EVALITA 2009
Recognizing Textual Entailment (RTE)
Italian Chapter

Johan Bos[1], Fabio Massimo Zanzotto[2], Marco Pennacchiotti[3]

[1]University of Rome "La Sapienza", Italy

[2]University of Rome "Tor Vergata", Italy

[3]Yahoo! Labs, Sunnyvale, CA, United States

Outline

- Introduction to the Recognizing Textual Entailment in Italian (iRTE)
- Building the corpus: Using Wikipedia
- Evaluation and results:
  - dataset and metrics
  - systems' results
- Conclusion and future works

What is Textual Entailment Recognition?

The textual entailment recognition task:

determine whether or not a text T implies an hypothesis H

$T_1 \Rightarrow H_1$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_1	*"At the end of the year, all solid insurance companies pay dividends."*

Why it is extremely difficult?

$T_1 \not\Rightarrow H_2$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_2	*"At the end of the year, all solid companies pay cash dividends."*

Using Wikipedia Revisions for Building a Corpus

It is well known that Wikipedia is an open encyclopedia:

- where every person can behave as an author
- inserting new entries or modifying existing ones.

If **S1** is a piece of text in Wikipedia before it is modified by an author, and **S2** is its **revision**

Our hypothesis is that (S1, S2) pairs extracted from the Wikipedia database, represent good candidate of both true and false entailment pairs (T,H)

Using Wikipedia Revisions for Building a Corpus

S_1	"Tutt'ora, nel 2008, e a 40 anni d'età, delizia seppur con qualche pausa a causa di qualche infortunio in più, i suoi tifosi."
S_2	"Tutt'ora, nel 2008, e a 40 anni d'età, delizia seppur con qualche infortunio in più, i suoi tifosi"

S_1	"In carcere si convertì al cattolicesimo, si sposò e visse fino al 1981, senza che di lei si sapesse mediaticamente più nulla."
S_2	"In carcere si convertì al cattolicesimo, si sposò e visse fino al 1981, senza che di lei si sapesse più nulla"

Using Wikipedia Revisions for Building a Corpus

Annotation Guidelines

- The annotators classified each pair into one of the following classes:
  - **bidirectional**: S1 entails S2 and viceversa
  - **left**: S1 entails S2, but not viceversa
  - **right**: S2 entails S1, but not viceversa
  - **no**: neither S1 entails S2, nor viceversa
  - **reject**: rejected pairs

iRTE Corpus Description

Selecting Relevant Wikipedia Entries

Target	Domain	Entries	Pairs
Training	Italian, German, Cuban, US, and Russian Revolutionary People; Polititians of Popolo della Libertà, Partito Democratico, and Forza Italia; Massons	881	7,651
Testing	Soccer Players (a–d)	1,198	7,720
	Soccer Referees	86	
	Important Cases	62	

Final Corpus

- Training: 400 balanced pairs
- Testing: 400 balanced pairs

Participants

We had two participants:

- The Alicante/Pisa Team
  - machine learning classifier
  - features derived from
    - lexical distances,
    - part-of-speech information
    - Semantic knowledge taken from SIMPLE-CLIPS, an Italian language resource
- The FBK Irst's team
  - EDITS (Edit Distance Textual Entailment Suite), a freely available open source tool for Recognizing Textual Entailment (RTE)

Participants' Results

Run	Correct	Accuracy
FBKirst_run1.txt	285	0.71
FBKirst_run2.txt	282	0.71
ofe_semTypes_1.txt	257	0.64
ofe_semTypes_2.txt	228	0.57
ofe_lexical_2.txt	230	0.58
ofe_lexical_1.txt	225	0.56
FBKirst_run4.txt	202	0.51
FBKirst_run3.txt	199	0.50

Conclusion

- Compared to the English edition, the Italian RTE shared task attracted relatively few participants.
- Most of the research on Natural Language Processing focusses on English.

Is it too early in the development of Italian NLP for organising a shared task that requires complete systems comprising several layers of linguistic analysis and having access to lexical resources?