

+     -     -     -     -     -     +

No mortal man can slay every dragon

No mortal Dutchman can slay every  
dragon

No mortal man can slay every animal

No mortal man can decapitate every  
dragon

# Extending the monotonicity calculus and embedding it in a textual inference environment

based on McCartney 2009

# Textual Inference

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

## **QA**

Aeschylus is often called the father of Greek tragedy; he wrote the earliest complete plays which survive from ancient Greece. He is known to have written more than 90 plays, though only seven survive. The most famous of these are the trilogy known as Orestia. Also well-known are The Persians and Prometheus Bound.

"The Persians" was written by Aeschylus.

## **SUM**

A Pentagon committee and the congressionally chartered Iraq Study Group have been preparing reports for Bush, and Iran has asked the presidents of Iraq and Syria to meet in Tehran.

Bush will meet the presidents of Iraq and Syria in Tehran.

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

## **NO**

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.



# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

YES

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

YES

## **QA**

Aeschylus is often called the father of Greek tragedy; he wrote the earliest complete plays which survive from ancient Greece. He is known to have written more than 90 plays, though only seven survive. The most famous of these are the trilogy known as Orestia. Also well-known are The Persians and Prometheus Bound.

"The Persians" was written by Aeschylus.

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

YES

## **QA**

Aeschylus is often called the father of Greek tragedy; he wrote the earliest complete plays which survive from ancient Greece. He is known to have written more than 90 plays, though only seven survive. The most famous of these are the trilogy known as Orestia. Also well-known are The Persians and Prometheus Bound.

"The Persians" was written by Aeschylus.

YES

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

YES

## **QA**

Aeschylus is often called the father of Greek tragedy; he wrote the earliest complete plays which survive from ancient Greece. He is known to have written more than 90 plays, though only seven survive. The most famous of these are the trilogy known as Orestia. Also well-known are The Persians and Prometheus Bound.

"The Persians" was written by Aeschylus.

YES

## **SUM**

A Pentagon committee and the congressionally chartered Iraq Study Group have been preparing reports for Bush, and Iran has asked the presidents of Iraq and Syria to meet in Tehran.

Bush will meet the presidents of Iraq and Syria in Tehran.

# Examples

## **IE**

At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the nation's government to investigate Sony over its use of anti-piracy software.

Italy's government investigates Sony.

NO

## **IE**

Parviz Davudi was representing Iran at a meeting of the Shanghai Co-operation Organisation (SCO), the fledgling association that binds Russia, China and four former Soviet republics of central Asia together to fight terrorism.

China is a member of SCO.

YES

## **IR**

Between March and June, scientific observers say, up to 300,000 seals are killed. In Canada, seal-hunting means jobs, but opponents say it is vicious and endangers the species, also threatened by global warming.

Hunting endangers seal species.

YES

## **QA**

Aeschylus is often called the father of Greek tragedy; he wrote the earliest complete plays which survive from ancient Greece. He is known to have written more than 90 plays, though only seven survive. The most famous of these are the trilogy known as Orestia. Also well-known are The Persians and Prometheus Bound.

"The Persians" was written by Aeschylus.

YES

## **SUM**

A Pentagon committee and the congressionally chartered Iraq Study Group have been preparing reports for Bush, and Iran has asked the presidents of Iraq and Syria to meet in Tehran.

Bush will meet the presidents of Iraq and Syria in Tehran.

NO

# Approaches

# Approaches

robust,  
but shallow



Jijkoun & de Rijke 2005

Imprecise,  
confused by  
negation,  
quantification,  
etc.



deep,  
but brittle



Bos & Markert 2006

Brittle,  
requires  
translation of  
NL into FOL

robust,  
but shallow



Jijkoun & de Rijke 2005

natural  
logic

FOL &  
theorem  
proving

Bos & Markert 2006

deep,  
but brittle



Imprecise  
confused by  
negation,  
quantification,  
etc.

Brittle,  
requires  
translation of  
NL into FOL



robust,  
but shallow



deep,  
but brittle

lexical/  
semantic  
overlap

Jijkoun & de Rijke 2005

patterned  
relation  
extraction

semantic  
graph  
matching

natural  
logic

FOL &  
theorem  
proving

Bos & Markert 2006

??

