



„Al. I. Cuza“ University, Iasi, Romania



Faculty of Computer Science

UAIC Participation at RTE4

Adrian Iftene
adiftene@info.uaic.ro



Overview

- 3-way RTE4 System
 - New added components
 - Positive and Negative rules
 - Results
 - Limit of the system
- Applications
 - Ranking answers in a QA system
 - Building a system for AVE competition
- Conclusions

RTE Competition

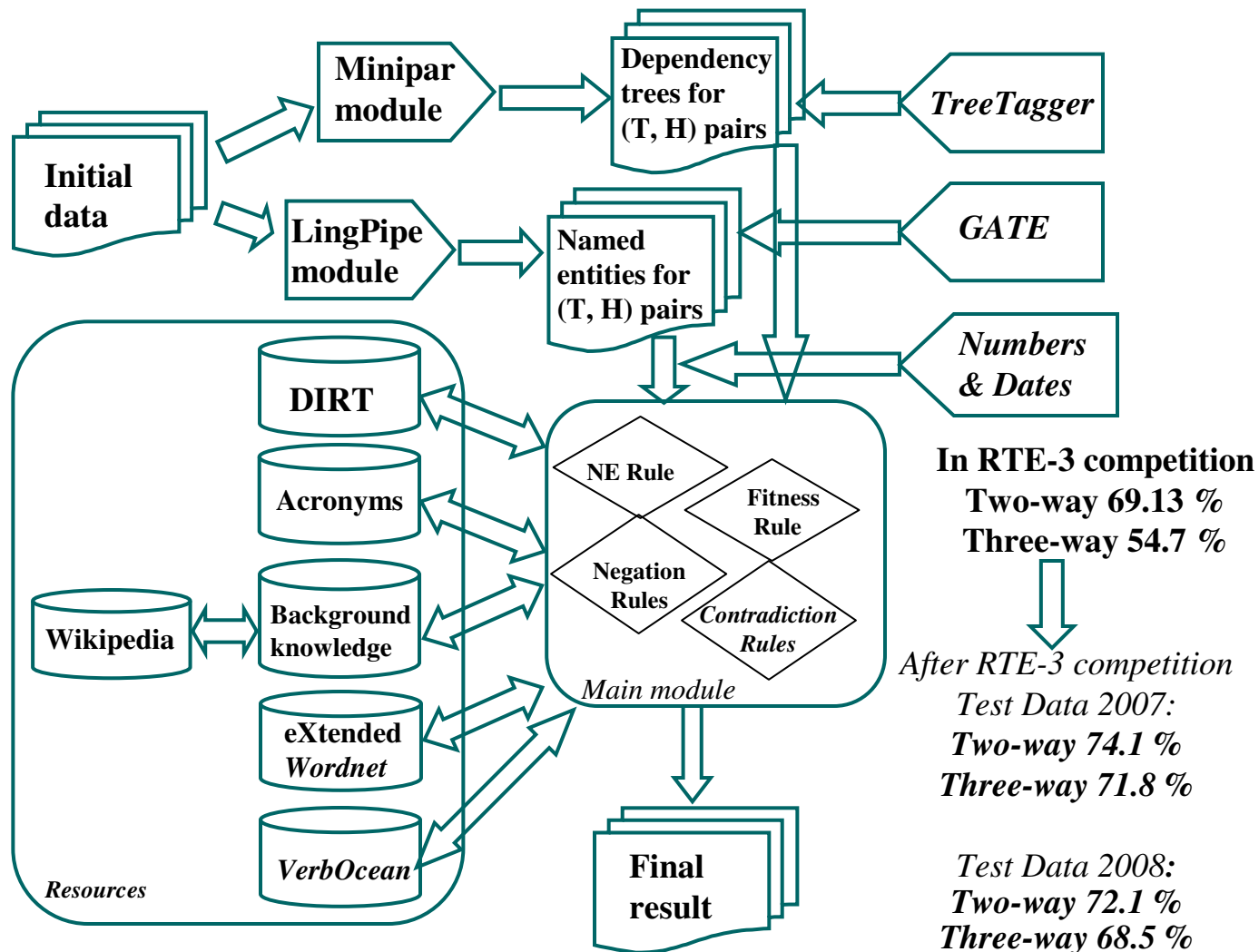
The **two-way RTE task** (2005-2008) is to decide whether:

- **T entails H - ENTAILMENT**
- **T does not entail H - NO ENTAILMENT**

○ The **three-way RTE task** (2007-2008) is to decide whether:

- **T entails H - ENTAILMENT**
- **T contradicts H - CONTRADICTION**
- The **truth of H cannot be determined on the basis of T - UNKNOWN**

System presentation

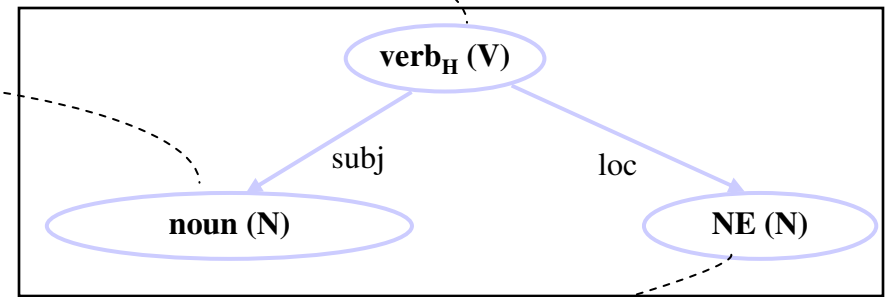
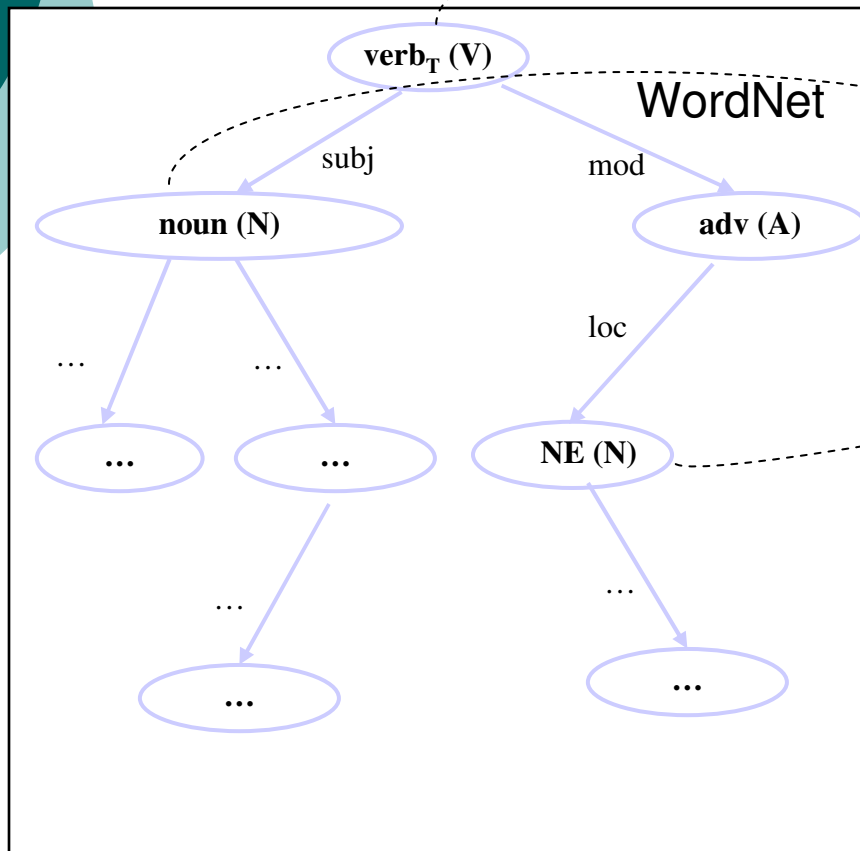


Example

Text

DIRT, VerbOcean

Hypothesis



Acronyms, BK

DIRT: solve=resolve 0.31453
 DIRT: convict=arrest 0.28895
 DIRT: **convict=acquit** 0.302455, OPPOSITE
 VerbOcean: **increase<>decrease**,
 VerbOcean: **leave<>stay**
 WordNet: trouble=problem
 WordNet: talk=discussion
 Acronym: EU=European Union
 BK: Buenos Aires [in] Argentine
 BK: 16 [is] sixteen



