

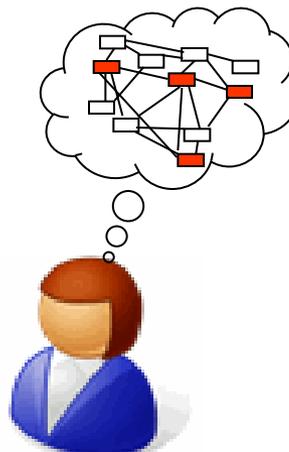
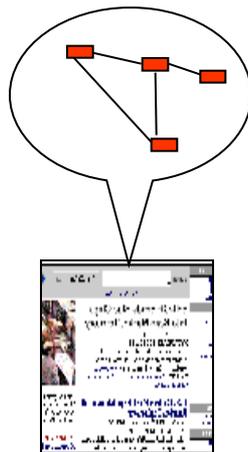
Textual Entailment, QA4MRE, and Machine Reading

Peter Clark
Vulcan Inc.

What is Machine Reading?

- Not (just) parsing + word senses
- Construction of a **coherent representation** of the scene the text describes
- Challenge: much of that representation is **not** in the text

“A soldier
was killed in
a gun battle”



The soldier died
The soldier was shot
There was a fight

...

What is Machine Reading?

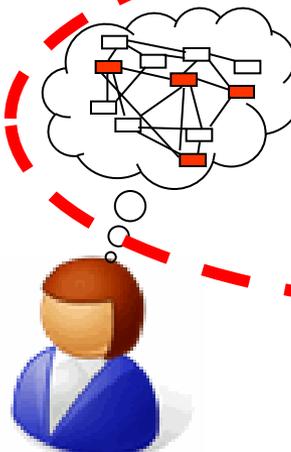
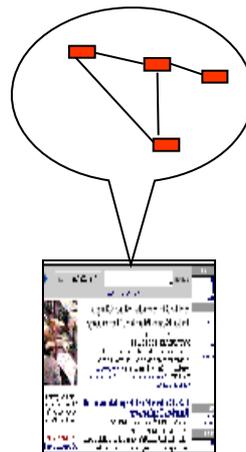
How do we get this knowledge into the machine?

Because

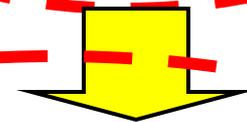
A battle involves a fight.
Soldiers use guns.
Guns shoot.
Guns can kill.
If you are killed, you are dead.

How do we exploit it?

“A soldier was killed in a gun battle”



The soldier died
The soldier was shot
There was a fight
...

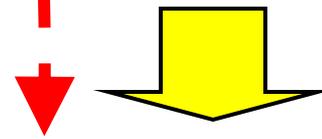


What is Machine Reading?

Because

A battle involves a fight.
Soldiers use guns.
Guns shoot.
Guns can kill.
If you are killed, you are
dead.

....



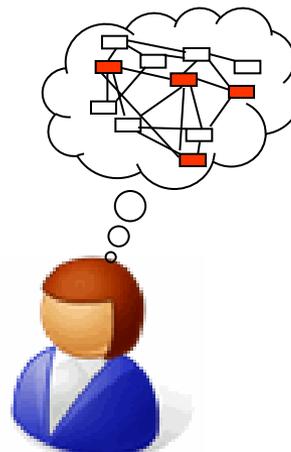
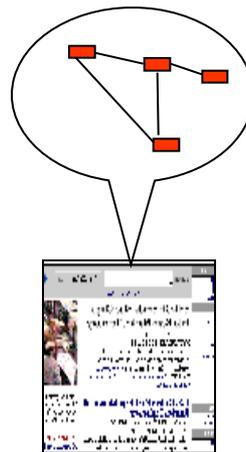
The soldier died

The soldier was shot
There was a fight

...

An entailment

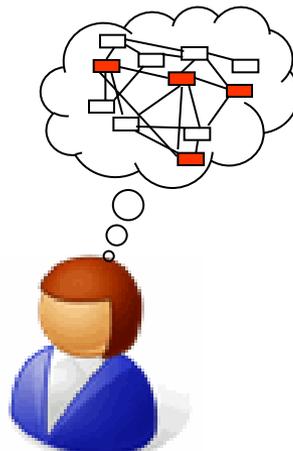
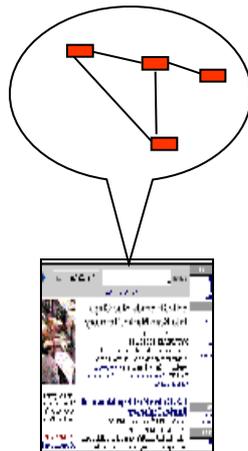
**“A soldier
was killed in
a gun battle”**



What is Machine Reading?

Another entailment

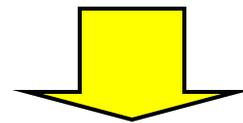
“A soldier was killed **in a gun battle**”



Because

A battle involves a fight.
Soldiers use guns.
Guns shoot.
Guns can kill.
If you are killed, you are dead.

....



The soldier died
The soldier was shot
There was a fight

...

Entailment and QA4MRE

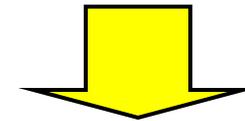
Because

If you teach an instrument then you play that instrument.

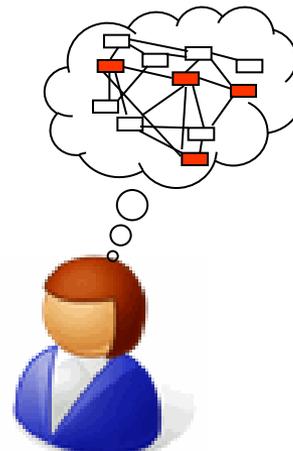
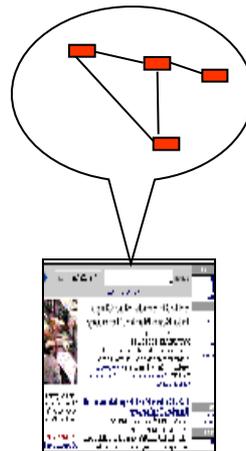
If X studies under Y then Y teaches X.

Studying an instrument involves playing it

....



“Corelli studied the violin under Bassani.”



Corelli played the violin.
Bassani taught Corelli.
Bassani taught the violin.
Bassani played the violin.

....

Entailment and QA4MRE

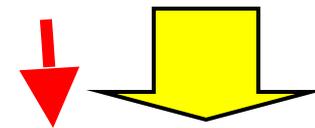
Because

If you teach an instrument then you play that instrument.

~~If X studies under Y then Y teaches X.~~

~~Studying an instrument involves playing it~~

....



Corelli played the violin.

Bassani taught Corelli.

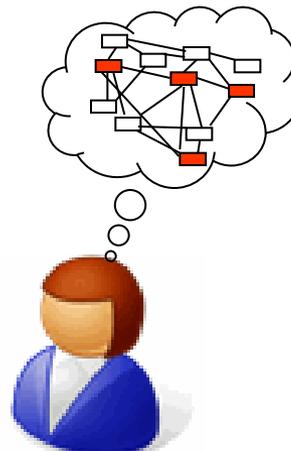
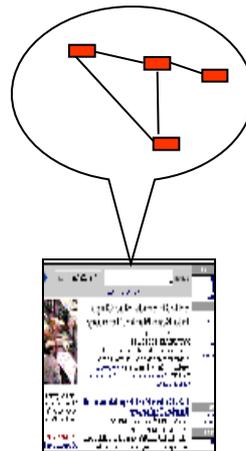
Bassani taught the violin.

Bassani played the violin.

....

