Textual Entailment Recognition Based on Dependency Analysis and WordNet

Jesús Herrera, Anselmo Peñas, Felisa Verdejo
Departmento de Lenguajes y Sistemas Informáticos
Universidad Nacional de Educación a Distancia
Madrid, Spain
{jesus.herrera, anselmo, felisa}@lsi.uned.es

Abstract

The UNED-NLP Group [1] Recognizing Textual Entailment System is based on the use of a broad-coverage parser to extract dependency relations and a module which obtains lexical entailment relations from *WordNet*. The work aims at comparing whether the matching of dependency trees substructures give better evidence of entailment than the matching of plain text alone.

1 Introduction

The system of the UNED-NLP Group which has taken part in the 2005 PASCAL [2] Recognizing Textual Entailment Challenge is a proposal towards the resolution of the Recognizing Textual Entailment (RTE) problem. The present approach explores the possibilities of matching between dependency trees of text and hypothesis. System's components are the following:

1. A dependency parser, based on Lin's *Minipar* (Lin, 1998), which normalizes data from the corpus of text and hypothesis pairs, accomplishes the dependency analysis and creates into memory appropriated structures to represent it.

2. A lexical entailment module, which takes the information given by the parser and returns hypothesis' nodes entailed by the text.

3. A matching evaluation module, which searches for paths into hypothesis' dependency tree, conformed by the lexically entailed nodes.

Section 2 shows how lexical entailment is accomplished. Section 3 presents the methodology followed to evaluate matching between dependency trees. Section 4 describes some experiments and their results. Finally, some conclusions are given.

2 Lexical Entailment

After the dependency parsing, a module of lexical entailment is applied over the nodes of both text and hypothesis. The output of this module is a list of pairs *(T,H)* where *T* is a node in the text's dependency tree whose lexical unit entails the lexical unit of the node *H* in the hypothesis' dependency tree. This entailment at the word level considers *WordNet* relations, detection of *WordNet* multiwords and negation, as follows:

2.1 Synonymy and Similarity

The lexical unit *T* entails the lexical unit *H* if they can be *synonyms* according to *WordNet* or if there is a relation of *similarity* between them. Some examples were found in the PASCAL Challenge training corpus such as, for example: *discover* and *reveal*, *obtain* and *receive*, *lift* and *rise*, *allow* and *grant*, etcetera.

[1] Natural Language Processing and Information Retrieval Group at the Spanish Distance Learning University. http://nlp.uned.es/

[2] Pattern Analysis, Statistical Modelling and Computational Learning Network of Excellence. http://www.pascal-network.org/

2.2 Hyponymy and WordNet Entailment

Hyponymy and *entailment* are relations between *WordNet synsets* having a transitive property. The entailment predicate between two *synsets* was implemented according to these relations as the searching of a path from *synset* S_T to *synset* S_H, in which *hyponymy* and *WordNet entailment* relations between intermediate *synsets* are considered in the direction from S_T to S_H. Then, the lexical unit *T* entails the lexical unit *H* if there is a path from one *synset* of *T* to one *synset* of *H*. Some examples after processing the training corpus of PASCAL Challenge are: *glucose* entails *sugar* (i.e. *glucose* is a hyponym of *sugar*), *crude* entails *oil*, *death* entails *kill*.

2.3 Multiwords

The recognition of multiwords cannot be a previous to lemmatization and parsing step, so a pre and a post processing must be performed in order to avoid errors in the processing. For example, the recognition of the multiword *came_down* requires the previous obtention of the lemma *come*, because the multiword present in *WordNet* is *come_down*.

The variation of multiwords is not due only to lemmatization. Sometimes there are some characters that change as, for example, a dot in an acronym or a proper noun with different wordings. For this reason, a fuzzy matching between candidate and *WordNet* multiwords was implemented using the Levenshtein's edit distance (1965). If the two strings differ in less than 10%, then the matching is permitted. For example, the multiword *Japanise_capital* in hypothesis 345 of the training corpus is translated into the *WordNet* multiword *Japanese_capital*, allowing the entailment between Tokyo and it. Some other examples of entailment after multiword recognition are, regarding synonymy, *blood_glucose* and *blood_sugar*, *Hamas* and *Islamic_Resistance_Movement*, *Armed_Islamic_Group* and *GIA* and, regarding hyponymy, *war_crime* entails *crime*, *melanoma* entails *skin_cancer*.

