



Twitter is a service to communicate and a quick, frequent and doing?

Sunday, August 30, 2006

Martians invade earth

Incredible as it may seem, it has been confirmed that a large martian invasion fleet has landed on earth tonight.

First vessels were sighted over Great Britain, Denmark and Norway already in the late evening from where, as further reports indicate, the fleet

headed towards the North Pole and Santa Claus was taken hostage by the invaders.

Afterwards they split apart in order to approach most major cities around the earth. The streets filled as thousands fled their homes, many only wearing their pajamas...

Ren
follo
imp

The
that
rela
the
beh
of a
exp
in li
ite

the carnival of feminists

24 HOURS OF LIVE IN THE COMPANY OF FEMINISTS. 480 HOURS OF BLOG BATH BATHING.

About the carnival

The Carnival of Feminists is held on the first and third Wednesday of each month. Hosted by a different blogger for each edition, it aims to showcase the finest feminist posts from around the blogosphere. Posts will usually have been made in the period since the last carnival. (Only one nomination per blog please!)

The sixth and last edition will be on [Inappropriate](#) on January 4. Send submissions to jen AT reappropriate DOT com by January 3 with 'feminist carnival' in the subject line.

The Carnival hopes to build the profile of feminist blogging, to direct extra traffic to all participating bloggers, but particularly newer bloggers, and to build networks among feminist bloggers.

How to define feminist? Well that is up to each host. I don't intend to get

DEFINABLES

- No. 11 will be on [Indian Writing](#) on March 8
- No. 12 will be on [Mind the Gap](#) on February 22
- No. 13 will be on [TBC](#) on February 8
- (See I'm shifting the weeks to spread the thing out...)
- No. 14 will be by [Laurie de Fermoy](#) on January 18
- No. 15 will be [Inappropriate](#) on January 4
- No. 16: [Inappropriate](#)
- No. 17: [The Happy Feminist](#)
- No. 18: [Blue Bird](#)

Empirical approaches to discourse

ESSLI 2012
Day 2
Jennifer Spender

Outline

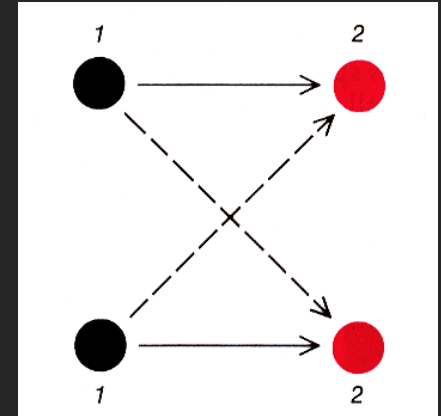
1. Mann & Thompson: Rhetorical Structure Theory
 - Aims and assumptions
 - types of relations
2. Is RST so different from G&S?
 - Moore and Moser
3. Problems with Coherence Relations so far
 - Can you always determine a dominance relationship?
distinguish Nuclei from Satellites?
 - The case of PURPOSE relations
 - Can two relations hold between the same two segments?
 - in particular: one semantic and one pragmatic?
 - What information should be handled by the theory of Coherence?
 - Can Hobbsian and RST relations be reduced to cognitive primitives?
 - How can you systematically study these relations psycholinguistically?