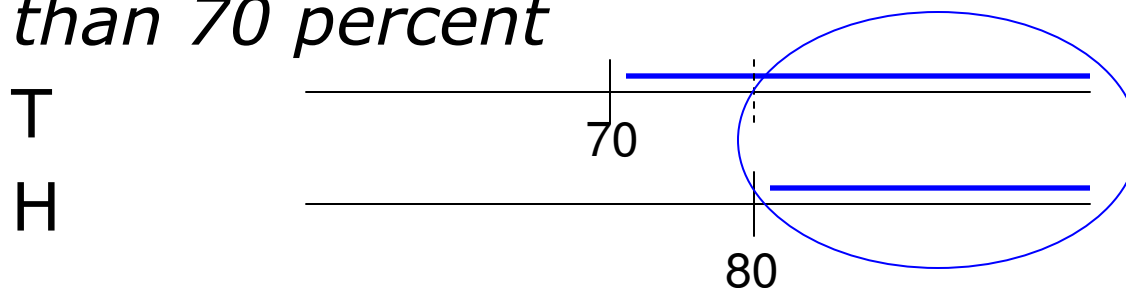
Rules

- For every type of possible answer we will present the rules that promote it
- Possible cases are:
 - Entailment
 - No entailment
 - Contradiction
 - Unknown

Entailment cases

- Every type of *mapping*: *direct* or *indirect* (using DIRT, WordNet, Acronym, Background Knowledge)
- *Positive rules for Numbers (context rules)*:
quantification words: at least, more than, less than, over, under, etc.

Pair 304: T: *at least 80 percent* \approx H: *more than 70 percent*





Entailment cases (cont...)

- *Considering of Additional numbers when numbers from T or H are separated by "and" or ","*

Pair 331:

T: *killing all **109** people on board **and four** workers on the ground*

H: *killed **113** people*

Because *109* and *four* are separated by "and" we consider their sum 113 belong to T



No entailment cases

- *Basic rule*: we don't consider in global fitness calculation the stop words ("the", "an", "a", "at", "to", "of", "in", "on", "by", etc.)
- *Negation rule*: for every verb we verify on tree branches to see if one or more of the following words are found: "not", "never", "may", "might", "cannot", etc.
- A special case is for particle "to" when it precedes a verb



Contradiction cases

- *Negations*: when verbs are negated with words like “never”, “not”, “no”, “cannot”, “unsuccessfully”, “false” etc.

Pair 660:

T: *Aquacell Water, Inc announced today that it has **not received** ...*

H: *Aquacell Water **receives** ...*

Contradiction cases (cont...)

- When before particle "to" are "refuse", "deny", "ignore", "plan", "intend", "proposal", "able", etc.

Pair 54:

T: **Plans to detain** terrorist suspects for up to 42 days without charge ...

H: Police **can detain** terror suspects for 42 days without charge.

Pair 354:

T: ... Shin was sacked from the school on June 20 after **refusing to resign** from his post as director of KBS.

H: Shin Tae-seop **resigned** from his post at Dong-eui University.

Contradiction cases (cont...)

- *Antonymy relation*: use [opposite-of] relation from VerbOcean resource (Chklovski and Pantel, 2004) and antonymy relation from WordNet

Pair 8:

T: *Europe, New Zealand and Australia were also beginning to report **decreases** in new HIV cases.*

H: *AIDS victims **increase** in Europe.*

- *Combination between synonyms from WordNet and antonymy relation from WordNet or opposite relation from VerbOcean*

verb1 [opposite-of] *verb2*

$\text{Synonyms}_{\text{verb1}} = \{vs1, vs2, \dots\}$

- Similar for $\text{Synonyms}_{\text{verb2}}$

=> *vs1* [opposite-of] *verb2*

vs2 [opposite-of] *verb2*

etc.



Contradiction cases (cont...)

- Extra verification for *similarity relation* from DIRT: in some cases it is *antonymy relation* (in WordNet or in VerbOcean)

Pair 167

T: R. Kelly **was acquitted** of child pornography after the star witness Van Allen was discredited after admitting she once stole Kelly's \$20,000 diamond-studded watch from a hotel.

