

Discovering *asymmetric* entailment relations between verbs using selectional preferences

Fabio Massimo Zanzotto
DISCo

University of Milano-Bicocca

Via Bicocca degli Arcimboldi 8, Milano, Italy

zanzotto@disco.unimib.it

Marco Pennacchiotti, Maria Teresa Pazienza
ART Group - DISP

University of Rome “Tor Vergata”

Viale del Politecnico 1, Roma, Italy

{pennacchiotti, pazienza}@info.uniroma2.it

Abstract

In this paper we investigate a novel method to detect *asymmetric* entailment relations between verbs. Our starting point is the idea that some point-wise verb selectional preferences carry relevant semantic information. Experiments using WordNet as a gold standard show promising results. Where applicable, our method, used in combination with other approaches, significantly increases the performance of entailment detection. A combined approach including our model improves the AROC of 5% absolute points with respect to standard models.

1 Introduction

Natural Language Processing applications often need to rely on large amount of lexical semantic knowledge to achieve good performances. Asymmetric verb relations are part of it. Consider for example the question “*What college did Marcus Camby play for?*”. A question answering (QA) system could find the answer in the snippet “*Marcus Camby won for Massachusetts*” as the question verb *play* is related to the verb *win*. The vice-versa is not true. If the question is “*What college did Marcus Camby won for?*”, the snippet “*Marcus Camby played for Massachusetts*” cannot be used. *Winnig* entails *playing* but not vice-versa, as the relation between *win* and *play* is asymmetric.

Recently, many automatically built verb lexical-semantic resources have been proposed to support lexical inferences, such as (Resnik and Diab, 2000; Lin and Pantel, 2001; Glickman and Dagan, 2003). All these resources focus on *symmetric* semantic relations, such as *verb similarity*. Yet,

not enough attention has been paid so far to the study of *asymmetric* verb relations, that are often the only way to produce correct inferences, as the example above shows.

In this paper we propose a novel approach to identify asymmetric relations between verbs. The main idea is that asymmetric entailment relations between verbs can be analysed in the context of class-level and word-level selectional preferences (Resnik, 1993). Selectional preferences indicate an entailment relation between a verb and its arguments. For example, the selectional preference $\{human\} win$ may be read as a *smooth* constraint: *if x is the subject of win then it is likely that x is a $human$, i.e. $win(x) \rightarrow human(x)$* . It follows that selectional preferences like $\{player\} win$ may be read as suggesting the entailment relation $win(x) \rightarrow play(x)$.

Selectional preferences have been often used to infer semantic relations among verbs and to build symmetric semantic resources as in (Resnik and Diab, 2000; Lin and Pantel, 2001; Glickman and Dagan, 2003). However, in those cases these are exploited in a different way. The assumption is that verbs are semantically related if they share similar selectional preferences. Then, according to the Distributional Hypothesis (Harris, 1964), verbs occurring in similar sentences are likely to be semantically related.

The Distributional Hypothesis suggests a generic *equivalence* between words. Related methods can then only discover symmetric relations. These methods can incidentally find verb pairs as $(win, play)$ where an asymmetric entailment relation holds, but they cannot state the direction of entailment (e.g., $win \rightarrow play$).

As we investigate the idea that a single relevant verb selectional preference (as $\{player\}$

win) could produce an entailment relation between verbs, our starting point can not be the Distributional Hypothesis. Our assumption is that some *point-wise assertions* carry relevant semantic information (as in (Robison, 1970)). We do not derive a semantic relation between verbs by comparing their selectional preferences, but we use point-wise corpus-induced selectional preferences.

The rest of the paper is organised as follows. In Sec. 2 we discuss the intuition behind our research. In Sec. 3 we describe different types of verb entailment. In Sec. 4 we introduce our model for detecting entailment relations among verbs. In Sec. 5 we review related works that are used both for comparison and for building combined methods. Finally, in Sec. 6 we present the results of our experiments.

