

Considering Discourse References in Textual Entailment Annotation

L. Bentivogli¹, I. Dagan²,
H. T. Dang³, D. Giampiccolo⁴,
M. Lo Leggio⁴, B. Magnini¹

¹FBK-irst, Trento, Italy

²Bar- Ilan University, Ramat Gan, Israel

³NIST, Gaithersburg, Maryland, USA

⁴CELCT, Trento, Italy

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

Textual entailment: directional relationship between pairs of text expressions, denoted by T - the entailing “Text”, and H - the entailed “Hypothesis”

T entails H if, typically, a human reading T would infer that H is most likely true

T: *A shootout at the Guadalajara airport in May, 1993, killed Cardinal Juan Jesus Posadas Ocampo and six others.*

H: *Juan Jesus Posadas Ocampo died in 1993.*

Recognizing Textual Entailment (RTE)

- annual evaluation campaign started in 2005, now at 5th edition
- the **traditional task**: given a manually created H-T pair, recognize whether T entails H
- **RTE-5**: a new pilot task has been introduced to answer the need of moving towards more realistic scenarios

- Introduction
- **The Entailment Search Task**
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

The RTE-5 Search Pilot Task



Consists in finding all the sentences that entail a given H in a given set of documents

Hypotheses Set

H1: The AS-28 mini-submarine was trapped underwater.
H2: Seven submariners were onboard the AS-28
H5: The AS-28 crew was rescued in satisfactory conditions.
...

Document 1

S1: Effort seen to raise stricken Russian sub

S2: The effort to attach the cables marked the start of an operation to raise the vessel trapped below the surface with seven crewmen on board.

S4: He was subsequently quoted by Interfax news agency as saying that attachment of the cables to the mini-submarine was not certain and that a video camera was being lowered to find out what the cables were attached to.

S5: "Then we can tell 100 percent that it's the submarine.

S6: And only then can we say that we carried out our task," Fyodorov said.

...

S8: The submarine, used in rescue, research and intelligence-gathering missions, became stuck at the bottom of the Bay of Berezovaya off Russia's Far East

S9: Fyodorov said earlier that the seven crewmen were in satisfactory condition...

...

Document 2

S0: Japan sends help to trapped Russian submarine

S1: Japan on Friday dispatched four military ships to help Russia rescue seven crew members aboard a small submarine trapped on the seabed in the Far East.

S3: The Russian Pacific Fleet said the crew had 120 hours of oxygen reserves on board when the submarine submerged at midday Thursday 300 GMT Wednesday).

S4: Japan dispatched a submarine-rescue vessel, and a depot ship from Yokosuka port and two minesweepers from Hokkaido, the defense spokesman said.

S5: "We will do our utmost efforts to rescue them.

S6: We are hopeful," he said.

S7: The assistance comes despite rocky relations between Japan and Russia, which have yet to formally end World War II amid Japan's claims to four islands off Hokkaido that Soviet troops seized in August 1945.

...

Document 3

S0: Russian sub snagged on undersea surveillance antenna: official

S1: Rescue of a submarine stuck on the seabed off Russia's east coast is complicated because it is snagged in an underwater surveillance antenna system as well as snared in a fishing net, a senior Russian naval officer said Friday.

S2: A remote-controlled device was lowered to the stricken vessel "to cut the flexible tubes and cables of the coastal surveillance antenna in which the AS-28 vehicle is caught," Rear Admiral Vladimir Pepelayev was quoted by RIA-Novosti news agency as saying.

S3: The vessel was spotted by Russian surveillance

S6: There are seven crew members aboard the vessel, stranded on the ocean floor in a bay off the coast of the Kamchatka peninsula in Russia's Far East region.

S4: The vessel was spotted by Russian surveillance

S7: There were widely varying accounts of how much oxygen was remaining for the crew.

...