**Imprecise  
confused by  
negation,  
quantification,  
etc.**

**Brittle,  
requires  
translation of  
NL into FOL**

# Shallow approaches

Bag-of-words alignment and deletion: based on the independent probabilities that each word in the premise supports the hypothesis and that each word in the hypothesis derives its support from one word in the premise.

For each word pair we calculate lexical similarity, e.g. string similarity, distributional similarity, WordNet-based distance metrics, ...

Some words are more important than others (different weights).

The first settlements on the site of Jakarta were established at the mouth of the Ciliwung, *perhaps* as early as the 5th century AD.

The first settlements on the site of Jakarta were established as early as the 5th century AD.

Several airlines polled *saw* costs grow more than expected even after adjusting for inflation.

Some of the companies in the poll *reported* cost increases.

Sharon denies Arafat could be targeted for assassination

Arafat targeted for assassination

# Graph-matching

Represent sentence as a (dependency) graph

Align p and h, score each possible alignment, take the best, evaluate strength.

But: finding the best alignment in this way is NP-complete, so we need heuristics.

Assumptions:

once a match is found the rest of the graph doesn't affect the validity of the match

- No dogs barked loudly
- No dogs barked
- Sharon denies Arafat could be targeted for assassination
- Arafat targeted for assassination

alignment and inference determination are done in one go

- 25 of the dead were members of the law enforcement agencies and the rest of the 67 were civilians
- 25 of the dead were civilians
- Sales fell/sales rose// Sales fell/sales did not rise.

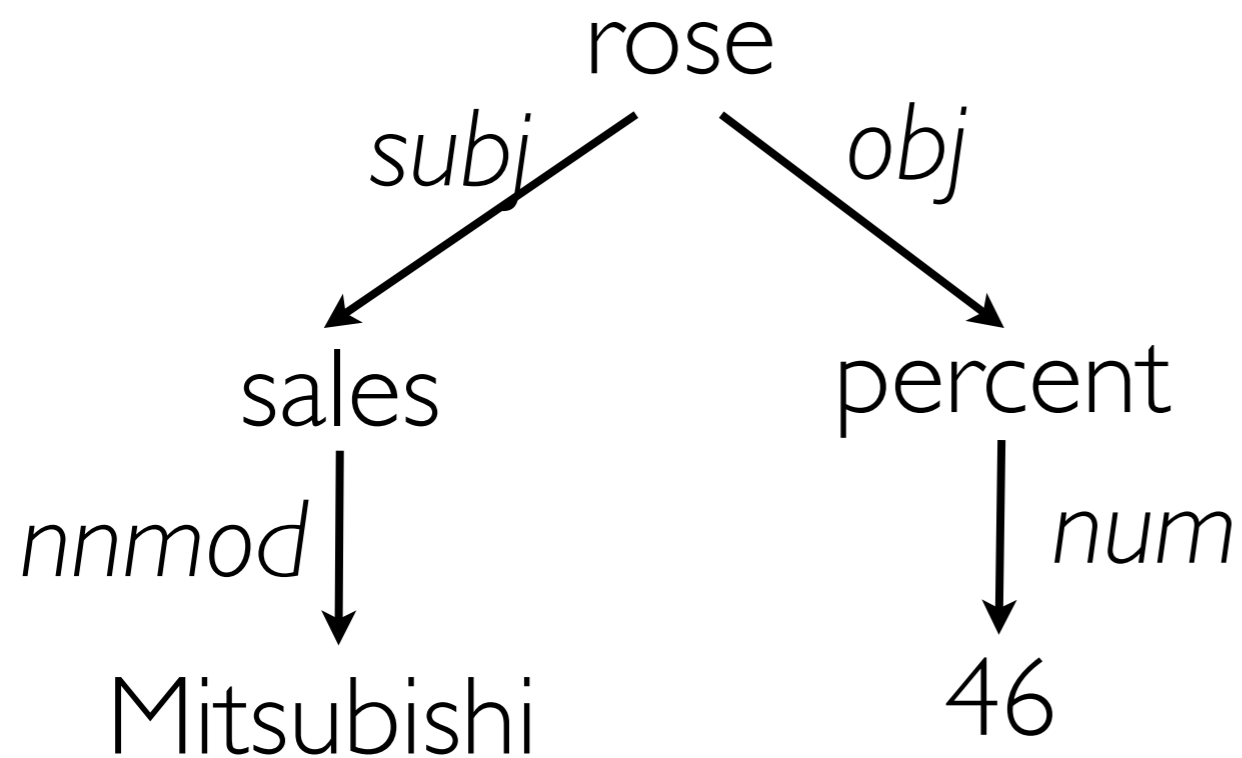
# More evolved RTE systems

Linguistic analysis: tokenization, phrase structure  
parse and/or dependency parse, ...

