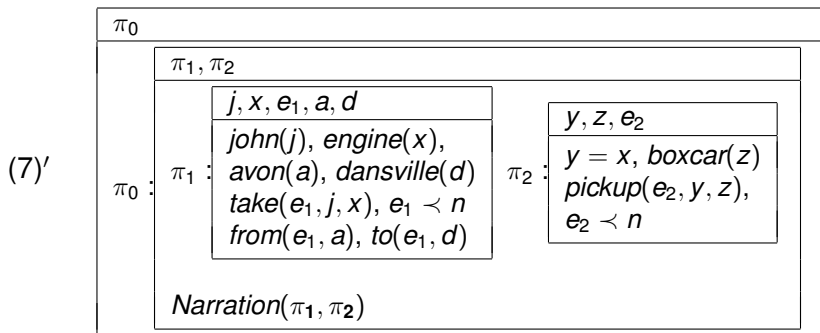
π_2	
	y, z, e_2
$\pi_2 :$	$y = ?,$ $boxcar(z)$ $pickup(e_2, y, z)$ $e_2 \prec n$

Discourse Update: Assume *coherence*!

- Only π_1 is available; so $\pi_0 : ?(\pi_1, \pi_2)$;
 so $y = x$ *whatever* the rhetorical relation.

The Final SDRS

- *Narration*(π_1, π_2) inferred on basis of various clues (more later).
- This has spatio-temporal consequences.



Truth Conditions

- 1 $f[K_{\pi_0}]g$ iff $f[Narration(\pi_1, \pi_2)]g$; iff there are h and k such that:
 - 1 $f[K_{\pi_1}]h$; and
 - 2 $h[K_{\pi_2}]k$; and
 - 3 $k[\phi_{Narr(\pi_1, \pi_2)}]g$
- 2 By Axiom on Narration; (3c) only if
 - 1 $k[e_1 \prec e_2]k$;
 - 2 $k[in(z, d)]k$

So (7)' entails more than the compositional semantics of the clauses: Implicatures!

Comparison with DRT

Flat Structure!

j, x, y, a, d, e_1, e_2

$john(j), engine(x), boxcar(y), avon(a), dansville(d)$
 $take(e_1, j, x), pickup(e_2, j, y), e_1 \prec e_2 \prec n$

Advantage of SDRT:

- Semantics of *Narration* models implicatures:
Boxcar is in Dansville.
- And it predicts incoherence.

(11) ??Max entered the room. Mary dyed her hair black.

- Better predictions about pronouns:

(7) John took an engine to Dansville. He picked up a boxcar.

??It had a broken fuel pump.

Constructing Logical Form: A Preview

The grammar

- Produces underspecified LFs for clauses (e.g., $x = ?$);
These are *partial descriptions* of logical forms (separate logic)

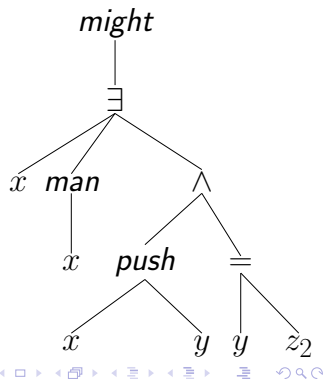
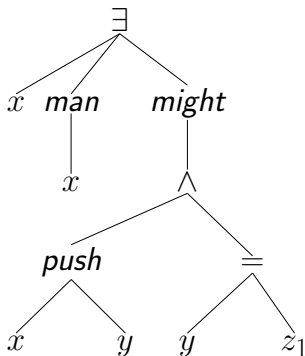
Glue Logic:

- Can only access ULFs;
- Performs the following co-dependent inferences:
 - 1 Infer (preferred) values of underspecified conditions generated by the grammar;
 - 2 Infer what's rhetorically connected to what;
 - 3 Infer the values of the rhetorical relations

Some Formal Details: Underspecification

(12) A man might push him.

Assuming only z_1 and z_2 available, there are four LFs. Here are two of them:



Strategy: Introduce \mathcal{L}_{ulf}

Want to describe just the four trees and no others. So:

- Reify nodes of the tree
- So you can talk about scope independently of predicates
- Introduce variables (written ?) to show where values of symbols are unknown.

(12) A man might push him.