Mann & Thompson (1988) Rhetorical Structure Theory

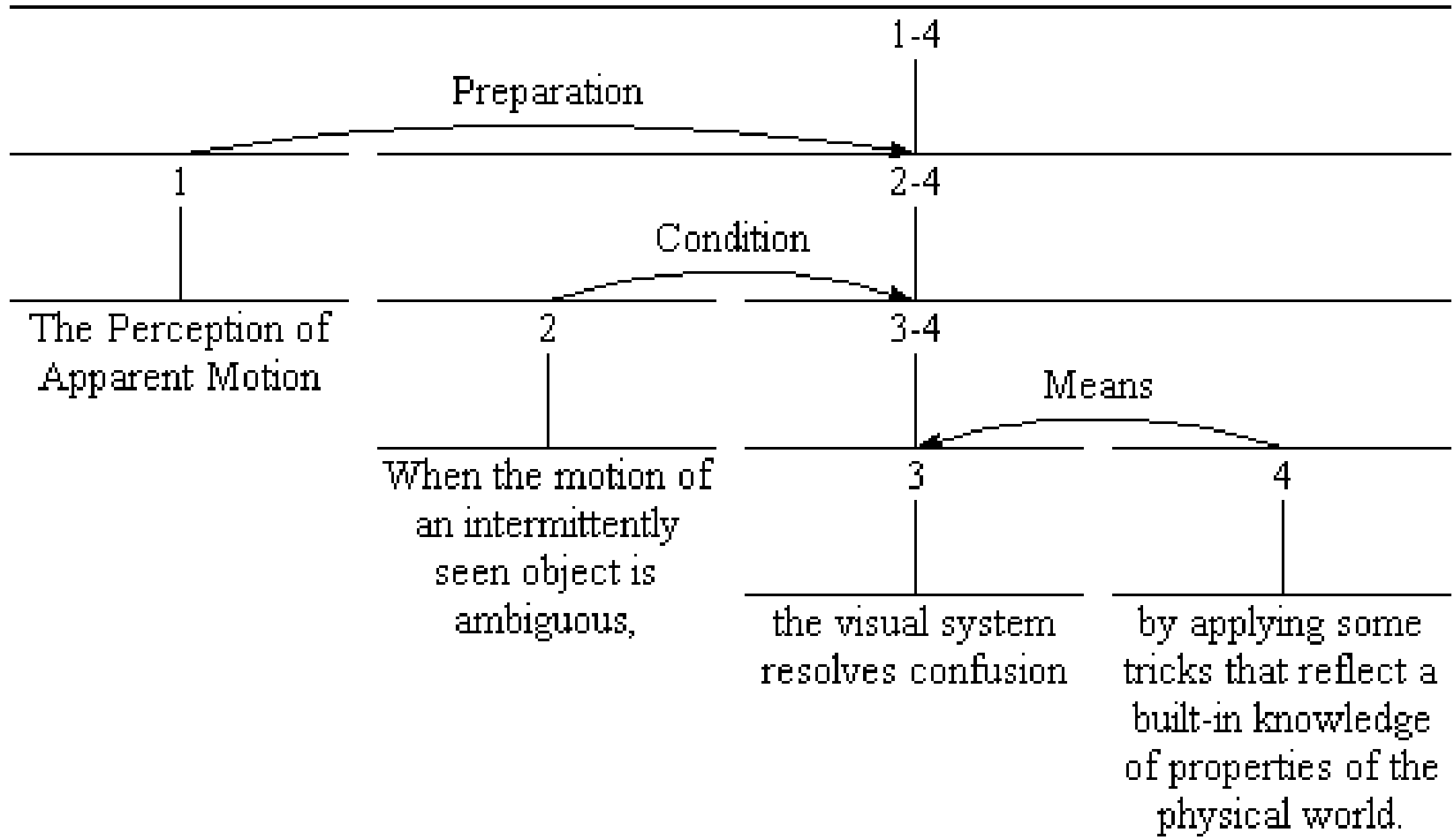


RST (Mann & Thompson 1988) was developed to be used in text generation

- RST tries to explain the coherence of a text, and describes the text itself, rather than the processing of the text
- Has been for text analysis, text generation, automatic summarization etc.



1. [Title:] The Perception of Apparent Motion
2. [Abstract:] When the motion of an intermittently seen object is ambiguous,
3. the visual system resolves confusion
4. by applying some tricks that reflect a built-in knowledge of properties of the physical world.



RST – important concepts

Objects in RST

1. Nucleus

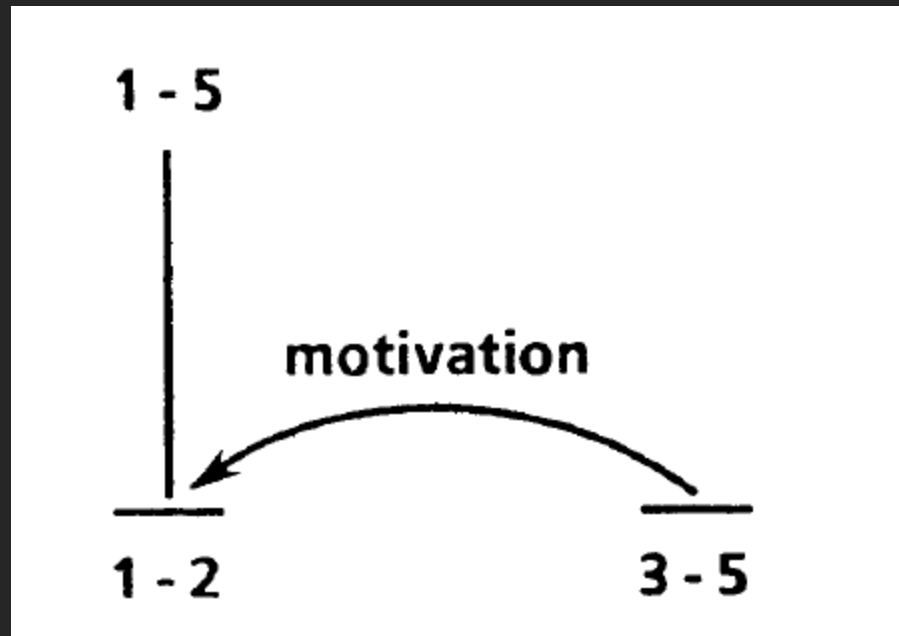
- Nucleus (N) “More central to the authors purposes”

2. Satellite

3. Relations

4. Schemes

Most relations N and S



Arrow points to Nucleus

Relations

1. Relations hold between segments (text spans)
2. Set of relations: open question
3. Relations are explicitly defined according to the requirements for the
 1. Nucleus
 2. Satellite
 3. Combination of nucleus and satellite
 4. And according to their **effect**