2 Selectional Preferences and Verb Entailment

Selectional restrictions are strictly related to entailment. When a verb or a noun *expects* a modifier having a predefined property it means that the truth value of the related sentences strongly depends on the satisfiability of these expectations. For example, “*X is blue*” implies the expectation that *X* has a colour. This expectation may be seen as a sort of entailment between “being a modifier of that verb or noun” and “having a property”. If the sentence is “*The number three is blue*”, then the sentence is false as the underlying entailment $blue(x) \rightarrow has_colour(x)$ does not hold (cf. (Resnik, 1993)). In particular, this rule applies to verb *logical* subjects: if a verb *v* has a selectional restriction requiring its logical subjects to satisfy a property *c*, it follows that the implication:

$$v(x) \rightarrow c(x)$$

should be verified for each logical subject *x* of the verb *v*. The implication can also be read as: if *x* has the property of doing the action *v* this implies that *x* has the property *c*. For example, if the verb is *to eat*, the selectional restrictions of *to eat* would imply that its subjects have the property of being *animate*.

Resnik (1993) introduced a smoothed version of selectional restrictions called *selectional preferences*. These preferences describe the *desired properties* a modifier should have. The claim is that if a selectional preference *holds*, it is more probable that *x* has the property *c* given that it

modifies *v* rather than *x* has this property in the general case, i.e.:

$$p(c(x)|v(x)) > p(c(x)) \quad (1)$$

The probabilistic setting of selectional preferences also suggests an entailment: the implication $v(x) \rightarrow c(x)$ holds with a given degree of certainty. This definition is strictly related to the probabilistic textual entailment setting in (Glickman et al., 2005).

We can use selectional preferences, intended as probabilistic entailment rules, to induce entailment relations among verbs. In our case, if a verb v_t expects that the subject “has the property of doing an action v_h ”, this may be used to induce that the verb v_t probably entails the verb v_h , i.e.:

$$v_t(x) \rightarrow v_h(x) \quad (2)$$

As for class-based selectional preference acquisition, corpora can be used to estimate these particular kinds of preferences. For example, the sentence “*John McEnroe won the match...*” contributes to probability estimation of the class-based selectional preference $win(x) \rightarrow human(x)$ (since *John McEnroe* is a *human*). In particular contexts, it contributes also to the induction of the entailment relation between *win* and *play*, as *John McEnroe* has the *property of playing*. However, as the example shows, classes relevant for acquiring selectional preferences (such as *human*) are explicit, as they do not depend from the context. On the contrary, properties such as “having the property of doing an action” are less explicit, as they depend more strongly on the context of sentences. Thus, properties useful to derive entailment relations among verbs are more difficult to find. For example, it is easier to derive that *John McEnroe* is a *human* (as it is a stable property) than that he has the property of *playing*. Indeed, this latter property may be relevant only in the context of the previous sentence.

However, there is a way to overcome this limitation: *agentive nouns* such as *runner* make explicit this kind of property and often play subject roles in sentences. Agentive nouns usually denote the “*doer*” or “*performer*” of some action. This is exactly what is needed to make clearer the relevant property $v_h(x)$ of the noun playing the logical subject role. The action v_h will be the one entailed by the verb v_t heading the sentence. As an example in the sentence “*the player wins*”, the action *play*

evocated by the agentive noun *player* is entailed by *win*.

3 Verb entailment: a classification

The focus of our study is on verb entailment. A brief review of the WordNet (Miller, 1995) verb hierarchy (one of the main existing resources on verb entailment relations) is useful to better explain the problem and to better understand the applicability of our hypothesis.

In WordNet, verbs are organized in synonymy sets (*synsets*) and different kinds of semantic relations can hold between two verbs (i.e. two synsets): *troponymy*, *causation*, *backward-presupposition*, and *temporal inclusion*. All these relations are intended as specific types of *lexical entailment*. According to the definition in (Miller, 1995) lexical entailment holds between two verbs v_t and v_h when the sentence *Someone v_t* entails the sentence *Someone v_h* (e.g. “*Someone wins*” entails “*Someone plays*”). Lexical entailment is then an asymmetric relation. The four types of WordNet lexical entailment can be classified looking at the temporal relation between the entailing verb v_t and the entailed verb v_h .