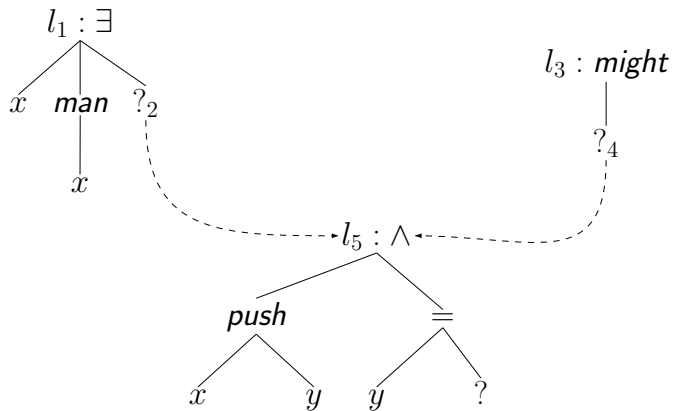
(12)' $l_1 : \exists(x, \text{MAN}(x), ?_2) \wedge$

$l_3 : \text{MIGHT}(?_4) \wedge$

$l_5 : \wedge(l_6, l_7) \wedge l_6 : \text{push}(x, y) \wedge l_7 : x = ? \wedge$

$\text{OUTSCOPES}(?_4, l_5) \wedge \text{OUTSCOPES}(?_2, l_5)$

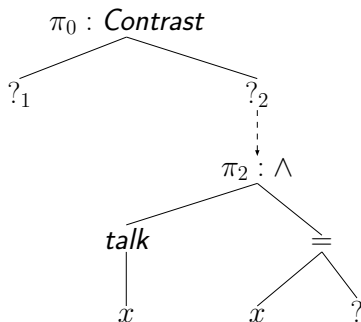
Graphically



Rhetorical Underspecification

(13) But he talks

(13)' $\pi_0 : \text{Contrast}(\text{?}_1, \text{?}_2) \wedge$
 $\pi_2 : \wedge(l_1, l_2) \wedge l_1 : \text{TALK}(x) \wedge l_2 : x = ? \wedge$
 $\text{OUTSCOPES}(\text{?}_2, \pi_2)$



Semantics of the ULF-Language \mathcal{L}_{ulf}

- Models are the trees.
- So each model corresponds to a unique SDRS.
- $M \models_{\mathcal{L}_{ulf}} \phi$ means
 ϕ is a (partial) description of the SDRS M .

Comparison of Semantics:

SDRSs: dynamic, first-order, modal (though not here)

ULFs: static, extensional, finite first-order

ULFs 'access' the *form* of LFs, but not their entailments
(according to the logic of LFs)

From Clauses to Discourse

Discourse update is used to perform three interdependent tasks:

Task 1: Attachment Sites:

- a Which π' in the context are possible attachment sites? **Done!**
- b Of these, which does π *actually* attach to?

Task 2: **Rhetorical Relations:** If π attaches to π' , then which rhetorical relation do we use?

Task 3: **Augment Content:** Apart from old and new information to be added to the update:

- a What underspecifications do we resolve; and
- b What else do we add?

- Rhetorical Relations aren't always linguistically marked.
- They depend on:
 - Compositional and lexical Semantics
 - World Knowledge
 - Cognitive states. . .
- We need to:
 - Encode knowledge used to infer rhetorical relations.
 - Use a logic that supports the inferences we need.

Temporal Relations & Defeasible Reasoning

(14) Max took an aspirin. He was sick.

Background and Explanation

(15) Max took an aspirin overdose. He was sick.

Result

- “states are backgrounds” applies to both.
- But this is overridden in (15).
- These are default rules!

Default guess can get Corrected

- (16)
- a. A: John went to jail. He was caught embezzling funds from the pension plan.
 - b. B: No! John was caught embezzling funds, but he went to jail because he was convicted of tax fraud.

Default Rules in the Glue Logic

- $A > B$ means “If A then normally B .”
- The nonmonotonic validity, \vdash_g , supports intuitive patterns of commonsense reasoning.

The glue logic axioms:

- $(\lambda : ?(\alpha, \beta) \wedge \text{some stuff}) > \lambda : R(\alpha, \beta)$

To make things computable:

- ‘some stuff’ rendered with *descriptions* of formulae from richer information sources (e.g., SDRSs, domain knowledge...).

Patterns of Common Sense Reasoning

Closure on the Right:

$A > B, B \rightarrow C \vdash A > C$

Lions walk

Things that walk must have legs

Lions have legs.

Defeasible Modus Ponens:

$A > B, A \sim_g B$

If Tweety is a bird,
then normally

Tweety flies

Tweety is a bird

Tweety flies

Knowledge Conflict

Penguin Principle:

If $C \vdash A$ then

$A > B, C > \neg B, C \vdash_g \neg B$

Nixon Diamond:

$A > B, C > \neg B, A, C \not\vdash_g B$
(or $\neg B$)

If Tweety is a penguin,
then Tweety is a bird
If Tweety is a bird,
then normally Tweety flies
If Tweety is a penguin,
then normally
Tweety doesn't fly

Tweety is a Penguin
Tweety doesn't fly

If Nixon is a Quaker,
then normally
he's a pacifist
If Nixon is a Republican,
then normally
he's a non-pacifist
Nixon is a Quaker
Nixon is a Republican