Alignment

Entailment determination

# Linguistic analysis



subj(rose,sales)  
nnmod(sales,Mitsubishi)  
obj(rose,percent)  
num(percent,46)

# Aside: why dependency grammar

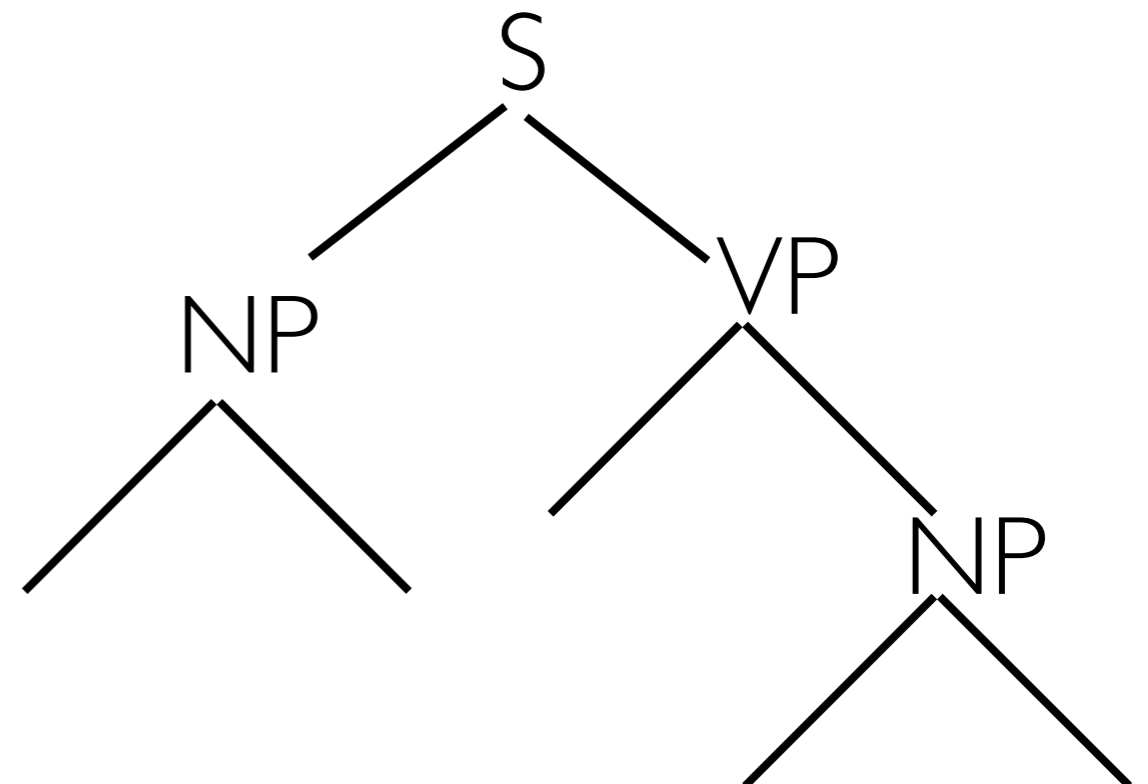
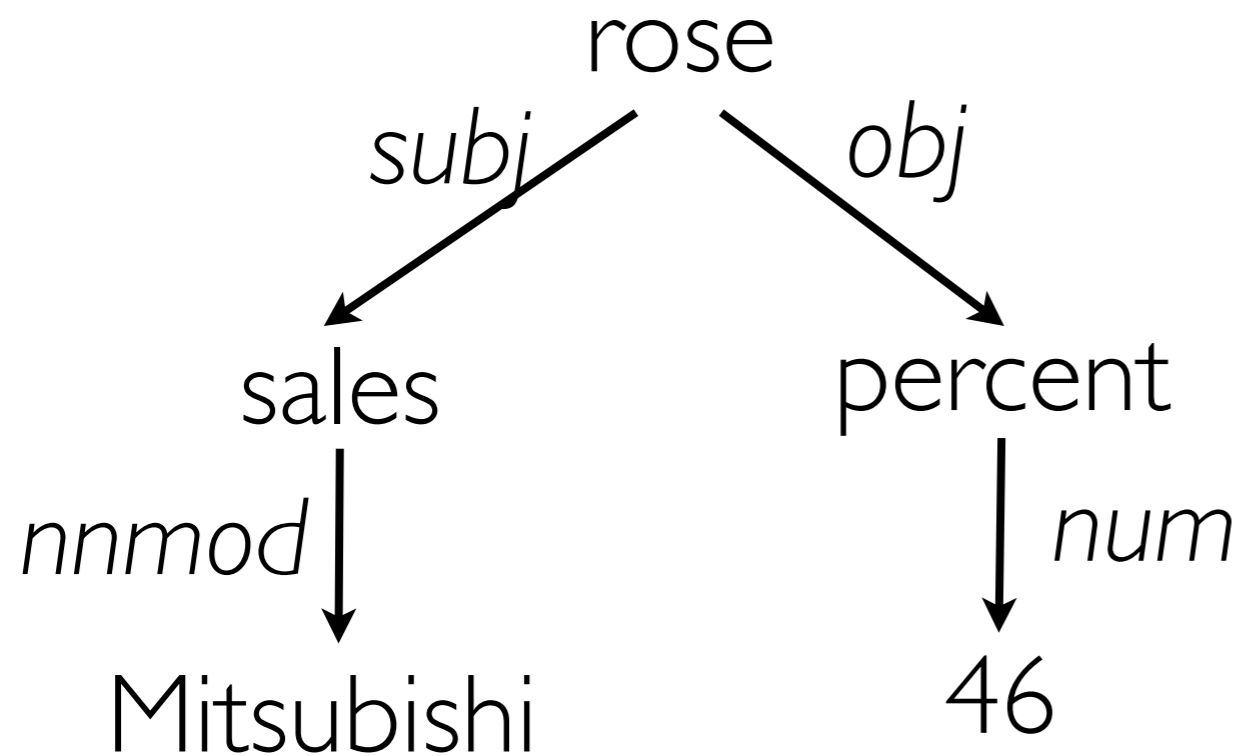
Closer to semantic relationships, but also more underspecified: no distinction between elements that modify the head of a phrase and those that modify the whole phrase.

