Semantic Annotation for Textual Entailment Recognition

Assaf Toledo[1], Sophia Katrenko[1], Stavroula Alexandropoulou[1], Heidi Klockmann[1], Asher Stern[2], Ido Dagan[2], and Yoad Winter[1]

[1] Utrecht University, Utrecht, the Netherlands
[2] Bar Ilan University, Ramat Gan, Israel

Abstract. We introduce a new semantic annotation scheme for the Recognizing Textual Entailment (RTE) dataset as well as a manually annotated dataset that uses this scheme. The scheme addresses three types of modification that license entailment patterns: *restrictive*, *appositive* and *conjunctive*, with a formal semantic specification of these patterns' contribution for establishing entailment. These inferential constructions were found to occur in 77.68% of the entailments in the RTE 1-3 corpora. They were annotated with cross-annotator agreement of 70.73% on average. A central aim of our annotations is to examine components that address these phenomena in RTE systems. Specifically, the new annotated dataset is used for examining a syntactic rule base within the BIUTEE recognizer, a publicly available entailment system. According to our tests, the rule base is rarely used to process the phenomena annotated in our corpus and most of the recognition work is done by other components in the system.[3]

Keywords: RTE Textual Entailment Semantic Annotation

1 Introduction

The *Recognizing Textual Entailment* (RTE) [9] task aims to automatically determine whether an entailment relation obtains between a naturally occurring **text** (T) and a **hypothesis** sentence (H). In the RTE corpus, which is currently the only available textual entailments resource, the entailment candidates are marked manually as valid/invalid.[4] This categorization contains no indication of the linguistic and informational processes that underlie entailment. Consider the following example from RTE 2:

- T: The widow of John Lennon, Yoko Ono, may take legal action against a new breakfast cereal called "Strawberry Fields" which she believes is too close in name to Lennon's song.

---

[3] The annotated corpus is freely downloadable.

[4] Pairs of sentences in RTE 1-3 are categorized in two classes: *yes-* or *no-entailment*; the data in RTE 4-5 are categorized in three classes: *entailment*, *contradiction* and *unknown*. We label the judgments *yes-entailment* from RTE 1-3 and *entailment* from RTE 4-5 as *valid*, and the other judgments as *invalid*.

– H: Yoko Ono is John Lennon's widow.

Here, the judgment on the validity of this entailment is based on the semantic properties of the appositive construction - *The widow of John Lennon, Yoko Ono* - whereby two noun phrases refer to the same entity.

Because linguistic information of this kind is not analyzed or even stated in the RTE corpus, it is difficult to develop new entailment recognizers or evaluate existing ones based on theoretical research that has been carried out in natural-language semantics or on supervised learning techniques. For previous work that yielded similar conclusions, see [28].

In this paper we describe a new scheme for annotating valid entailments in the RTE corpus and provide an RTE dataset annotated manually using this scheme.[5] The scheme addresses three types of modification that license entailment: *restrictive*, *appositive* and *conjunctive*; a formal semantic specification is provided for each of these, as well as for its role in establishing the entailment [24, 27, 29, 10]. The annotation proposed in this paper is *semantic* in the sense that it targets semantic relations that govern the inferences licensed in various expressions, rather than merely annotating the syntax of these expressions.

The rationale behind our decision to focus on the three semantic phenomena listed above is that they are common in the RTE dataset and linguistically intuitive enough to yield high inter-annotator agreement. The annotated corpus is designated to serve as a benchmark for a theoretically informed evaluation of entailment modules in RTE systems. In addition, it can be used in the development of entailment recognizers by facilitating automatic learning of the above constructions.

The results of this work show that, in the corpora of RTE 1-3, the inferential constructions marked according to the proposed scheme occur in 77.68% of the valid entailments, with an average inter-annotator agreement of 70.73% (see Section 3.3). As a use case for our corpus, we examined the processing of the annotated constructions by BIUTEE [30], a state of the art entailment recognizer which employs a rule base targeting a broad range of inferential patterns.[6] We found that, in its treatment of RTE 1-3, BIUTEE processes only 4.52% of the entailment cases based on rules that correspond to the annotated constructions. It seems, therefore, that the rule base examined rarely participates in the processing of the phenomena investigated.

