Recognize the Generality Relation between Sentences using Asymmetric Association Measures

Sebastião Pais
MINES ParisTech
Centre de Recherche en Informatique
77305 Fontainebleau, France
pais@cri.ensmp.fr

Gaël Dias and **Rumen Moraliyski**
Normandie University
UNICAEN, GREYC CNRS
F-14032 Caen, France
firstname.lastname@unicaen.fr

Abstract

In this paper we focus on a particular case of entailment, namely entailment by generality. We argue that there exist various types of implication, a range of different levels of entailment reasoning, based on lexical, syntactic, logical and common sense clues, at different levels of difficulty. We introduce the paradigm of Textual Entailment (TE) by Generality, which can be defined as the entailment from a specific statement towards a relatively more general statement. In this context, the Text T entails the Hypothesis H, and at the same time H is more general than T. We propose an unsupervised and language-independent method to recognize TE by Generality given a case of $Text - Hypothesis$ or $T - H$ where entailment relation holds.

1. Introduction

We introduce the paradigm of TE by Generality, which can be defined as the entailment from a specific sentence towards a more general sentence. For example, from sentences (1) and (2) extracted from RTE-1, we would easily state that $(1) \rightarrow (2)$ as their meaning is roughly the same and sentence (2) is more general than sentence (1).

(1) Mexico City has a very bad pollution problem because the mountains around the city act as walls and block in dust and smog.

(2) Poor air circulation out of the mountain-walled Mexico City aggravates pollution.

To understand how TE by Generality can be modeled for two sentences, we propose a new paradigm based on the Asymmetric InfoSimba Similarity (AIS) measure. Instead of relying on the exact matches of words between texts, we propose that one sentence entails the other one in terms of generality if two constraints hold: (a) if and only if many of the words in T are semantically similar to the words that make H, and (b) if most of the words of H are more general than the words of T. As far as we know, we are the first to propose an unsupervised, language-independent, threshold free methodology in the context of TE by Generality, although the approach of Glickman and Dagan (2005) is based on similar assumptions. This new proposal is exhaustively evaluated against the first five RTE datasets. In particular, the RTE-1 is the only dataset for which there exist comparable results with linguistic-free methodologies (Glickman and Dagan, 2005; Perez et al., 2005; Bayer et al., 2005).

In this paper we hypothesize the existence of a special mode of TE, namely TE by Generality. Thus, the main contribution of our study is to highlight the importance of this inference mechanism.

2. Variants of the Entailment

Pazienza et al. (2005) define three types of entailment:

1. *Semantic Subsumption* - T and H express the same fact, but the situation described in T is more specific than the situation in H. The specificity of T is expressed through one or more semantic operations. For example, in the sentential pair:

   > • H: The cat eats the mouse. | T: The cat devours the mouse.
   >
   > T is more specific than H, as eat is a semantic generalization of devour.

2. *Syntactic Subsumption* - T and H express the same fact, but the situation described in T is more specific than the situation in H. The specificity of T is expressed through one or more syntactic operations. For example, in the pair:

   > • H: The cat eats the mouse. | T: The cat eats the mouse in the garden.
   >
   > T contains a modifying prepositional phrase.

3. *Direct Implication* - H expresses a fact that is implied by a fact in T. For example:

   > • H:The cat killed the mouse. | T: The cat devours the mouse.
   >
   > H is implied by T, as it is supposed that killed is a precondition for devour. In Dagan and Glickman (2004) syntactic subsumption roughly corresponds to the restrictive extension rule, while direct implication and semantic subsumption correspond to the axiom rule.

We want to regard entailment by generality as a relation between utterances (that is, sentences in context), where the context is relevant to understand the meaning. In relation to the classification proposed by Pazienza et al. (2005), entailment by generality is comparable to *Semantic Subsumption* kind of TE. Thus, Entailment by Generality can be defined as the entailment from specific sentence towards a more general sentence.

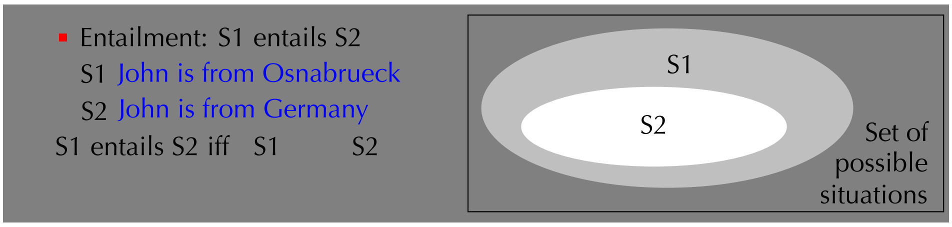

Figure 1: Venn diagram: entailment by generality.

2.1. Context Textual Entailment

Within TE framework, a text T is said to entail a textual hypothesis H if the truth of H can be inferred from T. This means that most people would agree that the meaning of T implies that of H. Somewhat more formally, we say that T entails H when some representation of H can be "matched" with some (or part of a) representation of T, at some level of granularity and abstraction.

Dagan and Glickman (2004) define TE as a relationship between a coherent textual fragment T and a language expression, which is considered as a hypothesis H. Entailment holds (i. e. $T \rightarrow H$) if the meaning of H can be inferred from the meaning of T, as interpreted by a typical language user. This relationship is directional and asymmetric since the meaning of one expression may usually entail the other while entailment in the other direction is less certain.

For instance, a Question Answering (QA) system has to identify texts that entail the expected answer. Given the question "Who painted the Mona Lisa?", the text "Among the works created by Leonardo da Vinci in the 16th century is the small portrait known as the Mona Lisa or la 'Gioconda'", entails the expected answer "Leonardo da Vinci painted the Mona Lisa". Similarly, in Information Retrieval (IR) relevant documents should entail the combination of semantic concepts and relations denoted by the query. In Information Extraction (IE), entailment holds between different text variants expressing the same target relation (Romano et al., 2006). In text summarization, an important processing stage is sentence extraction, which identifies the most important sentences of the texts to be summarized; especially when generating a single summary from several documents (Barzilay and McKeown, 2005), it is important to avoid selecting sentences that convey the same information as other sentences that have already been selected, i.e. ones that entail such sentences.

3. Recognizing Textual Entailment

Basically, Recognizing Textual Entailment (RTE) is the task of deciding, given two text fragments, whether the meaning of one of the texts is entailed (can be inferred) from the other text. Also, this task captures generically a broad range of inferences that are relevant for multiple applications. A necessary step in transforming textual entailment from a theoretical idea into an active empirical research field was the introduction of benchmarks and an evaluation forum for entailment systems.

3.1. Unsupervised and Language-Independent Methodologies

Different approaches have been proposed to recognize Textual Entailment: from unsupervised language-independent methodologies (Glickman and Dagan, 2005; Perez et al., 2005; Bayer et al., 2005) to deep linguistic analysis. We will particularly detail the unsupervised language-independent approaches, to which our work can be directly compared, at least to a certain extent.

One of the most simple proposals (Perez et al., 2005) explores the *BLEU algorithm* (Papineni et al., 2002). First, for several values of n (typically from 1 to 4), they calculate the percentage of n-grams from the text T, which appear in the hypothesis H. The frequency of each n-gram is limited to the maximum frequency with which it appears in any text T. Then, they combine the marks obtained for each value of n as a weighted linear average and finally apply a brevity factor to penalize short texts T. The output of BLEU is then taken as the confidence score. Finally, they perform an optimization procedure to choose the best threshold according to the percentage of success of correctly recognized entailment. This procedure achieves 0.495 accuracy in recognizing TE.

In Bayer et al. (2005) the entailment data is treated as an aligned translation corpus. In particular, they use the *GIZA++* toolkit (Och and Ney, 2003) to induce alignment models. However, the alignment scores alone were next to useless for the RTE-1 development data, predicting entailment correctly only slightly above chance. As a consequence, they introduced a combination of metrics intended to measure translation quality. Finally, they combined all the alignment information and string metrics with the classical K Nearest Neighbors (K-NN) classifier to choose for each test pair the dominant truth value among the five nearest neighbors in the development set. This method achieves 0.586 accuracy.

The most interesting work is certainly the one described in Glickman and Dagan (2005), who propose a general probabilistic setting that formalizes the notion of TE. Here, they focus on identifying when the lexical elements of a textual hypothesis H are inferred from a given text T. The probability of lexical entailment is derived from Equation 1 where $hits(.,.)$ is a function that returns the number of documents containing its arguments.

$$P(H|T) = \prod_{u \in H} max_{v \in T} \frac{hits(u,v)}{hits(v)} \tag{1}$$

The text and hypothesis of all pairs in the development and test sets were tokenized and stop words were removed to empirically tune a decision threshold, λ. Thus, for a pair $T-H$, they tagged an example as true (i.e. entailment holds) if $P(H|T) > \lambda$, and as false otherwise. The threshold was empirically set to 0.005. With this method accuracy of 0.586 is achieved. The best results from these three approaches are obtained by Glickman and Dagan (2005), who introduce the notion of asymmetry within their model. The underlying idea is based on the fact that for each word in H the best asymmetrically co-occurring word in T is chosen to evaluate $P(H|T)$. Although all three approaches show interesting properties, they all depend on tuned thresholds, which can not reliably be reproduced and need to be changed for each new application. Moreover, they need training data, which may not be available. Our idea aims at generalizing the hypothesis made by Glickman and Dagan (2005).

4. Asymmetric Word Similarities

Two different types of knowledge can be acquired depending on the basic textual unit under study. On the one hand, analyzing word similarities evidences intrinsic knowledge about the language (i.e. information about the language which is not explicitly encoded in texts). Traditional examples are collocations and word semantic relations such as hypernymy/hyponymy, meronymy/holonymy, synonymy or antonymy, which must be mined from texts. On the other hand, explicit knowledge about the language (i.e. information about the message conveyed by the texts) can be extracted from the evaluation of sentence, passage and text similarities[1]. There are obviously some exceptions.

4.1. Asymmetric Association Measures (AAMs)

In order to stay within the domain of language-independent and unsupervised methodologies, a number of asymmetric association measures have been proposed (Pecina and Schlesinger, 2006; Tan et al., 2004) and applied to the problems of taxonomy construction (Sanderson and Croft, 1999; Cleuziou et al., 2010), cognitive psycholinguistics (Michelbacher et al., 2007) and general-specific word order induction (Dias et al., 2008). Sanderson and Croft (1999) is certainly one of the first studies to propose the use of the conditional probability for taxonomy construction.

They assume that a term t_2 subsumes a term t_1 if the documents in which t_1 occurs are a subset of the documents in which t_2 occurs constrained by $P(t_2|t_1) \geq 0.8$ and $P(t_1|t_2) < 1$. By gathering all subsumption relations, they build the semantic structure of any domain, which corresponds to a directed acyclic graph. In Sanderson and Lawrie (2000), the subsumption relation is indicated by the following expressions $P(t_2|t_1) \geq P(t_1|t_2)$ and $P(t_2|t_1) > t$ where t is a given threshold and all term pairs found to have a subsumption relationship are passed through a transitivity module, which removes extraneous subsumption relationships in the way that transitivity is preferred over direct pathways, thus leading to a non-triangular directed acyclic graph.

Eight of the AAMs used in that work will be evaluated in the context of asymmetric similarity between sentences: the Added Value (Equation 2), the Braun-Blanket (Equation 3), the Certainty Factor (Equation 4), the Conviction (Equation 5), the Gini Index (Equation 6), the J-measure (Equation 7), the Laplace (Equation 8) and the Conditional Probability (Equation 9).

$$AV(x\|y) = P(x|y) - P(x). \tag{2}$$

$$BB(x\|y) = \frac{f(x,y)}{f(x,y) + f(\bar{x},y)}. \tag{3}$$

$$CF(x\|y) = \frac{P(x|y) - P(x)}{1 - P(x)}. \tag{4}$$

$$CO(x\|y) = \frac{P(x) \times P(\bar{y})}{P(x,\bar{y})}. \tag{5}$$

$$GI(x\|y) = P(y) \times (P(x|y)^2 + P(\bar{x}|y)^2) - P(x)^2 \times P(\bar{y}) \times (P(x|\bar{y})^2 + P(\bar{x}|\bar{y})^2) - P(\bar{x})^2. \tag{6}$$

$$JM(x\|y) = P(x,y) \times \log\frac{P(x|y)}{P(x)} + P(\bar{x},y) \times \log\frac{P(\bar{x}|y)}{P(\bar{x})}. \tag{7}$$

$$LP(x\|y) = \frac{N \times P(x,y) + 1}{N \times P(y) + 2} \tag{8}$$

$$P(x|y) = \frac{P(x,y)}{P(y)} \tag{9}$$

4.2. Asymmetric Attributional Word Similarities

The InfoSimba (IS) aims to measure the correlations between all the pairs of words in two word context vectors instead of just relying on their exact match as with the cosine similarity measure. Further, IS guarantees to catch similarity between pairs of words even when they do not share contexts, for example due to data sparseness. IS takes under account the fraction of similar contexts instead. It is defined in Equation 10 where $S(.,.)$ is any symmetric similarity measure and each W_{ik} corresponds to the attribute word at the k^{th} position in the vector X_i, p and q are the lengths of the vectors X_i and X_j respectively.

[1]From now on, we will refer to sentences, passages and texts simply as texts.

$$IS(X_i, X_j) = \frac{\sum_{k=1}^{p}\sum_{l=1}^{q} X_{ik} \times X_{jl} \times S(W_{ik}, W_{jl})}{\left(\begin{array}{l} \sum_{k=1}^{p}\sum_{l=1}^{p} X_{ik} \times X_{il} \times S(W_{ik}, W_{il}) + \\ \sum_{k=1}^{q}\sum_{l=1}^{q} X_{jk} \times X_{jl} \times S(W_{jk}, W_{jl}) - \\ \sum_{k=1}^{p}\sum_{l=1}^{q} X_{ik} \times X_{jl} \times S(W_{ik}, W_{jl}) \end{array}\right)}. \tag{10}$$

Although there are many asymmetric similarity measures, they evidence problems that may reduce their utility. On the one hand, asymmetric association measures can only evaluate the generality/specificity relation between words that are known to be in a semantic relation (Sanderson and Croft, 1999; Dias et al., 2008). Indeed, they generally capture the direction of association between two words based on document contexts and only take into account a loose semantic proximity between words. For example, it is highly probable to find that *Apple* is more general than *iPad*, which can not be considered to be an hypernymy/hyponymy or meronymy/holonymy relation. On the other hand, asymmetric attributional word similarities only take into account common contexts to assess the degree of asymmetric relatedness between two words. To leverage these issues, we propose the Asymmetric InfoSimba (AIS) measure whose underlying idea is to say that one word x is semantically related to word y and x is more general than y if x and y share as many similar contexts as possible and each context word of x is likely to be more general than most of the context words of y. The AIS is defined in Equation 11, where $AS(.\|.)$ is any asymmetric similarity measure, likewise for the IS in Equation 10 where $S(.,.)$ stands for any symmetric similarity measure. We also define its simplified version $AISs(.\|.)$ in Equation 12.

$$AIS(X_i \| X_j) = \frac{\sum_{k=1}^{p}\sum_{l=1}^{q} X_{ik} \times X_{jl} \times AS(W_{ik} \| W_{jl})}{\left(\begin{array}{l} \sum_{k=1}^{p}\sum_{l=1}^{p} X_{ik} \times X_{il} \times AS(W_{ik} \| W_{il}) + \\ \sum_{k=1}^{q}\sum_{l=1}^{q} X_{jk} \times X_{jl} \times AS(W_{jk} \| W_{jl}) - \\ \sum_{k=1}^{p}\sum_{l=1}^{q} X_{ik} \times X_{jl} \times AS(W_{ik} \| W_{jl}) \end{array}\right)}. \tag{11}$$

$$AISs(X_i \| X_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} X_{ik} \times X_{jl} \times AS(W_{ik} \| W_{jl}). \tag{12}$$

5. Asymmetry between Sentences

A number of ways to compute the similarity between two sentences were proposed in the literature. Most similarity measures determine the distance between two vectors associated to two sentences (i.e. the vector space model). However, when applying the classical similarity measures between two sentences, only the identical indexes of the row vector X_i and X_j are taken into account, which may result in misleading values. To deal with this problem, different methodologies have been proposed, but the most promising one is certainly the one proposed by Dias et al. (2007), the InfoSimba informative similarity measure, expressed in Equation 10.