Dependency trees contain one word per node, makes parsing more straightforward?

But the Stanford parser (and the XLE parser) are not pure dependency parsers.

# Linguistic analysis

subj(rose,sales)  
nnmod(sales,Mitsubishi)  
obj(rose,percent)  
num(percent,46)





# Alignment:MANLI

In most Pacific countries there are very few women in Parliament.

Women are poorly represented in Parliament.

Align via EQ, SUB, DEL, INS:

DEL(In), DEL(most), DEL(Pacific), DEL  
(countries), DEL(there), EQ(are, are), SUB(very  
few, poorly represented), EQ(women, Women),  
EQ(in,in),EQ(parliament,parliament),EG(.,.)

# MANLI Alignment

Scoring:

Edit type features: EQ higher than SUB and DEL  
higher than INS (why?)

Phrase features: size

Semantic relatedness feature for SUB based on  
string similarity, synonymy and the like, distributional  
similarity

Contextual features: difference in position, similarity  
of neighboring material

Insure best alignment for each phrase pair (simulated annealing technique).

Assigning weights to the features

# Determining inferences

Represent the data as features

Classification task

Statistical learning algorithm

Features:

- polarity: no, not, without, except
- adjunct: adverbs, prepositional phrases
- antonymy: derived from WordNet: fall, rise,... (scores high for alignment but gets a feature that insures that inference will be low)
- modality: possible, not possible, actual, not actual, nessary, not necessary
- factivity: try ~ manage
- quantifiers, numbers, dates, time, ...

RTE provides a rather small development/  
training set; lots of features leads to overfitting,  
too few features to errors.

Results in general rather imprecise.