* Nixon is a (non)-pacifist

Some Glue Logic Axioms

- Narration
 $(\lambda : ?(\alpha, \beta) \wedge \mathbf{occasion}(\alpha, \beta)) > \lambda : \mathbf{Narration}(\alpha, \beta)$
- Scripts for Occasion
 $(\lambda : ?(\alpha, \beta) \wedge \phi(\alpha) \wedge \psi(\beta)) > \mathbf{occasion}(\alpha, \beta).$
- Explanation
 $(\lambda : ?(\alpha, \beta) \wedge \mathbf{cause}_D(\beta, \alpha)) > \lambda : \mathbf{Explanation}(\alpha, \beta)$
- Causation and Change
 $(\mathbf{change}(e_\alpha, y) \wedge \mathbf{cause-change-force}(e_\beta, x, y)) \rightarrow \mathbf{cause}_D(\beta, \alpha)$

An Example of Narrative

The Logical Form of the Sentences

(3) Max fell. John helped him up.

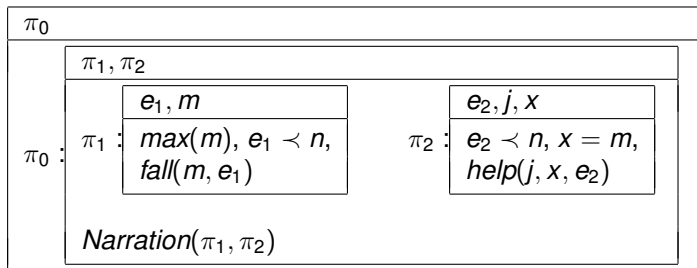
π_1 $max(m), e_1 \prec n, fall(m, e_1)$

π_2 $e_2 \prec n, x =?, help(j, x, e_2)$

Assume Coherence: $\pi_0 :?(\pi_1, \pi_2)$

- 1 $x =?$ resolves to $x = m$
- 2 scriptal information $\vdash \sim OCCASION(\pi_1, \pi_2)$
- 3 DMP on Narration yields $\pi_0 : NARRATION(\pi_1, \pi_2)$

Minimal SDRS Satisfying the \vdash_g -consequences



- By ‘minimal’ I mean minimum number of nodes.
- This entails $e_1 \prec e_2$; John and Max in the same ‘place’.

Another Narration

- (7)
- a. John took an engine from Avon to Dansville.
 - b. He picked up a boxcar. . .
- DMP on `Narration` gives $Narration(\alpha, \beta)$.
 - The spatial constraint on `Narration` means that John is in Dansville when he starts to pick up the boxcar.
 - So by the lexical semantics of *pick up*, this means that the boxcar is in Dansville (when it's picked up).
 - This is a bridging inference!
 - $e_\alpha \prec e_\beta$ is entailed too.

An Explanation

(4) Max fell. John pushed him.

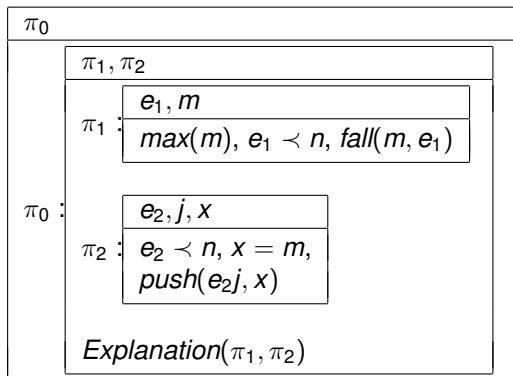
π_1 $max(m), e_1 \prec n, fall(m, e_1)$

π_2 $e_2 \prec n, x = ?, push(j, x, e_2)$

Assume coherence: $\pi_0 : ?(\pi_1, \pi_2)$

- MP on Causation and Change: $cause_D(\pi_2, \pi_1)$
- DMP on Explanation: $\pi_0 : Explanation(\pi_1, \pi_2)$ is inferred.

The SDRS



Entailments: Both clauses are true; $e_2 \prec e_1$

Constructing SDRSs: Simple Discourse Update +

- You update a set σ of SDRSs with $\lambda :?(\alpha, \beta)$, where \mathcal{K}_β is the ULF for β
- $\text{TH}(\sigma) =_{\text{def}} \{\phi : \forall \mathbf{s} \in \sigma, \mathbf{s} \models_{\mathcal{L}_{\text{ulf}}} \phi\}$
- The result is a set σ' of SDRSs
- + is monotonic: $\sigma' \subseteq \sigma$ (or $\text{TH}(\sigma) \subseteq \text{TH}(\sigma')$)

$$\sigma + \lambda :?(\alpha, \beta) = \{\tau : \text{if } \text{Th}(\sigma), \mathcal{K}_\beta, \lambda :?(\alpha, \beta) \vdash_g \phi \text{ then } \tau \models_{\mathcal{L}_{\text{ulf}}} \phi\}$$

So you just add glue-logic consequences to the ULFs,
and $\tau \in \sigma'$ must satisfy those.