Detailed relation

(from Mann & Thompson, 1988, p. 251)

Relation name:

EVIDENCE

Constraints on N:

R might not believe N to a degree satisfactory to W

Constraints on S:

R believes S, or will find it credible

Constraints on N+S combination:

R's comprehending S increases R's belief in S

The effect

R's belief of N is increased

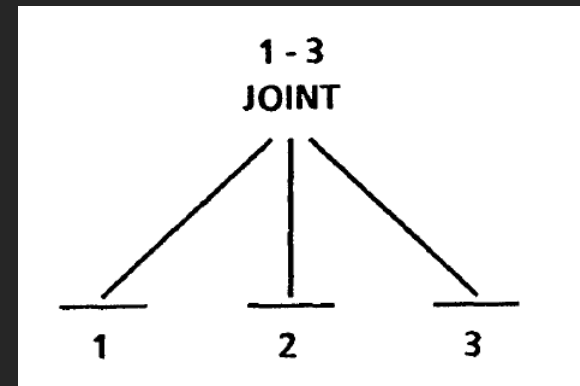
Locus of effect:

N

Other relations

| Relation | Nucleus | Satellite |
|-----------------|---|---|
| Background | Text whose understanding is being facilitated | Text for facilitating understanding |
| Elaboration | Basic information | Additional information |
| Preparation | Text to be presented | Text which prepares the reader to expect and interpret the text to be presented |

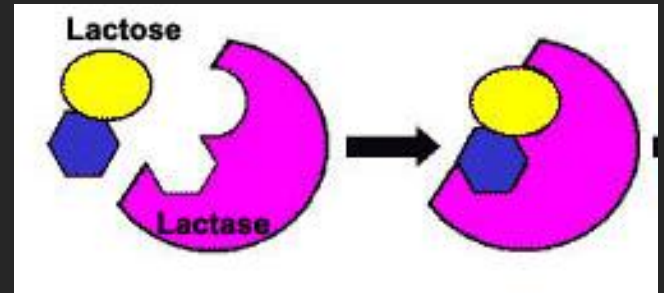
Multi-nuclear relations



Joint:

- Skies will be partly sunny in the New York metropolitan area today.
- It will be more humid, with temperatures in the mid-80's
- Tonight will be partly cloudy, with the low temperature between 70 and 80.

Lactose & Lactase

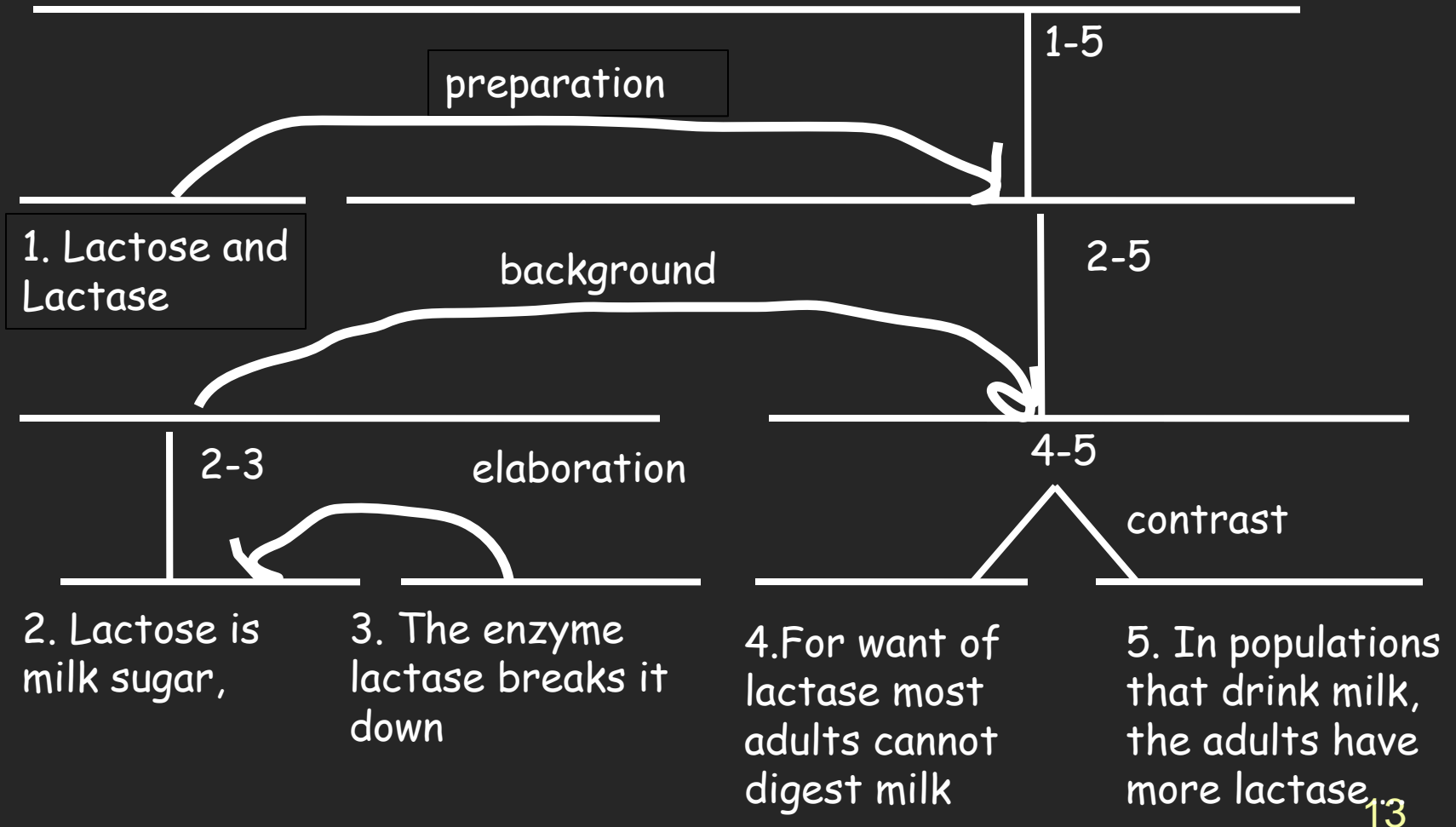
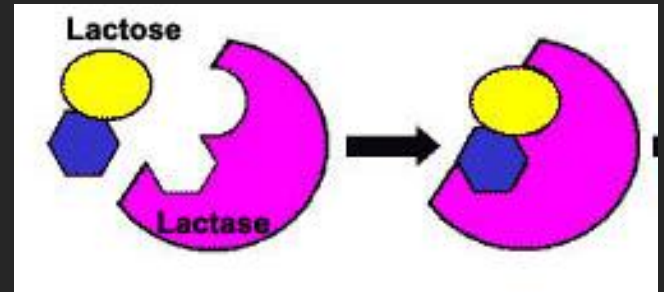


