Using Hypernymy Acquisition to Tackle (Part of) Textual Entailment

Elena Akhmatova
Centre for Language Technology
Macquarie University
Sydney, Australia
`elena@ics.mq.edu.au`

Mark Dras
Centre for Language Technology
Macquarie University
Sydney, Australia
`madras@ics.mq.edu.au`

Abstract

Within the task of Recognizing Textual Entailment, various existing work has proposed the idea that tackling specific subtypes of entailment could be more productive than taking a generic approach to entailment. In this paper we look at one such subtype, where the entailment involves hypernymy relations, often found in Question Answering tasks. We investigate current work on hypernymy acquisition, and show that adapting one such approach leads to a marked improvement in entailment classification accuracy.

1 Introduction

The goal of the Recognizing Textual Entailment (RTE) task (Dagan et al., 2006) is, given a pair of sentences, to determine whether a Hypothesis sentence can be inferred from a Text sentence. The majority of work in RTE is focused on finding a generic solution to the task. That is, creating a system that uses the same algorithm to return a *yes* or *no* answer for all textual entailment pairs. A generic approach never works well for every single entailment pair: there are entailment pairs that are recognized poorly by all the generic systems.

Some approaches consequently propose a component-based model. In this framework, a generic system would have additional special components that take care of special subclasses of entailment pairs. Such a component is involved when a pair of its subclass is recognized. Vanderwende and Dolan (2005), and subsequently Vanderwende et al. (2006), divide all the entailment pairs according to whether categorization could be accurately predicted based solely on syntactic cues. Related to this, Akhmatova and Dras (2007) present an entailment type where the relationship expressed in the Hypothesis is encoded in a syntactic construction in the Text.

Vanderwende et al. (2006) note that what they term *is-a* relationships are a particular problem in their approach. Observing that this encompasses hypernymy relations, and that there has been a fair amount of recent work on hypernymy acquisition, where ontologies containing hypernymy relations are extended with corpus-derived additions, we propose a HYPERNYMY ENTAILMENT TYPE to look at in this paper. In this type, the Hypothesis states a hypernymy relationship between elements of the Text: for example, *This was seen as a betrayal by the EZLN and other political groups* implies that *EZLN is a political group*. This subtype is of particular relevance to Question Answering (QA): in the RTE-2 dataset,[1] for example, all *is-a* Hypotheses were drawn from QA data.

In this paper we take the hypernymy acquisition work of Snow et al. (2005) as a starting point, and then investigate how to adapt it to an entailment context. We see this as an investigation of a more general approach, where work in a separate area of NLP can be adapted to define a related entailment subclass.

Section 2 of the paper discusses the relevant work from the areas of component-based RTE and hypernymy extraction. Section 3 defines the hypernymy entailment type and expands on the main idea of the paper. Section 4 describes the experimental set-up and the results; and Section 5 concludes the work.

2 Related Work

2.1 Component-based RTE

Vanderwende et al. (2006) use an approach based on logical forms, which they generate by the NLPwin parser. Nodes in the resulting syntactic dependency graphs for Text and Hypothesis are then heuristically aligned; then syntax-based heuristics

---

[1] http://pascallin.ecs.soton.ac.uk/Challenges/RTE2, (Bar-Haim et al., 2006)

are applied to detect false entailments. As noted above, *is-a* relations fared particularly badly. In our approach, we do not use such a heavy duty representation for the task, using instead the techniques of hypernym acquisition described in Section 2.2. Cabrio et al. (2008) proposed what they call a combined specialized entailment engine. They have created a general framework, based on distance between T and H (they measure the cost of the editing operations such as insertion, deletion and substitution, which are required to transform the text T into the hypothesis H) and several modular entailment engines, each of which is able to deal with an aspect of language variability such as negation or modal verbs. Akhmatova and Dras (2007) built a specific component from a subset of entailment pairs that are poorly recognized by generic systems participating in an RTE Challenge. These are the entailment pairs where a specific syntactic construction in the Text encodes a semantic relationship between its elements that is explicitly shown in the Hypothesis, as in example (1):

(1) *Text:* Japan's Kyodo news agency said the US could be ready to set up a liaison office—the lowest level of diplomatic representation—in Pyongyang if it abandons its nuclear program.
Hypothesis: Kyodo news agency is based in Japan.