Traditional vs Pilot Task

	<u>Traditional Task</u>	<u>Search Task</u>
Aim	<ul style="list-style-type: none">• Judging whether T entails H, for each manually created T-H pair	<ul style="list-style-type: none">• Retrieving all the sentences that entail a given H in a set of documents about a topic
Data Set	<ul style="list-style-type: none">• The distribution of entailing T's is determined a priori	<ul style="list-style-type: none">• Reflects the natural distribution of entailment in a corpus
References	<ul style="list-style-type: none">• Both T and H do not contain any textual reference outside the pair itself	<ul style="list-style-type: none">• Both T and H are to be interpreted within the context of the topic
Verb tenses	<ul style="list-style-type: none">• Verb tenses are ignored	<ul style="list-style-type: none">• Verb tenses must be interpreted in the context

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution: entities, time, space, events
- Conclusions and future work

Knowledge Required for the Entailment Judgment



The standard definition of entailment assumes prior knowledge:

Knowledge + Text \rightarrow Hypothesis

Three kinds of knowledge:

Traditional
and Search
Task

- **Linguistic Knowledge:** to interpret H and T
- **Common background World Knowledge:** to make additional inferences

Only
Search Task

- **Text and Corpus Knowledge:** to resolve local and cross-document co-references

Text and Corpus Knowledge

H: 2003 UB313 is bigger than Pluto.

Document 1

S0: US scientists announce **possible '10th planet'**
'10th planet'

S1: A US astronomer announced Friday he had discovered a **what could be a 10th planet in the solar system**

billion kilometers (nine billion miles) from earth.

S2: If confirmed, the discovery by Professor Mike Brown of the respected California Institute of Technology would be the first of a planet since Pluto was identified in 1930.

S3: **The body** is "the 10th planet of the solar system," said Brown, a professor of planetary astronomy in a phone conference from CalTech in the town of Pasadena.

S4: **It** the farthest object ever discovered to

S5: "It's definitely bigger than Pluto".

Corpus

Document 2

so-called 10th planet iting

S1: The astronomers who claim to **the 10th planet in the solar system**

S2: While observing the new **the new, so-called planet**

Brown of the California Institute of Technology spotted a faint object trailing next to it.

S4: The moon discovery is important **the new planet** scientists mass.

S5: It **an icy, rocky object**

larger than Pluto in the Kuiper Belt, a disc of icy bodies **the object**

S6: Brown lab **the object** net and nicknamed **it** Xena after the lead character in the former TV series "Xena: Warrior Princess."

S7: By determining the moon's distance and orbit around **Xena**, **Xena** in calculate how heavy

...

Document 3

S0: **10TH PLANET** D IN SKE

S1: Add a tenth planet to the solar system -- or possibly subtract one.

S2: Astronomers announced Friday that **a lump of rock and ice** and

S3: The discovery will likely rekindle debate over the definition of "planet" and whether Pluto still merits the designation.

S4: **The new obj** but temporarily **2003 UB313** UB313 -- is now 9 billion miles away from the sun and about three times Pluto's current distance from the sun.

S5: Its 560-year elliptical orbit brings **it** close as 3.3 billion miles.

Pluto's orbit ranges between 2.7 billion and 4.6 billion miles.

S7: The astronomers do not have an exact size **the new planet** brightness **it**

S8: **It** is guaranteed bigger than Pluto, said Michael E. Brown, a professor of planetary astronomy at Caltech.

...

Text 4, Text 5 ...

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- **The Data set**
- Creation of hypotheses
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

The Pilot Search Task of the 5th RTE campaign is situated in the Summarization application setting.