# Using Natural Logic for Entailment

# Entailment determination

premise

hypothesis (= thesis to be proven)

Functional view: input an ordered pair  $(p, h)$ ,  
output a Boolean value, 1 if  $p$  entails  $h$ , 0  
otherwise.

|                             |                            |                            |                             |                          |
|-----------------------------|----------------------------|----------------------------|-----------------------------|--------------------------|
| X is a couch<br>X is a sofa | X is a crow<br>X is a bird | X is a fish<br>X is a carp | X is a hippo<br>X is hungry | X is a cat<br>X is a dog |
|-----------------------------|----------------------------|----------------------------|-----------------------------|--------------------------|



# Which notion of entailment?

1. entailment as a two way classification: output labels (entailment, non-entailment) are interpreted as denoting sets of ordered pairs (relations) of declarative expressions (T):

entailment (def)  $\{(p,h) \in \text{Dom}_{T \times 2} : p \models h\}$

non-entailment (def)  $\{(p,h) \in \text{Dom}_{T \times 2} : p \not\models h\}$

X is a crow, X is a bird: yes

X is a crow, X is a canary: no

X is a crow, X is hungry: no

From MacCartney 2009

2. entailment as a three-way classification: difference between contradiction and compatibility

entailment (def)  $\{(p,h) \in \text{Dom}_{T \times 2} : p \models h\}$

contradiction (def)  $\{(p,h) \in \text{Dom}_{T \times 2} : p \models \neg h\}$

compatibility (def)  $\{(p,h) \in \text{Dom}_{T \times 2} : p \not\models h \wedge p \not\models \neg h\}$

X is a crow, X is a bird: yes

X is a crow, X is a canary: no

X is a crow, X is hungry: compatible

3. a. entailment as containment (monotonicity); output space

$\equiv$  (def)  $\{(p,h) \in \text{Dom}_{T \times 2}: p \models h \wedge h \models p\}$

$\sqsubset$  (def)  $\{(p,h) \in \text{Dom}_{T \times 2}: p \models h \wedge h \not\models p\}$

$\supset$  (def)  $\{(p,h) \in \text{Dom}_{T \times 2}: p \not\models h \wedge h \models p\}$

no-containment (def)  $\{(p,h) \in \text{Dom}_{T \times 2}: p \not\models h \wedge h \not\models p\}$

X is a crow, X is a bird:  $\sqsubset$

X is a bird, X is a crow:  $\supset$

X is a sofa, X is a coach:  $\equiv$

X is a crow, X is a canary: no containment

X is a crow, X is hungry: no containment

From MacCartney 2009

3. b. entailment as containment; input space: not just T but also E and mappings

if  $x, y \in \text{Dom}_T$  then  $x \sqsubseteq y$  iff  $x = \text{false}$  or  $y = \text{true}$

if  $x, y \in \text{Dom}_E$  then  $x \sqsubseteq y$  iff  $x = y$

if  $x, y \in \text{Dom}_{A \rightarrow B}$  then  $x \sqsubseteq y$  iff for all  $a \in \text{Dom}_A$   $x(a) \sqsubseteq y(a)$

(one function entails another if each of its outputs entails the corresponding output of the other function)

otherwise  $x \not\sqsubseteq y$  and  $y \not\sqsubseteq x$

From MacCartney 2009

# Entailment relations

|  | X is a couch<br>X is a sofa | X is a crow<br>X is a bird       | X is a fish<br>X is a carp       | X is a hippo<br>X is hungry | X is a cat<br>X is a dog |
|--|-----------------------------|----------------------------------|----------------------------------|-----------------------------|--------------------------|
| <b>2-way</b><br>RTE1,2,3               | Yes<br>entailment           |                                  | No<br>non-entailment             |                             |                          |
| <b>3-way</b><br>FraCaS,<br>PARC, RTE4  | Yes<br>entailment           | Unknown<br>compatibility         |                                  | No<br>contradiction         |                          |
| <b>containment</b><br>Sánchez-Valencia | $P = Q$<br>equivalence      | $P < Q$<br>forward<br>entailment | $P > Q$<br>reverse<br>entailment | $P \# Q$<br>non-entailment  |                          |

From MacCartney 2011

- Garfield is a cat
- Garfield is a mammal
- Garfield is not a fish
- Garfield is not a carp

Which of these entailments can the monotonicity calculus do?