Lactose and Lactase

Lactose is milk sugar; the enzyme lactase breaks it down.

For want of lactase most adults cannot digest milk.

In populations that drink milk, the adults have more lactase, perhaps through natural selection.



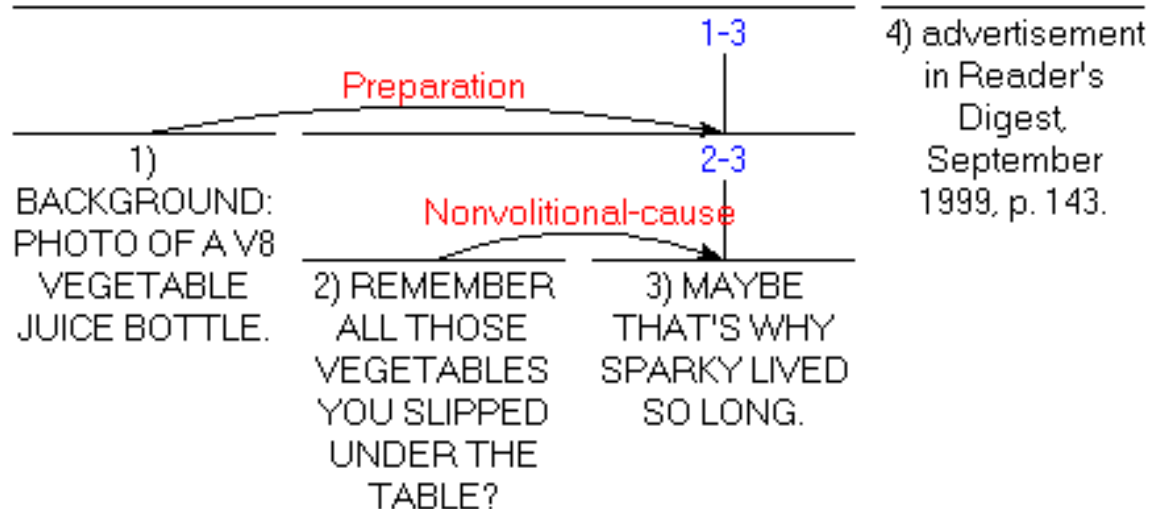


Remember all those vegetables you slipped under the table?



Maybe that's why Sparky lived so long .





Sparky lived! “(This text) is notable for the extreme feat of imagination that it requires in order to see it as coherent.

It is also notable for the lack of cohesive signals, especially that there is no clear signal of cause.

Two major types

Subject matter

“Subject matter relations are those whose intended effect is that the reader recognizes the relation in question.”

Presentational

“are those whose intended effect is to increase some inclination in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus.”

Two major types of relations

Subject matter

“Subject matter relations are those whose **intended effect is that the reader recognizes the relation in question.**”

not very exciting

Presentational

“are those whose intended effect is to increase some inclination in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus.”

Subject matter relations

NON-VOLITIONAL CAUSE

N: a situation

S: another situation which causes that one, but not by anyone's deliberate action

BACKGROUND

N: text whose understanding is being facilitated

S: text for facilitating understanding

Subject matter relations

- Condition
- Elaboration
- Means
- Non-volitional cause
- Non-volitional results
- Volitional Cause
- Volitional Result

Presentational relations

- Antithesis
- Background
- Concession
- Evidence
- Justify
- Motivation
- Preparation
- Restatement
- Summary

What are N and S?

*N and S stand for **the situations** presented by N and S; N and S never stand for the text of N or S. Situation is a broad cover term that ranges over **propositions or beliefs, actions whether realized or not, desires to act and approval for another to act.***

Analyze the whole text

“In RST, a text is coherent if you can identify an intended role for every unit. ...Negatively, coherence is the absence of non-sequiturs” (Tabaoda & Mann)

The deletion test

- The nucleus is the main communication point of a relation.
 - the nucleus connects to the rest of the text
- It is claimed that the satellites could be left out, and the main message of the text would not change.

No 'correct' RST analysis

- Mann & Thompson emphasize that it is possible to have multiple interpretations of the same text
 - evidence of their agnosticism ?
 - feature? because it permits individual interpretations?
 - bug? because there is no 'gold standard' ?

RST Annotation Procedure

Step 1: Segment text into *elementary discourse units*.

Step 2: Connect pairs of units and label their status as *nucleus(N)* or *satellite (S)*

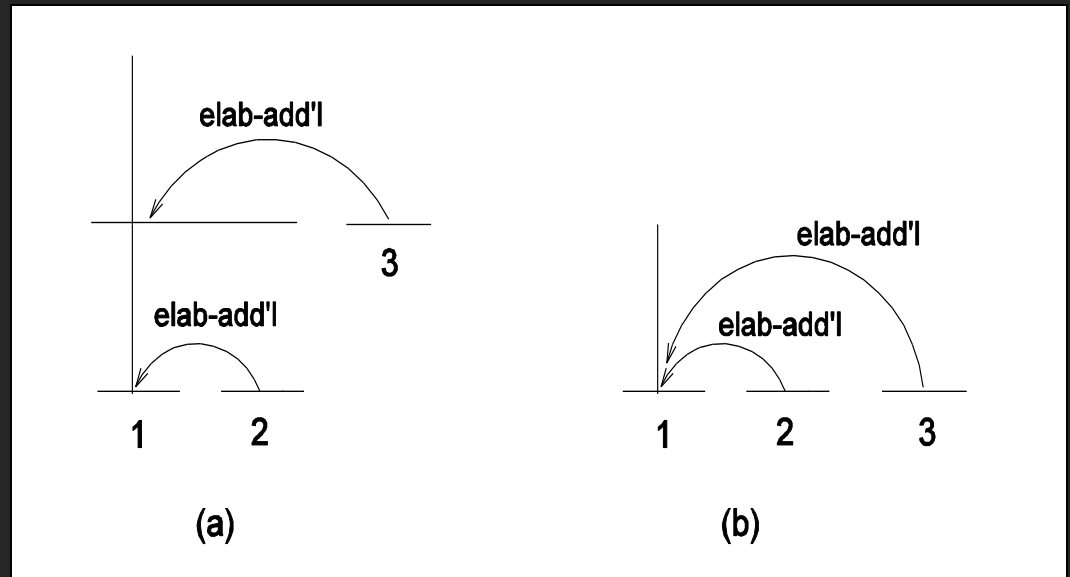
- He tried hard, but he failed.
- Although he tried hard, he failed.
- He tried hard, yet he failed.

Step 3: Assess which of 53 mono-nuclear and 25 multi-nuclear *relations* holds in each case.

- **Step 2** always precedes **Step 3**.
- The result must be a singly-rooted hierarchical cover of each text.

Resolving Ambiguities in RST Annotation

Attachment ambiguities:



Principle: Choose same level of embedding **(b)** if the units and their relations are independent of each other.

Labeling ambiguities: A protocol specifies the order in which to consider rhetorical relations. The first one to be satisfied is the one that is assigned.

RST annotation efforts

- 1. RST Discourse Treebank** (Carlson, Marcu & Okurowski, 2001)
 1. 385 articles from the Wall Street Journal portion of the Penn Treebank
- 2. The SFU Review Corpus** (Maite Taboada and Montana Hay)
 - 400 opinion texts from Eopions.com
 - annotated with RST relations at the sentence level (i.e., no full-text analysis; only those relations found within sentences)
- 3. The Sherlock Corpus (annotated with Relational Discourse Analysis)**

```
( Nucleus (leaf 53) (rel2par span) (prom 53) (text As a result,
industry operated out of small, expensive, highly inefficient
industrial units.) )
    ( Satellite (span 49 52) (rel2par result) (prom 49)
      ( Nucleus (leaf 49) (rel2par span) (prom 49)
        (text In the past, the socialist policies of the government
strictly limited the size of new steel mills, petrochemical
plants, car factories and other industrial concerns) )
      ( Satellite (span 50 52) (rel2par purpose) (prom
50 51)
        ( Nucleus (leaf 50) (rel2par List) (prom 50)
          (text to conserve resources)
            ( Nucleus (span 51 52) (rel2par List)
              (prom 51)
                ( Nucleus (leaf 51) (rel2par span) (prom
51) (text and restrict the profits ) )
                  ( Satellite (leaf 52) (rel2par elaboration-object-
attribute-e) (prom 52) (text businessmen could make.) )
                    )
                  )
                )
              )
            )
          )
        )
      )
    )
  )
```