2.4 Negation and Antonymy

Negation is detected after searching leaves with a negation relation in the dependency tree. This negation relation is then propagated to its ancestors until the head. For example, Figures 1 and 2 show an excerpt of the dependency trees for the training examples 74 and 78 respectivelly. Negation at node 11 of text 74 is propagated to node 10 (*neg(will)*) and node 12 (*neg(change)*). Negation at node 6 of text 78 is propagated to node 5 (*neg(be)*).

Entailment is not possible between a lexical unit and its negation. For example, before considering negation, node 5 in text 78 (*be*) entails node 4 in hypothesis 78 (*be*). Now, this entailment is not possible.

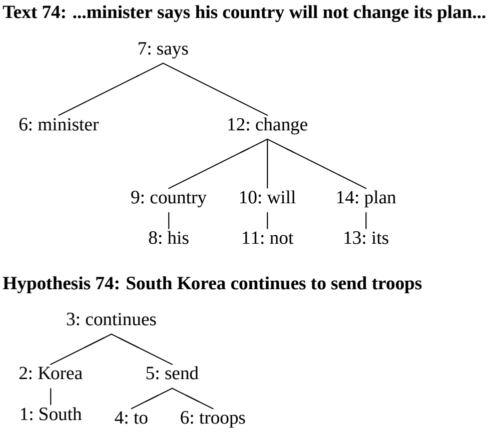

Figure 1: Dependency trees for pair 74 from training corpus.

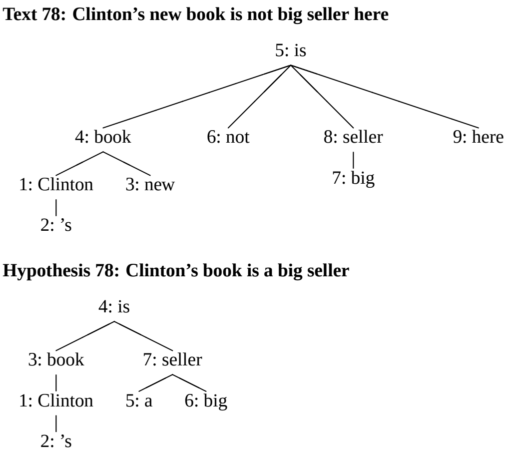

Figure 2: Dependency trees for pair 78 from training corpus.

The entailment between nodes affected by negation is implemented considering the antonymy relation of *WordNet*, and applying the previous processing to them (sections 2.1, 2.2, 2.3). For example, since node 12 in text 74 is negated (*neg(change)*), the antonyms of *change* are considered in the entailment relations between text and hypothesis. Thus, *neg(change)* in text entails *continue* in the hypothesis because the antonym of *change*, *stay*, is a synonym of *continue*.

3 Matching between Dependency Trees

Dependency trees give a structured representation for every text and hypothesis. Matching between dependency trees can give an idea about how semantically similar are two text snippets; this is because a certain semantic information is implicitly contained into dependency trees. The technique used to evaluate matching between dependency trees is inspired in Lin's proposal (Lin, 2001). The initial idea was to use a very simple matching algorithm, focused on searching for all the branches starting at any leaf from hypothesis' tree and showing a matching with any branch from text's tree. Hence, a hypothesis' matching branch is defined as the one whose all nodes show a lexical entailment with nodes from a branch of the correspondent text.

The existence or not of an entailment relation from a text to its correspondent hypothesis was determined by means of their similarity. Similarity between text and hypothesis is defined as the proportion of hypothesis' nodes pertaining to matching branches. From the results obtained against the training corpus, it was empirically determined a threshold for that similarity value. Best accuracy for the system was obtained when 50% was assigned as threshold value. Hence, it was said that a text entailed a hypothesis if hypothesis' dependency tree showed a percentage of matching nodes greater or equal than 50%. If that percentage was less than 50% it was said that no entailment existed from text to hypothesis.

4 Experiments

Along the development time of the proposed system some experiments were accomplished in order to obtain feedback about succesive improvements made to it. For this purpose, several baseline systems – whose results against the training corpus were compared – were developed.

4.1 Baselines

Two different baselines were generated in order to analyse the behaviour of the proposed system against the training corpus. Since lexical entailment is previous to matching between dependency trees, two more simple systems were developed to obtain the mentioned baselines:

- Baseline system I calculated the ratio of words from the hypothesis which appeared into the text.
- Baseline system II computed the ratio of lemmas from the hypothesis which are entailed by any lemma from the text.

In all cases the classification threshold was 50%, as explained in section 3.