The entailment pairs share a set of similar features: they have a very high word overlap regardless of being a *true* or *false* entailments, for example. High word overlap is one of the features for an RTE system for the majority of the entailment pair types, which presumably hints at *true*, but this is not useful in our case. Akhmatova and Dras (2007) described a two-fold probabilistic approach to recognizing entailment, that in its turn was based on the well-known noisy channel model from Statistical Machine Translation (Brown et al., 1990). In the work of this paper, by contrast, we look at only identifying a hypernymy-related Text, so the problem reduces to one of classification over the Text.

2.2 Hypernymy Extraction

The aim of work on hypernymy extraction is usually the enrichment of a lexical resource such as WordNet, or creation of specific hierarchical lexical data directly for the purpose of some appli-

cation, such as information extraction or question answering. There can be found several approaches to the task of hypernymy extraction: co-occurrence approaches, asymmetric association measures, and pattern-based methods.

Cooccurence Approaches Co-occurrence approaches first cluster words into similarity classes and consider the elements of a class to be siblings of one parent. Therefore the search for a parent for some members from the class gives a parent for the other members of the class. The first work that introduced co-occurrence methods to the field is that of Caraballo (1999). First she clusters nouns into groups based on conjunctive and appositive data collected from the Wall Street Journal. Nouns are grouped according to the similarity of being seen with other nouns in conjunctive and appositive relationships. In the second stage, using some knowledge about which conjuncts connect hypernyms reliably, a parent for a group of nouns is searched for in the same text corpora. Other co-occurrence methods can be found in works by Pantel et al. (2004) and Pantel and Ravichandran (2004).

Asymmetric Association Measures In Asymmetric Association (see Dias et al. (2008)) hypernymy is derived through the measure of how much one word 'attracts' another one. When hearing "fruit", more common fruits will be likely to come into mind such as "apple" or "banana". In this case, there exists an oriented association between "fruit" and "mango" (mango → fruit) which indicates that "mango" attracts "fruit" moreso than "fruit" attracts "mango". As a consequence, "fruit" is more likely to be a more general term than "mango".

Pattern-based Methods Pattern-based methods are based on the observation that hypernyms tend to be connected in the sentences by specific words or patterns, and that some patterns can predict hypernymy with very high probability, like the *X and other Y* pattern. Generally, some amount of manual work on finding the seed patterns is done first. Automated algorithms use these patterns for discovering more patterns and for the subsequent hypernymy extraction. The fundamental work for the pattern-based approaches is that of Hearst (1992). More recently, Snow et al. (2005) and Snow et al. (2006) have described a method of hypernymy extraction using machine learning of

patterns. Pattern-based methods are known to be successfully used for the creation of hierarchical data for other languages as well, such as Dutch; for example, see Tjong Kim Sang and Hofmann (2007). For our purposes, pattern-based methods are particularly suitable, as we have as context two words and a single pattern connecting them; we thus describe these approaches in more detail.

In her early work on pattern-based hypernymy extraction Hearst (1992) noticed that a particular semantic relationship between two nouns in the sentence can be indicated by the presence of certain lexico-syntactic patterns linking those nouns. Hypernymy (*is-a*, *is a kind of* relation) is one such relationship.

Linking two noun phrases via the patterns *such NP_y as NP_x* often implies that NP_x is a hyponym of NP_y, that is NP_x *is a kind of NP_y*. She gives the following example to illustrate the patterns

(2) The bow lute, such as the Bambara ndang, is plucked and has an individual curved neck for each string.