RST analyses span entire texts

```
( Root (span 1 19) (prom 2)
  ( Nucleus (span 1 15) (rel2par span) (prom 2)
    ( Nucleus (span 1 5) (rel2par span) (prom 2)
      ( Nucleus (span 1 4) (rel2par span) (prom
2)
        ( Nucleus (span 1 3) (rel2par span)
(prom 2)
          ( Nucleus (span 2 3) (rel2par span)
(prom 2)
```

The PURPOSE problem

what is the purpose of a dragonball?



- Nuclearity is assigned to certain functions in a given relation
- But this leaves out the possibility that nuclearity might differ depending on the context in which the relation is used, e.g.

PURPOSE

Jim bought a red sports car_N to impress his new girl-friend_S.

- Text could continue with both

The PURPOSE problem



- Bateman & Rondhuis (1997) and Stede (2008) suggest that PURPOSE may be special in that Nuclearity is not consistently assigned to the first argument of PURPOSE examples.
- Stede(2008) argues however that the entire concept of nuclearity in RST is problematic because it takes too many different ideas and realizes them as one construct
 - The Purpose problem is just one symptom of this

Summary different approaches

Hobbs: similar relations to RST, no nuclei-satellite distinction, emphasis on reasoning

G&S: Two main relations: dominance and satisfaction precedes, incorporates handling of focus with coherence

RST: Many binary relations, functional definitions, nucleus-satellite relations, agnostic as to psychological validity/processing

Summary different approaches

Hobbs: similar relations to RST, no nuclei-satellite distinction, emphasis on reasoning

G&S: Two main relations: dominance and satisfaction precedes, incorporates handling of focus with coherence

RST: Many binary relations, functional definitions, nucleus-satellite relations, agnostic as to psychological validity/processing

RST vs. G&S: are they so different?

Moore & Pollack (date):

The nucleus-satellite distinction is the same as the dominance relationships between DSPs in G&S

from Sherlock corpus, Example from Poesio & Di Eugenio

(S1) Before troubleshooting inside the test station,

(S2) it is always best to eliminate both the UUT and TP.

(S3) Since the test package is moved frequently,

(S4) it is prone to damage.

(S5) Also, testing the test package is much easier and faster

(S6) than opening up test station drawers.

from Sherlock corpus, Example from Poesio & Di Eugenio

(S1) Before troubleshooting inside the test station,

(S2) it is always best to eliminate both the unit under test and the test package.

(S3) Since the test package is moved frequently,

(S4) it is prone to damage.

(S5) Also, testing the test package is much easier and faster

(S6) than opening up test station drawers.

from Sherlock corpus, Example from Poesio & Di Eugenio

(S1) Before troubleshooting inside the test station,

(S2) it is always best to eliminate both the unit under test and the test package.