Although there exsist many asymmetric similarity measures between words, there does not exist any attributional similarity measure capable to assess whether a sentence is more specific/general than another one. To overcome this issue, we introduce the asymmetric InfoSimba similarity measure (AIS), which underlying idea is to say that a sentence T is semantically related to sentence H and H is more general than T, if H and T have many related words in common and each word of H is likely to be more general than most of the words of T. The AIS is defined in Equation 11.

As AIS is computationally expensive, we also define its simplified version $AISs(.\|.)$ in Equation 12, which we will specifically use in our experiments.

As a consequence, entailment by generality ($T \overset{G}{\to} H$) will hold if and only if

$$AISs(T \| H) < AISs(H \| T).$$

Due to its asymmetric definition, in contrast to existing methodologies, we do not need to define or tune thresholds.

6. Three Levels of Pre-Processing

We consider three approaches for selecting the words for the calculation of the asymmetry between sentences. Thus, we can assess which approach performs best to identify entailment by generality. In the first approach, we chose to do the calculations without preprocessing, i.e., do the calculations with all the words. The next approach was to use a list of Stop Words[2].

Finally, in the last approach, we used the Software for the Extraction of N-ary Textual Associations (SENTA) (Dias et al., 1999), in order to extract important Multiword Units (MWU). This system is parameter free and language independent, thus allowing to extract MWU from raw text.

In summary, our experiments are based on three approaches to the calculations to which we refer bellow as *With All Words*, *Without Stop Words* and *With MWU*.

7. Evaluation

In order to evaluate our methodology against well known test data used to compare a number of methodologies our evaluation is based on analysis of Confusion Matrix and values calculated from it. An important performance measure is classification Accuracy (AC) and Precision (P). More specifically, in our work we used the following performance measures – *Average Accuracy*, *Average Precision* and *Weighted Average Accuracy*, *Weighted Average Precision*. Although the obtained results are not excellent, they are promising and encouraging.

Averages ACCURACY by RTE Challenges — Measures versus Approach			
AAM	**Arithmetic Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.54	0.52	0.53
BRAUN-BLANKET	**0.55**	**0.53**	**0.54**
CERTAINTY FACTOR	0.53	**0.53**	0.53
CONDITIONAL PROBABILITY	0.53	**0.53**	0.53
CONVICTION	0.52	0.51	0.51
GINI INDEX	0.54	0.51	0.53
J-MEASURE	0.53	0.51	0.52
LAPLACE	0.53	**0.53**	0.53
AAM	**Weighted Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.53	0.52	0.53
BRAUN-BLANKET	**0.54**	0.52	**0.56**
CERTAINTY FACTOR	0.53	**0.53**	0.52
CONDITIONAL PROBABILITY	0.53	**0.53**	0.53
CONVICTION	0.52	0.50	0.51
GINI INDEX	0.53	0.52	0.53
J-MEASURE	**0.54**	0.51	0.52
LAPLACE	0.52	0.52	**0.56**

Table 1: Accuracy Averages | Measures versus Approach

Regarding the **Arithmetic Average** (Table 1), the combination that has the best performance is the *Braun-Blanket* measure on *All Words*. Best **Weighted Average** is achieved on *With WMU* approach by *Braun-Blanket* and *Laplace* measures. Overall, the worst result was obtained with the measure *Conviction* in the approach *Without Stop Words*.

Accuracy values of our experiments on RTE Challenges span a relatively short range between 0.50 and 0.56.

---

[2]Obtained using `http://www.microsoft.com/en-us/download/confirmation.aspx?id=10024` [Last access: 14th December, 2013]

Average PRECISION - ENTAILMENT by RTE Challenges — Measures versus Approach			
AAM	**Arithmetic Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.66	0.65	**0.78**
BRAUN-BLANKET	0.53	0.60	0.63
CERTAINTY FACTOR	0.63	0.62	0.46
CONDITIONAL PROBABILITY	0.64	0.64	0.48
CONVICTION	0.60	0.54	0.54
GINI INDEX	0.67	0.55	0.53
J-MEASURE	**0.81**	**0.75**	0.63
LAPLACE	0.65	0.64	0.47
AAM	**Weighted Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.60	0.59	**0.74**
BRAUN-BLANKET	0.49	0.54	0.58
CERTAINTY FACTOR	0.58	0.57	0.46
CONDITIONAL PROBABILITY	0.58	0.58	0.48
CONVICTION	0.59	0.53	0.53
GINI INDEX	0.62	0.52	0.53
J-MEASURE	**0.73**	**0.66**	0.63
LAPLACE	0.59	0.57	0.48

Table 2: PRECISION – ENTAILMENT Averages | Measures versus Approach

Table 1 points out the approach *Without Stop Words* as the one with worst performance in terms of accuracy, while *All Words* achieves slightly better accuracy compared to *With MWU*.

In Table 2, the combination with the best performance on the **Arithmetic Average Precision** is the *J-measure* with approach *All Words*. For the **Weighted Average Precision**, the *Added Value* shows the best result *With MWU*. The worst result is obtained with the measure *Certainty Factor With MWU* – 0.46.

With respect to the **Precision – Entailment** criterion, the approach that achieves the best results is *With All Words*.

In contrast to the results for **Precision – Entailment**, our method shows unsatisfactory behavior when considered from the perspective of **Precision – No Entailment** (see Table 3). For **Arithmetic Average** the best combination is *Certainty Factor*, *Conditional Probability* and *Laplace With MWU*. For **Weighted Average**, *Laplace* has the best performance *With MWU* approach. Note the low results obtained by the *J-measure* and *Added Value*. In Table 3 the approach with the best performance is *With MWU*, and the worst performing approach is *Without Stop Words*.

After an exhaustive analysis of the results obtained, we can compare our results with the results of the methodologies presented in Section 3.1. Precisely, Bayer et al. (2005), Glickman and Dagan (2005) and Perez et al. (2005) obtained accuracy of 0.586, 0.586 and 0.495, respectively. We prove that our methodology has better performance compared to what was possible in previous works. On RTE-1 Challenge *With MWU* approach, our methodology achieved its best results. The measures *Braun-Blanket* and *Laplace* achieve good results in **Weighted Average Accuracy**, namely 0.61.

8. Conclusion

We study the behavior of our methodology for recognizing TE by Generality. Also, we provide a thorough comparison to related works. This is done taking into account the limitations of typical language-independent and unsupervised learning techniques. In order to obtain fair comparison, we used a well known dataset studied in the RTE Challenge as our test-bed. Further, as we are interested in a special kind of TE, we built a suitable corpus.

Average PRECISION - NO ENTAILMENT by RTE Challenges — Measures versus Approach			
AAM	**Arithmetic Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.40	0.39	0.28
BRAUN-BLANKET	**0.56**	0.46	0.45
CERTAINTY FACTOR	0.43	0.44	**0.59**
CONDITIONAL PROBABILITY	0.42	0.41	**0.59**
CONVICTION	0.44	**0.49**	0.48
GINI INDEX	0.39	0.48	0.53
J-MEASURE	0.26	0.27	0.42
LAPLACE	0.40	0.41	**0.59**
AAM	**Weighted Average by Approach**		
	With All Words	**Without Stop Words**	**With MWU**
ADDED VALUE	0.49	0.48	0.32
BRAUN-BLANKET	**0.62**	**0.53**	0.51
CERTAINTY FACTOR	0.50	0.51	0.60
CONDITIONAL PROBABILITY	0.50	0.49	0.61
CONVICTION	0.46	0.49	0.52
GINI INDEX	0.46	**0.53**	0.55
J-MEASURE	0.37	0.38	0.43
LAPLACE	0.47	0.48	**0.62**

Table 3: PRECISION – NO ENTAILMENT Averages | Measures versus Approach

In this process we learned that detecting entailment between sentences is not an exact science. We saw that each new RTE Challenge required different approach to the problem. Thus, we do not provide a measure or an approach that pretends to solve the problem. We can only conclude, based on evidences from Table 2 that for some combinations of measure and preprocessing approach our method shows good precision in recognizing TE.

Comparing our results, with the results of other relevant methodologies, presented in Section 3.1., we prove that our methodology achieves higher performance figures. The measures *Braun-Blanket* and *Laplace* achieve better results for **Weighted Average Accuracy**, namely 0.61.

With this paper, we contribute an original proposal to RTE. Our methodology is unsupervised and language-independent, and accounts for the asymmetry of the studied phenomena by means of asymmetric similarity measures.

References

Barzilay, R. and McKeown, K. R. (2005). Sentence Fusion for Multidocument News Summarization. *Computational Linguistics*, 31(3):297–328.

Bayer, S., Burger, J., Ferro, L., Henderson, J., and Yeh, E. (2005). Mitre's Submission to the EU Pascal RTE Challenge. In *PASCAL. Proc. of the First Challenge Workshop. Recognizing Textual Entailment*, pages 41–44.

Cleuziou, G., Dias, G., and Levorato, V. (2010). Modélisation Prtopologique pour la Structuration Sémantico-Lexicale. In *Proceedings of the 17èmes Rencontres de la Société Francophone de Classification (SFC 2010).*

Dagan, I. and Glickman, O. (2004). Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. In *Learning Methods for Text Understanding and Mining*.

Dias, G., Guilloré, S., and Lopes, J. (1999). Language Independent Automatic Acquisition of Rigid Multiword Units from Unrestricted Text Corpora. In *Proceedings of the 6ème Confrence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN 1999)*, pages 333–339.

Dias, G., Alves, E., and Lopes, J. (2007). Topic Segmentation Algorithms for Text Summarization and Passage Retrieval: An Exhaustive Evaluation. In *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI 2007)*, pages 1334–1340.

Dias, G., Mukelov, R., and Cleuziou, G. (2008). Unsupervised Graph-Based Discovery of General-Specific Noun Relationships from Web Corpora Frequency Counts. In *Proceedings of the 12th International Conference on Natural Language Learning (CoNLL 2008)*.

Glickman, O. and Dagan, I. (2005). Web Based Probabilistic Textual Entailment. In *Proceedings of the 1st Pascal Challenge Workshop*, pages 33–36.

Michelbacher, L., Evert, S., and Schtze, H. (2007). Asymmetric Association Measures. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2007)*, pages 1–6.

Och, F. J. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL 2002, pages 311–318.

Pazienza, M. T., Pennacchiotti, M., and Zanzotto, F. M. (2005). A Linguistic Inspection of Textual Entailment. In *Proceedings of the 9th Conference on Advances in Artificial Intelligence*, AI*IA 2005, pages 315–326.

Pecina, P. and Schlesinger, P. (2006). Combining Association Measures for Collocation Extraction. In *Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING/ACL 2006)*, pages 651–658.

Perez, D., Alfonsecaia, E., and Rodríguez, P. (2005). Application of the Bleu Algorithm for Recognising Textual Entailments. In *Proceedings of the Recognising Textual Entailment Pascal Challenge*.

Romano, L., Kouylekov, M., and Szpektor, I. (2006). Investigating a Generic Paraphrase-Based Approach for Relation Extraction. In *EACL 2006*.

Sanderson, M. and Croft, B. (1999). Deriving Concept Hierarchies from Text. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 1999, pages 206–213.

Sanderson, M. and Lawrie, D. (2000). Building, Testing, and Applying Concept Hierarchies. *Advances in Information Retrieval*, 7:235–266.

Tan, P.-N., Kumar, V., and Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. *Information Systems*, 29(4):293–313.