Hearst comments that most fluent readers of English who have never before encountered the term *Bambara ndang* will nevertheless from this sentence infer that a *Bambara ndang* is a kind of *bow lute*. This is true even if the reader has only a fuzzy conception of what a bow lute is. The complete set of patterns semi-automatically found by Hearst are:

1. NP_y *and other* NP_x
2. NP_y *or other* NP_x
3. NP_y *such as* NP_x
4. *such* NP_y *as* NP_x
5. NP_y *including* NP_x
6. NP_y, *especially* NP_x

Snow et al. (2005) had the aim of building upon Hearst's work in order to extend the WordNet semantic taxonomy by adding to it hypernym-hyponym pairs of nouns that are connected by a wider set of lexico-syntactic pairs. They developed an automatic approach for finding hypernym-hyponym pairs of nouns in the text corpus without a set of predefined patterns.

The work was carried out on a corpus of 6 million newswire sentences. Every pair of nouns (n_i, n_j) in the sentence was extracted. The pairs were labelled as Known Hypernym pair if n_j is

an ancestor of the first sense of n_i in the WordNet hypernym taxonomy (Fellbaum, 1998). A noun pair might have been assigned to the second set of Known Non-Hypernym pairs if both nouns are contained within WordNet, but neither noun is an ancestor of the other in the WordNet hypernym taxonomy for any senses of either noun. Each sentence was parsed using MINIPAR. The dependency relations between n_i and n_j constituted the lexico-syntactic patterns connecting Known Hypernyms or Known Non-Hypernyms. The main idea of their work was then to collect all the lexico-syntactic patterns that may indicate the hypernymy relation and use them as the features for a decision tree to classify NP pairs as hypernym-hyponym or not-hypernym-hyponym pairs.

Snow et al. (2005) state in their work that the dependency paths acquired automatically contained all the patterns mentioned in Hearst (1992). The comparison of the results of a classifier whose vectors were created from all the patterns seen with the Known Hypernyms in their corpus, and a classifier whose vectors contained only the patterns of Hearst (1992), showed that the results of the former classifier are considerably better than that of the latter one. In an RTE context where the entailment recognition relies on recognising hypernymy, an approach like this, where patterns acquired from a corpus are used, could be useful; but how it should best be adapted is not clear. That is then the goal of this paper.

3 Hypernymy Entailment Type

3.1 Definition

We define *Hypernymy Entailment* to be an entailment relationship where the *is-a* relationship between two nouns in the hypothesis is 'hidden behind' the lexico-syntactic pattern connecting them in the text. Being more precise, the Text-Hypothesis pairs of interest have the following characteristics:

1. The Hypothesis is a simple sentence. That is a sentence that consists of a subject, a 3rd person form the verb *to be*, and a direct object, and that contains no subordinate clauses.

2. Both subject and object of the Hypothesis (or in some cases their morphological variants) are found in the text.

Thus, the hypernymy relationship is not stated in the Text, but is hidden in the way the subject and

object of the Hypothesis are connected to each other in the Text. Examples of the *true* hypernymy entailment pairs are as follows:[2]

(3) *Text:* Soon after the EZLN had returned to Chiapas, Congress approved a different version of the COCOPA Law, which did not include the autonomy clauses, claiming they were in contradiction with some constitutional rights (private property and secret voting); this was seen as a betrayal by the EZLN and other political groups.
Hypothesis: EZLN is a political group.

Both *EZLN* and *political groups* are present in the text sentence, and are connected by an *is-a* relation in the hypothesis. The pattern *and other* and the syntactical connection between the noun phrases give a good indication that the noun phrases are in the hypernym-hyponym relationship. An example of a *false* hypernymy entailment pair is as follows:

(4) *Text:* Laboring side by side on the outer hull of the station's crew quarters, Vladimir Dezhurov and Mikhail Turin mounted science packages and two Eastman Kodak Co. placards while U.S. astronaut Frank Culbertson looked on from inside the complex.
Hypothesis: Vladimir Dezhurov is a U.S. astronaut.

3.2   Idea

In the case of Snow et al. (2005) the main accent is on automatic extraction of all the patterns that might, even if not reliably on their own, predict the hypernymy relation between two nouns. Their task is, given a previously unseen pair of nouns, to determine whether they are in a hypernymy relationship, using a classifier whose feature values are derived from many occurrences of acquired patterns in a corpus.