The structure of this paper is as follows. Section 2 elucidates the connection between this and previous annotation work on the RTE data. In Section 3 we introduce the annotation scheme, elaborate on the methods employed, and present some quantitative data on the targeted constructions and inter-annotator agreement. Section 4 describes a use case of the annotated corpus with reference to the rule base employed by the BIUTEE recognizer. Section 5 is a conclusion.

---

[5] The annotated corpus is available at: `http://logiccommonsense.wp.hum.uu.nl/papers/annotatedrte`

[6] `http://cs.biu.ac.il/~nlp/downloads/biutee`; version 2.4.1

2 Related Work

During the past few years, several methodological proposals and annotation projects have been attempted with the purpose of uncovering and documenting entailment processes in the RTE. The main assumption underlying most of the work in this direction is that decomposing the complex entailment problem would improve the performance of RTE systems.

In [5], a methodology is described for creating specialized entailment data sets by isolating linguistic phenomena relevant for entailment. Pairs of text and hypothesis from the existing corpus were used to generate a set of mono-thematic pairs, each containing one specific phenomenon that takes part in the original entailment. As part of a feasibility study, this methodology was applied to a sample of 90 pairs randomly extracted from RTE 5. The phenomena were divided into broad categories (e.g. lexical, syntactic, etc.) and then into more fine-grained categories (e.g. synonymy, active-passive, etc.), and a general rule of inference was defined for each of these categories (e.g. *argument realization*: "x's y" → "y of x").

This methodology allows a detailed analysis of the entailments in the corpus, but a full analysis of all entailment patterns in the corpus would necessarily involve complex judgments, and this, in turn, would make high cross-annotator consistency very hard to achieve. Moreover, as discussed in Section 3.3, our experience shows that efficient annotation with high cross-annotator consistency is hard to obtain even in more restricted cases which involve less complex judgments.

Our annotation work is to a large extent in line with the proposal described in [28], whose authors appeal to the NLP community to contribute to entailment recognition work by incrementally annotating the phenomena that underlie the inferences in the RTE corpus. However our annotation scheme does not require a full understanding of the phenomena involved in the entailment in a given T-H pair. Nor does it aim to uncover all the connections between these phenomena, and accordingly, it does not specify the order by which the inferential steps are carried out in each particular case.

3 Corpus Description

This section details our annotation strategy, the inference principles it is based on, the annotation work-flow, and some quantitative data on the targeted constructions and inter-annotator agreement.

3.1 Phenomena Annotated

The choice of phenomena for annotation is based on the following criteria:

1. Phenomena that are commonly involved in entailments.

2. Phenomena that are well understood in the semantic literature and that lend themselves readily to linguistic intuitions as well as to an analysis that is likely to yield high annotation consistency.
3. Phenomena that do not require sophisticated abstract representations and which therefore are easy to classify.

Based on these considerations, three types of modifications were selected for annotation: restrictive modification, appositive modification and intersective modification (conjunction).

An important feature of our annotation is that it marks inferences by aligning strings in the text and the hypothesis. This is done by pairing each annotation in the text with a corresponding annotation in the hypothesis that marks the output of the inferential process of the phenomenon in question. In the next subsections we describe the annotated phenomena in detail and in each example we underline the annotated part in the text with its correspondence in the hypothesis.

Restrictive modification (RMOD) RMOD is an instance of two adjacent expressions in which one expression (the *modifier*) restricts the semantic class of objects denoted by the other (the *modifiee*). RMOD licenses an entailment pattern of modifier subsumption:

- T: *A* $\underline{Cuban_{Modifier}}$ $\underline{American_{Modifiee}}$ *who is accused of espionage pleads innocent.*
- H: $\underline{American}$ *accused of espionage.*

In this case, *Cuban* modifies *American* and restricts the set of Americans to Cuban Americans. This instance of RMOD validates the inference from *Cuban Americans* to *Americans* which is required for establishing the entailment.

The following examples contain additional syntactic configurations in which RMOD licenses inferences:

- A verb phrase restricted by a prepositional phrase:
  - T: *The watchdog International Atomic Energy Agency* $\underline{meets\ in}$ $\underline{Vienna_{Modifiee}}$ *on September* $\underline{19_{Modifier}}$.
  - H: *The International Atomic Energy Agency* $\underline{holds\ a\ meeting\ in\ Vienna}$.
- A noun phrase restricted by a prepositional phrase:
  - T: *U.S. officials have been warning for weeks of* $\underline{possible\ terror}$ $\underline{attacks_{Modifiee}}$ *against* $\underline{U.S.\ interests_{Modifier}}$.
  - H: *The United States has warned a number of times of* $\underline{possible\ terrorist}$ $\underline{attacks}$.

Appositive modification (APP) Apposition involves two adjacent expressions whereby one part designates an entity and the other supplements its description by additional information. APP licenses three main entailment patterns:

- Appositive subsumption (left part):

- T: *Mr. Conway, Iamgold's chief executive officer, said the vote would be close.*
- H: *Mr. Conway said the vote would be close.*
  - Appositive subsumption (right part):
    - T: *The country's largest private employer, Wal-Mart Stores Inc., is being sued by a number of its female employees who claim they were kept out of jobs in management because they are women.*
    - H: *Wal-Mart sued for sexual discrimination.*
  - Identification of the two parts of the apposition as referring to one another:
    - T: *The incident in Mogadishu, the Somali capital, came as U.S. forces began the final phase of their promised March 31 pullout.*
    - H: *The capital of Somalia is Mogadishu.*

In addition to appositive constructions as illustrated above, APP appears in several more syntactic constructions:

- Non-Restrictive Relative Clauses:
  - T: *A senior coalition official in Iraq said the body, which was found by U.S. military police west of Baghdad, appeared to have been thrown from a vehicle.*
  - H: *A body has been found by U. S. military police.*
- Title Constructions:
  - T: *Prime Minister Silvio Berlusconi was elected March 28 with a mandate to reform Italy's business regulations and pull the economy out of recession.*
  - H: *The Prime Minister is Silvio Berlusconi.*

Conjunction (CONJ) CONJ is an instance of two or more adjacent expressions that are interpreted intersectively. Typically CONJ licenses an entailment pattern of conjunct subsumption:

- T: *Nixon was impeached and became the first president ever to resign on August 9th 1974.*
- H: *Nixon was the first president ever to resign.*

In this example the conjunction intersects the two verb phrases *was impeached* and *became the first president ever to resign*. The entailment relies on subsumption of the full construction to the second conjunct.

In addition to canonical conjunctive constructions, CONJ appears also in Restrictive Relative Clauses whereby the relative clause is interpreted intersectively with the noun being modified:

- T: *Iran will soon release eight British servicemen detained along with three vessels.*
- H: *British servicemen detained.*

3.2 Marking Annotations

Given a pair from the RTE in which the entailment relation obtains between the text and hypothesis, the task for the annotators is defined as follows:

1. Read the data and verify the entailment.
2. Describe informally why the entailment holds.
3. Annotate all the instances of the phenomena described in Section 3.1 that play a role in the inferential process.

The annotations were performed using GATE Developer [8] and recorded above the original RTE XML files. The annotators use GATE annotation schemes that were defined to correspond to RMOD, APP and CONJ as shown in Table 1. [7]

Table 1. GATE Annotation Schemes

Phenomenon	Annotation Schemes
RMOD	r_modification
APP	apposition, title, rel_clause
CONJ	conjunction, rel_clause

The work was performed in two steps: (1) marking the relevant string in the text using one of GATE annotation schemes that had been defined for the purpose (e.g. *apposition*), and (2) - marking a string in the hypothesis that corresponds to the output of the inferential process. The annotation in the hypothesis is done using a dedicated *reference_to* scheme.

Example

Consider the following pair from RTE 2:

- T: *The anti-terrorist court found two men guilty of murdering Shapour Bakhtiar and his secretary Sorush Katibeh, who were found with their throats cut in August 1991.*
- H: *Shapour Bakhtiar died in 1991.*

The entailment patterns in this example can be explained by appealing to the semantics of APP, CONJ and RMOD, as follows:

---

[7] The scheme *rel_clause* appears twice in this table because it is used for annotating non-restrictive relative clauses, expressing appositive modification (APP) and also restrictive relative clauses, expressing intersective modification (CONJ). The phenomena APP and CONJ are annotated using several annotation schemes in order to capture the different syntactic expressions that they allow.

- APP: The appositive modification in *Shapour Bakhtiar and his secretary Sorush Katibeh, who were found with their throats cut in August 1991* licenses the inference that *Shapour Bakhtiar and his secretary Sorush Katibeh were found with their throats cut in August 1991*.
- RMOD: The restrictive modification in *August 1991* licenses a subsumption of this expression to *1991*.
- CONJ: The conjunction in *Shapour Bakhtiar and his secretary Sorush Katibeh* licenses a subsumption of this expression to *Shapour Bakhtiar*.