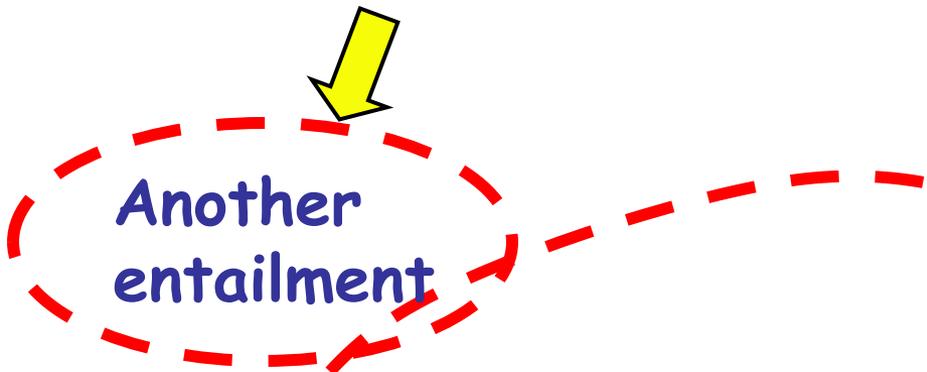
Another entailment

“**Corelli studied** the violin **under Bassani.**”



Entailment and QA4MRE

Entailment is *part of machine reading*

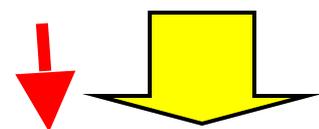


If you teach an instrument then you play that instrument.

~~If X studies under Y then Y teaches X.~~

Studying an instrument involves playing it

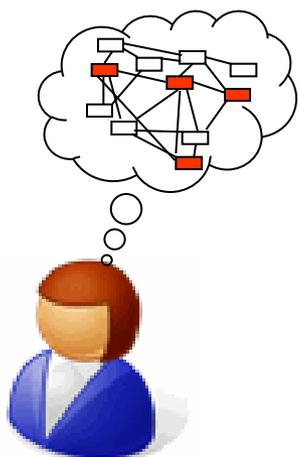
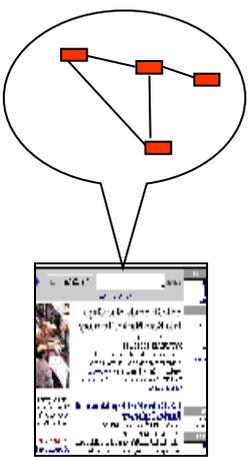
....

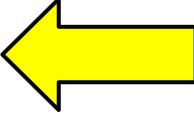


Corelli played the violin.
Bassani taught Corelli.
Bassani taught the violin.
Bassani played the violin.

...

“**Corelli studied** the violin **under Bassani.**”



- Introduction
- **The RTE Competitions** 
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**

Recognizing Textual Entailment (RTE)

T: A soldier was killed in a gun battle.
H: A soldier died.

- Annual RTE competition for 7 years

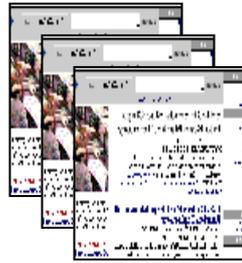
RTE 1-5: does H “reasonably” follow from T?

- Is **very difficult**, and largely unsolved still
 - most problems require lexical and world knowledge
 - typical scores ~50%-70% (baseline is 50%)
 - RTE4 (2008): Mean score was 57.5%

RTE5 (pilot), RTE6, and RTE7

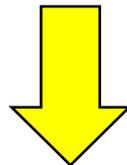
- Find which sentences, in context, entail a hypothesis

TDocuments:



Search: A soldier died.

S1: During the battle, the ...
S2:reported that a soldier was killed...
... ..
S100: Then they left, and returned...



Which S's entail...?

H: A soldier died.

Recognizing Textual Entailment (RTE)

- Clearly very closely related to QA4MRE

Q[11.3] Why were **transistor radios** a significant development?

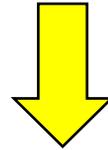
A2 **young people could listen to pop outside**

H₂ Transistor **radios** were a significant development because young people **could listen to pop outside**

Recognizing Textual Entailment (RTE)

- Clearly very closely related to QA4MRE

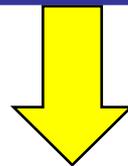
Document:



S1: During the new era of music...

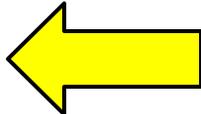
S27: **transistor radios** meant that **teenagers could listen to music outside** of the home.

S100:pop music also affected...



Do any S's entail...?

H₂ Transistor **radios** were a significant development because young people **could listen to pop outside**

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE 
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**

A “Natural Logic” Approach to RTE

- A “Deep” Approach to RTE:
 - Convert text T to a full logical meaning representation T_{logic}
 - See if $T_{\text{logic}} \rightarrow H_{\text{logic}}$. But: very hard to do 😞
- **“Natural Logic”** (MacCartney and Manning)
 - Reason at the *textual level*
 - T: “Airlines proceeded to raise ticket prices”
 - H: “Some ticket prices increased” ?

A “Natural Logic” Approach to RTE

- A “Deep” Approach to RTE:
 - Convert text T to a full logical meaning representation T_{logic}
 - See if $T_{\text{logic}} \rightarrow H_{\text{logic}}$. But: very hard to do 
- **“Natural Logic”** (MacCartney and Manning)
 - Reason at the ***textual level***
 - T: “Airlines proceeded to raise ticket prices”
 - \rightarrow “Airlines raised ticket prices”
 - \rightarrow “Airlines increased ticket prices”
 - \rightarrow “Airlines increased some ticket prices”
 - \rightarrow H: “Some ticket prices increased”

A “Natural Logic” Approach to RTE

1. Interpret T and H sentences individually
 - Generate dependency tree-based representation
2. See if:
 - H **subsumes** (is implied by) T
 - H:“An animal eats a mouse” ← T:“A black cat eats a mouse”
 - H subsumes an **elaboration** of T
 - H:“An animal digests a mouse” ← T:“A black cat eats a mouse”
 - via **IF** X eats Y **THEN** X digests Y

Two sources of World Knowledge

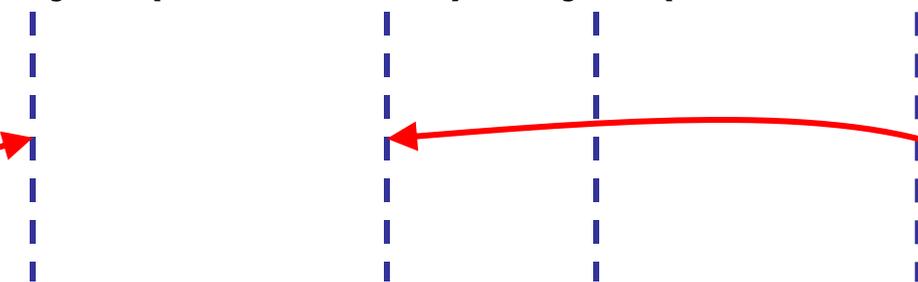
- **WordNet glosses** converted to textual rules
- **DIRT paraphrases**

“Lexico-semantic inference”

- Subsumption

T: A black cat ate a mouse

subject(eat01,cat01), object(eat01,mouse01), mod(cat01,black01)



“by”(eat01,animal01), object(eat01,mouse01)

H: A mouse was eaten by an animal

predicates match if:

- same
- subject(), by() match
- of(), modifier() match anything

arguments match
if same/more
general word

With Inference...

T: A black cat ate a mouse

IF X isa cat_n1 THEN X has a tail_n1

IF X eats Y THEN X digests Y

T': A black cat ate a mouse. The cat has a tail.
The cat digests the mouse. The cat chewed the
mouse. The cat is furry.

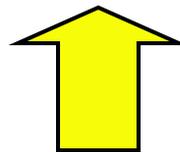
With Inference...

T: A black cat ate a mouse

IF X isa cat_n1 THEN X has a tail_n1

IF X eats Y THEN X digests Y

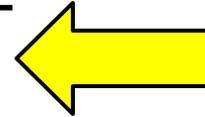
T': A black cat ate a mouse. The cat has a tail.
The cat **digests** the mouse. The cat chewed the
mouse. The cat is furry.



Subsumes

H: An animal **digested** the mouse.

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



WordNet's Glosses as a source of knowledge

- WordNet: A *lot* of “almost accessible” knowledge

Airplane (definition):

- fixed-wing aircraft powered by propellers or jets

But also: 110 uses in *other* definitions, e.g. airplanes...

- can stall, take off, crash, land
- airplane propellers rotate to push against air
- pilots fly airplanes
- can carry passengers
- can be hijacked by hijackers
- passengers can pay money to fly on an airplane
- have wings, fuselage, tail, airfoils, flaps, ...