Troponymy represents the hyponymy relation between verbs. It stands when v_t and v_h are temporally co-extensive, that is, when the actions described by v_t and v_h begin and end at the same times (e.g. *limp*→*walk*). The relation of *temporal inclusion* captures those entailment pairs in which the action of one verb is temporally included in the action of the other (e.g. *snore*→*sleep*). *Backward-presupposition* stands when the entailed verb v_h happens before the entailing verb v_t and it is necessary for v_t . For example, *win* entails *play* via backward-presupposition as it temporally follows and presupposes *play*. Finally, in *causation* the entailing verb v_t necessarily causes v_h . In this case, the temporal relation is thus inverted with respect to backward-presupposition, since v_t precedes v_h . In causation, v_t is always a causative verb of change, while v_h is a resultative stative verb (e.g. *buy*→*own*, and *give*→*have*).

As a final note, it is interesting to notice that the Subject-Verb structure of v_t is generally preserved in v_h for all forms of lexical entailment. The two verbs have the same subject. The only exception is *causation*: in this case the subject of the entailed verb v_h is usually the object of v_t (e.g., *X give Y* → *Y have*). In most cases the subject of v_t carries

out an action that changes the state of the object of v_t , that is then described by v_h .

The intuition described in Sec. 2 is then applicable only for some kinds of verb entailments. First, the *causation* relation can not be captured since the two verbs should have the same subject (cf. eq. (2)). Secondly, troponymy seems to be less interesting than the other relations, since our focus is more on a *logic* type of entailment (i.e., v_t and v_h express two different actions one depending from the other). We then focus our study and our experiments on *backward-presupposition* and *temporal inclusion*. These two relations are organized in WordNet in a single set (called *ent*) parted from troponymy and causation pairs.

4 The method

Our method needs two steps. Firstly (Sec. 4.1), we translate the verb selectional expectations in specific Subject-Verb lexico-syntactic patterns $\mathcal{P}(v_t, v_h)$. Secondly (Sec. 4.2), we define a statistical measure $\mathcal{S}(v_t, v_h)$ that captures the verb preferences. This measure describes how much the relations between target verbs (v_t, v_h) are *stable* and commonly agreed.

Our method to detect verb entailment relations is based on the idea that some point-wise assertions carry relevant semantic information. This idea has been firstly used in (Robison, 1970) and it has been explored for extracting semantic relations between nouns in (Hearst, 1992), where lexico-syntactic patterns are induced by corpora. More recently this method has been applied for structuring terminology in *isa* hierarchies (Morin, 1999) and for learning question-answering patterns (Ravichandran and Hovy, 2002).

4.1 Nominalized textual entailment lexico-syntactic patterns

The idea described in Sec. 2 can be applied to generate Subject-Verb textual entailment lexico-syntactic patterns. It often happens that verbs can undergo an *agentive nominalization*, e.g., *play* vs. *player*. The overall procedure to verify if an entailment between two verbs (v_t, v_h) holds in a point-wise assertion is: *whenever it is possible to apply the agentive nominalization to the hypothesis v_h , scan the corpus to detect those expressions in which the agentified hypothesis verb is the subject of a clause governed by the text verb v_t .*

Given a verb pair (v_t, v_h) the assertion is for-

Lexico-syntactic patterns	
<i>nominalization</i>	$\mathcal{P}_{nom}(v_t, v_h) = \{ \text{"agent}(v_h) _{num:sing} \ v_t _{person:third,t:pres} \},$ $\{ \text{"agent}(v_h) _{num:plur} \ v_t _{person:nothird,t:pres} \},$ $\{ \text{"agent}(v_h) _{num:sing} \ v_t _{t:past} \},$ $\{ \text{"agent}(v_h) _{num:plur} \ v_t _{t:past} \} \}$
<i>happens-before</i> (Chklovski and Pantel, 2004)	$\mathcal{P}_{hb}(v_t, v_h) = \{ \text{"v}_h _{t:inf} \ \text{and then} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:inf} \ * \ \text{and then} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:past} \ \text{and then} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:past} \ * \ \text{and then} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:inf} \ \text{and later} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:past} \ \text{and later} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:inf} \ \text{and subsequently} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:past} \ \text{and subsequently} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:inf} \ \text{and eventually} \ v_t _{t:pres} \},$ $\{ \text{"v}_h _{t:past} \ \text{and eventually} \ v_t _{t:pres} \} \}$
<i>probabilistic entailment</i> (Glickman et al., 2005)	$\mathcal{P}_{pe}(v_t, v_h) = \{ \text{"v}_h _{person:third,t:pres} \ \wedge \ \text{"v}_t _{person:third,t:pres} \},$ $\{ \text{"v}_h _{t:past} \ \wedge \ \text{"v}_t _{t:past} \},$ $\{ \text{"v}_h _{t:pres,cont} \ \wedge \ \text{"v}_t _{t:pres,cont} \},$ $\{ \text{"v}_h _{person:nothird,t:pres} \ \wedge \ \text{"v}_t _{person:nothird,t:pres} \} \}$
<i>additional sets</i>	$\mathcal{F}_{agent}(v) = \{ \text{"agent}(v) _{num:sing} \}, \text{"agent}(v) _{num:plur} \}$ $\mathcal{F}(v) = \{ \text{"v} _{person:third,t:present} \},$ $\{ \text{"v} _{person:nothird,t:present} \}, \text{"v} _{t:past} \}$ $\mathcal{F}_{all}(v) = \{ \text{"v} _{person:third,t:pres} \}, \text{"v} _{t:pres,cont} \},$ $\{ \text{"v} _{person:nothird,t:present} \}, \text{"v} _{t:past} \} \}$

Table 1: Nominalization and related textual entailment lexico-syntactic patterns

malized in a set of textual entailment lexico-syntactic patterns, that we call *nominalized patterns* $\mathcal{P}_{nom}(v_t, v_h)$. This set is described in Tab. 1. $agent(v)$ is the noun deriving from the agentification of the verb v . Elements such as $l|_{f_1, \dots, f_N}$ are the tokens generated from lemmas l by applying constraints expressed via the feature-value pairs f_1, \dots, f_N . For example, in the case of the verbs *play* and *win*, the related set of textual entailment expressions derived from the patterns are $\mathcal{P}_{nom}(win, play) = \{ \text{"player wins"}, \text{"players win"}, \text{"player won"}, \text{"players won"} \}$. In the experiments hereafter described, the required verbal forms have been obtained using the publicly available morphological tools described in (Minnen et al., 2001). Simple heuristics have been used to produce the agentive nominalizations of verbs¹.

Two more sets of expressions, $\mathcal{F}_{agent}(v)$ and $\mathcal{F}(v)$ representing the single events in the pair, are needed for the second step (Sec. 4.2). This two *additional sets* are described in Tab. 1. In the example, the derived expressions are $\mathcal{F}_{agent}(play) = \{ \text{"player"}, \text{"players"} \}$ and $\mathcal{F}(win) = \{ \text{"wins"}, \text{"won"} \}$.

4.2 Measures to estimate the entailment strength

The above textual entailment patterns define *point-wise* entailment assertions. If pattern instances are found in texts, the related verb-subject pairs suggest but not confirm a verb selectional preference.

¹Agentive nominalization has been obtained adding “-er” to the verb root taking into account possible special cases such as verbs ending in “-y”. A form is retained as a correct nominalization if it is in WordNet.

The related entailment can not be considered commonly agreed. For example, the sentence “*Like a writer composes a story, an artist must tell a good story through their work.*” suggests that *compose* entails *write*. However, it may happen that these correctly detected entailments are accidental, that is, the detected relation is only valid for the given text. For example, if the text fragment “*The writers take a simple idea and apply it to this task*” is taken in isolation, it suggests that *take* entails *write*, but this could be questionable.

In order to get rid of these wrong verb pairs, we perform a statistical analysis of the verb selectional preferences over a corpus. This assessment will validate point-wise entailment assertions.

Before introducing the statistical entailment indicator, we provide some definitions. Given a corpus C containing samples, we will refer to the absolute frequency of a textual expression t in the corpus C with $f_C(t)$. The definition can be easily extended to a set of expressions T .

Given a pair v_t and v_h we define the following *entailment strength indicator* $\mathcal{S}(v_t, v_h)$. Specifically, the measure $\mathcal{S}_{nom}(v_t, v_h)$ is derived from point-wise mutual information (Church and Hanks, 1989):

$$\mathcal{S}_{nom}(v_t, v_h) = \log \frac{p(v_t, v_h | nom)}{p(v_t) p(v_h | pers)} \quad (3)$$

where nom is the event of having a nominalized textual entailment pattern and $pers$ is the event of having an agentive nominalization of verbs. Probabilities are estimated using maximum-likelihood:

$$p(v_t, v_h | nom) \approx \frac{f_C(\mathcal{P}_{nom}(v_t, v_h))}{f_C(\bigcup \mathcal{P}_{nom}(v'_t, v'_h))},$$

$p(v_t) \approx f_C(\mathcal{F}(v_t))/f_C(\cup \mathcal{F}(v))$, and $p(v_h|pers) \approx f_C(\mathcal{F}_{agent}(v_h))/f_C(\cup \mathcal{F}_{agent}(v))$. Counts are considered useful when they are greater or equal to 3.

The measure $\mathcal{S}_{nom}(v_t, v_h)$ indicates the relatedness between two elements composing a pair, in line with (Chklovski and Pantel, 2004; Glickman et al., 2005) (see Sec. 5). Moreover, if $\mathcal{S}_{nom}(v_t, v_h) > 0$ the verb selectional preference property described in eq. (1) is satisfied.

5 Related “non-distributional” methods and integrated approaches

Our method is a “non-distributional” approach for detecting semantic relations between verbs. We are interested in comparing and integrating our method with similar approaches. We focus on two methods proposed in (Chklovski and Pantel, 2004) and (Glickman et al., 2005). We will shortly review these approaches in light of what introduced in the previous sections. We also present a simple way to combine these different approaches.

The lexico-syntactic patterns introduced in (Chklovski and Pantel, 2004) have been developed to detect six kinds of verb relations: *similarity*, *strength*, *antonymy*, *enablement*, and *happens-before*. Even if, as discussed in (Chklovski and Pantel, 2004), these patterns are not specifically defined as entailment detectors, they can be useful for this purpose. In particular, some of these patterns can be used to investigate the *backward-presupposition* entailment. Verb pairs related by backward-presupposition are not completely temporally included one in the other (cf. Sec. 3): the entailed verb v_h precedes the entailing verb v_t . One set of lexical patterns in (Chklovski and Pantel, 2004) seems to capture the same idea: the *happens-before* (*hb*) patterns. These patterns are used to detect not temporally overlapping verbs, whose relation is semantically very similar to entailment. As we will see in the experimental section (Sec. 6), these patterns show a positive relation with the entailment relation. Tab. 1 reports the *happens-before* lexico-syntactic patterns (\mathcal{P}_{hb}) as proposed in (Chklovski and Pantel, 2004). In contrast to what is done in (Chklovski and Pantel, 2004) we decided to directly count patterns derived from different verbal forms and not to use an estimation factor. As in our work, also in (Chklovski and Pantel, 2004), a mutual-information-related measure is used as statistical

indicator. The two methods are then fairly in line.

The other approach we experiment is the “quasi-pattern” used in (Glickman et al., 2005) to capture lexical entailment between two sentences. The pattern has to be discussed in the more general setting of the probabilistic entailment between texts: the *text* T and the *hypothesis* H . The idea is that the implication $T \rightarrow H$ holds (with a degree of truth) if the probability that H holds knowing that T holds is higher than the probability that H holds alone, i.e.:

$$p(H|T) > p(H) \quad (4)$$

This equation is similar to equation (1) in Sec. 2. In (Glickman et al., 2005), words in H and T are supposed to be mutually independent. The previous relation between H and T probabilities then holds also for word pairs. A special case can be applied to verb pairs:

$$p(v_h|v_t) > p(v_h) \quad (5)$$

Equation (5) can be interpreted as the result of the following “quasi-pattern”: the verbs v_h and v_t should co-occur in the same document. It is possible to formalize this idea in the *probabilistic entailment* “quasi-patterns” reported in Tab. 1 as \mathcal{P}_{pe} , where verb form variability is taken into consideration. In (Glickman et al., 2005) point-wise mutual information is also a relevant statistical indicator for entailment, as it is strictly related to eq. (5).

For both approaches, the strength indicator $\mathcal{S}_{hb}(v_t, v_h)$ and $\mathcal{S}_{pe}(v_t, v_h)$ are computed as follows:

$$\mathcal{S}_y(v_t, v_h) = \log \frac{p(v_t, v_h|y)}{p(v_t)p(v_h)} \quad (6)$$

where y is *hb* for the *happens-before* patterns and *pe* for the *probabilistic entailment* patterns. Probabilities are estimated as in the previous section.

Considering independent the probability spaces where the three patterns lay (i.e., the space of subject-verb pairs for *nom*, the space of coordinated sentences for *hb*, and the space of documents for *pe*), the combined approaches are obtained summing up \mathcal{S}_{nom} , \mathcal{S}_{hb} , and \mathcal{S}_{pe} . We will then experiment with these combined approaches: *nom + pe*, *nom + hb*, *nom + hb + pe*, and *hb + pe*.

6 Experimental Evaluation

The aim of the experimental evaluation is to establish if the nominalized pattern is useful to help

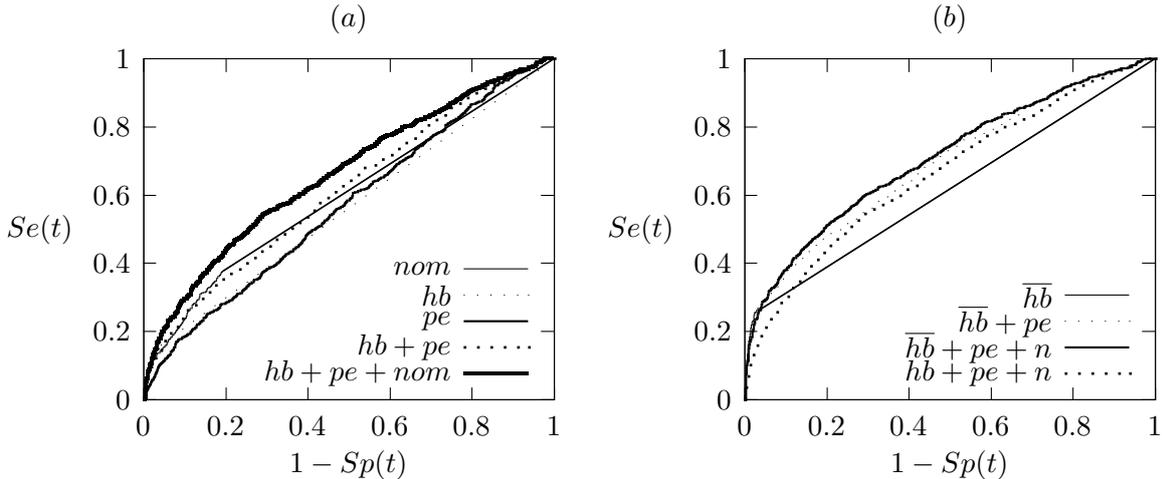


Figure 1: ROC curves of the different methods

in detecting verb entailment. We experiment with the method by itself or in combination with other sets of patterns. We are then interested only in verb pairs where the nominalized pattern is applicable. The best pattern or the best combined method should be the one that gives the highest values of \mathcal{S} to verb pairs in entailment relation, and the lowest value to other pairs.

We need a corpus C over which to estimate probabilities, and two dataset, one of verb entailment pairs, the *True Set* (TS), and another with verbs not in entailment, the *Control Set* (CS). We use the web as corpus C where to estimate S_{mi} and GoogleTM as a count estimator. The web has been largely employed as a corpus (e.g., (Turney, 2001)). The findings described in (Keller and Lapata, 2003) suggest that the count estimations we need in our study over *Subject-Verb* bigrams are highly correlated to corpus counts.