The Development Set has been created from the TAC 2008 Update Summarization Task Data Set:

- **Documents:** 10 topic newswire collections randomly selected from the 48 TAC topics
- **Hypotheses:** based on Summary Content Units (SCUs) that have been created from human-authored summaries of the topic collections

- **Description:**

- 10 Topics:
 - 10 newspaper articles manually splitted into numbered sentences
 - Between 6 and 10 Hypotheses
- Three annotators

- **Numbers:**

- 80 H's
- 2,538 sentences
- 20,104 sentence annotations (annotation of each sentence against each of the related hypotheses)
- 810 entailing sentences

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- **Creation of hypotheses**
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

Criteria:

- Explicitness:

to reduce ambiguities and facilitate the correct interpretation of H's

- Conciseness:

to maintain linguistic fluency

- Time anchoring:

if not otherwise specified, the content of H is conventionally fixed at the day after the publication of the last document in the topic collection

Examples of Hypotheses



- Dennis Rader was arrested on February 25, 2005

ENTITY (person):
complete proper name

TIME:
explicit absolute date

- The ice is melting in the Arctic

TIME:
conventional anchoring
→ the day after the last
document: 2005/08/05

SPACE:
explicit mention
of location

- The Kansas Bureau of Investigation collected hundreds of DNA swabs related to the BTK case

EVENT:
use of a short phrase

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution:
entities, time, space, events
- Conclusions and future work

H: Michael Brown discovered 2003 UB313

Michael Brown mentions:

“Michael Brown”
“Michael A. Brown”
“Michael E. Brown”
“Mike Brown”
“Brown”
“he”
“the astronomer”
“the Caltech professor”
...

2003 UB313 mentions:

“2003 UB313”
“UB313”
“Xena” (nick-name)
“it”
“the planet”
“the new planet”
“the possible new planet”
“the object”
“a lump of rock and ice”
“the 10th planet of the solar system”
...

Entity References/2

H: 2003 UB313 has a moon.

...2003 UB313...

- ✓ **T1:** The astronomers who claim to have discovered the 10th planet in the solar system have made another intriguing announcement. it has a moon.

...EL61...

- ✗ **T2:** It is different from the previously discovered Kuiper bodies in that it has a moon that circles it every 49 days in a highly elliptical orbit.

Time References/1

- Conventional anchorage in H / Deictic expression in T

H: *The ice is melting in the Antarctic*

[Time anchor: 2005/08/15]

- ✓ **T1:** Last month, scientists have sounded an alarm bell on the effect of global warming on Antarctica. **May 2005** (date of the article: 2005/06/06) at more than 200 coastal glaciers are in retreat.
- ✓ **T2:** (approximately) second half of 20th century (date of the article: 2005/06/06) drain inland ice on the Antarctic peninsula back over the last half century.
- ✗ **T3:** Since **9,000 BC** (date of the article: 2005/08/03) Ice Age, some 11,000 years ago, the iceshelf had slowly thinned, by several dozen metres.

Time References/2

- Past tense in H / Future tense in T: modality

H: Orhan Pamuk went to court on 16 December 2005

✓ **T1:** "It is not Orhan Pamuk who *will stand trial tomorrow* but Turkey."

16 December 2005
(date of the article:
2005/12/15)

✗ **T2:** Pamuk, 53, Turkey's best-known novelist, *is expected to go on trial Friday* [...].

16 December 2005
(date of the article:
2005/12/15)

Space References

Ha: The ice is melting in the Arctic.

Hb: The ice is melting in the Antarctic.

... the Exit Glacier in the Arctic ...

- **T1:** [...] people moved ever closer to the rope line near the glacier as it shied away, practically groaning and melting before their eyes.

✓ **Entailment for Ha: YES**

✗ **Entailment for Hb: NO**

... in the Antarctic Peninsula...

- **T2:** Of the 244 glaciers that drain inland ice and feed these shelves, 87 percent have fallen back since the mid-1950s, according to a British study published in April.

✗ **Entailment for Hb: NO**

✓ **Entailment for Ha: YES**

H: The AS-28 was at a depth of about 600 feet when it was trapped.

... the AS-28 mini-submarine was freed on Sunday from its undersea entanglement...