In our own work we are put in the situation where there is only one pattern that is available to judge if two words are in a hypernym/hyponym relation, not the whole text corpus as in the case of Snow et al. (2005). Thus, we are mostly interested in the prediction of the hypernymy using this pattern that is available for us. The fact that the named entities we are working with, such as person, organization, location, are not that frequently

seen in any text corpora also shifts the accent onto the pattern rather than on the word pair itself. As well as the fact that even in the case when two words are hypernym-hyponym, that may not follow at all from the sentence that they are seen in; and non hypernym-hyponym pair can be used as such in a metaphoric expression or just in a particular sentence we are dealing with. To illustrate, consider example (5):

(5) *Text:* Note that the auxiliary verb function derives from the copular function; and, depending on one's point of view, one can still interpret the verb *as a* copula and the following verbal form as being adjectival.
Hypothesis: A copular is a verb.

Snow et al. (2005) aim to determine whether *copular* and *verb* are in a hypernymy relation; to this end they use the *as a* pattern as in this example, along with all others throughout the corpus. The reliability of the *as a* pattern (which as it turns out is quite high) adds weight to the accumulated evidence, but is not the sole evidence. In the individual case, however, it can be incorrect, as in example (6):

(6) *Text:* In the 1980s, Minneapolis took its place *as a* center of the arts, with the Walker Arts Center leading the nation in appreciation of pop and postmodern art , and a diverse range of musicians, from Prince to Hüsker Dü to the Replacements to the Suburbs to Soul Asylum keeping up with the nation in musical innovation.
Hypothesis: A centre is a place.

Example (6) has a similar structure to example (5), but *center* governs a preposition *of* after it, that seem to make the hypernymy more doubtful in this context. Taking into account all of the above, the major focus of the work has shifted for us from the word pair to the environment it has occurred in. Thus, we use the major ideas from the work of Snow et al. (2005), but as we show below, it is necessary to develop a more complex set of counts in order to apply this to our entailments type. In particular, we expect that the division of patterns into lexical and syntactic parts, in order to score them separately, is beneficial for entailment. Again, it is a result of scarcity of information: we have only one text sentence, not the whole text corpus to make the entailment decision.

---

[2]Examples (3) - (4) are taken from the RTE2 test corpus.

4 Experimental Setup

4.1 Data

Our goal is to build a classifier that will detect whether a given potential hypernymy entailment pair is true or false; we first need to construct sets of such pairs for training and testing. As our basic data source, we use 500 000 sentences from the Wikipedia XML corpus (Denoyer and Gallinari, 2006); this is the corpus used by Akhmatova and Dras (2007), and related to one used in one set of experiments by Snow et al. (2005). These sentences were parsed with the MINIPAR parser.

We identified Known Hypernym pairs as did Snow et al. (2005) (see Section 2.2); of our basic corpus, 13310 sentences contained Known Hypernyms. From these sentences we extracted the dependency relations between the Known Hypernyms, of which there were 166 different types; we refer to these as syntactic patterns hereafter.

We reserved 259 of these sentences to construct a test set for our approach, as described below. These sentences were selected randomly in proportion to the syntactic patterns occurring in the overall set. The remaining sentences constituted our SYNTACTIC PATTERN TRAINING SET. For the test set, these sentences constituted the Texts; to derive the Hypotheses, we extracted the Known Hypernyms and connected them by *is a*. These sentences were annotated with *yes* if they entail hypernymy, and *no* otherwise; the resulting annotated data has 2:1 ratio of *no* to *yes*. The main annotation was carried out by the first author, with the second author carrying out a separate annotation to evaluate agreement. The number of items where there was agreement was 206, giving a κ of 0.54. This is broadly in line with the κ found in construction of the RTE datasets ($\kappa = 0.6$) (Glickman, 2006) where it is characterized as "moderate agreement", based on Landis and Koch (1977). Results later are presented for both the overall set of 259 (based on the first author's original annotations) and for the subset with agreement of 206.