Converting the Glosses to Logic

- Sometimes we get good interpretations:

restrict#v2: place limits on

isa(restrict01,restrict#v2), object(restrict01,X)

→ isa(place01,place#v3), object(place01,limit01), on(place01,X)

- **But** often not. Primary problems:

1. **Errors** in the language processing

2. **“flowery” language**, many gaps, metonymy, ambiguity;
If logic closely follows syntax → “logico-babble”

“hammer#n2: tool used to deliver an impulsive force by striking”

isa(hammer01,hammer#n2) →

isa(hammer01,tool#n1), subject(use01,hammer01),

to (use01,deliver01), subject(deliver01,force01),

mod(force01,impulsive01), manner(deliver01,strike01).

→ Hammers hit things??

Successful Examples with the Glosses

- Good example

14.H4

T: Britain puts curbs on immigrant labor from Bulgaria and Romania.

H: Britain restricted workers from Bulgaria.

Successful Examples with the Glosses

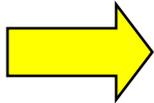
- Good example

14.H4

T: Britain puts curbs on immigrant labor from Bulgaria and Romania.

H: Britain **restricted** workers from Bulgaria.

WN: limit_v1:"restrict": **place limits on**.



T: Britain puts curbs on immigrant labor from Bulgaria and Romania.

H: Britain **placed limits on** workers from Bulgaria.

→ ENTAILED (correct)

Successful Examples with the Glosses

- Another (somewhat) good example

56.H3

T: The administration managed to track down the perpetrators.

H: The perpetrators were being chased by the administration.

Successful Examples with the Glosses

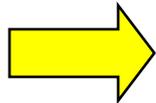
- Another (somewhat) good example

56.H3

T: The administration managed to **track down** the perpetrators.

H: The perpetrators were being chased by the administration.

WN: hunt_v1 “hunt” “track down”: **pursue for food or sport**



T: The administration managed to **pursue** the perpetrators [**for food or sport!**].

H: The perpetrators were being chased by the administration.

→ ENTAILED (correct)

Unsuccessful Examples with the Glosses

- Bad example

29.H

T: Foodstuffs are being blocked from entry into Iraq.

H*: Food goes into Iraq. [NOT entailed]

Unsuccessful Examples with the Glosses

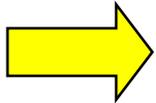
- Bad example

29.H

T: Foodstuffs are being blocked from entry into Iraq.

H*: Food **goes** into Iraq. [NOT entailed]

WN: go_v22:"go": **be contained in**; How many times does 18 go into 54?



T: Foodstuffs are being blocked from entry into Iraq.

H: Food is **contained in Iraq**.

→ ENTAILED (**incorrect**)

Unsuccessful examples with the glosses

- More common: Being “tantalizingly close”

16.H3

T: Satomi Mitarai bled to death.

H: His blood flowed out of his body.

Unsuccessful examples with the glosses

- More common: Being “tantalizingly close”

16.H3

T: Satomi Mitarai **bled** to death.

H: His blood flowed out of his body.

WordNet:

bleed_v1: "shed blood", "**bleed**", "hemorrhage": lose blood from one's body

So close! (but no cigar...)

Need to also know:

“lose *liquid* from *container*” $\xrightarrow{\text{usually}}$ “*liquid* flows out of *container*”

Unsuccessful examples with the glosses

- More common: Being “tantalizingly close”

20.H2

T: The National Philharmonic orchestra draws large crowds.

H: Large crowds were drawn to listen to the orchestra.

Unsuccessful examples with the glosses

- More common: Being “tantalizingly close”

20.H2

<p>T: The National Philharmonic orchestra draws large crowds. H: Large crowds were drawn to listen to the orchestra.</p>
--

WordNet:

WN: orchestra = collection of musicians

WN: musician: plays musical instrument

WN: music = sound produced by musical instruments

WN: listen = hear = perceive sound

So close!

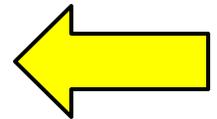
Gloss-Based Axioms: Some Reflections

- In practice, only a little leverage
 - RTE4: ~30 of 1000 entailments with WordNet glosses
- **Very noisy**
 - short, simple glosses work best
- In many cases is **“tantalizingly close”**
- 0.1M axioms is actually quite a small number (!)

RULEBASE # RULES FIRING ON A SENTENCE

WordNet:	15, 3, 10, 3, 41, 18, 7, 6, 24, 10, 13, 7, 15, 2
DIRT:	0, 0, 1138, 0, 2550, 1896, 476, 72, 933, 394, 521, 195, 7096

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



Paraphrases

- Can say the same thing in multiple ways:
 - “Pete works for Vulcan”
 - “Vulcan hires Pete”
 - “Pete goes to work for Vulcan”
 - Pete is a Vulcan employee”
 - ...
- Can we **learn** such equivalences?
- DIRT: An impressive, (partially) successful attempt
 - 12 million rules: **IF** *X relation* Y **THEN** *X relation*’ Y

For Example...

N:subj:V<work>V:for:N

1 N:as:V<work>V:for:N

2 N:obj:V<hire>V:subj:N

3 N:obj:V<hire>V:by:N

4 N:subj:V<work at>V:obj:N

5 N:subj:V<work>V:with:N

6 N:obj:V<employ>V:by:N

7 N:obj:V<sentence>V:to:N

8 N:subj:V<tell>V:obj:N

9 N:subj:V<join>V:obj:N

10 N:subj:V<work>V:in:N

11 N:obj:V<employ>V:subj:N

12 N:subj:V<work>V:at:N

13 N:subj:V<do>V:obj:N>work>N:for:N

14 N:subj:V<go to work>V:for:N

15 N:subj:V<sue>V:obj:N

16 N:obj:V<detain>V:for:N

17 N:obj:V<recruit>V:subj:N

18 N:subj:V<retire>V:from:N

19 N:obj:V<interview>V:by:N

20 N:subj:N<employee>N:nn:N

IF X works for Y THEN:

Y hires X

X is employed by Y

X is sentenced to Y

etc

Some selected paraphrases from DIRT

IF Anselmo organizes a lab **THEN:**

Anselmo promotes a lab.

Anselmo participates in a lab.

Anselmo makes preparations for a lab.



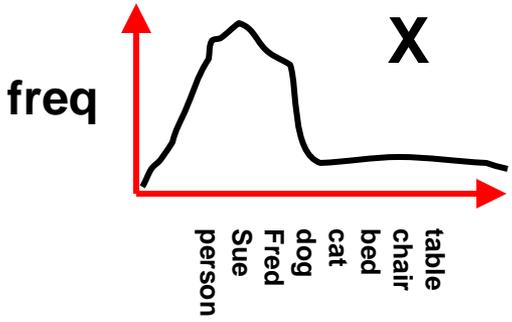
Anselmo intensifies a lab.

Anselmo denounces a lab.

Anselmo urges a boycott of a lab.

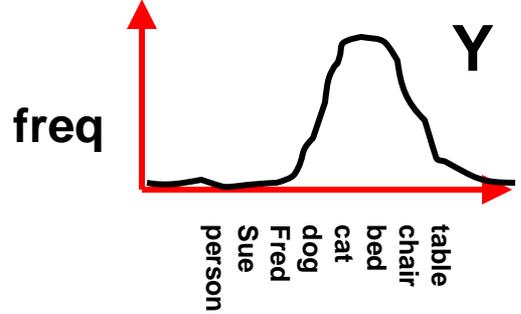


(Approximately) how DIRT learns rules

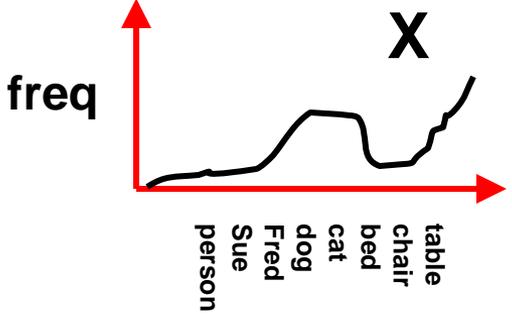
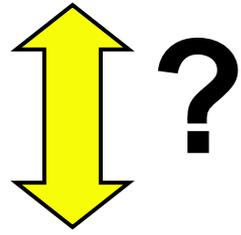


