Textual Entailment
Guidelines for Evalita 2009

Johan Bos, Marco Pennacchiotti, Fabio Massimo Zanzotto

February 27, 2009

Contents

1 Introduction

In this evaluation exercise systems have to detect inferential relationships between pairs of short texts. We will focus on the entailment relation. In terminology and evaluation method we follow, by and large, the well-known Pascal RTE exercises for English.

A pair of texts consists of T (for text) and H (hypothesis). Textual entailment is defined as a directional relationship between such pairs. Systems have to decide whether T entails H, which is the case when the meaning of H can be inferred from T within the context induced by T. Another way of thinking about this is whether H contains any new information with respect to T: if it doesn't, H is entailed by T. The hypothesis must be fully entailed by the text. When the inference is very probable (but not completely certain) the entailment relation still holds (i.e., in cases where T is true, H is 'most likely' true as well).

Note that similar referring expressions in text and hypothesis are assumed to have the same referent (including zero-pronouns, that occur frequently in Italian by subject-drop, as in Example 1 below). Different name expansions in text and hypothesis are taken to refer to the same individual or organisation. Furthermore, the use of presupposition of common linguistic and world knowledge is permitted for drawing inferences. For instance, in Example 3 below, the fact that Napolitano is the Italian President is a commonly known fact, and therefore the entailment relation holds for this pair.

> Example 1: entailed
>
> **T**: *Parla di attività nei panni di direttore commerciale e, dopo sei mesi, di direttore generale.*
> **H**: *Parla di attività di direttore commerciale e, dopo sei mesi, di direttore generale.*
>
> Example 2: not entailed
>
> **T**: *Il primo acquisto immobiliare fu un terreno in via Alciati a Milano, per 190 milioni di lire.*
> **H**: *Il primo acquisto è un terreno in via Alciati a Milano.*
>
> Example 3: entailed
>
> **T**: *Napolitano non ha apposto la firma sul decreto.*
> **H**: *Il Presidente non ha apposto la firma sul decreto.*

The development and test data will consist of several hundred examples of such pairs, equally divided into positive and negative examples. Performance is measured by accuracy (number of correctly recognised pairs divided by the total number of pairs), thereby setting the baseline score at 50% for a coin-flipping approach.

2 Dataset Description

The development and test data will consists of newly annotated data (because it is the first time that the RTE challenge is organised for Italian). Pairs of texts will be taken from Italian Wikipedia articles, and are constructed by manually annotating contrasting texts taken from the version history as provided by Wikipedia. Texts are expected to be relatively short, and cover a phrase, a sentence, or a couple of sentences.

3 Data Format

The development and test data will be provided to the participants in XML marked-up files, using UTF-8 encoding for Unicode. Here is an example of this format:

```
<?xml version='1.0' encoding='UTF-8'?>
<entailment-corpus>
<pair entailment="YES" id="0001" task="WIKI">
   <t>Sposato con due figli, diplomato all'Istituto per Geometri
      'Giacomo Quarenghi' di Bergamo, è funzionario al Genio
      civile di Bergamo.</t>
   <h>Sposato, due figli, è funzionario al Genio civile di
      Bergamo.</h>
</pair>
<pair entailment="NO" id="0002" task="WIKI">
   <t>Alle elezioni politiche 2006 è candidato al Senato, ma non
      risulta eletto.</t>
   <h>Dopo le elezioni politiche 2006 conferma il suo seggio a
      Montecitorio, risultando vincitore nella circoscrizione Calabria
      con la lista dell'L'Ulivo.</h>
</pair>
</entailment-corpus>
```

The test data will follow the format of the development data, but obviously without the `entailment` attribute. The development and test data will be made available via the EVALITA website (which URL is `http://aclweb.org/aclwiki/index.php?title=Resources_for_Italian`). Expected dates: 1st of April 2009 (development data) and 10th of September 2009 (test data).

4 Further Linguistic Annotation

In addition to the raw data format, the organisers are planning to make linguistic analyses of the development and test data available. As of now we don't know what is feasible, but we're thinking of tokenisation, part of speech tagging, and dependency parsing. In any case, this won't be gold-standard data but rather be produced by already available NLP tools.

5 Run Submission

The format of the run is a text file with on each line a pair id, and one of the following three predictions YES (entailment), NO (no entailment), or NIL (in case the system is unable to come up with an answer). A single white space should be used to separate the pair id from the prediction. Here is a snapshot example of this format:

```
0001 YES
0002 NO
0003 NIL
0004 NO
```

Up to four runs can be submitted by each participant. The name of each submission file should include an acronym of the participant plus an identifier for the run (each run should have a different filename, obviously). All runs should be submitted in a single tar file and be sent by email to `bos@di.uniroma1.it`, with "EVALITA 2009 submission" in the subject header, and brief information of the participant (organisation(s), names of participant(s), contact person) in the body of the mail. The deadline for run submissions is the last minute of 20th of September, 2009 (Italian time). Soon after submission a short notice is sent to the participant with information on the number of runs received. Results will be made available to the participants on or before the 30th of September, 2009.

6 Pointers to Resources

Here are some pointers to resources that could be helpful in building an RTE system. First, you might want to consider some of the links to NLP tools for Italian:

ACL Wiki: `http://aclweb.org/aclwiki/index.php?title=Resources_for_Italian`

EVALITA 2007 tools: `http://evalita.fbk.eu/2007/results.html`

And here are some links to Recognising Textual Entailment in general:

PASCAL RTE: `http://pascallin.ecs.soton.ac.uk/Challenges/RTE3/`

RTE at TAC: `http://www.nist.gov/tac/2009/RTE/index.html`

Textual Entailment Resource Pool: `http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool`