A Lexical Alignment Model for Probabilistic Textual Entailment

Oren Glickman, Ido Dagan and Moshe Koppel

Bar Ilan University, Ramat Gan 52900, ISRAEL,
{glikmao, dagan, koppel}@cs.biu.ac.il,
WWW home page: http://cs.biu.ac.il/~{glikmao, dagan, koppel}/

Abstract. This paper describes the Bar-Ilan system participating in the Recognising Textual Entailment Challenge. The paper proposes first a general probabilistic setting that formalizes the notion of textual entailment. We then describe a concrete alignment-based model for lexical entailment, which utilizes web co-occurrence statistics in a bag of words representation. Finally, we report the results of the model on the *Recognising Textual Entailment* challenge dataset along with some analysis.

1 Introduction

Many Natural Language Processing (NLP) applications need to recognize when the meaning of one text can be expressed by, or inferred from, another text. Information Retrieval (IR), Question Answering (QA), Information Extraction (IE), text summarization and Machine Translation (MT) evaluation are examples of applications that need to assess this semantic relationship between text segments. The Recognising Textual Entailment (RTE) task ([8]) has recently been proposed as an application independent framework for modeling such inferences. Within the applied textual entailment framework, a text t is said to entail a textual hypothesis h if the truth of h can be most likely inferred from t.

Textual entailment indeed captures generically a broad range of inferences that are relevant for multiple applications. For example, a QA system has to identify texts that entail a hypothesized answer. Given the question "*Does John Speak French?*", a text that includes the sentence "*John is a fluent French speaker*" entails the suggested answer "*John speaks French.*" In many cases, though, entailment inference is uncertain and has a probabilistic nature. For example, a text that includes the sentence "*John was born in France.*" does not strictly entail the above answer. Yet, it is clear that it does increase substantially the likelihood that the hypothesized answer is true.

The uncertain nature of textual entailment calls for its explicit modeling in probabilistic terms. We therefore propose a general generative probabilistic setting for textual entailment, which allows a clear formulation of probability spaces and concrete probabilistic models for this task. We suggest that the proposed setting may provide a unifying framework for modeling uncertain semantic inferences from texts.

An important sub task of textual entailment, which we term *lexical entailment*, is recognizing if the lexical concepts in a hypothesis h are entailed from a given text t, even if the relations which hold between these concepts in h may not be entailed from t. This is typically a necessary, but not sufficient, condition for textual entailment. For example, in order to infer from a text the hypothesis "*Chrysler stock rose*," it is a necessary that the concepts of *Chrysler*, *stock* and *rise* must be inferred from the text. However, for proper entailment it is further needed that the right relations would hold between these concepts. In this paper we demonstrate the relevance of the general probabilistic setting for modeling lexical entailment, by devising a preliminary alignment-based model that utilizes document co-occurrence probabilities in a bag of words representation.

Although our proposed lexical system is relatively simple, as it doesn't rely on syntactic or other deeper analysis, it nevertheless achieved an overall accuracy of 59% and an average precision of 0.57. The system did particularly well on the Comparable Documents (CD) task achieving an accuracy of 83%. These results may suggest that the proposed probabilistic framework is a promising basis for improved implementations that incorporate richer information.

2 A Probabilistic setting for Textual Entailment

2.1 Motivation

example	text	hypothesis
1	*John is a French Speaker*	*John speaks French*
2	*John was born in France*	*John speaks French*
3	*Harry's birthplace is Iowa*	*Harry was born in Iowa*
4	*Harry is returning to his Iowa hometown*	*Harry was born in Iowa*

Table 1. example sentence pairs

A common definition of entailment in formal semantics ([5]) specifies that a text t entails another text h (hypothesis, in our terminology) if h is true in every circumstance (*possible world*) in which t is true. For example, in examples 1 and 3 from Table 1 we'd assume humans to agree that the hypothesis is necessarily true in any circumstance for which the text is true. In such intuitive cases, textual entailment may be perceived as being certain, or, taking a probabilistic perspective, as having a probability of 1. In quite many other cases, though, entailment inference is uncertain and has a probabilistic nature. In example 2, the text doesn't contain enough information to infer the hypothesis' truth. And in example 4, the meaning of the word hometown is ambiguous and therefore one cannot infer for certain that the hypothesis is true. In both of these cases there are conceivable circumstances for which the text is true and the hypothesis is false. Yet, it is clear that in both examples, the text does increase substantially the likelihood of the correctness of the hypothesis, which naturally extends the classical notion of certain entailment. Given the text, we expect the probability that the hypothesis is indeed true to be relatively high, and significantly higher than its probability of being true without reading the text. Aiming to model application needs, we suggest that the probability of the hypothesis being true given the text reflects an appropriate confidence score for the correctness of a particular textual inference. In the next subsections we propose a concrete generative probabilistic setting that formalizes the notion of truth probabilities in such cases.

2.2 A Probabilistic Setting and Generative Model

Let T denote a space of possible texts, and $t \in T$ a specific text. Let H denote the set of all possible hypotheses. A hypothesis $h \in H$ is a propositional statement which can be assigned a truth value. It is assumed here that h is represented as a textual statement, but in principle it could also be expressed as a text annotated with additional linguistic information or even as a formula in some propositional language.

A semantic state of affairs is captured by a mapping from H to {0=false, 1=true}, denoted by $w : H \rightarrow \{0, 1\}$, called here *possible world* (following common terminology). A possible world w represents a concrete set of truth value assignments for all possible propositions. Accordingly, W denotes the set of all possible worlds.

We assume a probabilistic generative model for texts and possible worlds. In particular, we assume that texts are generated along with a concrete state of affairs, represented by a possible world. Thus, whenever the source generates a text t, it generates also corresponding hidden truth assignments that constitute a possible world w. The probability distribution of the source, over all possible texts and truth assignments $T \times W$, is assumed to reflect inferences that are based on the generated texts. That is, we assume that the distribution of truth assignments is not bound to reflect the state of affairs in a particular "real" world, but only the inferences about propositions' truth which are related to the text. The probability for generating a true hypothesis h that is not related at all to the corresponding text is determined by some prior probability $P(h)$. For example, h="*Paris is the capital of France*" might have a prior smaller than 1 and might well be false when the generated text is not related at all to Paris or France. In fact, we may as well assume that the notion of textual entailment is relevant only for hypotheses for which $P(h) < 1$, as otherwise (i.e. for tautologies) there is no need to consider texts that would support h's truth. On the other hand, we assume that the probability of h being true (generated within w) would be higher than the prior when the corresponding t does contribute information that supports h's truth.

2.3 Probabilistic textual entailment definition

We define two types of events over the probability space for $T \times W$:

I) For a hypothesis h, we denote as Tr_h the random variable whose value is the truth value assigned to h in a given world. Correspondingly, $Tr_h = 1$ is the event of h being assigned a truth value of 1 (true).

II) For a text t, we use t itself to denote also the event that the generated text is t (as usual, it is clear from the context whether t denotes the text or the corresponding event).

We say that a text t *probabilistically entails* a hypothesis h (denoted as $t \Rightarrow h$) if t increases the likelihood of h being true, that is, if $P(Tr_h = 1|t) > P(Tr_h = 1)$, or equivalently if the pointwise mutual information, $I(Tr_h = 1, t)$, is greater then 0. Once knowing that $t \Rightarrow h$, $P(Tr_h = 1|t)$ serves as a probabilistic confidence value for h being true given t.

Application settings would typically require that $P(Tr_h = 1|t)$ obtains a high value; otherwise, the text would not be considered sufficiently relevant to support h's truth (e.g. a supporting text in QA or IE should entail the extracted information with high confidence). Finally, we ignore here the case in which t contributes negative information about h, leaving this relevant case for further investigation.

2.4 Model Properties

It is interesting to notice the following properties and implications of our probabilistic setting:

A) Textual entailment is defined as a relationship between texts and propositions whose representation is typically based on text as well, unlike logical entailment which is a relationship between propositions only. Accordingly, textual entailment confidence is conditioned on the actual generation of a text, rather than its truth. For illustration, we would expect that the text "*His father was born in Italy*" would logically entail the hypothesis "*He was born in Italy*" with high probability - since most people who's father was born in Italy were also born there. However we expect that the text would actually not probabilistically textually entail the hypothesis since most people for whom it is specifically reported that their father was born in Italy were not born in Italy [1].

B) We assign probabilities to propositions (hypotheses) in a similar manner to certain probabilistic reasoning approaches (e.g. [1], [13]). However, we also assume a generative model of text, similar to probabilistic language models and statistical machine translation, which supplies the needed conditional probability distribution. Furthermore, since our conditioning is on texts rather than propositions we do not assume any specific logic representation language for text meaning, and only assume that textual hypotheses can be assigned truth values.

[1] This seems to be the case when analyzing the results of entering the above text in a web search engine.

C) Our framework does not distinguish between textual entailment inferences that are based on knowledge of language semantics (such as *murdering* $\Rightarrow$ *killing*) and inferences based on domain or world knowledge (such as *live in Paris* $\Rightarrow$ *live in France*). Both are needed in applications and it is not clear where and how to put such a borderline.

D) An important feature of the proposed framework is that for a given text many hypotheses are likely to be true. Consequently, for a given text t, $\sum_h P(Tr_h = 1|t)$ does not sum to 1. This differs from typical generative settings for IR and MT (e.g. [4], [19]), where all conditioned events are disjoint by construction. In the proposed model, it is rather the case that $P(Tr_h = 1|t) + P(Tr_h = 0|t) = 1$, as we are interested in the probability that a single particular hypothesis is true (or false).

E) An implemented model that corresponds to our probabilistic setting is expected to produce an estimate for $P(Tr_h = 1|t)$. This estimate is expected to reflect all probabilistic aspects involved in the modeling, including inherent uncertainty of the entailment inference itself (as in example 2 of Table 1), possible uncertainty regarding the correct disambiguation of the text (example 4), as well as uncertain probabilistic estimates that stem from the particular model structure and implementation.

3 A Lexical Entailment Model

We suggest that the proposed setting above provides the necessary grounding for probabilistic modeling of textual entailment. Since modeling the full extent of the textual entailment problem is clearly a long term research goal, in this paper we rather focus on the above mentioned subtask of *lexical entailment* - identifying when the lexical elements of a textual hypothesis are inferred from a given text.

To model lexical entailment we first assume that the meaning of each individual content word u in a hypothesis can be assigned a truth value. One possible interpretation for such truth values is that lexical concepts are assigned existential meanings. For example, for a given text t, $Tr_{book} = 1$ if it can be inferred in t's state of affairs that a book exists. Our model does not depend on any such particular interpretation, though, as we only assume that truth values *can* be assigned for lexical items but do not explicitly annotate or evaluate this subtask.

Given this lexically-projected setting, a hypothesis is assumed to be true if and only if all its lexical components are true as well. This captures our target perspective of lexical entailment, while not modeling here other entailment aspects. When estimating the entailment probability we assume that the truth probability of a term u in a hypothesis h is independent of the truth of the other terms in h, obtaining:

$$\begin{array}{lcl} P(Tr_h = 1|t) & = & \prod_{u \in h} P(Tr_u = 1|t) \\ P(Tr_h = 1) & = & \prod_{u \in h} P(Tr_u = 1) \end{array} \tag{1}$$

In order to estimate $P(Tr_u = 1|t)$ for a given word u and text $t = \{v_1, \ldots, v_n\}$, we further assume that the majority of the probability mass comes from a specific entailing word in t, allowing the following approximation:

$$P(Tr_u = 1|t) = \max_{v \in t} P(Tr_u = 1|T_v) \tag{2}$$

where T_v denotes the event that a generated text contains the word v. This corresponds to expecting that each word in h will be entailed from a specific word in t (rather than from the accumulative context of t as a whole). One can view Equation 2 as inducing an alignment between terms in the hypothesis and terms in the text, somewhat similar to alignment models in statistical MT (e.g. [4]).

Thus we obtain an estimate for the entailment probability based on lexical entailment probabilities from (1) and (2) as follows:

$$P(Tr_h = 1|t) = \prod_{u \in h} \max_{v \in t} P(Tr_u = 1|T_v) \tag{3}$$

3.1 Web-based Estimation of Lexical Entailment Probabilities

We perform unsupervised empirical estimation of the lexical entailment probabilities, $P(Tr_u = 1|T_v)$, based on word co-occurrence frequencies from the web. Following our proposed probabilistic model (cf. Section 2.2), we assume that the web is a sample generated by a language source. Each document represents a generated text and a (hidden) possible world. Given that the possible world of the text is not observed we do not know the truth assignments of hypotheses for the observed texts. We therefore further make the simplest assumption that all hypotheses stated verbatim in a document are true and all others are false and hence $P(Tr_u = 1|T_v) \approx P(T_u|T_v)$, the probability that u occurs in a text given that v occurs in that text. The lexical entailment probability estimate is thus derived from (3) as follows:

$$P(Tr_h = 1|t) \approx \prod_{u \in h} \max_{v \in t} P(T_u|T_v) \tag{4}$$

The co-occurrence probabilities are easily estimated based on maximum likelihood counts:

$$P(T_u|T_v) = \frac{n_{u,v}}{n_v} \tag{5}$$

where n_v is the number of documents containing word v and $n_{u,v}$ is the number of documents containing both u and v. In the experiments we obtained the corresponding counts by performing queries to a web search engine, since the majority of RTE examples were based on web snippets.

4 Experimental Setting

The text and hypothesis of all pairs in the RTE development and test sets were tokenized by the following simple heuristic - split at white space and remove any preceding or trailing of the following punctuation characters: ([{)]}"'`.,;:-!?. A standard stop word list was applied to remove frequent tokens. Counts were obtained using the AltaVista search engine[2], which supplies an estimate for the number of results (web-pages) for a given one or two token query.

We empirically tuned a threshold, λ, on the the estimated entailment probability to decide if entailment holds on not. For a $t-h$ pair, we tagged an example as true (i.e. entailment holds) if $p = P(Tr_h = 1|t) > \lambda$, and as false otherwise. We assigned a confidence of p to the positive examples ($p > \lambda$) and a confidence of $1 - p$ to the negative ones.

The threshold was tuned on the 567 annotated text-hypothesis example pairs in the development set for optimal *confidence weighted score* (*cws*). The optimal threshold of $\lambda = 0.005$ resulted in a *cws* of 0.57 and *accuracy* of 56% on the development set. This threshold was used to tag and assign confidence scores to the 800 pairs of the test set.

5 Analysis and Results

The resulting accuracy on the test set was of 59% and the resulting confidence weighted score was of 0.57. Both are statistically significantly better then chance at the 0.01 level.

Table 2 lists the accuracy and cws when computed separately for each task. As can be seen by the table the system does well on the CD and MT tasks, and quite poorly (not significantly better than chance) on the RC, PP, IR and QA tasks. It seems as if the success of the system is attributed almost solely to

task	accuracy	cws
Comparable Documents (CD)	0.8333	0.8727
Machine Translation (MT)	0.5667	0.6052
Information Extraction (IE)	0.5583	0.5143
Reading Comprehension (RC)	0.5286	0.5142
Paraphrase (PP)	0.5200	0.4885
Information Retrieval (IR)	0.5000	0.4492
Question Answering (QA)	0.4923	0.3736

Table 2. accuracy and cws by task

its success on the CD and MT tasks. Indeed it seems as if there is something common to these two tasks, which differentiates them from the others - in both tasks high overlap of content words (or their meanings) tends to correlate with entailment.

[2] http://www.av.com/

5.1 Success and failure cases

The system misclassified 331 out of the 800 test examples. The vast majority of these mistakes (75%) were false positives - pairs the system classified as true but were annotated as false. It is also interesting to note that the false negative errors were more common among the MT and QA tasks while the false positive errors were more typical to the other tasks. An additional observation from the recall-precision curve (Figure 3) is that high system confidence actually corresponds to false entailment. This is attributed to an artifact of this dataset by which examples with high word overlap between the text and hypothesis tend to be biased to negative examples (see [8]).

In an attempt to 'look under the hood' we examined the underlying alignment obtained by our system on a sample of examples. Figure 1 illustrates a typical alignment. Though some of the entailing words correspond to what we believe to be the correct alignment (e.g. voter → vote, Japan's → Japanese), the system also finds many dubious lexical pairs (e.g. turnout → half, percent → less).