H: R. Kelly **was convicted** for child pornography.

- Initial we use DIRT relation between *convict* and *acquit* with score 0.302455, but because we found in WordNet that *convict* and *acquit* are antonyms, we apply the *Contradiction rule*



Unknown cases

- *Negations*: when verbs are negated with words like “may”, “can”, “should”, “could”, “must”, “might”, “infrequent”, “rather”, “probably”, etc. At pair 198 T: “... **could** also **be linked** to ...” and H: “... **is linked** to ...”
- Related to the particle “to” we will consider the cases which are not included in contradiction cases. At pair 391 T: “***It is hard to like Will Carling ...***” H: “***Nobody likes Will Carling***” we insert a penalty
- In these cases, inserted penalties are not decisive in establishing the final answer, which is obtained only after the calculation of global fitness



Unknown cases (cont...)

- *Name entities rule*: if we cannot map a NE from H direct or using acronyms database and background knowledge the final result for the current pair is: “*Unknown*”

Pair 454:

T: In 1977 to wide media fanfare, Polanski was charged with a host of sexual crimes for his involvement with a 13-year-old girl. He was subsequently convicted of unlawful intercourse with a minor, but fled the country in 1978 before final sentencing.

*H: Polanski fled from the **U.S. to Russia** in 1978.*



Unknown cases (cont...)

- *Numbers context*: the rule presented above when the same numbers from T and H have different “*unit measures*”

Pair 441:

T: *Britain deployed troops to Afghanistan shortly after the attacks of **11 September**, 2001...*

H: *Britain has **11 troops** that take part in Nato's International Security and Assistance Force.*



Unknown cases (cont...)

- An exception from the named entity rule: when the type of name entity is *first name* (Tatu & Moldovan, 2007)

Pair 122:

T: **Mr Brown's** *courage and determination are not in doubt: he soaks up punishment as if he believes it is good for him. But week after week he gives no sign that he knows how to seize the initiative and dictate the course of the fight.*

H: **Gordon Brown** *is the UK Prime Minister.*

- In this case we only insert a penalty in the global fitness.

Results in RTE4

| Answer Type | In Gold | Correctly offered by our system | Total offered by our system | Precision | Recall | F-measure |
|---------------|-------------|---------------------------------|-----------------------------|----------------|---------------|---------------|
| Entailment | 500 | 466 | 712 | 65.45 % | 93.20% | 76.90% |
| Contradiction | 150 | 69 | 85 | 81.18 % | 46.00% | 58.72% |
| Unknown | 350 | 150 | 203 | 73.89 % | 42.86% | 54.25% |
| Total | 1000 | 685 | 1000 | 68.50 % | | |

| Answer Type | In Gold | Correctly offered by our system | Total offered by our system | Precision | Recall | F-measure |
|--------------|-------------|---------------------------------|-----------------------------|---------------|---------------|---------------|
| Yes | 500 | 466 | 712 | 65.45% | 93.20% | 76.90% |
| No | 500 | 255 | 288 | 88.54% | 51.00% | 64.72% |
| Total | 1000 | 721 | 1000 | 72.10% | | |

Comparison with RTE-3

To be able to see each component's relevance, the systems were run in turn with each component removed

| System Description | RTE-3 | | RTE-4 | |
|----------------------------|-----------|----------------|-----------|---------------|
| | Precision | Relevance | Precision | Relevance |
| Without DIRT | 68.76 % | 0.54 % | 71.40 % | 0.97 % |
| Without WordNet | 68.00 % | 1.63 % | 69.10 % | 4.16 % |
| Without Acronyms | 68.38 % | 1.08 % | 71.80 % | 0.42 % |
| Without BK | 67.75 % | 2.00 % | 70.40 % | 2.36 % |
| Without Negation rule | 67.63 % | 2.17 % | 68.70 % | 4.72 % |
| Without NE rule | 57.58 % | 16.71 % | 66.90 % | 7.21 % |
| Without Contradiction rule | - | - | 68.10 % | 5.55 % |



Limits of the system

- Resource limits:

- Paraphrase missing: pair 137 in T we have: “are raised in ways that are ethically and environmentally unsound” and in H we have “are reared cruelly”
- Rules for obtain extra information: at pair 13 the “Becker has never played tennis in his life” is in contradiction with hypothesis “Becker was a tennis champion” if we use the rule:
 - if (*X is SPORT champion*) then *X play SPORT*
- Resource missing: at pair 64 in T we have the sentence “All the victims are adults”. Because we don’t have a resource where to have that “adults” are different by “children” and to deduce that “Children were not killed”



Limits of the system (cont...)