As our additional, much larger data source for deriving purely lexical patterns and associated scores, we use the Web1T n-gram corpus (Brants and Franz, 2006), which provides n-grams and their counts for up to 5-grams inclusive. We use these n-grams to get the lexical patterns of length 1, 2 and 3 that connect Known Hypernyms and Known Non-Hypernyms correspondingly. The length is up to 3 as we need 2 slots for the nouns

from the pair itself. The counts are extracted with the help of the software *get1t* written by Hawker et al. (2007). We refer to this as our LEXICAL PATTERN TRAINING SET.

4.2 Baselines

We use two baselines. The first is a simple most-frequent one, choosing always false (noting from Section 4.1 that this is more common by a ratio of approximately 2:1). For the second one, we attempt to use the idea of Snow et al. (2005) in a straightforward way. We note again that the fixed context for a given Known Hypernym pair that we have, unlike Snow et al. (2005), is the single Text; we therefore cannot apply the classifier from that work directly. Our second baseline based on their approach is as follows. For each sentence we look at all nouns it contains. If a pair of nouns from the sentence is a Known-Hypernym pair we save the lexical pattern connecting the nouns and the syntactic pattern between the nouns in a pattern list. We take into account only those syntactic patterns that have been seen in the corpus at least three times. We then consider that a test entailment pair is a true entailment if both the lexical pattern between the nouns in question and the syntactic connection between them is found in the list.

4.3 Two-Part Model

We now propose a two-component model to compensate for the fixed context. The first component, $score_{lex}$, involves the use of the lexical pattern to predict hypernymy. Unless we know something else about the structure of the text sentence, the pattern (a sequence of words) that connects two entities in question is the only evidence of the possible hypernym-hyponym relation between them. It does not guarantee the relation itself, but the more probable it is that the pattern predicts hypernymy, the more probable it is that the entailment relation between the Text and Hypothesis holds. To motivate the second component, we take as an example the pattern NP_y and other NP_x, the first of the Hearst (1992) patterns and a good predictor of hypernymy, and consider the following examples:

(7) *Text:* Mr. Smith and other employees stayed in the office.
Hypothesis: Mr. Smith is an employee.

(8) *Text:* I talked to Mr. Smith and other

employees stayed in the office.
Hypothesis: Mr. Smith is an employee.

Mr. Smith and *an employee* are connected in both cases by *and other*. We know that the pattern *and other* is a good indicator of the hypernymy relation. The probability of the pattern *and other* to predict the hypernymy relation is the prior probability of the entailment relation in a text-hypothesis pair. As can be seen in examples (7) and (8), there is an entailment relationship only in example (7); in example (8) entailment does not hold.

The second component $score_{synt}$ is an indicator of the syntactic possibility of the entailment relationship. Hypernym-hyponyms tend to be in certain syntactic relations in the sentence, such as being subjects of the same verb, for example, in the cases where we can decide on the relation of the hypernymy between them. Other syntactic relationships, even though they may connect hypernym and hyponym, do not allow us to conclude that there is a hypernymy relation between the words. As it can be seen from examples (7) and (8), every syntactical relation has its own level of certainty about the hypernym relation between *Mr. Smith* and *an employee*, and therefore about the fact that the Text entails the Hypothesis.

4.3.1 Lexical Patterns

From our lexical pattern training corpus, we derived for both Known Hypernym and Known Non-Hypernym pairs, the counts of both tokens (total number of pairs connected) and types (number of different pairs connected). To illustrate, we take two example pairs, $w_1 = rock$ and $w_2 = material$, and $w_1 = rice$ and $w_2 = grain$. We find *rock , and other material* occurs 47 times, and *rice , and other grain* 166 times. Totalling these, that would give us the following statistics for the pattern *, and other*: seen with the Known Hypernyms 213 times (total of tokens), connecting 2 different pairs (total of types). We hypothesize that knowing the number of different types of patterns will be important as a way of compensating for the more limited context relative to Snow et al. (2005) which used only the number of pattern tokens.