(S3) Since the test package is moved frequently,

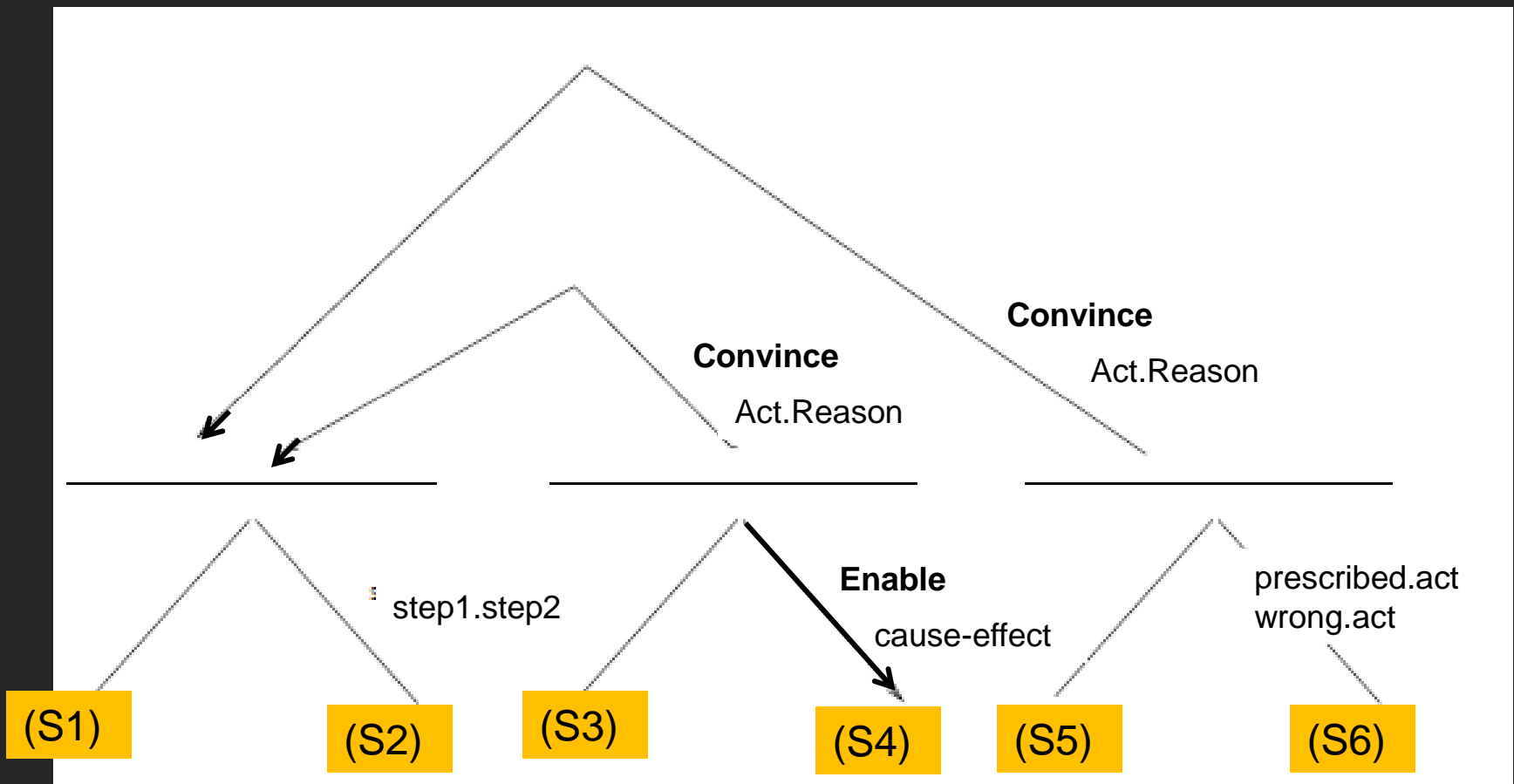
(S4) **it** is prone to damage.

(S5) Also, testing **it** much easier and faster

(S6) than opening up test station drawers.

RDA style analysis

- (S1) Before troubleshooting inside the test station,
- (S2) it is always best to eliminate both the UUT and TP.
- (S3) Since the test package is moved frequently,
- (S4) it is prone to damage.
- (S5) Also, testing the test package is much easier and faster
- (S6) than opening up test station drawers.



Informational vs. intentional

- Seems that there is a very big difference between subject matter relations and 'presentational relations'.
- Relations like: **Motivation**, **Justify** and **Evidence** all refer to the speaker's statement doing something to the hearer: (a real perlocutionary effect), e.g.

Evidence:

R's comprehending S increases R's belief in S

The effect R's belief of N is increased

Subject matter relations

- “Subject matter relations are those whose **intended effect is that the reader recognizes the relation in question.**”
- Condition
- Elaboration
- Means
- Non-volitional cause
- Non-volitional results
- Volitional Cause
- Volitional Result

All have a fairly trivial function of informing about the state of the world

Not really **rhetorical**.

(see Nicholson 1994 for a nice discussion.)

Should these relations be part of RST?

Most discourse theorists would want to keep

CAUSE-EFFECT

as part of a theory of discourse.



Let's not throw the baby out with the bath water!





semantic
relations

rhetorical
relations

**Keep both, but keep
them separate...**

Moore & Pollack (1992)

- Intentional and informational analyses are both present in many relations
- It is strange to force a choice between them.

