BLUE-Lite:
A Knowledge-Based
Lexical Entailment System for RTE6

Peter Clark* and Phil Harrison

Boeing Research and Technology

Seattle, WA

* Current address: Vulcan Inc, Seattle, WA

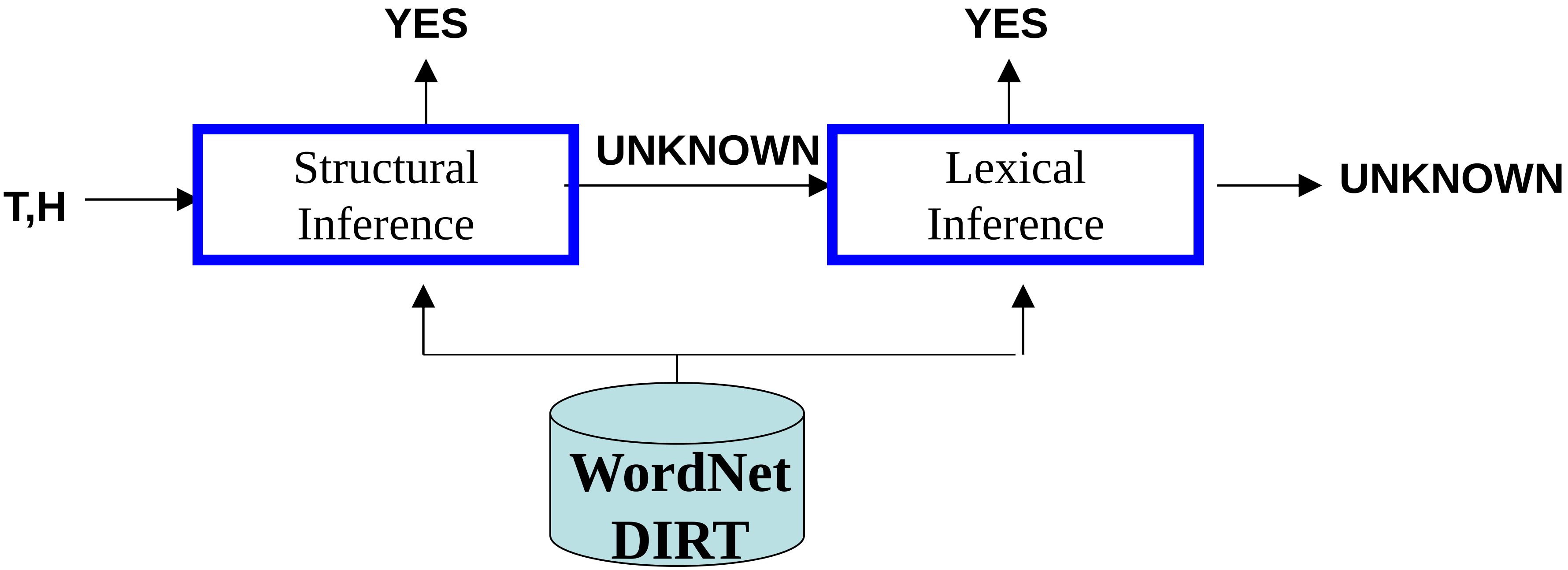
Overview

- System Description (BLUE-Lite)
- Results
- Analysis
- Discussion and Ways Forward

Overview

- **System Description (BLUE-Lite)**
- Results
- Analysis
- Discussion and Ways Forward

BLUE – from RTE5



T,H

YES

Structural Inference

UNKNOWN

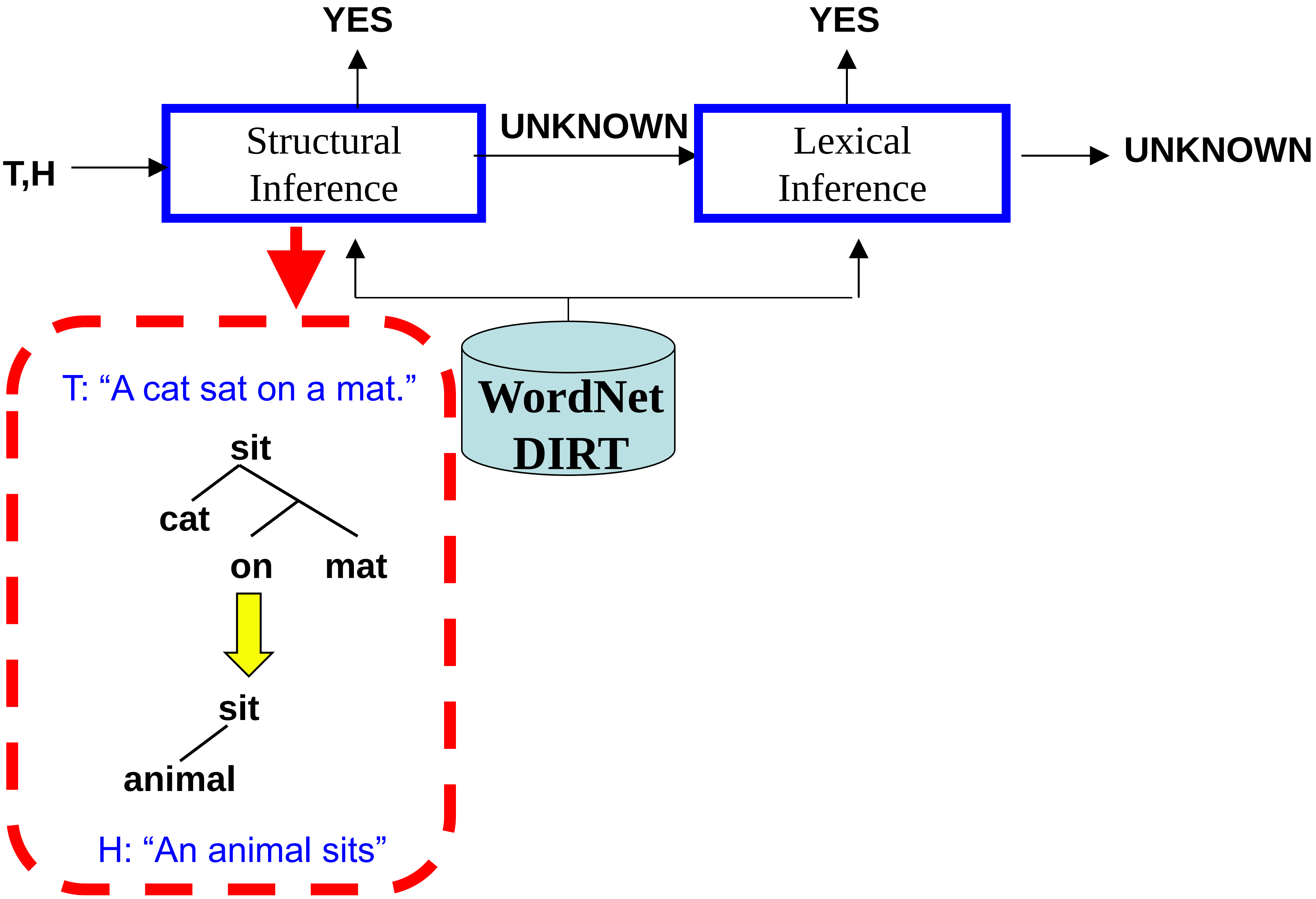
YES

Lexical Inference

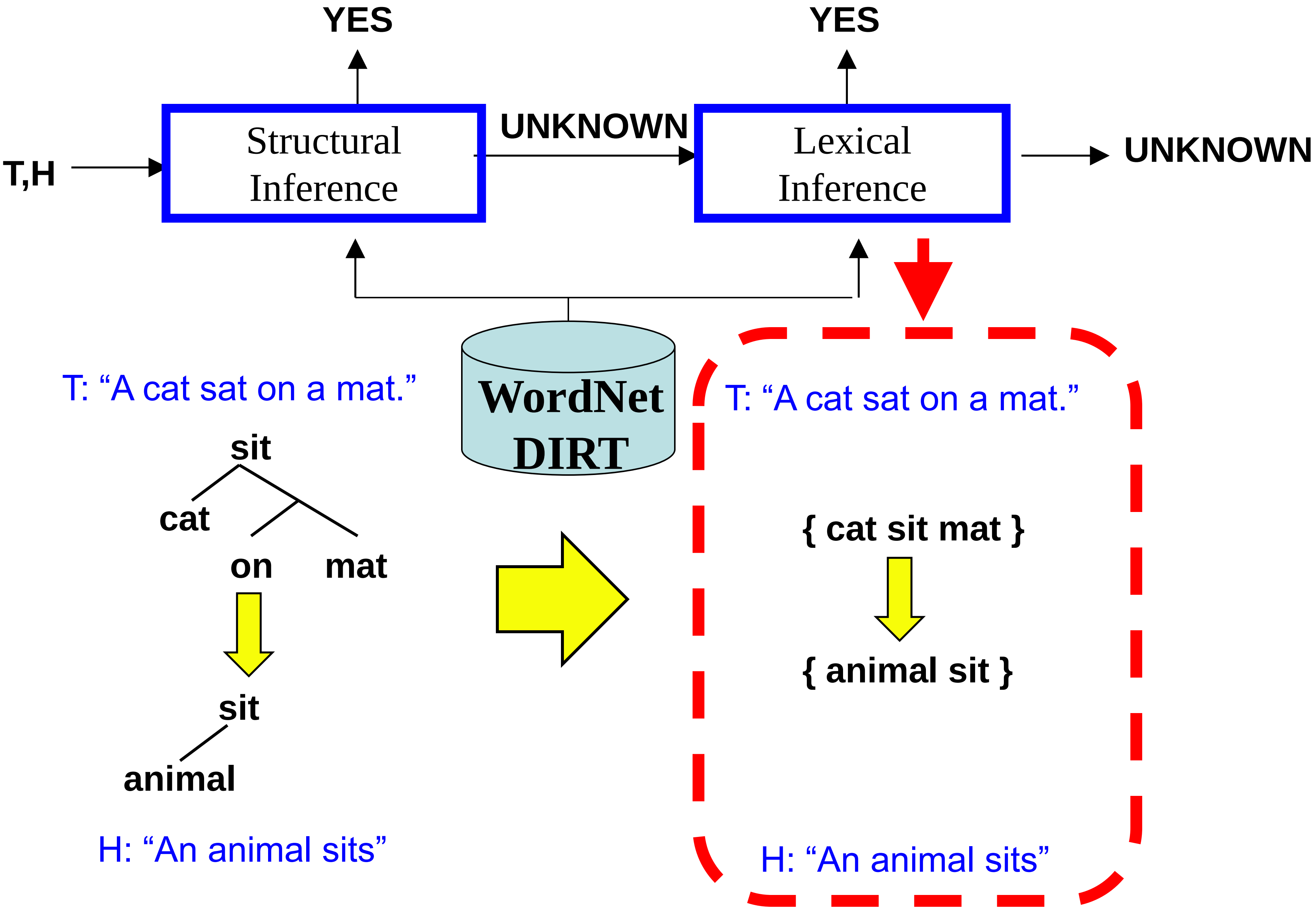
UNKNOWN

WordNet DIRT

BLUE – from RTE5



BLUE – from RTE5



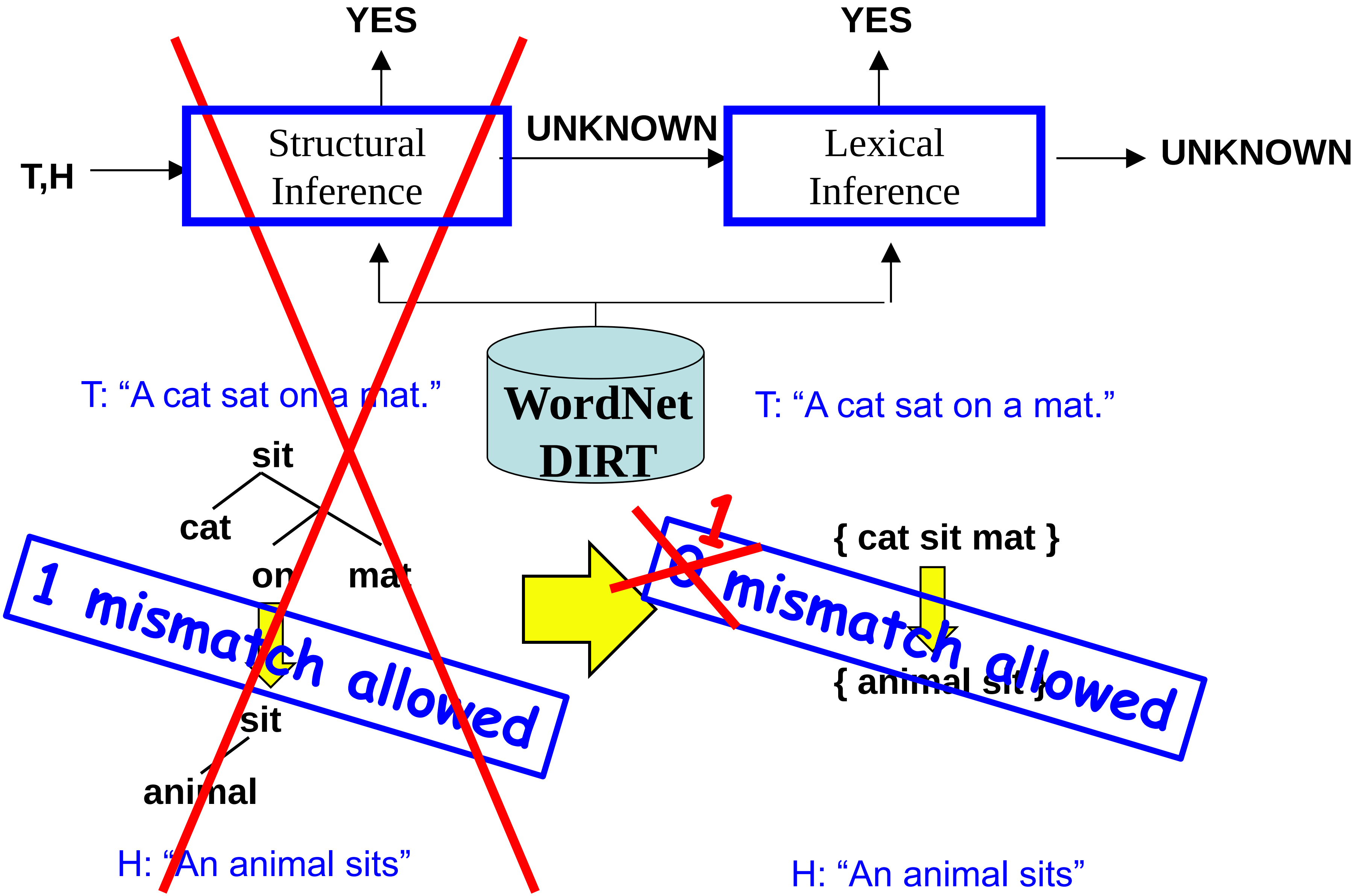
T: “A cat sat on a mat.”