4.2 Results over the Training Corpus

The proposed system and the baselines show similar results. Accuracy, calculated for every type of application setting, ranges between 46.67% and 55.56%, except for comparable documents (CD), showing 76.29%, 71.13% and 80.41% accuracy for baseline system I, baseline system II and proposed system, respectively. The overall results are 54.95%, 55.48% and 56.36% accuracy for baseline system I, baseline system II and proposed system, respectively.

4.3 Official Results at the Challenge

Since up to two runs were admitted for submission, it was decided to prepare a third baseline to compare the system against the test corpus. For this baseline system III, queries to *WordNet* were not used but only coincidence between lemmas from text and hypothesis. Hence, one of the submitted runs was generated by this latter baseline system.

The proposed system was refined for its run against the test corpus. This last implementation searched for *subject* or *object* relations along hypothesis' matching branches, requiring also a matching between these relations.

Accuracy, calculated for every type of application setting, ranges between 42.55% and 55.83%, except for CD, showing 79.33% and 78.67% accuracy for baseline system III and proposed system, respectively. The overall results are 55.75% and 54.75%

accuracy for baseline system III and proposed system, respectively.

The behaviour of both systems is similar to the ones executed against the training corpus. However, consideration of *subject* and *object* relations cause a slight decrease of accuracy.

5 Analysis and Conclusions

Results show that a matching-based approach (as shown here) is not enough to tackle appropriately the problem except, perhaps, for CD tasks.

The analysis of cases shows that a high lexical overlap does not mean a semantic entailment and a low lexical overlap does not mean different semantics. Both lexical and syntactic issues to be improved have been detected.

Some kind of paraphrasing between n-grams would be useful; for example, in pair 96[3] of the training corpus is necessary to detect the equivalence between *same-sex* and *gay* or *lesbian*; or, in pair 128[4], *come into conflict with* and *attacks* must be detected as equivalent. Previous work has been developed; for example, Szpektor et al. (2004) propose a web-based method to acquire entailment relations; Barzilay and Lee (2003) use multiple-sentence alignment to learn paraphrases in an unsupervised way; Hermjakob et al. (2002) show how *WordNet* can be extended as a reformulation resource; Pang et al. (2003) represent paraphrases as word lattices; Tomuro (2003) studies the case of question paraphrases.

Other problem is that, in certain cases, a high matching between hypothesis' nodes and text's nodes is given but, simultaneously, hypothesis' branches match with disperse text's branches; then, syntactic relations between subestructures of the text and the hypothesis must be analyzed in order to determine the existence of an entailment. This fact suggests to accomplish an in-depth treatment of syntactic relations.

[3] *Text 96:* The Massachusetts Supreme Judicial Court has cleared the way for lesbian and gay couples in the state to marry, ruling that government attorneys "failed to identify any constitutionally adequate reason" to deny them the right.
Hypothesis 96: U.S. Supreme Court in favor of same-sex marriage

[4] *Text 128:* Hippos do come into conflict with people quite often.
Hypothesis 128: Hippopotamus attacks human.

Hence, it is observed that for RTE is necessary to tackle a wide set of linguistic phenomena in a specific way, at the lexical level and at the syntactic level.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Science and Technology within the following project: TIC-2003-07158-C04-02 Multilingual Answer Retrieval Systems and Evaluation, SyEMBRA.

References

Regina Barzilay and Lillian Lee. 2003. *Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment*. NAACL-HLT-2003.

U. Hermjakob, A. Echihabi, and D. Marcu. 2002. *Natural Language Based Reformulation Resource and Web Exploitation for Question Answering*. Proceedings of TREC-2002.

V. I. Levenshtein. 1965. *Binary Codes Capable of Correcting Deletions, Insertions, and Reversals*. Soviet Physics - Doklady, Vol. 10 No. 8 pp. 707-710, February 1966. Translated from Doklady Akademii Nauk SSSR, Vol. 163 No. 4 pp. 845-848, August 1965.

Dekang Lin. 1998. *Dependency-based Evaluation of MINIPAR*. Workshop on the Evaluation of Parsing Systems, Granada, Spain, May, 1998.

Dekang Lin and Patrick Pantel. 2001. *Discovery of Inference Rules for Question Answering*. Natural Language Engineering 7(4):343-360.

B. Pang, K. Knight, and D. Marcu. 2003. *Sintax-based Alignment of Multiple Translations: Extracting Paraphrases and Generating New Sentences*. NAACL-HLT-2003.

I. Szpektor, H. Tanev, I. Dagan, and B. Coppola. 2004. *Scaling Web-based Acquisition of Entailment Relations*. Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics, Barcelona, 2004.

N. Tomuro. 2003. *Interrogative Reformulation Patterns and Acquisition of Question Answering Paraphrases.*. Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003).