By combining these three patterns, we can infer that *Shapour Bakhtiar was found with his throat cut in 1991*.[8] In figure 1 we show the annotation tags of CONJ which were added in the text body and serve as boundary markers for the phenomenon and as pointers to the annotation content tags. Figure 2 presents the annotation content tags that detail the conjunction.

- Text:
  The anti-terrorist court found two men guilty of murdering **<Node id= "11404"/>Shapour Bakhtiar and his secretary Sorush Katibeh <Node id="11453"/>**, *who were found with their throats cut in August 1991.*
- Hypothesis:
  <Node id="11511"/>Shapour Bakhtiar <Node id="11527"/> *died in 1991.*

Fig. 1. Annotation tags in the text body

3.3 Annotation Statistics

The annotated corpus is based on the scheme described above, applied to the datasets of RTE 1-4 [9, 2, 14, 13]. The statistics in Table 2 are based on analyzing the annotation work of RTE 1-3 (development and test sets).

We performed two cross-annotator consistency checks. In each check we picked a number of entailment pairs that both annotators worked on independently and compared the phenomena that they annotated. We reached cross-annotator consistency on 70.73% of the annotations on average, as reported in Table 3.

4 Corpus Use Case

Our annotations reflect human judgments on phenomena that take part in inferential processes in the RTE corpus. A straightforward use case of these annotations is to examine the coverage of the annotated phenomena by specific modules

---

[8] Additional world knowledge is required to infer that *found with his throat cut* entails *died*; given that, the entailment can be validated.

Table 2. Counters of annotations in RTE 1-3 separated into development and test sets. $A_\#$ indicates the number of annotations, $P_\#$ indicates the number of entailment pairs containing an annotation and $P_\%$ indicates the portion of annotated pairs relative to the total amount of entailment pairs.

Ann.	RTE 1						RTE 2						RTE 3					
	Dev set			Test set			Dev set			Test set			Dev set			Test set		
	$A_\#$	$P_\#$	$P_\%$	$A_\#$	$P_\#$	$P_\%$	$A_\#$	$P_\#$	$P_\%$	$A_\#$	$P_\#$	$P_\%$	$A_\#$	$P_\#$	$P_\%$	$A_\#$	$P_\#$	$P_\%$
APP	97	87	31	161	134	34	178	149	37	155	135	34	162	128	31	166	136	33
CONJ	90	79	28	126	112	28	140	118	30	161	144	36	157	121	29	162	134	33
RMOD	180	124	44	243	167	42	311	204	51	394	236	59	262	176	43	306	193	47
Any	367	210	74	530	297	74	629	316	79	710	350	88	581	293	71	635	328	80

Table 3. Results of Consistency Checks - In Check 1, 50 entailment pairs were analyzed and 66% of the annotations that were marked were identical; in Check 2, 70 pairs were analyzed and 74.11% of the annotations that were marked were identical. On average, 70.73% of the annotations we checked were identical. The rubric *Ambiguities* presents cases of structural or modifier-attachment ambiguity in the text that led to divergent annotations. *Major mistakes* are cases of missing annotations, and *Minor mistakes* are cases of divergent annotations (not stemming from ambiguity) or incorrect scope.

Measure	Check 1	Check 2
Entailment Pairs	50	70
Sources	RTE 1	RTE 1 + 2
Total Annotations	93	112
Identical Annotations	62	83
Ambiguities	9	18
Major mistakes	2	7
Minor mistakes	10	4
Consistency (%)	66.67	74.11

Table 4. Sample of Syntactic Rules in BIUTEE

Type	Description
Passive-Active	Transform "X $Verb_{active}$ Y" to "Y is $Verb_{passive}$ by X"
Apposition	Extract an NP and its apposition to an independent IS-A construction
Genitive	Substitute an "X's Y" construction with a "the Y of X" construction

Table 5. Accuracy of BIUTEE in two configurations tested on RTE 1-3

Config	RTE 1	RTE 2	RTE 3
C1	55.38%	61.38%	66.13%
C2	55.88%	61.75%	65.75%

– Text:
<Annotation Id="1833" Type="conjunction" StartNode="11404"
EndNode="11453">
<Feature><Name>E1</Name><Value>Shapour Bakhtiar</Value>
</Feature>
<Feature><Name>E2</Name><Value>his secretary Sorush Katibeh</Value>
</Feature>
<Feature><Name>construction_id</Name><Value>2</Value></Feature>
</Annotation>
– Hypothesis:
<Annotation Id="1834" Type="reference_to" StartNode="11511"
EndNode="11527">
<Feature><Name >construction_id</Name><Value >2</Value></Feature>
</Annotation>

Fig. 2. Annotations tags of *conjunction* in the Text and Hypothesis

Table 6. Accuracy of top three recognizers that participated in the RTE 1-3 challenges. The data presented in this table is based on the results reported in [19].