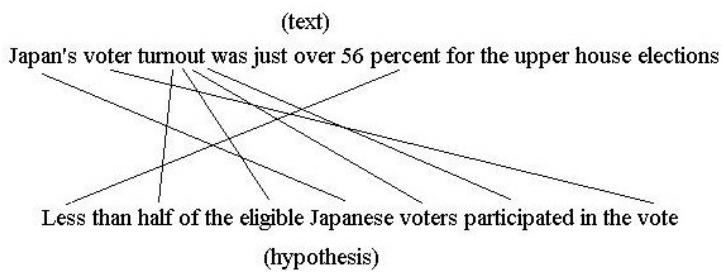

Fig. 1. system's underlying alignment for example 1026 (RC). gold standard - false, system - false

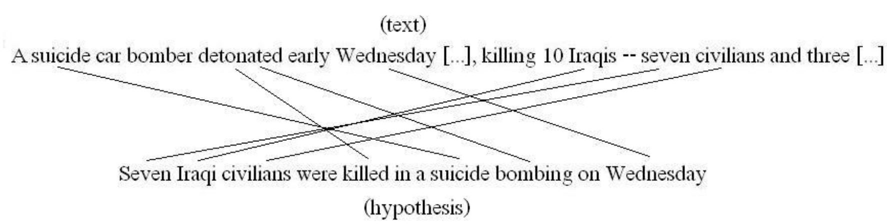

Fig. 2. system's underlying alignment for example 1095 (RC). gold standard - true, system - true

Furthermore, the induced alignments do not always correspond to the "expected" alignment. For example, in Figure 2 - based on the web co-occurrence statistics, *detonated* is a better trigger word for both *killed* and *bombing* even though one would expect to align them with the words *killing* and *bomber* respectively. Obviously, co-occurrence within documents is only one factor in estimating the entailment between words. This information should be combined with other statistical criteria that capture complementary notions of entailment, such as lexical distributional evidence as addressed in ([9], [10], [11]), or with lexical resources such as WordNet ([17]).

5.2 Comparison to baseline

As a baseline model for comparison we use a heuristic score proposed within the context of text summarization and Question Answering ([18], [20]). In this score semantic overlap between two texts is modeled via a word overlap measure, considering only words that appear in both texts weighted by *inverse document frequency* (*idf*). More concretely, this directional entailment score between two texts, denoted here by $entscore(t, h)$, is defined as follows:

$$entscore(t, h) = \frac{\sum_{w \in t \wedge h} idf(w)}{\sum_{w \in h} idf(w)} \tag{6}$$

where $idf(w) = log(N/n_w)$, N is the total number of documents in the corpus and n_w the number of documents containing word w. We have tested the performance of this measure in predicting entailment on the RTE dataset. Tuning the classification threshold on the development set (as done for our system), *entscore* obtained a somewhat lower accuracy of 56%.

To further investigate the contribution of the co-occurrence probabilities we extended the *entscore* measure by incorporating lexical co-occurrence probabilities in a somewhat analogous way to their utilization in our model. In this extended measure, termed $entscore_2$, we compute a weighted average of the lexical probabilities, rather than their product in our model (Equation 3), where the weights are the *idf* values, following the rational of the *entscore* measure. More concretely, $entscore_2$ is defined as follows:

$$entscore_2(t, h) = \frac{\sum_{u \in h} idf(u) * max_{v \in t} P(Tr_u = 1|v)}{\sum_{u \in h} idf(u)} \tag{7}$$

$P(Tr_u = 1|v)$ is approximated by $P(T_u|T_v)$ and estimated via co-occurrence counts, as in our model (equations 4 and 5). Note that when using this approximation, $P(Tr_u = 1|v) = 1$ when $u = v$ and thus the max value in (7) is obtained as 1 for hypothesis words that appear also in the text, naturally extending the rational of *entscore*.

Figure 3 compares the recall-precision curves for our system and the two baseline entailment scores. The different recall points are obtained by varying a threshold over the entailment score (or probability), considering all examples

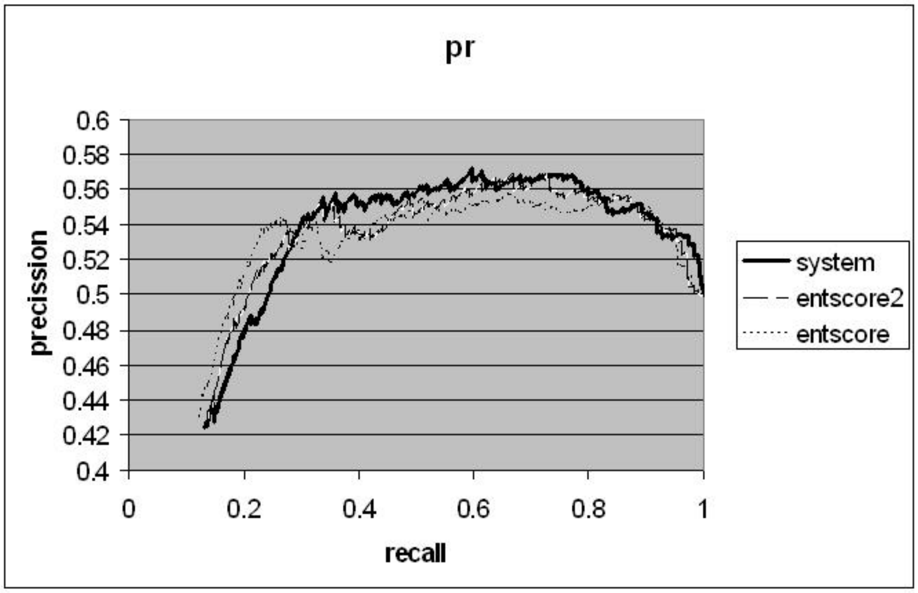

Fig. 3. Comparison to baselines (*system* refers to our probabilistic model)

with a score higher than the threshold as positive classifications. The figures show that on this dataset our system has higher precision over most recall ranges. [3] In addition, $entscore_2$, which incorporates lexical co-occurrence probabilities, performs somewhat better than the baseline *entscore* which considers only literal lexical overlap. These results demonstrate the marginal contribution of (i) utilizing lexical co-occurrence probabilities and (ii) embedding them within a principled probabilistic model.

5.3 Working at the lexical level

The proposed lexical model is quite simple and makes many obviously wrong assumptions. Some of these issues were addressed in another work by the authors ([12]), which was tested in a different setting. That model views lexical entailment as a text classification task. Entailment is derived from the entire context in the sentence (rather than word-to-word alignment) and Naïve Bayes classification is applied in an unsupervised setting to estimate the hidden lexical truth assignments. It would be interesting for future work to thoroughly compare and possibly combine the two models and thus capture entailment from a

[3] Note the anomaly that high lexical overlap, which yields high system confidence, actually correlates with false entailment (as noted in [8]). This anomaly explains the poor precision of all systems at the lower recall ranges, while the generally more accurate models are effected more strongly by this anomaly.

specific text term as well as the impact of the entire context within the given text.

Clearly, there is an upper bound of performance one would expect from a system working at the lexical level (see the analysis in [2]). Incorporating additional linguistic levels into the probabilistic entailment model, such as syntactic matching, co-reference resolution and word sense disambiguation, becomes a challenging target for future research.

6 Related Work

Modeling semantic overlap between texts is a common problem in NLP applications. Many techniques and heuristics were applied within various applications to model such relations between text segments. Within the context of Multi Document Summarization, [18] propose modeling the directional entailment between two texts t, h via the entailment score of Equation 6 to identify redundant information appearing in different texts. A practically equivalent measure was independently proposed in the context of QA in [20]. This baseline measure captures word overlap, considering only words that appear in both texts and weighs them based on their inverse document frequency.

Different techniques and heuristics were applied on the RTE-1 dataset to specifically model textual entailment. Interestingly, a number of works (e.g. [3], [6], [14]) applied or utilized a lexical based word overlap measure similar to Equation 7. The measures vary in the word-to-word similarity used and the weighting scheme. Distributional similarity (such as [16]) and WordNet based similarity measures (such as [15]) were applied. In addition, the different works vary in the preprocessing done (tokenization, lemmatization, etc.) and in the corpora used to collect statistics. For this reason it is difficult to compare the performance of the different measure variants of different systems. Nevertheless the reported results were all comparable, which may suggest that these lexical techniques are somewhat close to exhausting the potential of lexical based systems.