✓ **T:** The vessel rose to the surface at 4:26 p.m. local time, ending the crew's ordeal in the cold and darkness more than 600 feet below the surface off the Kamchatka Peninsula.

- Introduction
- The Entailment Search Task
- Knowledge required to perform entailment annotation
- The Data set
- Creation of hypotheses
- Entailment annotation and co-reference resolution: entities, time, space, events
- **Conclusions and future work**

- Preliminary qualitative evaluation of the impact of corpus knowledge and reference resolution on the textual entailment annotation in a natural occurring corpus
- Task: innovative way of performing textual entailment recognition as a search task against a corpus
- Data set: useful resource that reflects the “natural” distribution of entailment

- Realization of an enriched data-set containing all the entailing sentences annotated with:
 - Type(s) of co-reference involved in the entailment
 - Resolution of implicit/deictic/phoric references, relevant for the entailment
- Contradiction detection in the future RTE campaigns

Thank you for your attention

The Traditional Task



Given a couple H-T, the task consists in deciding if

- T entails H
- T contradicts H
- the veracity of H can't be inferred from T

T: Internet media company Yahoo Inc. announced Monday it is buying Overture Services Inc. in a \$1.63-billion (U.S.) cash-and-stock deal that will bolster its on-line search capabilities.

H1: *Yahoo bought Overture* → **Entailment**

H2: *Overture bought Yahoo* → **Contradiction**

H3: *Yahoo manufactures cars* → **Unknown**

Knowledge required for entailment annotation/2



- Linguistic Knowledge

H: Mine accidents cause deaths in China.

T: Four mine disasters in China have claimed the lives of at least 60 workers.

to claim the lives = to cause the death of some people

- World Knowledge

H: The ice is melting in the Arctic.

T: The scene at the receding edge of the Exit Glacier in Alaska was part festive gathering, part nature tour with an apocalyptic edge.

Alaska is in the Arctic

The edge of the glacier is receding because the ice is melting

Creation of Hypotheses/2



- **Entities:** *Michael Brown discovered 2003 UB313*
- **Time:**
 - explicit date:
Dennis Rader was arrested on February 25, 2005
 - conventional anchorage:
The ice is melting in the Arctic
(anchored on the day after the last document,
2005/08/15)
- **Locations:** *Mine accidents cause deaths in China*
- **Events:** *The Kansas Bureau of Investigation collected hundreds of DNA swabs related to the BTK case*
About 50 people were killed in the attack

- Absolute expression in H / Deictic expression in T

H: BTK resurfaced in 2004 after a period of silence.

- ✓ **T:** The killer has sent several letters to police since resurfacing last March after years of silence.

Time Reference: publication date of the article: 2005/01/06

→ *last March* = March 2004

Entailment: YES

- Past tense in H / Present tense in T

**H: The conditions in subway tunnels after the attack
hampered rescue operations**

[Time anchor: 2005/07/12]

- ✓ **T:** "It is extremely hot and very dusty and it is a great challenge for them to continue their work to recover the remaining bodies from the train underground," British Transport Police Chief told.

Time Reference: publication date of the article:
2005/07/09

H: Each bomb used in the attack contained less than 10 pounds of explosives.

- ✓ **T:** Forensic evidence indicates that the bombs each contained less than 10 pounds of high explosives – the Madrid bombs weighed 17 to 22 pounds - which could have kept them small enough to fit in rucksacks, Commissioner Blair said.

Event co-reference (implicit): the first mention of bombs refers to the (London) attack.

Entailment: YES

Annotation and References: Difficult cases



H: Sea levels are rising.

- **T:** But what has happened since did.

Discourse analysis: previous sentence:

“Because an ice "shelf" already floats on the sea, displacing its weight in water, Larsen B's disintegration - and that of the smaller, nearby Larsen A in 1995 - didn't raise ocean levels.”

→ *did* = raise ocean levels.

Entailment: ann1: NO, ann2: NO, ann3: YES