Rhetorical Relations are in the Update!

Suppose:

$$Th(\sigma), \mathcal{K}_\beta, \lambda : ?(\alpha, \beta) \vdash_g \lambda : R(\alpha, \beta)$$

Then:

$$\forall \tau \in \text{update}, \mathcal{F}_\tau(\lambda) \rightarrow R(\alpha, \beta)$$

This justified putting *Narration*(π_1, π_2) in SDRS for (3).

Constructing SDRs: Discourse Update

$update_{\text{sdr}}$ abstracts over choices about what attaches to what:

- 1 Make a new choice about what β attaches to (you can choose more than one label).
- 2 Compute the results of $+$ with your choice.
- 3 Go back to step 1 and repeat. . .
- 4 $update_{\text{sdr}}(\sigma, \mathcal{K}_\beta)$ is the union of all the results from step 2

Conservative!

- $update_{\text{sdr}}(\sigma, \mathcal{K}_\beta)$ doesn't pick what β *actually* attaches to;
- Nor does it pick which underspecifications to resolve

So How do We Make Remaining Choices?

Go for as many connections as possible:

- (17)
 - a. Max had a lovely evening.
 - b. He had a fantastic meal.
 - c. He ate salmon

- (6)
 - a. A: Did you buy the apartment?
 - b. B: Yes, but we rented it.

Prefer discourse relations higher in the (discourse) ranking:

- (18)
 - a. John annoys Fred.
 - b. He calls all the time/never calls/
calls on Fridays.

Maximise Discourse Coherence (MDC)

An SDRS is better if it:

- 1 Contains relations higher in the 'ranking'
- 2 Contains more rhetorical relations
- 3 Contains fewer underspecifications
- 4 Has a minimal number of labels.

*Always interpret discourse so that coherence is maximised!
I.e., Prefer highest-ranked SDRSs in $\text{update}_{\text{sdrts}}$.*

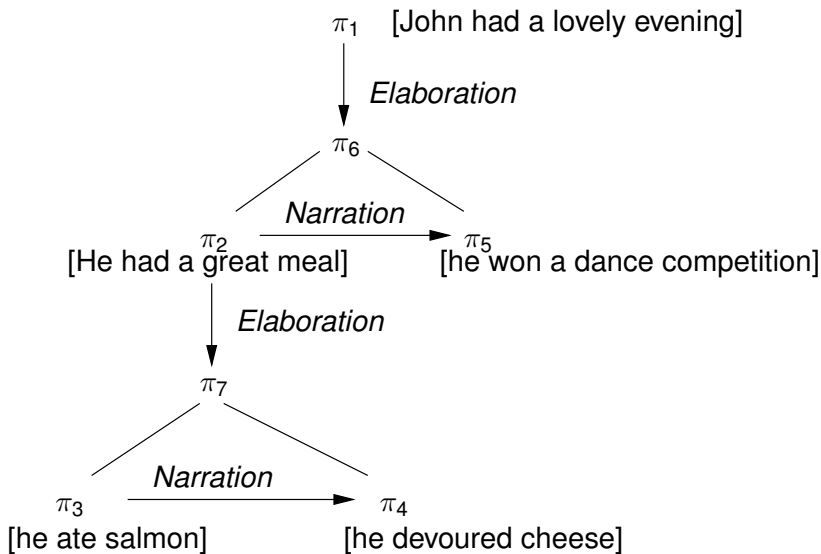
Discourse Popping

- (2)
- π_1 Max had a lovely evening last night.
 - π_2 He had a fantastic meal.
 - π_3 He ate salmon.
 - π_4 He devoured lots of cheese.
 - π_5 He won a dancing competition.

Attaching π_5 :

- Alternative choices of attachment sites would not have maximised rhetorical connections or minimised underspecification

A Diagram



Word Senses

- (6) a. A: Did you buy the apartment?
b. B: Yes, but we rented it.

If *rent* is rent-from:

- Get *Contrast*, but nothing else.

If *rent* is rent-to:

- Get *Contrast* and *Narration*

MDC: update resolves *rent* to rent-to sense,
because this gets more connections.

Summary

- There are problems with DRT's account of anaphora:
 - 1 Needs discourse structure given by rhetorical relations.
 - 2 LF construction should involve reasoning with non-linguistic information.
- There are also problems with the unmodular way AI-theories like Hobbs *et al* tackle task 2.
- SDRT attempts to combine 'best practices' of both:
 - 1 Improves constraints on anaphora for both frameworks.
 - 2 Maintains a separation between the logic of LF construction and the logic of LF interpretation.
 - 3 Choices modelled within the logic rather than via weights.