RTE 1		RTE 2		RTE 3	
System	Accuracy	System	Accuracy	System	Accuracy
Autonma de Madrid [26]	70%	LCC [16]	75.38%	LCC [17]	80%
Ca Foscari Venice [11]	60.6%	LCC [31]	73.75%	LCC [32]	72.25%
MITRE [4]	58.6 %	DISCo [33]	63.88%	"Al. I. Cuza" [18]	69.13%

that are designed to address them in entailment recognizers. For clarity, we treat these modules of recognizers as rule bases that assign a rule - an inferential operation - for every phenomenon that the system handles. Thus, measuring the coverage of annotated phenomena by rule applications is useful for discovering to what extent rule bases participate in the recognition of entailment at places where the phenomena exist according to human judgments.

4.1 Using the Annotations

Measuring the coverage of annotated phenomena by the rule base of an entailment recognizer is done in five steps:

1. Mapping between the rule base and the annotation scheme. This is done manually by pairing each rule with an annotation that corresponds to the same inferential phenomenon.[9]
2. Extracting data from the annotations by parsing the annotated RTE XML file. This includes information on the span of text marked in each annotation and the type of annotation (e.g. *conjunction*).
3. Extracting data from the log file of the recognizer which indicates the rules that the recognizer applied in processing the RTE corpus.

---

[9] Rules that correspond to unannotated phenomena are ignored.

4. Counting the entailment pairs that include an annotation whose span of text was processed by the recognizer using a rule corresponding to the same phenomenon (according to the mapping done in step 1).

5. Dividing the number obtained in step 4 by the total number of entailment pairs in the corpus.

Section 4.3 describes an execution of this procedure on the Bar Ilan University Textual Entailment Engine (BIUTEE) [30].

4.2 The BIUTEE System

Given a text T and a hypothesis H, BIUTEE is designed to seek for a proof - a sequence of textual inferential steps applied on T that turns T into a text that equals or contains H. Following the design first introduced in [3], BIUTEE employs a syntactic rule base comprised of hand-crafted rules [22] that represent syntactic transformations. The rules in the system are defined based on dependency structures [23]; in Table 4 several of these rules are illustrated in text.

Ideally, a sequence of such transformations should be able to turn T into H. In practice, however, the rule coverage is limited, in that many inferential paradigms that appear in the RTE dataset are not covered by the rule base. Consequently, in addition to the rule base, BIUTEE uses a set of on-the-fly operations aimed at maximizing the similarity between T and H by manipulating T through adding / moving / replacing words.[10]. Furthermore, BIUTEE utilizes lexical and lexical-syntactic rules learned from knowledge resources such as, inter alia, WordNet [12, 25], FrameNet [1], VerbOcean [7], Catvar [15], Lin Similarity [20], and DIRT [21].

In this architecture, the system always reaches a proof from T to H. In order to identify proofs that capture valid entailments, the system uses a confidence model that assigns a confidence value to each rule and on-the-fly operations that are performed as part of a proof. If the computed total confidence value of a proof is high enough, it is assumed to represent a valid entailment, and vice versa. The function that assigns confidence values to rules and on-the-fly operations, and the threshold for predicting a valid/invalid entailment are learned automatically based on the training set (see [30] for full details).

4.3 Examination of BIUTEE's rule base

We executed the procedure described in 4.1 on BIUTEE. Information on all instances of rule application was extracted from the log file of the system, and all the rules that had been applied were either mapped to annotation labels or classified as relevant to unannotated (ignored) phenomena. We ran the system in two comparable configurations:

---

[10] The goal of the transformations BIUTEE applies to the graph that represents T's dependency structure is to turn it into a graph that contains H as a subgraph.

- C1 - Basic configuration with only syntactic rules included.
- C2 - Resource-based configuration with lexical and lexical/syntactic rules based on WordNet, FrameNet, VerbOcean, Catvar and DIRT, as well as the syntactic rules.