The above can be illustrated by the counts obtained for patterns of Hearst (1992); see the first five rows of Table 1. One can see from the first three examples that in all cases the number of times the pattern has been seen with Known Hypernyms is overwhelmingly higher than with that of Known Non-Hypernyms. Even more extremely, in the next two examples in Table 1, Known Non-Hypernyms were not seen with these patterns at all. We contrast these with the non-Hearst patterns (extracted from our lexical pattern corpus) in the last two rows. As one can see, the patterns *and detailed travel* and *online game caribbean* have been seen only with the Known Hypernyms, and the frequency counts are very close to that of the pattern *, especially*. Both patterns however have connected the constituents of only one Known Hypernyms pair. That puts some doubt on the general reliability of the pattern to make hypernymy judgements.

We then define our scoring metric, based on the following quantities: $C(h\text{-}tok)$, the number of times the pattern has been seen with Known Hypernyms; $C(nh\text{-}tok)$, the number of times the pattern has been seen with Known Non-Hypernyms; $C(h\text{-}type)$, the number of times the pattern has been seen with different Known Hypernym patterns; $C(nh\text{-}type)$, the number of times the pattern has been seen with different Known Non-Hypernym patterns. We then define our lexical scoring function as follows:

$$score_{lex} = \frac{C(h\text{-}tok)}{C(h\text{-}tok) + C(nh\text{-}tok)} \times$$
$$\frac{C(h\text{-}type)}{C(h\text{-}type) + C(nh\text{-}type)}$$

We use it to score patterns where the number of times the pattern has been seen with different Known Hypernyms ($C(h\text{-}type)$) is greater than a threshold, here 5; for patterns below this threshold, the score is 0. We determined on this scoring function in comparison to others (notably using only token proportions, the first term in the scoring function above) by using them to rank patterns and then assess the relative ranking of the Hearst patterns among all others. Under the scoring function above, the Hearst patterns were ranked highest, with patterns *or other, such as* and *and other* taking the first, second and third positions respectively.

4.3.2 Syntactic Patterns

To estimate the probability of various syntactic patterns from our syntactic pattern training corpus, ideally we would annotate every sentence as

Table 1: Counts for the patterns of Hearst (1992) obtained from *the Web1T corpus*

Pattern	seen with			
	Hypernyms	Non-Hypernyms	Different Hypernyms	Different Non-Hypernyms
NP_y and other NP_x	172036	1716	486	3
NP_y or other NP_x	421083	1016	965	11
NP_y such as NP_x	86158	384	355	4
NP_y including NP_x	68098	0	251	0
NP_y, especially NP_y	10236	0	80	0
NP_y and detailed travel NP_x	9870	0	1	0
NP_y online game caribbean NP_x	9874	0	1	0

true or *false* according to whether the hypernymy is entailed from the sentence or not. The annotation would allow the calculation of the likelihood for every syntactical relation to indicate the entailment relationship.

It is quite a time-consuming task to annotate enough data to get reliable counts for all the syntactical patterns. Therefore, as an approximate first step we have divided all the sentences into three groups according to the type of a lexical patterns that connects a pair of Known Hypernyms: Hearst patterns; the patterns that were found from our lexical pattern training corpus; and all other patterns. We have assumed that Hearst patterns, as being a good indication of hypernymy, may in most cases predict entailment as well; the automatically derived lexical patterns may still sometime predict entailment, but less well than the Hearst patterns; and the unknown patterns are not considered to be good predictors of the entailment at all. Thus, for the initial estimate of the syntactical probabilities of the entailment we have employed a very coarse approximation of the maximum likelihood estimate of the probability of a syntactic pattern implying an entailment, weighting these three groups with the values 1, 0.5 and 0 respectively. This leads to a score as follows:

$$score_{synt\text{-}basic} = 0.5 \times \frac{C(automatic\ lexical\ pattern)}{C(all\ patterns)}$$
$$+ 1.0 \times \frac{C(Hearst\ pattern)}{C(all\ patterns)}$$

where $C(X)$ represents the count of occurrences of the pattern type X.

As a more refined scoring metric, we identified the set of the most frequent syntactic patterns

Table 2: Syntactic Pattern Probabilities

Pattern	Basic P	Improved P
obj	0.34	0.0
pcomp-n_ mod	0.40	0.038
appo	0.73	0.90
conj	0.76	0.10
mod_ pcomp-n	0.64	0.38
mod_ pcomp-n_ mod	0.45	0.023
mod_ conj	0.97	0.10

Table 3: Model Evaluation (full set of 259 / agreed subset of 206)

Model	Accuracy
Baseline (most frequent)	69% / 70%
Baseline (Snow)	71% / 72%
Lexical component only	60% / 60%
Improved syntactic component only	67% / 69%
Lexical and Basic Syntactic Component	76% / 73%
Lexical and Improved Syntactic Component	82% / 83%

and annotated data for them, in order to improve their probability estimates. Taking the seven most frequent, we annotated 100 randomly chosen sentences for each of the syntactical patterns containing them from the syntactic pattern training corpus. As a result of the annotation the probabilities of the syntactical patterns to indicate entailment has changed. The basic probabilities and the revised probabilities for these seven syntactic patterns can be found in Table 2.

4.4 Results and Discussion

We combine the lexical and syntactic scores as features to the J48 decision tree of WEKA (Wit-

ten and Frank, 1999). Our evaluation is a 10-fold cross-validation on the test set. Results are as in Table 3, presented for both the full test set of 259 and for the subset with agreement of 206.

We note first of all that the simple approach derived from Snow et al. (2005), as described in Section 4.2, does perform marginally better than the baseline of choosing always false. The lexical or syntactic components alone do not perform better than the most-frequent baseline approach. This is expected, as that approach includes both lexical and syntactic components. The lexical combined with the basic syntactic component does improve over the baselines. However, the lexical combined with the improved syntactic component experiences a much higher improvement. Overall, the results for the full set and for the subset are broadly the same, showing the same relative behaviour.

The lexical only component falsely recognizes examples such as example (9) as true, as it has no support of syntax. Just a comma by itself sufficiently frequently indicates entailment in case of apposition, so the lexical component is misled.

(9) *Text:* There were occasional outbreaks of violence, but most observers considered it remarkable that such an obvious breakdown of the capitalist system had not led to a rapid growth of socialism, communism, or fascism (as happened for example in Germany). *Hypothesis:* Communism is a socialism.

Syntax only, even though it prevents the mistakes of the lexical-only component for the examples above, introduces its own mistakes. Knowing that the subject and object in the Hypothesis are linked by direct dependency relations to a preposition in the Text is useful, but without a lexical pattern can be too permissive, as in example (10):

(10) *Text:* However, Griffin attracted criticism for writing in the aftermath of the bombing of the Admiral Duncan pub bombing (which killed three people, including a pregnant woman) that the gay people protesting against the murders were "flaunting their perversion in front of the world's journalists, and showed just why so many ordinary people find these creatures disgusting". *Hypothesis:* Criticism is a writing.

Both baseline and the final hypernymy entailment engine work well in the cases where the counts for or against entailment are very high, as in examples (11) and (12), which are correctly recognized as a true and a false entailment by both systems.

(11) *Text:* Carbon compounds form the basis of all life on Earth and the carbon-nitrogen cycle provides some of the energy produced by the sun and other stars. *Hypothesis:* Sun is a star.

(12) *Text:* In 1792 British explorer George Vancouver set up a small settlement near the village of Yerba Buena (later downtown San Francisco) which became a small base for English, Russian, and other European fur traders, explorers, and settlers. *Hypothesis:* Village is a settlement.

The final hypernymy system works better for more marginal cases, such as example (13).

(13) *Text:* The trials were held in the German city of Nuremberg from 1945 to 1949 at the Nuremberg Palace of Justice. *Hypothesis:* Nuremberg is a city.

The pattern *of* can not be called a good hint for hypernymy, but in some special cases, like that of the city and its name, the hypernymy is obvious. Division into lexical and syntactic parts helped in discovering the pattern and adjusting better its probability of entailing hypernymy. All this supports our idea that to compensate for the lack of information in the case of RTE the lexico-syntactic patterns should be divided into their lexical and syntactic components.

5 Conclusion

In this paper we have shown how work in hypernymy acquisition can be adapted to tackle a specific subtype of related entailment problem. Following work by Snow et al. (2005), we have defined an obvious first adaptation which nonetheless marginally improves over the baseline. We have then shown that by separating lexical and syntactic patterns we can obtain a significant improvement on the entailment classification accuracy. In our future work we aim to construct a baseline generic RTE engine and test its performance with and without this and other components in order to analyse the work of a component-based model as a whole. The approach also suggests that adapting work from other areas of NLP for entailment subclasses is promising.

References

Elena Akhmatova and Mark Dras. 2007. Entailment due to syntactically encoded semantic relationships. In *Proceedings of ALTA-2007*, pages 4–12.

Roy Bar-Haim, Ido Dagan, Bill Dolan, Lisa Ferro, Danilo Giampiccolo, Bernardo Magnini, and Idan Szpektor. 2006. The second pascal recognising textual entailment challenge. In *The second PASCAL Recognising Textual Entailment Challenge*, pages 3–11, Venice, Italy.

Thorsten Brants and Alex Franz. 2006. Web 1t 5-gram corpus version 1. Technical report, Google Research.

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. In *Computational Linguistics*, volume 16, pages 79 – 85.

Elena Cabrio, Milen Kouylekov, and Bernardo Magnini. 2008. Combining specialized entailment engines for RTE-4. In *Proceedings of TAC-2008*.

Sharon Caraballo. 1999. Automatic acquisition of a hypernym-labeled noun hierarchy from text. In *Proceedings of ACL-99*, pages 120–126.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The pascal recognising textual entailment challenge. In *Quionero-Candela, J.; Dagan, I.; Magnini, B.; d'Alch-Buc, F. (Eds.) Machine Learning Challenges. Lecture Notes in ComputerScience*, volume 3944, pages 177 – 190. Springer.

Ludovic Denoyer and Patrick Gallinari. 2006. The Wikipedia XML Corpus. In *SIGIR Forum, 40(1)*, pages 64–69.

Gaël Dias, Raycho Mukelov, and Guillaume Cleuziou. 2008. Unsupervised learning of general-specific noun relations from the web. In *Proceedings of FLAIRS Conference*, pages 147–152.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Oren Glickman. 2006. *Applied Textual Entailment*. Ph.D. thesis, Bar Ilan University.

Tobias Hawker, Mary Gardiner, and Andrew Bennetts. 2007. Practical queries of a massive n-gram database. In *Proceedings of ALTA-2007*, pages 40–48.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING*, pages 539–545.

Richard J. Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Patrick Pantel and Deepak Ravichandran. 2004. Automatically labeling semantic classes. In *Proceedings of HLT/NAACL-04*, pages 321–328.

Patrick Pantel, Deepak Ravichandran, and Eduard Hovy. 2004. Towards terascale semantic acquisition. In *Proceedings of Coling 2004*, pages 771–777, Geneva, Switzerland, Aug 23–Aug 27.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 1297–1304, Cambridge, MA. MIT Press.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2006. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of ACL-2006*, pages 801–808.

E.F. Tjong Kim Sang and K. Hofmann. 2007. Automatic extraction of dutch hypernym-hyponym pairs. In *Proceedings of CLIN-2006*, Leuven, Belgium. LOT, Netherlands Graduate School of Linguistics.

Lucy Vanderwende and William B. Dolan. 2005. What syntax can contribute in the entailment task. In *Proceedings of MLCW*, pages 205–216.

Lucy Vanderwende, Arul Menezes, and Rion Snow. 2006. Microsoft research at RTE-2: Syntactic contributions in the entailment task: an implementation. In *Proceedings of 2nd PASCAL Challenges Workshop on Recognizing Textual Entailment*.

Ian H. Witten and Eibe Frank. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.