- Inefficient exploit of Geographical Resources:

Pair 190:

T: A strong earthquake struck off the southern tip of Taiwan...

H: An earthquake strikes Japan

Possible rule:

if (the geographical regions are sufficiently closed) then “accordingly with earthquake power, it is possible like the earthquakes to be feel in both regions”

Variables are the *distance between regions* and the *power of the earthquake*



Limits of the system (cont...)

- Missing of Semantic Role Labeling:

Pair 551:

T: United Kingdom flag carrier British Airways (BA) has entered into merger talks with Spanish airline Iberia Lineas Aereas de Espana SA. BA is already Europe's third-largest airline.

H: The Spanish airline Iberia Lineas Aereas de Espana SA is Europe's third-largest airline.

- For verb “be” we have the same argument A2 “Europe's third-largest airline” but we have different A1 arguments: “BA” in text and “SA” in the hypothesis.

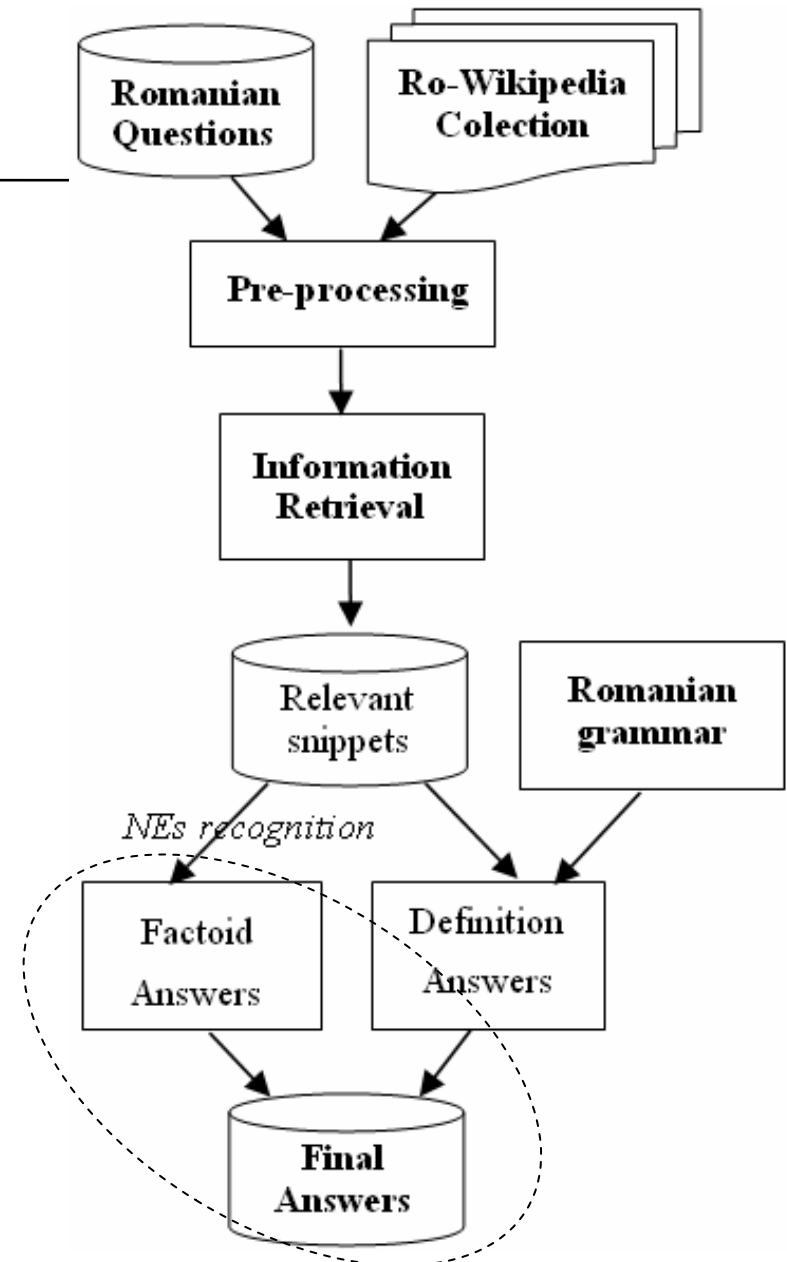


Applications

- At RTE-3 we presented the speed improvements using P2P networks
- This year we present how we use TE system in:
 - Question Answering (QA)
 - Answer Validation Exercise (AVE)

QA: System architecture

- Pre-processing
- Information retrieval (Lucene)
- Snippet extraction
- Answer extraction **and ranking**





Idea from (Bar-Haim et al., 2006)

- To associate a pattern to every question and to replace variable from it with all possible answer candidates extracted from relevant snippets => *Hypotheses*
- And to consider relevant snippets extracted by IR component => *Texts*



Example

- **Question:** How many passengers does the ship the "Canadian Empress" carry?
- **Pattern:** MEASURE passengers were carrying by the ship the "Canadian Empress".
- **Snippet:** The Empress Canadian (66 p) sails spring through fall on St. Lawrence and Ottawa River cruises to scenic and historic areas; shore excursions are included and most passengers are seniors. (800) 267-7868. Seabourn Cruise Line
- H1: **66** passengers were carrying by the ship the "Canadian Empress".
- Similar H2, H3, H4



Results in QA Competitions

- QA competition Ro-En 2006:
 - With QA ranking - **9.47** % on first answer and around **35 %** on all provided answers (9th place from 13 competitors)
 - With **English TE ranking** - **19** % on first answer (2nd place)