word

X loves Y

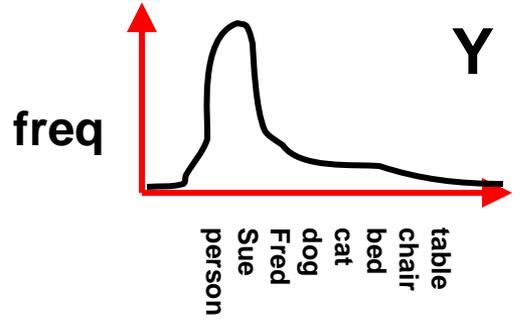


word



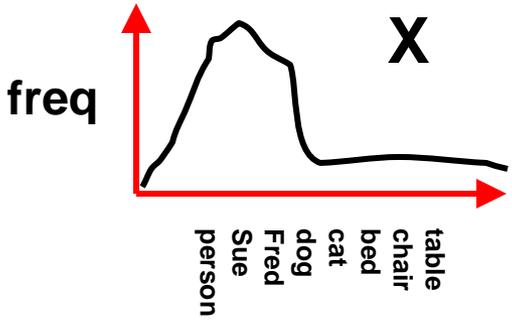
word

X falls to Y



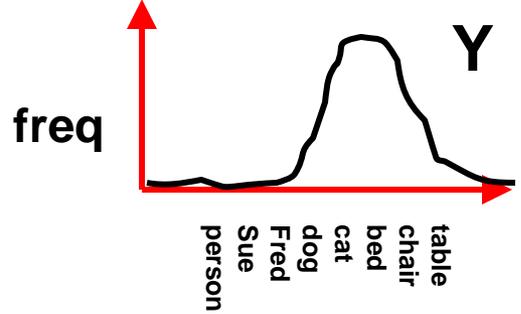
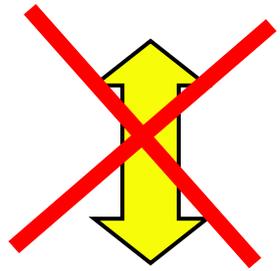
word

(Approximately) how DIRT learns rules

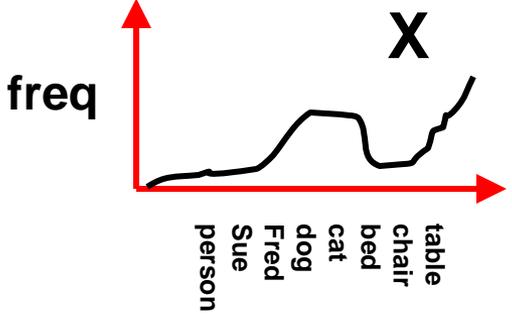


word

X loves Y

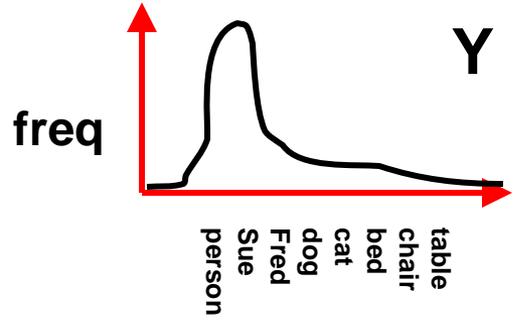


word



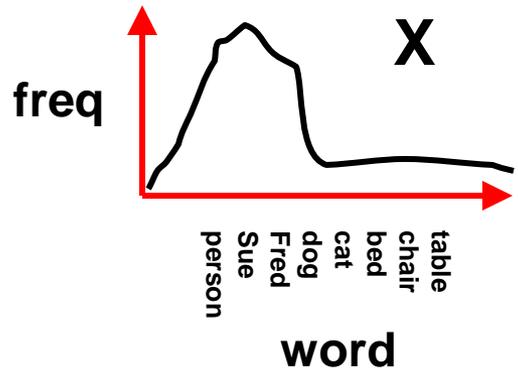
word

X falls to Y

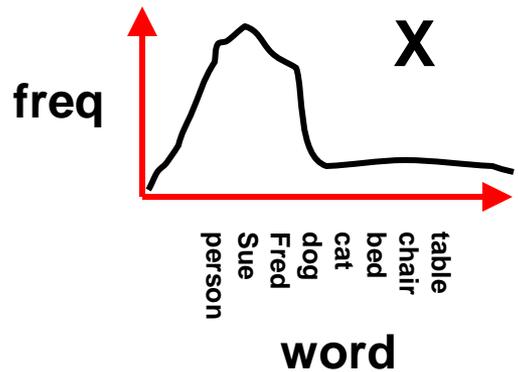
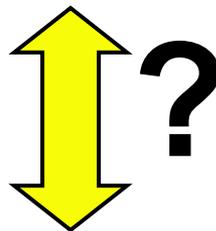
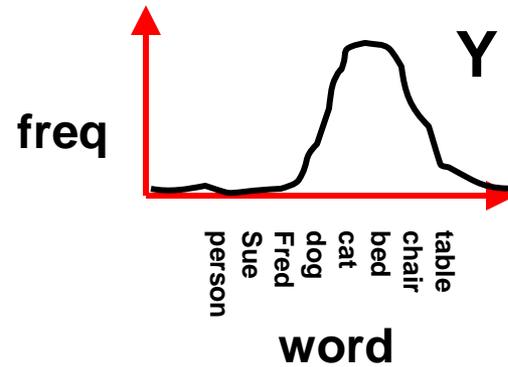


word

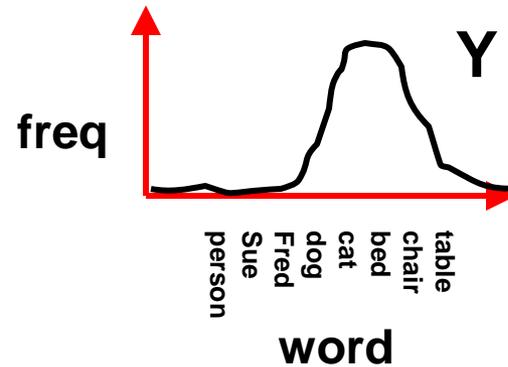
(Approximately) how DIRT learns rules



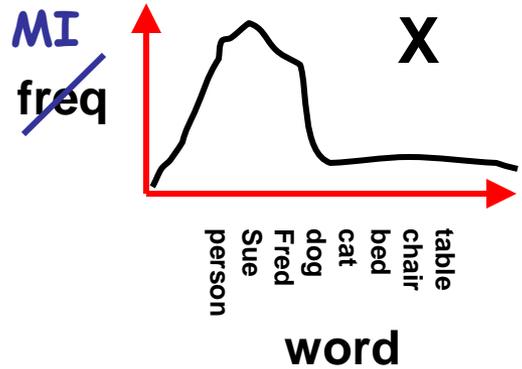
X loves Y



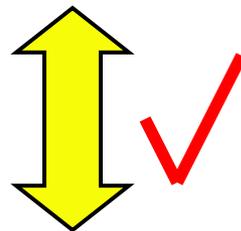
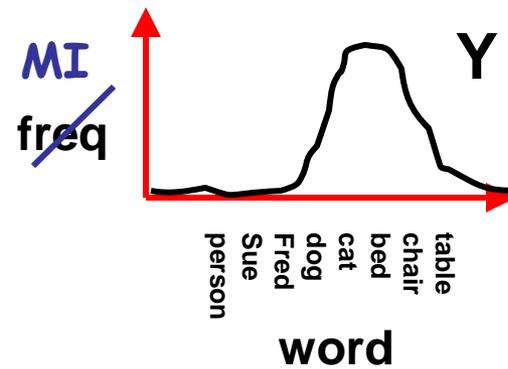
X likes Y



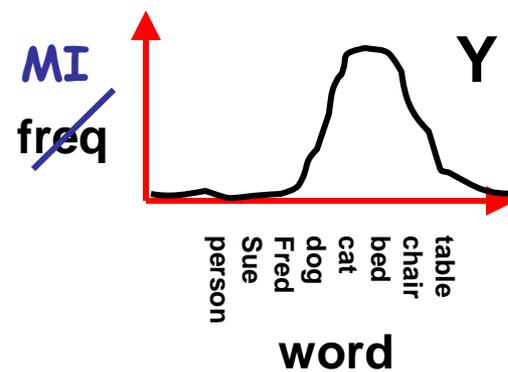
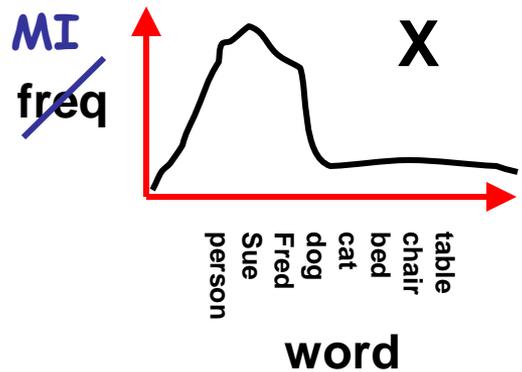
(Approximately) how DIRT learns rules



X loves Y



X likes Y



Successful Examples with DIRT

- Good example

24.H101

T: William Doyle works for an auction house in Manhattan.

H: William Doyle goes to Manhattan.

Successful Examples with DIRT

- Good example

24.H101

T: William Doyle **works** for an auction house **in** Manhattan.

H: William Doyle **goes to** Manhattan.

Yes! I have general knowledge that:

IF Y **works in** X THEN Y **goes to** X

Here: X = Manhattan, Y = Doyle

Thus, here:

We are told in T: Doyle works in Manhattan

Thus it follows that: Doyle goes to Manhattan

Successful Examples with DIRT

- Good(ish) example

54.H1

T: The president visited Iraq in September.

H: The president traveled to Iraq.

Successful Examples with DIRT

- Good(ish) example

54.H1

T: The president **visited** Iraq in September.

H: The president **traveled to** Iraq.

Yes! I have general knowledge that:

IF Y is **visited** by X THEN X **flocks to** Y

Here: X = the president, Y = Iraq

Thus, here:

We are told in T: Iraq is visited by the president

Thus it follows that: the president flocks to Iraq

In addition, I know:

"flock" is a type of **"travel"**

Hence: The president traveled to Iraq.

Unsuccessful Examples with DIRT

- Bad rule

55.H100

T: The US troops stayed in Iraq although the war was over.

H*: The US troops left Iraq when the war was over. [NOT entailed]

Unsuccessful Examples with DIRT

- Bad rule

55.H100

T: The US troops **stayed in** Iraq although the war was over.

H*: The US troops **left** Iraq when the war was over. [NOT entailed]

Yes! I have general knowledge that:

IF Y **stays in** X THEN Y **leaves** X

Here: X = Iraq, Y = the troop

Thus, here:

We are told in T: the troop stays in Iraq

Thus it follows that: the troop leaves Iraq

Hence: The US troops left Iraq when the war was over. (*wrong*)

Unsuccessful Examples with DIRT

- Misapplied rule

RTE4 797

T: In May, **Denver** underwent quadruple bypass surgery.
H*: **Denver** died in May. [NOT entailed]

Unsuccessful Examples with DIRT

- Misapplied rule

RTE4 797

T: In May, **Denver** underwent quadruple bypass surgery.
H*: **Denver** died in May. [NOT entailed]

Yes! I have general knowledge that:

IF Y **occurs in** X THEN **someone dies of** Y in X

Here: X = May, Y = Denver

Thus, here:

I can see from T: Denver occurs in May

(because "undergo" is a type of "occur")

Thus it follows that: someone dies of Denver in May

Hence: Denver died in May.

(wrong)

Results with DIRT

Mismatches allowed?	RTE4 DIRT-based entailments	
	COVERED	ACCURACY
0 (= deductive reasoning)	6.2%	67%
1 mismatch	18.3%	54%
2 mismatches	19%	49%



Helps a little bit

Reflections on DIRT

- Potentially powerful, goes beyond just definitional knowledge

X marries Y \rightarrow Y marries X
X lives with Y
X kisses Y
X's wife Y
X has a child with Y
X loves Y
X is murdered by Y (!)

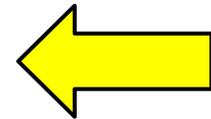
- But:
 - **Noisy** (but still useful)
 - Only **one rule type** (*can't* do “X buys Y \rightarrow X pays money”)
 - Helped with ~6% of the entailments (\rightarrow 250M needed?)

Overall Results

Respectable and gradually improving performance....

- RTE4:
 - **56.5%** (2 way) / mean performance 57.5%
- RTE5:
 - Main: **61.5%** (2 way) / mean performance 61.0%
 - Search: **F = 0.29** [pilot] / mean performance F = 0.22
- RTE6:
 - **F = 0.44** / mean performance F = 0.32 / max F = 0.48

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



- One obvious approach: $H (=Q+A_i)$ entailed by a sentence S ?

Q[11.3] Why were **transistor radios** a significant development?

A2 **young people could listen to pop outside**

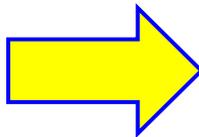
H₂ Transistor **radios** were a significant development because young people **could listen to pop outside**

H₂ entailed by any sentences S?

S27

...**transistor radios** meant that **teenagers could listen to music outside** of the home.

S27 → H2



Answer = A2

QA4MRE

- One obvious approach: $H (=Q+A_i)$ entailed by a sentence S ?

Q[11.3] Why were **transistor radios** a significant development?

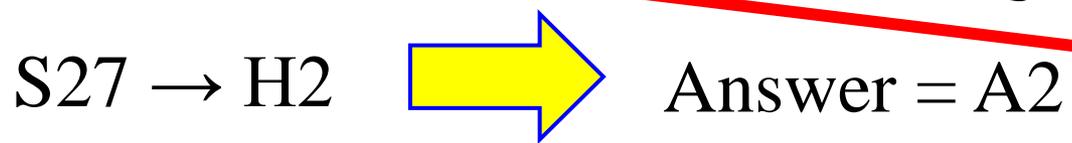
A2 **young people could listen to pop outside**

▪ **But:**

significant development because

- H s are hard to construct from $Q+A$
- H s are complex, and rarely fully entailed by S s
- Information about Q and A distributed in the document
- Multiple choice: *Relative entailment strength important*

S27 **music outside**



An Alternative Approach

1. Find sentences that entail the (target of) Q and A *independently*

S96 Public Law 106-264... earmarked 150 million dollars for each of the fiscal years 2001 and 2002, **for a Trust Fund.**

S98 **The Trust Fund will** also fund the implementation of specific HIV/AIDS programs in Africa.

S97 **The Trust Fund will** be used to leverage funds from multilateral development banks like the World Bank...



“entails” (to some degree)

Q[2.5] What is **the purpose of the Trust fund**...?

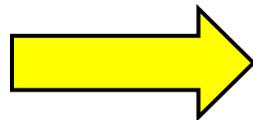
An Alternative Approach

1. Find sentences that entail the (target of) Q and A *independently*

S95 The Constituency for Africa proposed **a HIV/AIDS Marshall Plan for Africa with significant funds** to fight the disease.

S98 The Trust Fund will also **fund** the implementation of specific **HIV/AIDS programs in Africa**.

S2 More than 25 million **Africans live with HIV/AIDS**, and 17 million have already died.



“entails” (to some degree)

A2: ...**financing HIV/AIDS programs for Africa**

An Alternative Approach

1. Find sentences that entail the (target of) Q and A *independently*
2. Find evidence that the relation between the sentences = the relation between Q and A
≈ are the sentences close together?

IF $S1 \rightarrow Q$
AND $S2 \rightarrow A$
AND S1 and S2 are close
THEN $S1+S2 \rightarrow Q+A$

An Alternative Approach

1. Find sentences that entail the (target of) Q and A ***independently***
2. Find evidence that the relation between the sentences = the relation between Q and A
 \approx are the sentences close together?

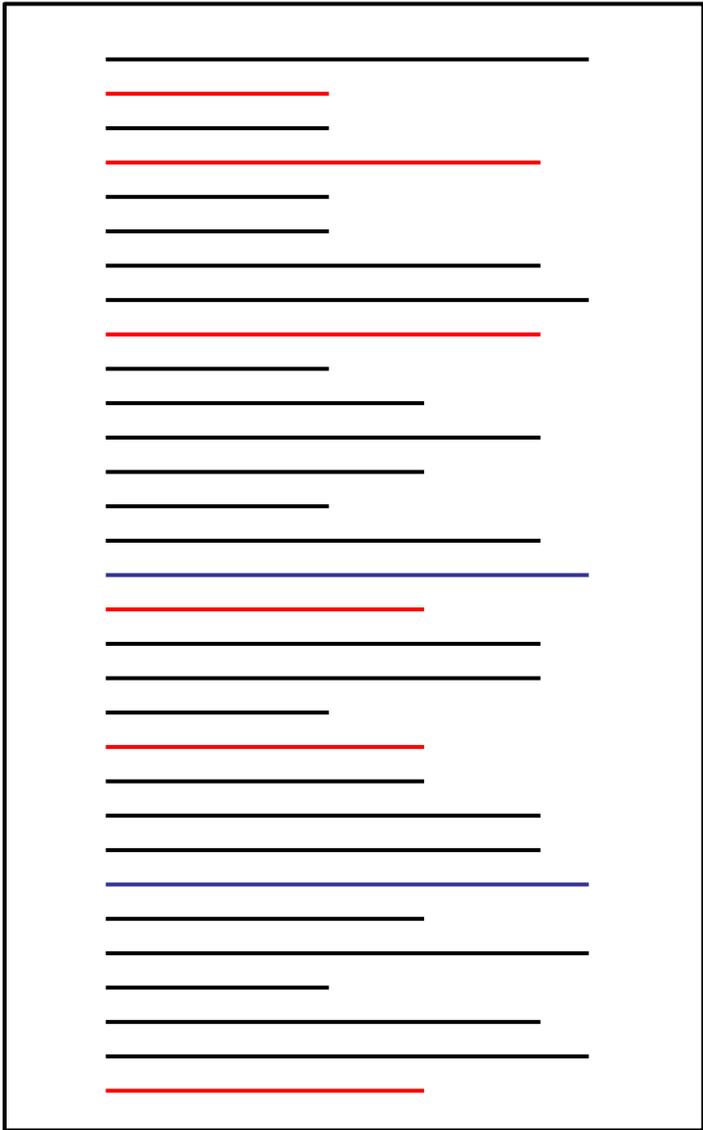
S98 **The Trust Fund will** also fund the implementation of specific HIV/AIDS programs in Africa.

 Q[2.5] What is **the purpose of the Trust fund**...?

S98 The Trust Fund will also **fund** the implementation of specific **HIV/AIDS programs in Africa.**

 A2: ...**financing HIV/AIDS programs for Africa**

Therefore, A2 strongly entailed



→ A2

→ A3

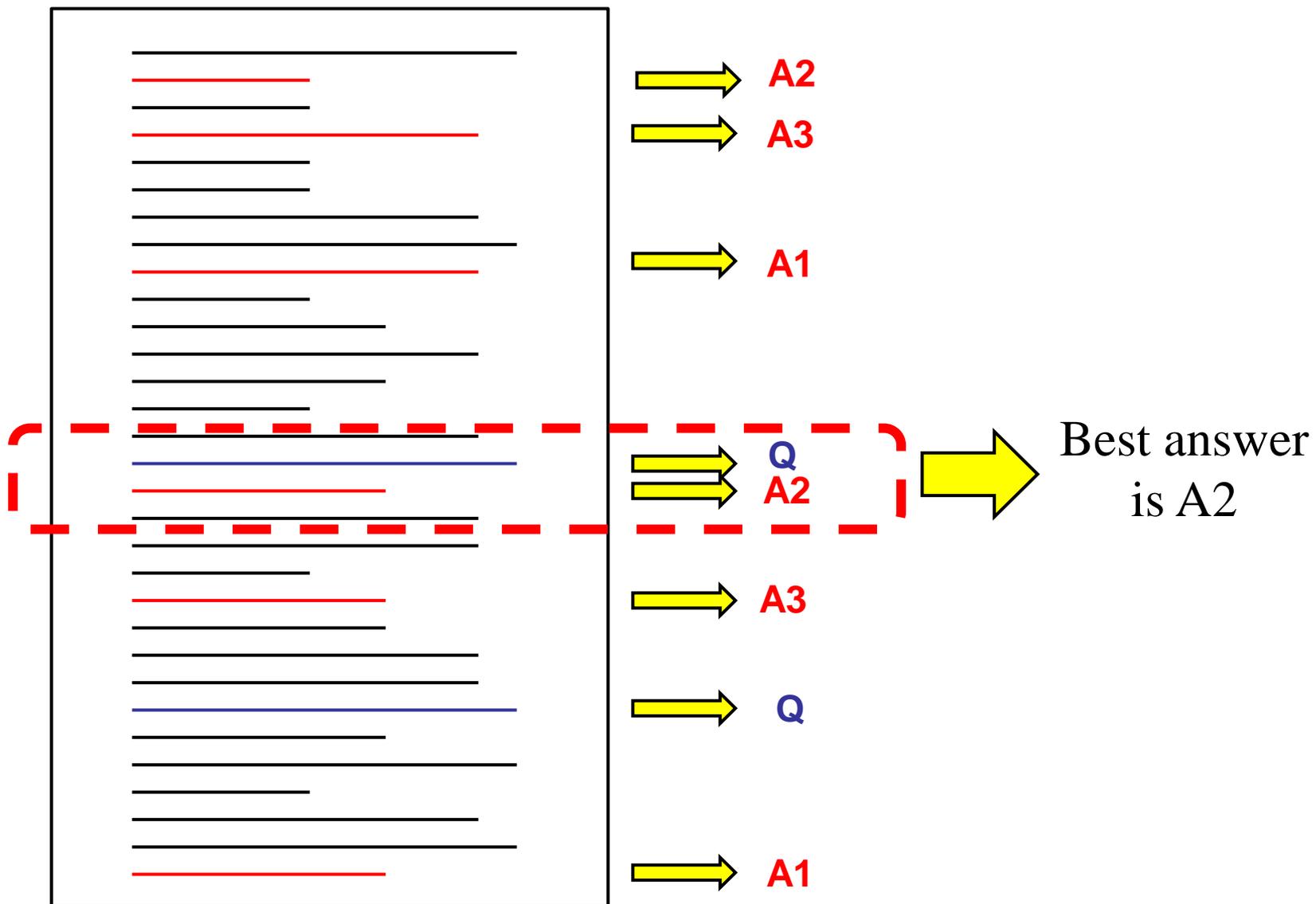
→ A1

→ Q
→ A2

→ A3

→ Q

→ A1

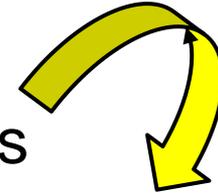


Assessing Entailment

- Hard to fully “prove” entailment
- Rather, we look for evidence of entailment
 - *Parts* of T entail (contain) *parts* of H
- Evidence of Entailment:
 1. Word (lexical) matches and entailments
 2. Parse-tree fragments
 3. Paraphrases

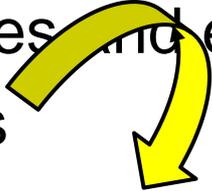
Assessing Entailment

- Hard to fully “prove” entailment
- Rather, we look for evidence of entailment
 - *Parts* of T entail (contain) *parts* of H
- Evidence of Entailment:
 1. Word (lexical) matches and entailments



- **Unusual words** carry more weight (“drawback” vs. “of”)
- **Topical words** carry more weight (“biofuel” vs. “particularly”)
- Two statistical measures of this:
 - $\text{Salience}(w) = \text{“how rare is } w\text{?”}$
 $= \log[1/p(w|\text{topical-docs})]$
 - $\text{Topicality}(w) = \text{“how unusually frequent is } w \text{ in topical docs?”}$
 $= \log[p(w|\text{topical-docs})/p(w|\text{general-docs})]$
- Use machine learning to combine:
 - $\text{Weight}(w) = \lambda.\text{salience}(w) + (1 - \lambda).\text{topicality}(w)$

Assessing Entailment

- Hard to fully “prove” entailment
- Rather, we look for evidence of entailment
 - *Parts* of T entail (contain) *parts* of H
- Evidence of Entailment:
 1. Word (lexical) matches and entailments
 2. Parse-tree fragments
 3. 