H: “An animal sits”

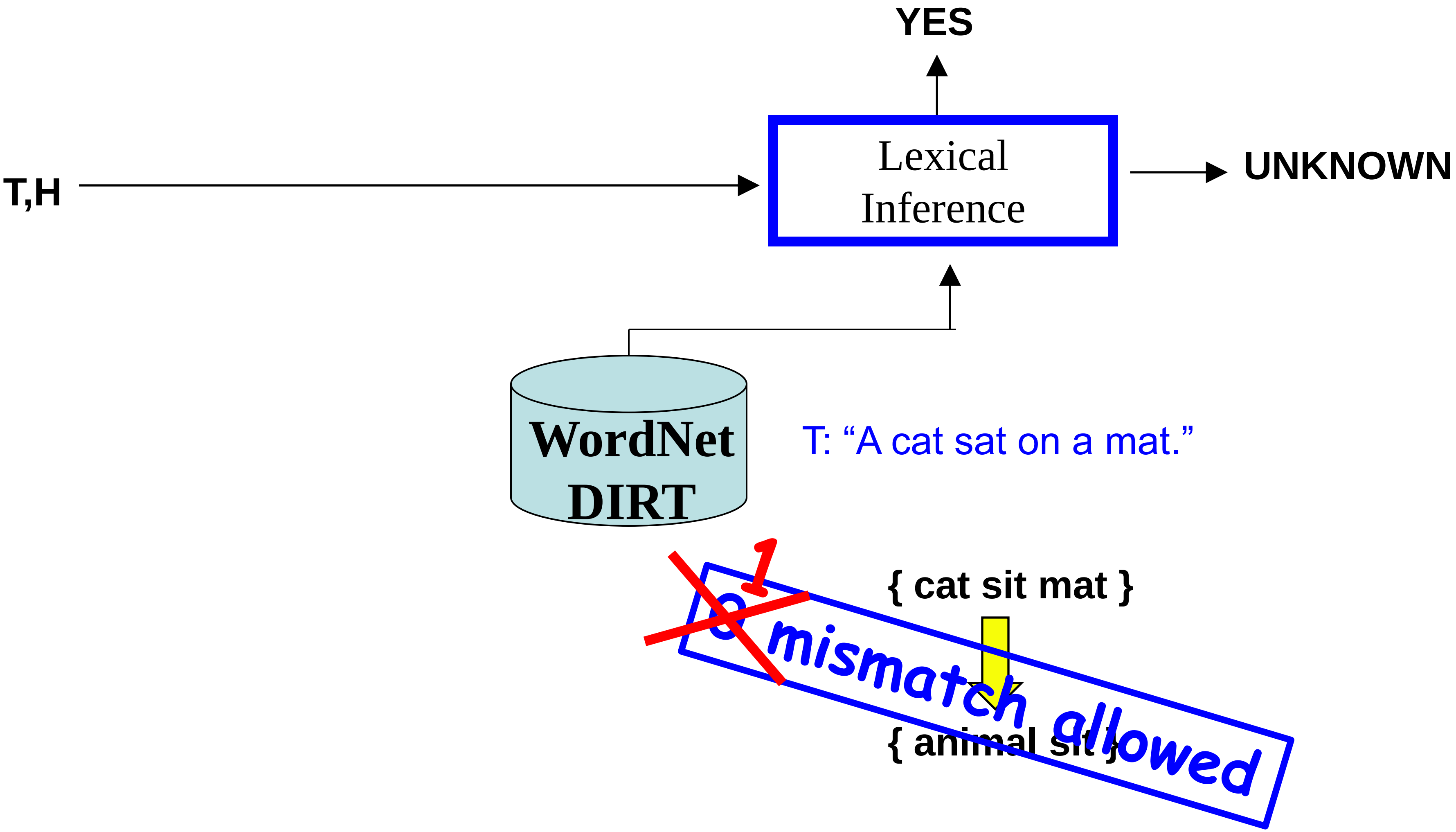
T: “A cat sat on a mat.”

H: “An animal sits”

BLUE – from RTE5



BLUE-Lite



YES

T,H

Lexical Inference

UNKNOWN

WordNet DIRT

T: “A cat sat on a mat.”

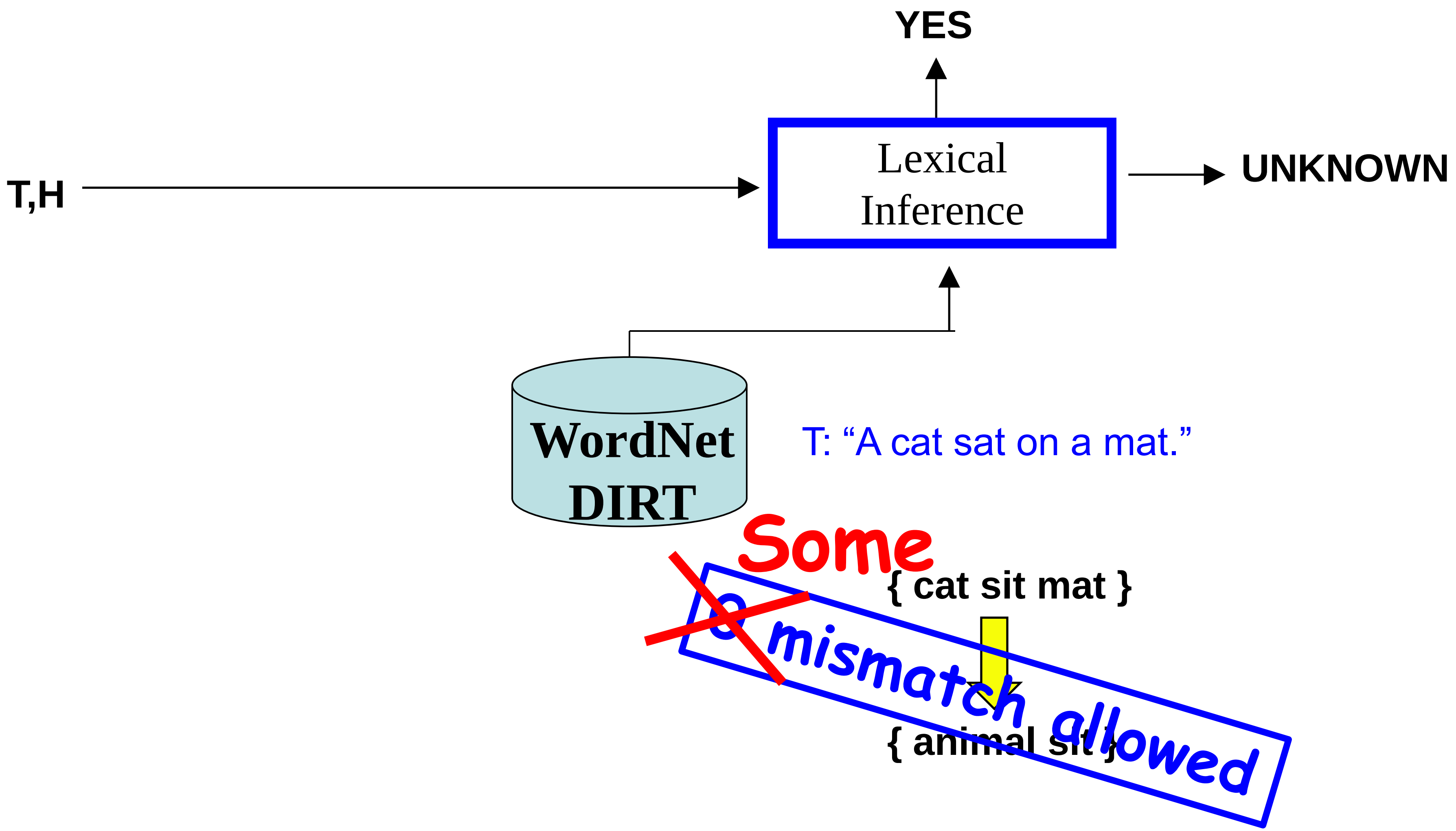
{ cat sit mat }

{ animal sit }

~~0~~ 1 mismatch allowed

H: “An animal sits”

BLUE-Lite



YES

T,H

Lexical Inference

UNKNOWN

WordNet DIRT

T: “A cat sat on a mat.”

Some ~~0~~ mismatch allowed

{ cat sit mat }

{ animal sit }

H: “An animal sits”

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison
2. Knowledge-driven (WordNet and DIRT)
3. Context: Use of previous sentence
4. Variable entailment threshold

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison

T: The Christian Science Monitor newspaper on Monday pleaded for the release of American reporter Jill Carroll...
H: Jill Carroll was abducted in Iraq.

T: "Christian Science" "Monitor" "newspaper" "Monday" "plead" "release" "American" "reporter" "Jill" "Carroll" …
H: "Jill" "Carroll" "abduct" "Iraq"

Main features:

- Normalize words
- Ignore stop words
- Use multiwords from WordNet
- Names are compared in a special way

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison
2. **Knowledge-driven** (WordNet and DIRT)
   - Exploit multiple WordNet relations

WordNet Equivalences:

“speedily” - -> rapidly#r1 ←**pertains-to**→ quick#a1 - - - -> “quick”

nice#a1 ←**similar-to**→ pleasant#s2

build#v1 -**equal**→ construction#n1

WordNet Implications:

car#n1 −**WN−isa**→ vehicle#n1

Baghdad#n1 –**WN-part-of**→ Iraq#n1

DIRT Equivalences:

“loves” ↔ “adores” <- - -

“mark” ↔ “symbolize”

“shoot” ↔ “get” ☹

Derived from DIRT rule:
X loves Y ↔ X adores Y

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison
2. Knowledge-driven (WordNet and DIRT)
3. **Context:** Use of previous sentence
   - If an H word is not entailed by T, look in T-1

T: Merck...pulled the...pain drug...
H: Vioxx is a pain drug.

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison
2. Knowledge-driven (WordNet and DIRT)
3. **Context:** Use of previous sentence
   - If an H word is not entailed by T, look in T-1

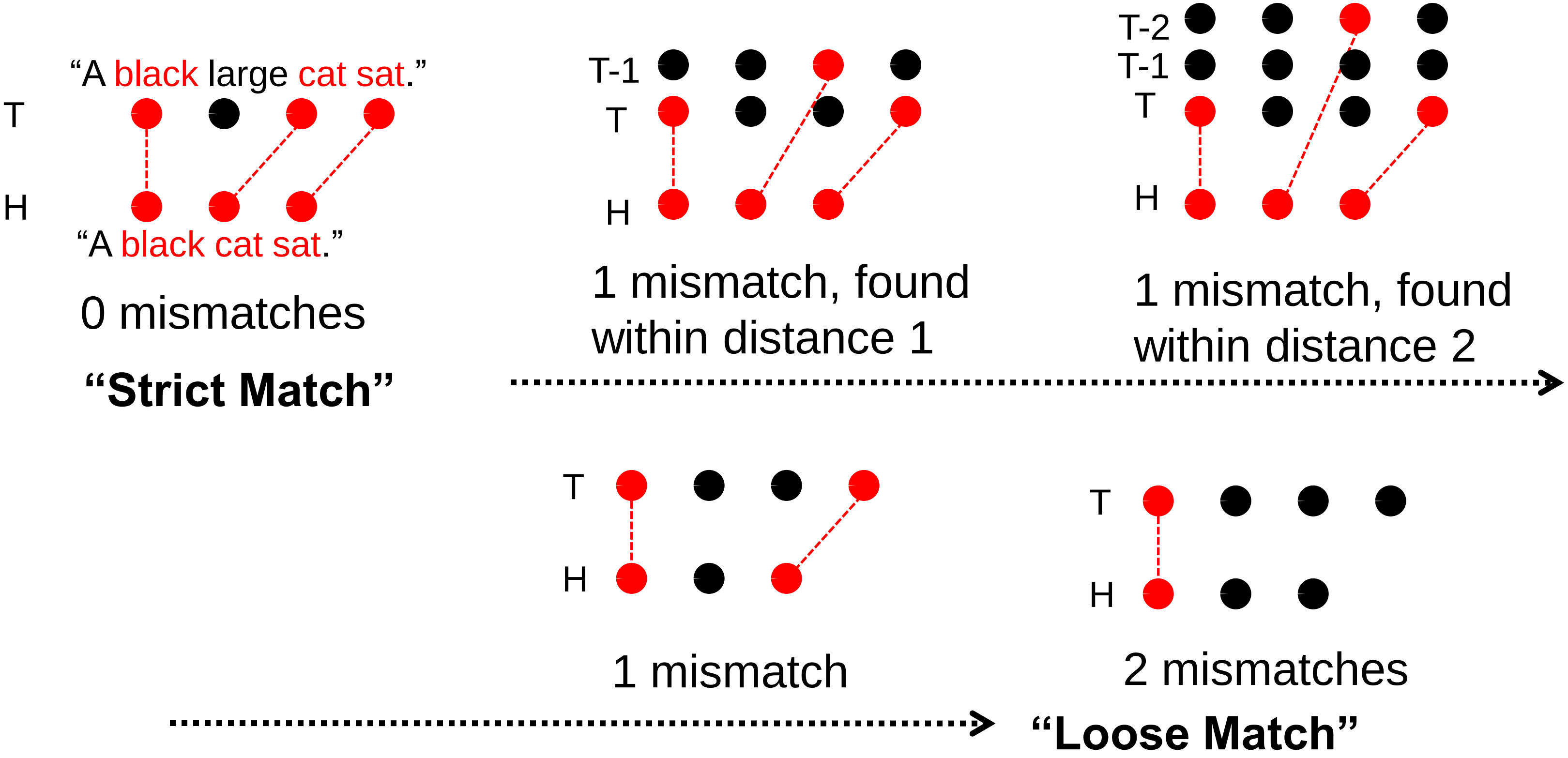
T-1: ...the drug **Vioxx**...