Table 5 displays the performance of the system in configurations C1 and C2. Table 6 reports the performance of the top three recognizers that participated in the RTE 1-3 challenges. In Table 7 we present our analysis of the coverage of annotated phenomena by the rule base.[11]

Table 7. Coverage of annotated phenomena by BIUTEE's rule base in two configurations tested on RTE 1-3. P_{Ann} stands for pairs containing an annotation of APP or CONJ, and $P_{Rule-Cx}$ stands for pairs that BIUTEE, running in configuration Cx, processed by applying a rule of APP or CONJ on an annotated phenomenon. $P_{\#}$ indicates the number of entailment pairs in each set (P_{Ann} or $P_{Rule-Cx}$ respectively) and $P_{\%}$ indicates the portion of a set in the total amount of entailment pairs (in percents). On average, BIUTEE processes only 4.52% of the entailment cases based on rules that correspond to the phenomena annotated in the corpus.

Pairs set	RTE 1		RTE 2		RTE 3	
	$P_{\#}$	$P_{\%}$	$P_{\#}$	$P_{\%}$	$P_{\#}$	$P_{\%}$
P_{Ann}	210	53	241	60	239	58
$P_{Rule-C1}$	2	1	37	9	3	1
$P_{Rule-C2}$	4	1	47	12	15	4

4.4 Summary of Results

In both configurations, our results indicate that the coverage of annotated phenomena by BIUTEE's rule base is rather low: 1% on RTE 1, 9-12% on RTE 2 and 1-4% on RTE 3. This means that the rule base is only rarely used for processing the entailment patterns annotated in our corpus. In practice, in order to process these linguistic phenomena and to predict the entailment, the system uses mostly on-the-fly operations.[12]

5 Conclusions

We have presented a semantic annotation of inferential phenomena in the RTE corpus for the purpose of evaluating and boosting RTE systems. We have chosen to focus on the annotation of semantic phenomena which are predominant in

---

[11] BIUTEE handles constructions that manifest RMOD internally, without rule application. The system does not report on instances of this phenomenon, and due to that we do not include RMOD in this examination.

[12] The marginal effect of the knowledge resources used in C2 on the overall accuracy of BIUTEE is in line with the results of ablation studies that were performed on various RTE systems (see [6] for more information).

the RTE and can be annotated with high consistency, but which may have several syntactic expressions and therefore allow us to generalize regarding abstract entailment patterns. A use case of the corpus is presented, based on an examination of the rule base employed by the BIUTEE recognizer. This study found that the rule base was rarely applied to the targeted phenomena. While we did not analyze the implications of these results for the architecture of BIUTEE, we believe that they show that the annotated corpus is useful for further research on entailment phenomena that are treated by rule bases, as in BIUTEE and other RTE systems (e.g. [17] and [18]). In addition, we think that the annotated data is also a useful benchmark for automatic learning of inferential phenomena recognition and processing.

References

1. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The berkeley FrameNet project. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1. pp. 86–90. ACL '98, Association for Computational Linguistics, Stroudsburg, PA, USA (1998)
2. Bar Haim, R., Dagan, I., Dolan, B., Ferro, L., Giampiccolo, D., Magnini, B., Szpektor, I.: The second pascal recognising textual entailment challenge. In: In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment (2006)
3. Bar-Haim, R., Dagan, I., Greental, I., Szpektor, I., Friedman, M.: Semantic inference at the lexical-syntactic level for textual entailment recognition. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 131–136 (2007)
4. Bayer, S., Burger, J., Ferro, L., Henderson, J., Yeh, E.: Mitres submission to the eu pascal rte challenge. In: In PASCAL. Proc. of the First Challenge Workshop. Recognizing Textual Entailment. pp. 41–44 (2005)
5. Bentivogli, L., Cabrio, E., Dagan, I., Giampiccolo, D., Leggio, M.L., Magnini, B.: Building textual entailment specialized data sets: a methodology for isolating linguistic phenomena relevant to inference. Proceedings of LREC 2010 (2010)
6. Bentivogli, L., Dagan, I., Dang, H.T., Giampiccolo, D., Magnini, B.: The fifth pascal recognizing textual entailment challenge. Proceedings of TAC 9, 14–24 (2009)
7. Chklovski, T., Pantel, P.: Verbocean: Mining the web for fine-grained semantic verb relations. In: Proceedings of EMNLP. vol. 4, pp. 33–40 (2004)
8. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., Heitz, T., Greenwood, M.A.,