- QA competition Ro-Ro 2008:
 - With QA ranking - **28.5** % on first answer
 - With **Romanian TE ranking** - **31** % on first answer



AVE competition

- Participants receive triplets (Question, Answer, Snippet) and they must specify the answer correctness:
 - VALIDATED – the answer is correct and supported
 - SELECTED – the answer is VALIDATED and it is the most probable answer
 - REJECTED – the answer is incorrect



The AVE system

- The steps are very similar with answer ranking using TE system except part with candidates extraction
- 2007 on English

| Group | QA accuracy |
|------------------------------------------|--------------------|
| DFKI, Germany | 70 % |
| "Al. I. Cuza" University, Romania | 70 % |
| Alicante University, Spain | 65 % |

- 2008 on English

| Group | QA accuracy |
|------------------------------------------|--------------------|
| DFKI, Germany | 70.37 % |
| "Al. I. Cuza" University, Romania | 70.37 % |
| Alicante University, Spain | 62.96 % |



Conclusions

- Main idea of our TE system is to map every node from hypothesis to a node from text *direct or indirect*
- In RTE-4 the new rules for *contradiction* and *unknown* cases were more clearly specified
- The TE system can be used with success in answer ranking for a QA system
- Based on TE system we got good results in AVE competition in 2007 and in 2008



Acknowledgments

- Students:
 - Alexandra Balahur-Dobrescu, Daniel Matei
- NLP group of Iasi:
 - Supervisor: Prof. Dan Cristea
 - Maria Husarciuc, Alex Moruz, Ionut Pistol, Marius Raschip, Diana Trandabat
- Support from SIR-RESDEC project



THANK YOU!



References

- R. Bar-Haim, I. Dagan, B. Dolan, L. Ferro, D. Giampiccolo, B. Magnini, and I. Szpektor. The second pascal recognising textual entailment challenge. *In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, Venice, Italy, 2006.*
- T. Chklovski and P. Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. *Proceedings of EMNLP 2004*, pages 33–40, 2004.
- A. Iftene and A. Balahur-Dobrescu. Hypothesis transformation and semantic variability rules used in recognizing textual entailment. *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 125–130, June 2007.
- M. Tatu and D. Moldovan. Cogex at RTE 3. *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 22–27, June 2007.