T: Merck...pulled **the...pain drug**...

H: Vioxx is a pain drug.

BLUE-Lite: “Knowledge-Based Lexical Entailment”

1. Lexical comparison
2. Knowledge-driven (WordNet and DIRT)
3. Context: Use of previous sentence
4. **Variable entailment threshold**
   - How good a match implies entailment?



0 mismatches

“Strict Match”

1 mismatch, found within distance 1

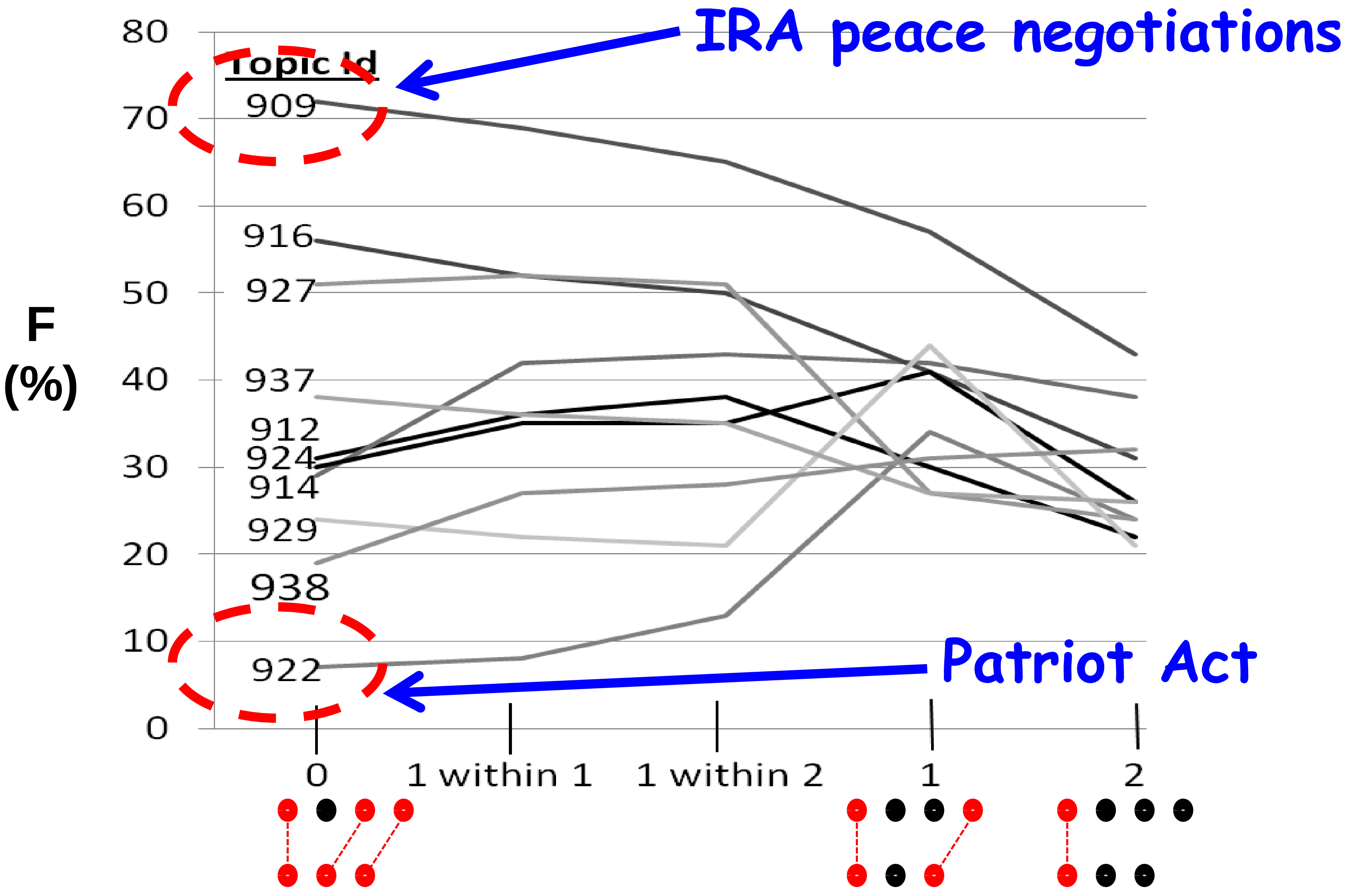
1 mismatch, found within distance 2

1 mismatch

2 mismatches

“Loose Match”

Some topics are more difficult than others…



IRA peace negotiations

Patriot Act

Strict Match ⇢ **Loose Match**

Overview

- System Description (BLUE-Lite)
- **Results**
- Analysis
- Discussion and Ways Forward

Results (F-Measure)

	Knowledge Sources ↓	**Entailment Threshold** (number of mismatches allowed)			
		0	1 within 1	1	Variable
Training Data	none	21.81	26.24	39.72	39.72
	DIRT	25.07	31.06	39.72	40.45
	WN	31.81	36.36	39.69	43.27
	WN+DIRT	37.59	40.30	35.29	42.68
Test Data	none	23.35	27.66	40.35	40.56
	DIRT	25.47	30.72	40.55	39.57
	WN	35.44	39.41	38.68	40.02
	WN+DIRT	37.20	41.56	38.74	43.99

Simple strategy (all but one H words in T) does well, even with no knowledge!

Results (F-Measure)

	Knowledge Sources ↓	Entailment Threshold (number of mismatches allowed)			
		0	1 within 1	1	Variable
Training Data	none	21.81	26.24	39.72	39.72
	DIRT	25.07	31.06	39.72	40.45
	WN	31.81	36.36	39.69	43.27
	WN+DIRT	37.59	40.30	35.29	42.68
Test Data	none	23.35	27.66	40.35	40.56
	DIRT	25.47	30.72	40.55	39.57
	WN	35.44	39.41	38.68	40.02
	WN+DIRT	37.20	41.56	38.74	43.99

WordNet + DIRT + Variable thresholding together adds 3%-4% improvement

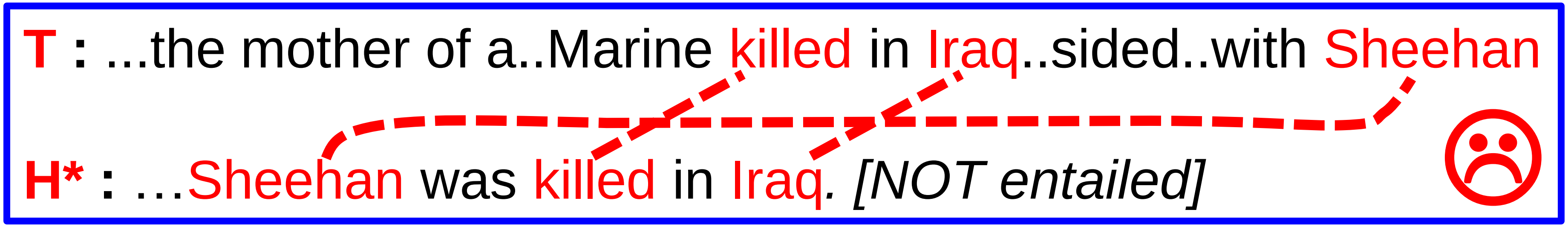
(Close to overall best system 48%)

Overview

- System Description (BLUE-Lite)
- Results
- **Analysis**
- Discussion and Ways Forward