Score for shared parse fragment

$$= [\sum \text{scores for shared words in fragment}] \cdot k$$

S98 The Fund will also ...specific **HIV/AIDS programs in Africa.**

A2: ...financing **HIV/AIDS programs in Africa**

**“in”(“program”,“Africa”)
mod(“program”,“HIV/AIDS”)**

Assessing Entailment

- Hard to fully “prove” entailment
- Rather, we look for evidence of entailment
 - *Parts* of T entail (contain) *parts* of H
- Evidence of Entailment:
 1. Word (lexical) matches and entailments
 2. Parse-tree fragments
 3. Paraphrases



Use WordNet, DIRT, and ParaPara

Assessing Entailment

- Hard to fully “prove” entailment
- Rather, we look for evidence of entailment
 - *Parts* of T entail (contain) *parts* of H

T S8 Being threatened with a pulmonary affection he [Burney] went in 1751 to Lynn in Norfolk.

paraphrase

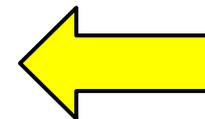
word match

synonym

H A5 He [Burney] suffered from a disease.

S8 → **A5** with strength w

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



Good Examples with DIRT

Q[2.5] What is the purpose of the Trust fund established by the US Congress?

A2: leveraging financial funds and financing HIV/AIDS programs for Africa **[Yes]**

S98: The Trust Fund will **fund** the implementation of ...HIV/AIDS programs in Africa.



IF X funds Y THEN X finances Y

A2: leveraging financial funds and **financing** HIV/AIDS programs for Africa

Good Examples with DIRT

Q[2.5] What is the purpose of the Trust fund established by the US Congress?

A2: leveraging financial funds and financing HIV/AIDS programs for Africa [Yes]

S98: The Trust **Fund** will **fund** the implementation of ...**HIV/AIDS programs** in **Africa**.

A2: leveraging financial **funds** and **financing HIV/AIDS programs** for **Africa**

- S98 is one of the top 3 sentences likely entailing most of A2 (**S98 → A2**)
- S98 is one of the top 3 sentences likely entailing most of Q (**S98 → Q**)
- S98 and S98 are close (actually, the same)
- So S98 likely entails most of Q+A2 (**S98 → Q+A2**)

Therefore **A2 is the answer**

CORRECT



Bad Example with DIRT

Q[12.7] How did Lulli conduct?
A2 He lived in Paris.
A4 He used a cane. [correct]

S3 He is said to have **visited** Paris, where Lulli exhibited such jealousy...that Corelli withdrew.

IF X **visits** Y THEN X **lives in** Y

A2 He **lived in** Paris.

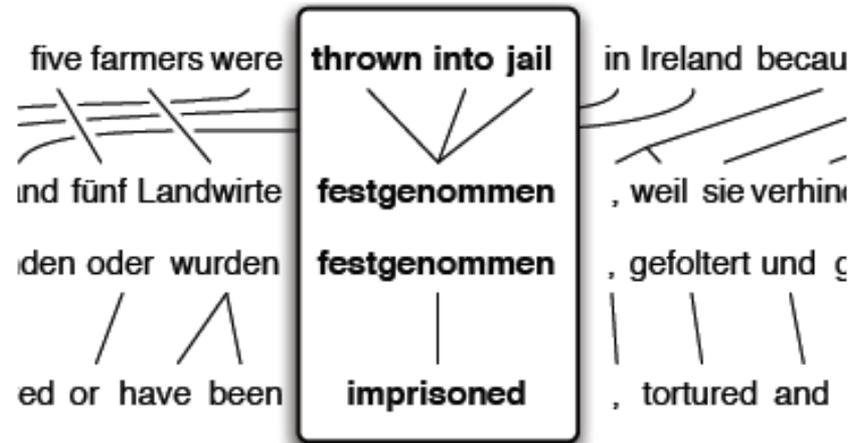
- S3 is one of the top 3 sentences likely entailing most of A2 (**S3 → A2**)
- S3 is one of the top 3 sentences likely entailing most of Q (**S3 → Q**)
- S3 and S3 are close (actually, the same)
- So S3 likely entails most of Q+A2 (**S3 → Q+A2**)

Therefore **A2 is the answer** **INCORRECT**



The ParaPara Paraphrase Database

- Paraphrases learned via **bilingual pivoting**
- Then filtered by **distributional similarity** against Google N-Grams



Good Example with ParaPara

Q[3.5] What is one of the MCP goals in Third World countries?
A5 separation of family planning from HIV prevention

S30 ...U.S. funding for global HIV programs will be...**aimed at** separating family planning from HIV prevention in **developing countries**.

"aimed at" → "goals in"

"developing country"
→ "third world country"

Q[3.5] What is one of the MCP **goals in Third World countries**?

- S30 is one of the top 3 sentences likely entailing most of Q (**S30 → Q**)
- S30 is one of the top 3 sentences likely entailing most of A5 (**S3 → A5**)
- S30 and S30 are close (actually, the same)
- So S30 likely entails most of Q+A5 (S30 → Q+A5)

Therefore **A5 is the answer** **CORRECT**



Bad Example with ParaPara

Q[3.10] Who considers HIV as a gay disease?

A2 President Bush [correct]

A4 intimate partners

S30 Now comes the **announcement** that...funding will be...aimed at separating family planning from HIV prevention in developing **countries**.

"announcement" → "intimate"

"country" → "partner"

A4 **intimate partners**

- S30 is one of the top 3 sentences likely entailing most of Q (**S30 → A4**)
- S28 is one of the top 3 sentences likely entailing most of Q (**S28 → Q**)
- S28 and S30 are close (2 sentences apart)
- So S28+S30 likely entails most of Q+A4 (S28+S30 → Q+A4)

Therefore **A4 is the answer**

INCORRECT



Results and Ablation Studies

Best run: 40.0 (submitted), 42.5 (subsequent version of the system)

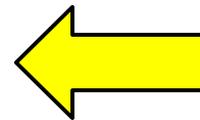
Subtractive ablations

- 42.5** Main system (all resources)
- 41.9** minus WordNet (only)
- 38.1** minus ParaPara (only)
- 41.9** minus DIRT (only)
- 38.1** baseline (none of the resources)

Additive ablations

- 38.1** baseline (none of the resources)
- 41.9** add WordNet (only)
- 39.4** add ParaPara (only)
- 41.9** add DIRT (only)
- 42.5** Main system (all resources)

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



Knowledge Limitations

- Huge amounts of knowledge still needed, e.g.,

Q[2.7] What is the **external debt** of all African countries?

S61 Africa **owes foreign banks and governments** about 350 billion.

- The amount you owe is your debt
- External debt is debt to outside groups
- If something is foreign, it is outside your country
- Africa is made up of all African countries

Knowledge Limitations

- Huge amounts of knowledge still needed, e.g.,

Q7.9: What solution has been applied in places suffering from water-stress?

A4: to **encourage the use of** groundwater [correct]

S73 the UN has ...a program **to give them access to** groundwater sources.

- UN sets up programs to help people
- Access is a prerequisite for use.
- If you give someone a prerequisite for X, then you encourage X.

Q[2.1] When did the rate of AIDS started to **halve** in Uganda?

A1: the 1990s [correct]

S73 The rate of AIDS in Uganda is **down to about 8, from a high of 16** in the early 1990s.

Reasoning Limitations

- Deductive reasoning is inappropriate

T: the Bush administration has **blocked** the sale of affordable generic drugs

H*: The administration **approved** the sale of drugs. [NOT entailed]

- Our system incorrectly inferred this, via:
 - X **blocks** $Y \rightarrow X$ **approves** Y
- Problem: ignoring evidence **against** H:
 - WordNet: “block” and “approve” are antonyms
 - World: “block” \rightarrow “unavailable” (mentioned later)

Better (and towards Machine Reading):

1. Look at multiple implications
2. Find “best”, consistent subset of facts

2. Identify Conflicts

"T" text:

..the Bush administration has blocked the sale of affordable generic drugs...
...many generic drugs are still unavailable.....