7 Conclusions

This paper described the Bar-Ilan system participating in the First Recognising Textual Entailment Challenge. We proposed a general probabilistic setting that formalizes the notion of textual entailment. In addition we described an alignment-based model in a bag of words representation for lexical entailment, which was applied using web co-occurrence statistics. Although our proposed lexical system is relatively simple, as it does not rely on syntactic or other deeper analysis, it nevertheless achieved competitive results. These results may suggest that the proposed probabilistic framework is a promising basis for improved implementations that would incorporate deeper types of information.

8 Acknowledgments

This work was supported in part by the IST Programme of the European Community, under the *PASCAL Network of Excellence*, IST-2002-506778. This publication only reflects the authors' views.

References

1. Bacchus, F.: Representing and Reasoning with Probabilistic Knowledge, M.I.T. Press (1990).
2. Bar-Haim, R., Szpektor, I., Glickman O.: Definition and Analysis of Intermediate Entailment Levels. ACL-05 Workshop on Empirical Modeling of Semantic Equivalence and Entailment. (2005)
3. Bos, J., Markert, K.: Recognising Textual Entailment with Robust Logical Inference. This Volume.
4. Brown, P.F., Della Pietra, V.J., Della Pietra, S.A., Mercer, R.L.: The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, 19(2):263-311 (1993).
5. Chierchia, G., McConnell-Ginet, S.: Meaning and grammar: An introduction to semantics, 2nd. edition. Cambridge, MA: MIT Press (2001).
6. Corley, C., Mihalcea, R.: Measuring the Semantic Similarity of Texts. ACL 2005 Workshop on Empirical Modeling of Semantic Equivalence and Entailment.
7. Dagan, I., Glickman, O.: Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. PASCAL workshop on Learning Methods for Text Understanding and Mining (2004).
8. Dagan, I., Glickman, O., Magnini, B.: The PASCAL Recognising Textual Entailment Challenge. This Volume.
9. Geffet, M., Dagan, I.: Feature Vector Quality and Distributional Similarity. Coling 2004.
10. Geffet, M., Dagan, I.: The Distributional Inclusion Hypotheses and Lexical Entailment. ACL 2005.
11. Glickman, O., Dagan, I.: Identifying Lexical Paraphrases From a Single Corpus: A Case Study for Verbs. Recent Advantages in Natural Language Processing (RANLP). 2003.
12. Glickman, O., Dagan, I., Koppel, M.: A Probabilistic Classification Approach for Lexical Textual Entailment, Twentieth National Conference on Artificial Intelligence (AAAI) 2005.
13. Halpern, J.Y.: An analysis of first-order logics of probability. Artificial Intelligence 46:311-350 (1990).
14. Jijkoun, V., de Rijke, M.: Recognizing Textual Entailment: Is Lexical Similarity Enough?. This Volume.
15. Leacock, C., Chodorow, M., and Miller, G.: Using corpus statistics and wordnet relations for sense disambiguation. Computational Linguistics 24, 1 1998, 147–165.
16. Lin, D.: Automatic Retrieval and Clustering of Similar Words. COLING-ACL, 1998.
17. Miller, G., A.: WordNet: A Lexical Databases for English. CACM, 38(11):39–41, 1995.
18. Monz, C., de Rijke, M.: Light-Weight Entailment Checking for Computational Semantics. In Proc. of the third workshop on inference in computational semantics (ICoS-3) (2001).

19. Ponte, J.M., Croft W.B.: A Language Modeling Approach to Information Retrieval. SIGIR conference on Research and Development in Information Retrieval (1998).
20. Saggion, H., Gaizauskas, R., Hepple, M., Roberts, I., Greenwood, M.: Exploring the Performance of Boolean Retrieval Strategies for Open Domain Question Answering. Proceedings of the Workshop on Information Retrieval for Question Answering. SIGIR 2004.