Analysis: 1. Ignoring structure

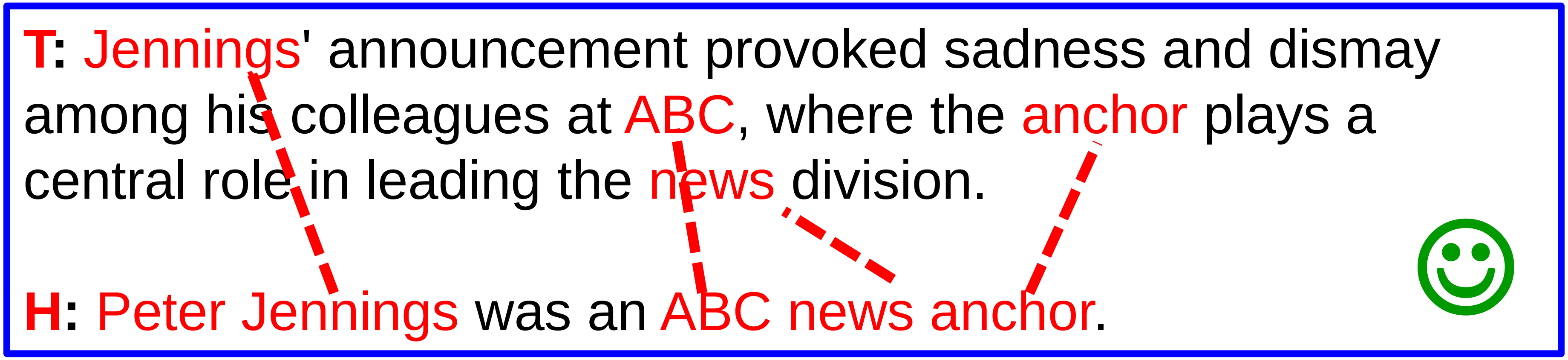
- Surprising that ignoring structure works at all…



T : ...the mother of a..Marine killed in Iraq..sided..with Sheehan

H* : …Sheehan was killed in Iraq. *[NOT entailed]* ☹

- …but this kind of example is common…

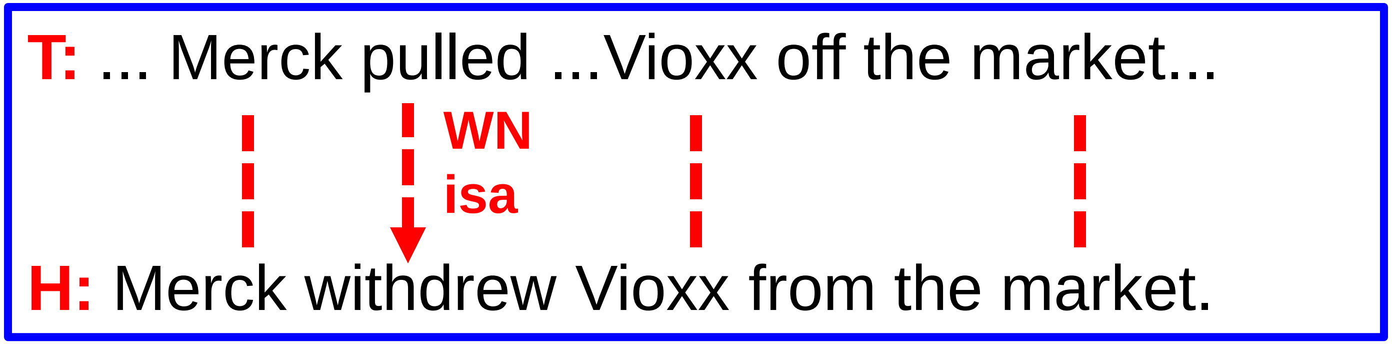


T: Jennings' announcement provoked sadness and dismay among his colleagues at ABC, where the anchor plays a central role in leading the news division.

H: Peter Jennings was an ABC news anchor. ☺

Analysis: 2. WordNet

- WordNet gives us a small overall advantage (3%-4%)…

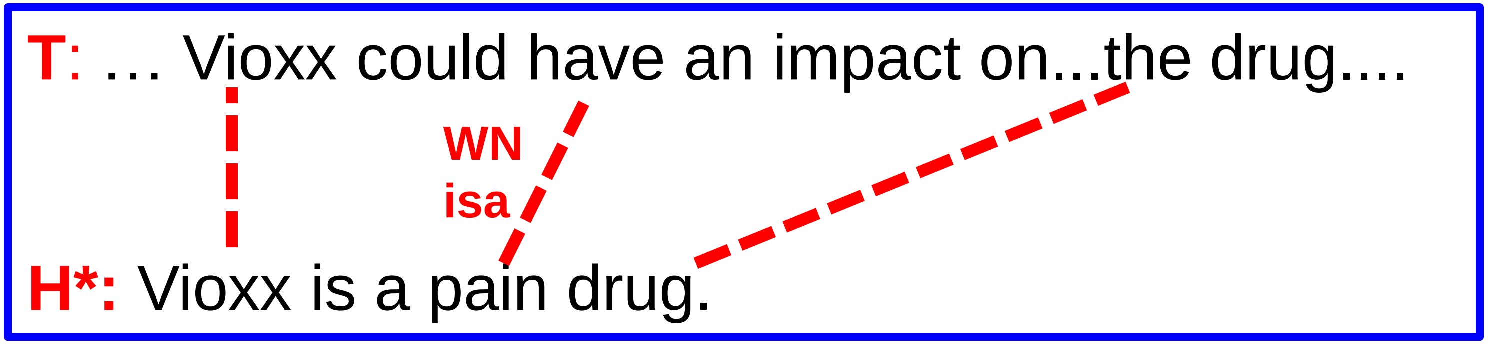


T: ... Merck pulled ...Vioxx off the market...

H: Merck withdrew Vioxx from the market.

☺

- …but can go wrong…



T: … Vioxx could have an impact on...the drug....

H*: Vioxx is a pain drug.

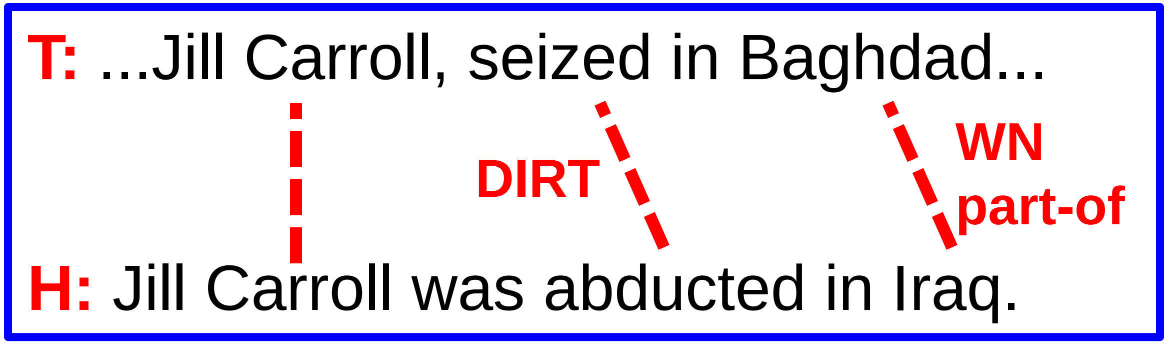
☹

have#v12 [isa suffer#v6 = pain#n1]: "have": suffer from; be ill with; as in "She has arthritis".