Bush blocked the drug sales

Drugs are unavailable

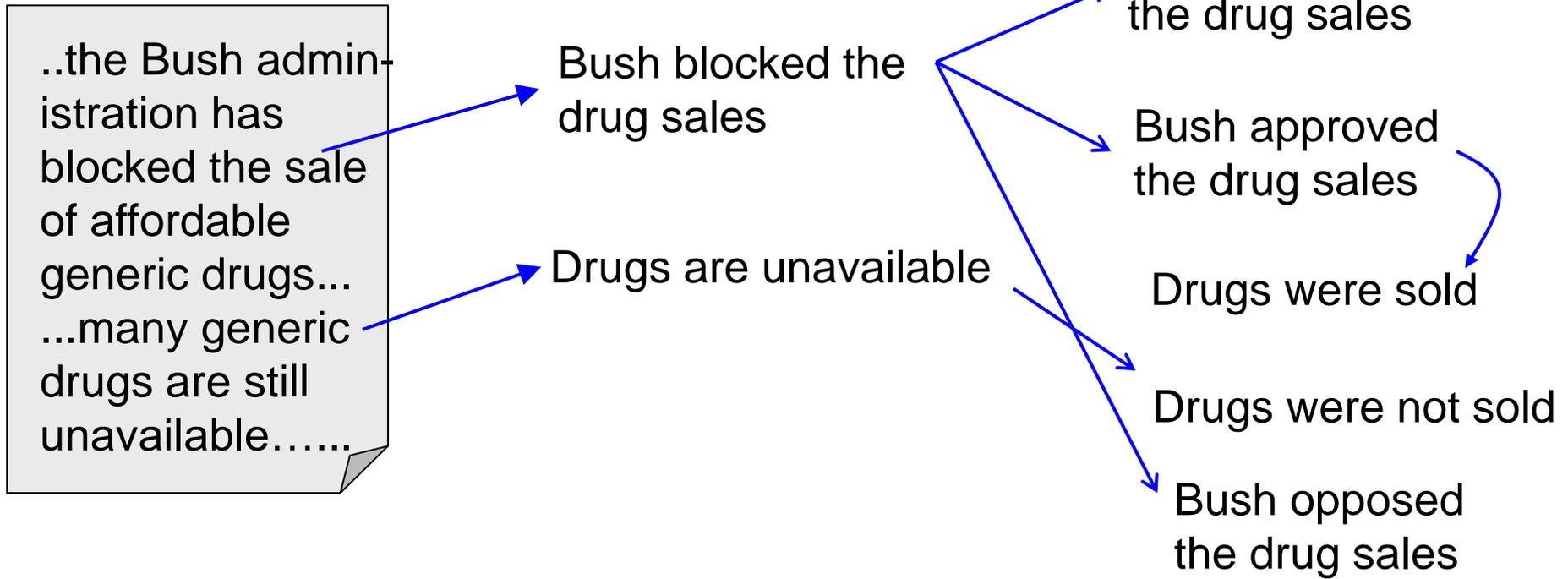
Bush prevented the drug sales

Bush approved the drug sales

Drugs were sold

Drugs were not sold

Bush opposed the drug sales



2. Identify Conflicts

"T" text:

..the Bush administration has blocked the sale of affordable generic drugs...
...many generic drugs are still unavailable.....

Bush blocked the drug sales

Drugs are unavailable

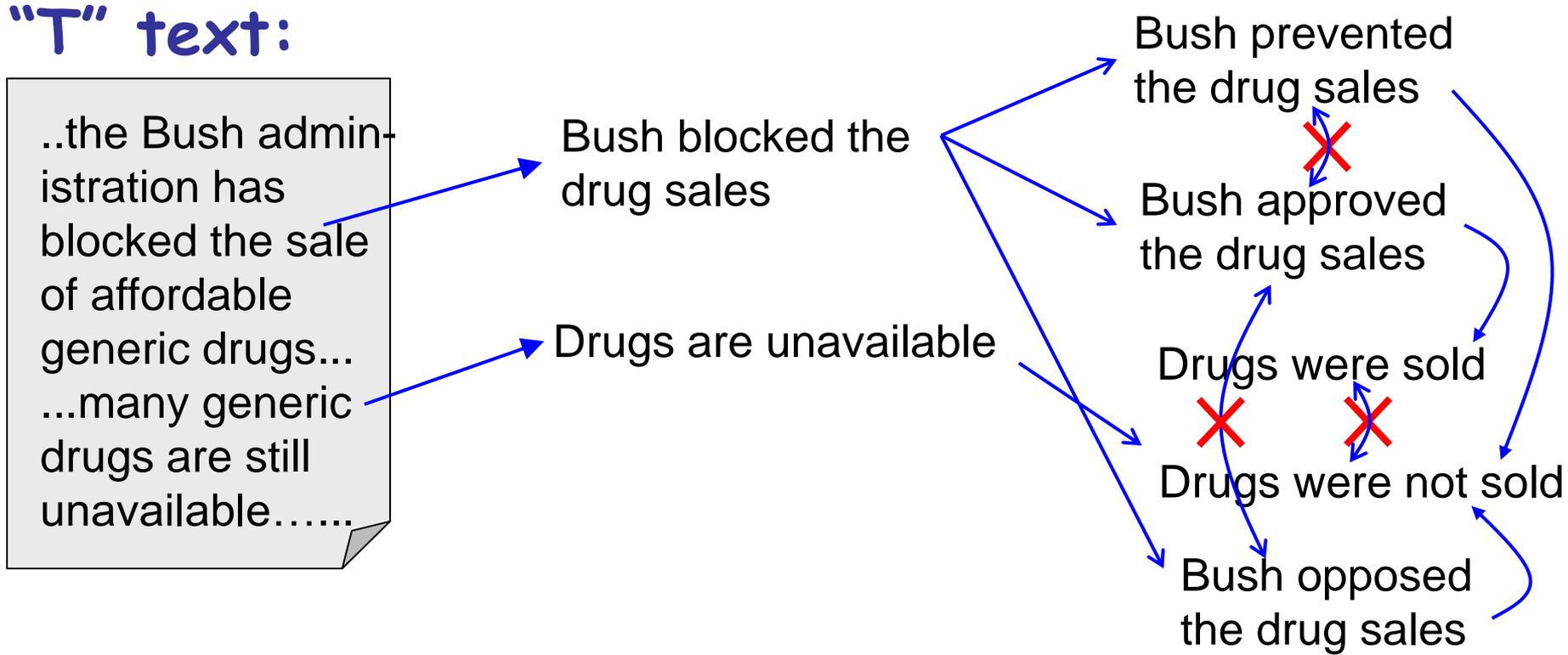
Bush prevented the drug sales

Bush approved the drug sales

Drugs were sold

Drugs were not sold

Bush opposed the drug sales



2. Identify Conflicts

"T" text:

..the Bush administration has blocked the sale of affordable generic drugs...
...many generic drugs are still unavailable.....

Bush blocked the drug sales

Drugs are unavailable

Bush prevented the drug sales

~~Bush approved the drug sales~~

~~Drugs were sold~~

Drugs were not sold

Bush opposed the drug sales

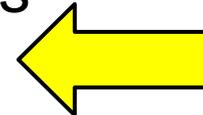
Can answer questions:

were affordable drugs sold? No

Forming a "picture" of the scene

Getting towards text understanding!

- Introduction
- **The RTE Competitions**
 - Overview
 - Our attempts at Natural Logic and RTE
 - WordNet as a Knowledge Source
 - DIRT Paraphrases as a Knowledge Source
- **QA4MRE**
 - A modified textual entailment approach
 - The role of paraphrases
 - The knowledge and reasoning problems
- **Reflections**



Reflections on QA4MRE

- Challenging
- Tests deeper understanding, while still simplifying
- Pushes beyond simple lexical methods

My main takeaways:

- Even with “natural logic” you can rarely “prove” an answer
 - Partly lack of world knowledge
 - Partly need a “leap of faith”
 - **IF** most pieces fit **THEN** assume the remainder also fit

Q[11.3] Why were **transistor radios** a **significant development**?

A2 **young people could listen to pop outside**

...**transistor radios** meant that **teenagers could listen to music outside** of the home.

Reflections on QA4MRE

- Challenging
- Tests deeper understanding, while still simplifying
- Pushes beyond simple lexical methods

My main takeaways:

- Even with “natural logic” you can rarely “prove” an answer
 - Partly lack of world knowledge
 - Partly need a “leap of faith”
 - **IF** most pieces fit **THEN** assume the remainder also fit
 - Less “finding a proof”, more “finding coherent evidence”

Conclusions

- **Machine Reading**
 - ≠ just parsing and disambiguation
 - = forming a coherent model of the text
- **Recognizing Textual Entailment**
 - A fundamental operation in Machine Reading
 - Our approach: Natural logic + paraphrases
- **QA4MRE**
 - Moved from “find a full proof” to “seek coherent evidence”
 - Outstanding challenges: knowledge and reasoning
 - QA4MRE is a great challenge for the community!

Thank you!