Saggion, H., Petrak, J., Li, Y., Peters, W.: Text Processing with GATE (Version 6) (2011)

9. Dagan, I., Glickman, O., Magnini, B.: The pascal recognising textual entailment challenge. Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment pp. 177–190 (2006)

10. Del Gobbo, F.: On the syntax and semantics of appositive relative clauses. In: Dehe, N., Kavalova, Y. (eds.) Parentheticals, pp. 173–201 (2007)

11. Delmonte, R., Tonelli, S., Boniforti, M.A.P., Bristot, A., Pianta, E.: Vensesa linguistically-based system for semantic evaluation. In: Proceedings of the Pattern Analysis, Statistical Modelling, and Computational Learning (PASCAL) Challenges Workshop on Recognising Textual Entailment (2005)

12. Fellbaum, C.: WordNet: An Electronic Lexical Database. Language, Speech, and Communication, Mit Press (1998)

13. Giampiccolo, D., Dang, H.T., Magnini, B., Dagan, I., Cabrio, E.: The fourth pascal recognising textual entailment challenge. In: In TAC 2008 Proceedings (2008)

14. Giampiccolo, D., Magnini, B., Dagan, I., Dolan, B.: The third pascal recognizing textual entailment challenge. In: In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 1–9. RTE '07, Association for Computational Linguistics, Stroudsburg, PA, USA (2007)

15. Habash, N., Dorr, B.: A categorial variation database for english. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. pp. 17–23 (2003)

16. Hickl, A., Williams, J., Bensley, J., Roberts, K., Rink, B., Shi, Y.: Recognizing textual entailment with LCCs GROUNDHOG system. In: In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment (2006)

17. Hickl, A., Bensley, J.: A discourse commitment-based framework for recognizing textual entailment. In: In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 171–176 (2007)

18. Iftene, A., Balahur-Dobrescu, A.: Hypothesis transformation and semantic variability rules used in recognizing textual entailment. In: In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 125–130 (2007)

19. Iftene, A.: Textual Entailment. Ph.D. thesis, "Al. I. Cuza" University, Iasi, Romania (Mar 2009)

20. Lin, D.: Automatic retrieval and clustering of similar words. In: Proceedings of the 17th international conference on Computational linguistics-Volume 2. pp. 768–774 (1998)

21. Lin, D., Pantel, P.: DIRT@ SBT@ discovery of inference rules from text. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 323–328 (2001)

22. Lotan, A.: A Syntax-based Rule-base for Textual Entailment and a Semantic Truth Value Annotator. MA thesis, Department of Linguistics, Tel Aviv University (2012)

23. de Marneffe, M.C., Manning, C.D.: The stanford typed dependencies representation. In: Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation. pp. 1–8 (2008)

24. McCawley, J.D.: Parentheticals and discontinuous constituent structure. Linguistic Inquiry 13(1), 91–106 (1982)

25. Miller, G.A.: WordNet: a lexical database for english. Communications of the ACM 38(11), 39–41 (1995)

26. Perez, D., Alfonseca, E.: Application of the bleu algorithm for recognising textual entailments. In: Proceedings of the First Challenge Workshop Recognising Textual Entailment. pp. 9–12 (2005)
27. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A Comprehensive grammar of the English language. Longman, London and New York (1985)
28. Sammons, M., Vydiswaran, V.G., Roth, D.: Ask not what textual entailment can do for you... In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. pp. 1199–1208 (2010)
29. Sells, P.: Restrictive and non-restrictive modification. No. 28 in CLSI Report, Center for the Study of Language and Information, Stanford University (1985)
30. Stern, A., Dagan, I.: A confidence model for syntactically-motivated entailment proofs. In: Proceedings of RANLP 2011 (2011)
31. Tatu, M., Iles, B., Slavick, J., Novischi, A., Moldovan, D.: Cogex at the second recognizing textual entailment challenge. In: In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment. pp. 104–109 (2006)
32. Tatu, M., Moldovan, D.: COGEX at RTE3. In: In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 22–27 (2007)
33. Zanzotto, F.M., Moschitti, A., Pennacchiotti, M., Pazienza, M.T.: Learning textual entailment from examples. In: In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment (2006)