☹

Analysis: 3. DIRT

- DIRT was sometimes helpful, e.g.,



T: ...Jill Carroll, seized in Baghdad...

H: Jill Carroll was abducted in Iraq.

☺

- …but inconclusive impact (+/- 1%) overall, e.g.,:

run ↔ oversee
mark ↔ symbolize
say ↔ report
shoot ↔ injure

☺

shoot ↔ get
withdraw ↔ back
remember ↔ expect
deliver ↔ make

☹

Analysis: 4. Lack of General Knowledge

- Several examples failed due to lack of general knowledge

T: Local schools have already closed...amid fears the hurricane could strike...
H: Texas braced for Hurricane Rita.

☹

T: Jennings anchored ABC's evening news for two years...
H: Peter Jennings delivered the news to Americans each night.

☹

T: …the foundation will be a fund-raising organization.
H: The …Foundation was created to raise money.

☹

Analysis: 5. Other Challenges

- **Negation, modals**
  - T: …did not need to be withdrawn… H:…withdrew…
- **Arithmetic**
  - T: …after..blast…two..explosions… H: Three blasts…
- **Calendrics**
  - T: …last Thursday… H: …September 30, 2004.
- **Geography**
  - T: …in Shitani and Ras Soltan… H: …in Sinai
- **Idioms**
  - T: …tower outlines… H: …footprints of the towers…
- **Cataphora**

  T: The angry mother of a fallen U.S. soldier...
  T+2: Cindy Sheehan told reporters...
  T+6: Her son, Casey, 24, was killed in…Iraq...

Overview

- System Description (BLUE-Lite)
- Results
- Analysis
- **Discussion and Ways Forward** ⇦

Coherence and Structure

- Why does ignoring structure work at all?
  - Clearly structure affects entailment:

T: Casey Shehan was in Iraq
H*: Iraq was in Casey Sheehan.

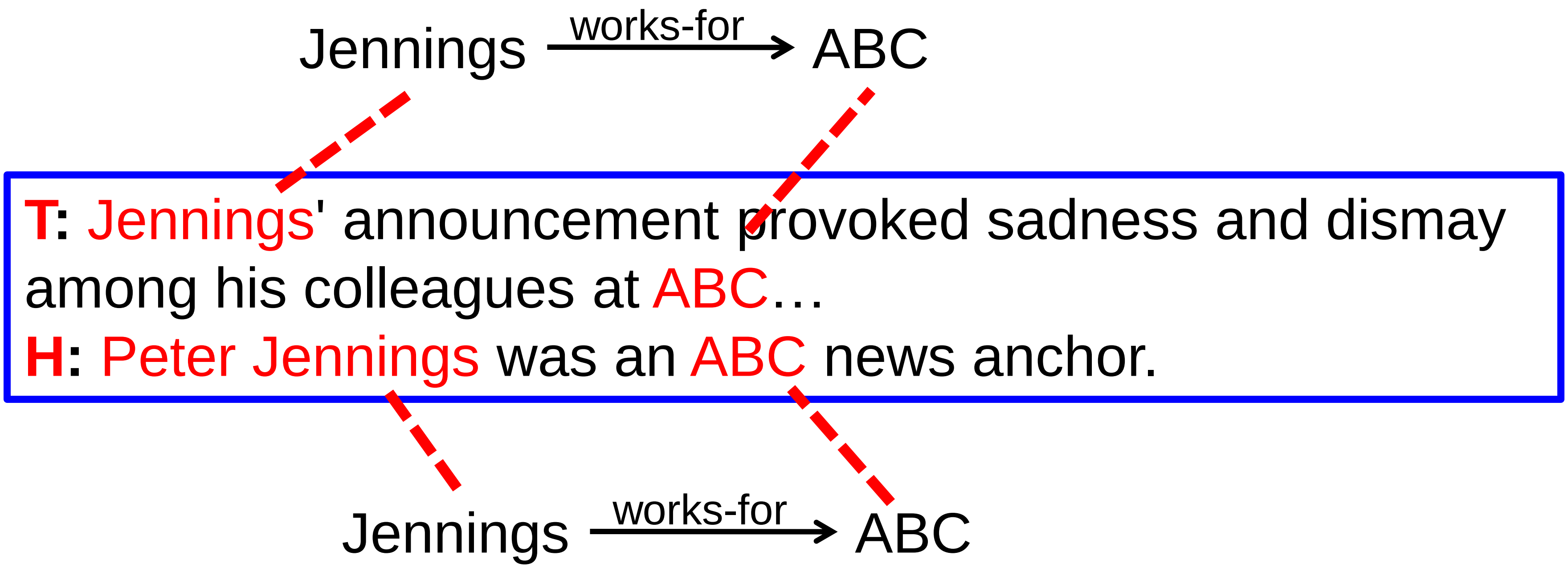
- **BUT:** most restructurings are non-sensical, so will not be seen in “coherent” datasets

A Coherence Conjecture

IF T and H are coherent (not non-sensical)
AND two words in H also occur in T
THEN **it is likely that the semantic relationship between the pair are the same in T and H**

Coherence and Structure

- For example:



T: Jennings' announcement provoked sadness and dismay among his colleagues at ABC…

H: Peter Jennings was an ABC news anchor.

Other Entailment Tasks

- **True/False Question-Answering**
  - Less constrained (query is coherent, but may be false)

H*: Gerry Adams is the prime minister of Ireland? (No)

  - **Structure needed more**

- **Find-a-value Question-Answering**
  - Even less constrained

H: Someone is the prime minister of Ireland?

  - **Structure is critical!**

What are the Ways Forward?

1. Re-introduce use of **structure**
2. **Clean and improve** existing resources

X reviews Y ↔ X approves Y	should be directional
X rejects Y ↔ X approves Y	remove antonyms
X estimates Y ↔ X increases Y	doesn’t “make sense”

3. **Add more rules** (Mechanical Turk?)

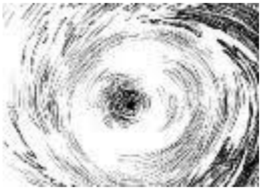
“close school” --suggests⟶ “brace for hurricane”

But: These all seem like small tweaks…

What are the Ways Forward? (cont)

- Each “rule” is a tiny manifestation of deeper knowledge

“close school” --suggests→ “brace for hurricane”



Threat of hurricane →

evacuate people
board up buildings
close schools
assemble supplies
response preparations →

Hurricane hits
Buildings damaged
Trees knocked down
Electricity out →

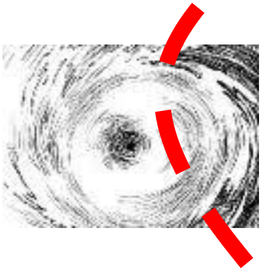
Clear roads
Repair buildings
People return
Restore power

Danger approaching → Prepair for impact → Danger arrives → Damage caused → Recovery

What are the Ways Forward? (cont)

- Each “rule” is a tiny manifestation of deeper knowledge

“close school” --suggests→ “brace for hurricane”



Threat of hurricane

evacuate people
board up buildings
close schools
assemble supplies
response preparations