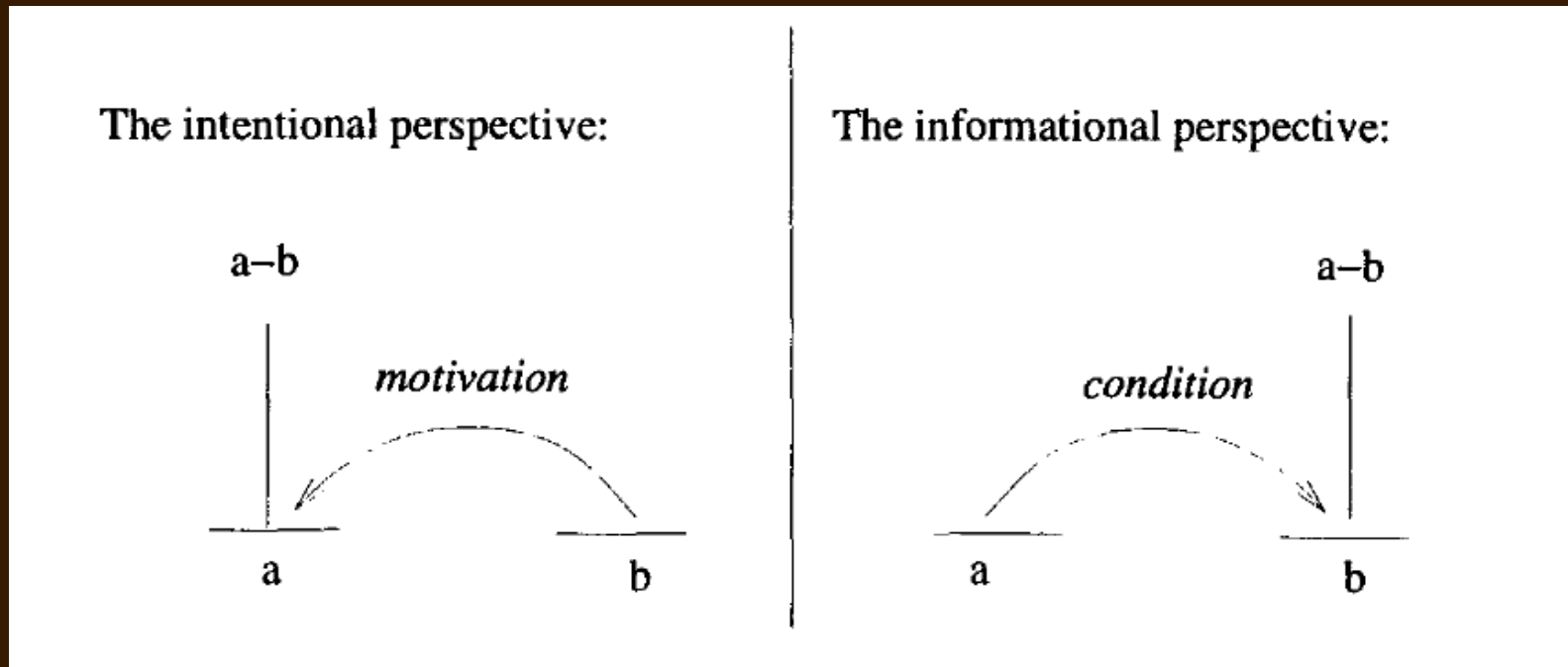


**Come home by
5:00.**

**Then we can
go to the store
before it
closes.**

Moore & Pollack (1992)

- Intentional and informational analyses exist simultaneously



- a) Come home by 5:00.
- b) Then we can go to the store before it closes.

**semantic
relations**

*linguistic
form*

**rhetorical
relations**



The UNTIL problem.

- Hovy: There is no problem with adding the relations you need to RST.
- **UNTIL:**
- Heat the pot until it starts frothing. (from Nichols)
- Rösner & Stede (1992) added an UNTIL relations to cover such relations.
- But this can be done with temporal logic.
 - Why should this type of information be in a theory of coherence?

No clear constraints on relations

1. Mann and Thompson made no clear constraints on relations
2. In fact, they said that certain text types may need different ones.
 1. Agnostic and theory neutral approach
 2. FEATURE, not BUG
.....or is it.

Elaboration

Knott et al. 2011 proposed removing Elaboration

- Definition has six subtypes
- if all cases were clear, there could be six relations
- Carlson and Marcu(2001) had annotators try to distinguish between subtypes
 - Generally, Elaboration was clear, but sub type was not.
- If within and between annotator agreement is required for psychologically valid categories, then Elaboration subtypes are not psychologically valid.





d
d
d

rhet

rhet
rhet

world.

world.

sem
sem
sem
sem

C.S.

world.

.t.l.

.t.l.

.t.l.

.t.l.

Keep everything! Add them all to the aquarium! It's beautiful (but not very principled....)

Actually:two problems

What should be a coherence relation?

1. Should intentional and semantic relations be conflated ?
 - or should they be seperated
2. Should we limit semantic coherence relations to ones that are not handled by other analyses?

Taxonomy of relations

- Knott and colleagues created a taxonomy of coherence relations based on cue phrases
 - also cross-linguistically, English and Dutch (Knott, 1996; Knott and Dale, 1994, 1996; Knott and Mellish, 1996).
- About 150 candidate connectives classified by using
- substitution tests used to classify which connectives were related
- `the taxonomy is already quite complex and sophisticated`
 - What does it tell us though?
 - is complex and sophisticated good things

Sanders et al. (1992)

- Reduced possible relations to combinations of a number of primitive concepts
 - basic operation (causal or additive)
 - source of coherence (semantic or pragmatic)
 - order of segments (basic or non-basic; basic order is antecedent-consequent in causal relations)
 - polarity (negative relations link the content of one of the spans to the negation of the other, otherwise relations are positive).
- 12 classes of relations by combining the four primitive concepts
- Sanders et al. (1992) present 34 relations