Hurricane hits
Buildings damaged
Trees knocked down
Electricity out

Clear roads
Repair buildings
People return
Restore power

Danger approaching

Prepair for impact

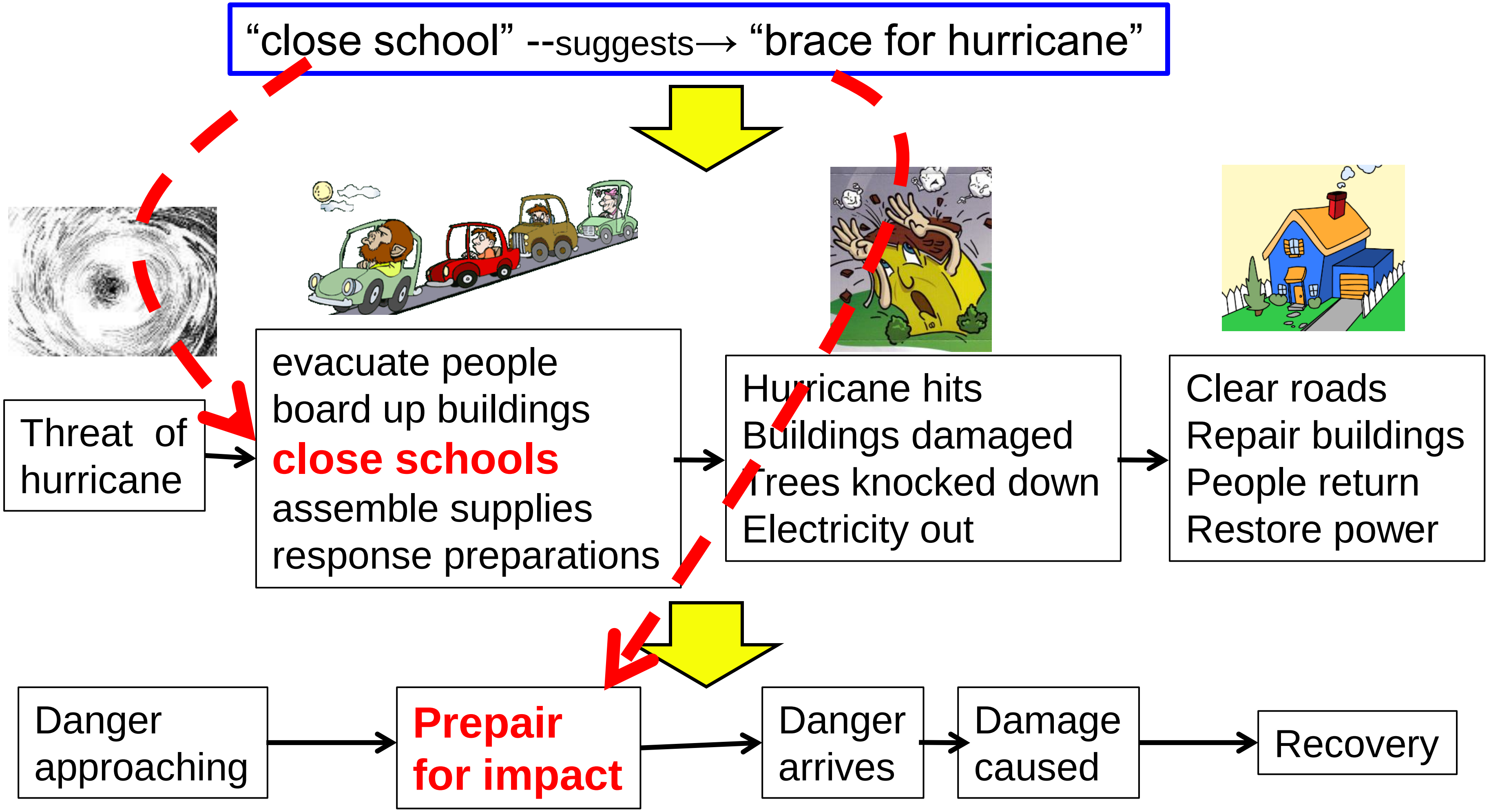
Danger arrives

Damage caused

Recovery

What are the Ways Forward? (cont)

- Each “rule” is a tiny manifestation of deeper knowledge



“close school” --suggests→ “brace for hurricane”

Threat of hurricane

evacuate people
board up buildings
close schools
assemble supplies
response preparations

Hurricane hits
Buildings damaged
Trees knocked down
Electricity out

Clear roads
Repair buildings
People return
Restore power

Danger approaching

Prepair for impact

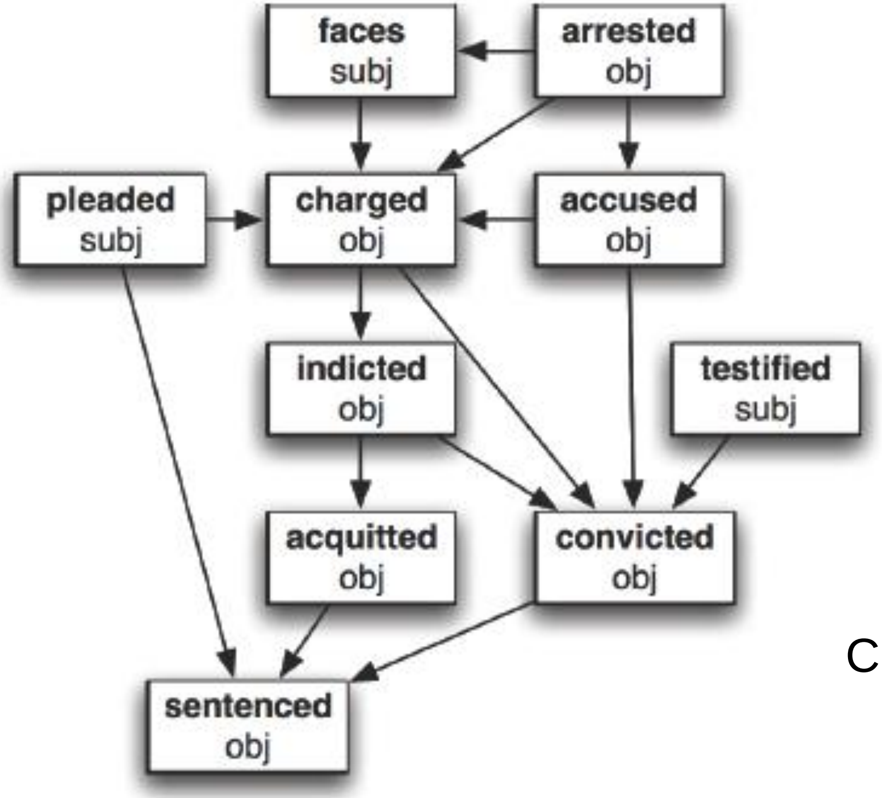
Danger arrives

Damage caused

Recovery

What are the Ways Forward? (cont)

- Can we build/mine such things?



Chambers & Jurafsky, 2008

Summary

- **BLUE-Lite:** Knowledge-Based Lexical Entailment
  - Match lexical items
  - WordNet + DIRT for comparison
  - Use of previous sentence for context
  - Topic-specific entailment threshold
- All help (a bit) to **get improved entailment**
- **But** to do really well (F > 0.50) need…
  - Use of structural information
  - Cleaner knowledge
  - (Lots!) more knowledge

Thank you!