6.1 Experimental settings

Since we have a predefined (but not exhaustive) set of verb pairs in entailment, i.e. ent in WordNet, we cannot replicate a natural distribution of verb pairs that are or are not in entailment. Recall and precision lose sense. Then, the best way to compare the patterns is to use the ROC curve (Green and Swets, 1996) mixing *sensitivity* and *specificity*. ROC analysis provides a natural means to check and estimate how a statistical measure is able to distinguish positive examples, the *True Set* (TS), and negative examples, the *Control Set* (CS). Given a threshold t , $Se(t)$ is the probability of a candidate pair (v_h, v_t) to belong to *True Set* if

the test is positive, while $Sp(t)$ is the probability of belonging to *Control Set* if the test is negative, i.e.:

$$Se(t) = p((v_h, v_t) \in TS | \mathcal{S}(v_h, v_t) > t)$$

$$Sp(t) = p((v_h, v_t) \in CS | \mathcal{S}(v_h, v_t) < t)$$

The ROC curve ($Se(t)$ vs. $1 - Sp(t)$) naturally follows (see Fig. 1). Better methods will have ROC curves more similar to the step function $f(1 - Sp(t)) = 0$ when $1 - Sp(t) = 0$ and $f(1 - Sp(t)) = 1$ when $0 < 1 - Sp(t) \leq 1$.

The ROC analysis provides another useful evaluation tool: the *AROC*, i.e. the total area under the ROC curve. Statistically, *AROC* represents the probability that the method in evaluation will rank a chosen positive example higher than a randomly chosen negative instance. *AROC* is usually used to better compare two methods that have similar ROC curves. Better methods will have higher *AROCs*.

As *True Set* (TS) we use the controlled verb entailment pairs ent contained in WordNet. As described in Sec. 3, the entailment relation is a semantic relation defined at the synset level, standing in the verb sub-hierarchy. That is, each pair of synsets (S_t, S_h) is an oriented entailment relation between S_t and S_h . WordNet contains 409 entailed synsets. These entailment relations are consequently stated also at the lexical level. The pair (S_t, S_h) naturally implies that v_t entails v_h for each possible $v_t \in S_t$ and $v_h \in S_h$. It is possible to derive from the 409 entailment synset a test set of 2,233 verb pairs. As *Control Set* we use two sets: *random* and \overline{ent} . The *random* set

is randomly generated using verb in *ent*, taking care of avoiding to capture pairs in entailment relation. A pair is considered a control pair if it is not in the *True Set* (the intersection between the *True Set* and the *Control Set* is empty). The \overline{ent} is the set of pairs in *ent* with pairs in the reverse order. These two *Control Sets* will give two possible ways of evaluating the methods: a *general* and a more *complex* task.

As a pre-processing step, we have to *clean* the two sets from pairs in which the hypotheses can not be nominalized, as our pattern \mathcal{P}_{nom} is applicable only in these cases. The pre-processing step retains 1,323 entailment verb pairs. For comparative purposes the *random Control Set* is kept with the same cardinality of the *True Set* (in all, 1400 verb pairs).

\mathcal{S} is then evaluated for each pattern over the *True Set* and the *Control Set*, using equation (3) for \mathcal{P}_{nom} , and equation (6) for \mathcal{P}_{pe} and \mathcal{P}_{hb} . The best pattern or combined method is the one that is able to most neatly split entailment pairs from random pairs. That is, it should in average assign higher \mathcal{S} values to pairs in the *True Set*.

6.2 Results and analysis

In the first experiment we compared the performances of the methods in dividing the *ent* test set and the *random* control set. The compared methods are: (1) the set of patterns taken alone, i.e. *nom*, *hb*, and *pe*; (2) some combined methods, i.e. *nom + pe*, *hb + pe*, and *nom + hb + pe*. Results of this first experiment are reported in Tab. 2 and Fig. 1.(a). As Figure 1.(a) shows, our nominalization pattern \mathcal{P}_{nom} performs better than the others. Only \mathcal{P}_{hb} seems to outperform nominalization in some point of the ROC curve, where \mathcal{P}_{nom} presents a slight concavity, maybe due to a consistent overlap between positive and negative examples at specific values of the \mathcal{S} threshold t . In order to understand which of the two patterns has the best discrimination power a comparison of the AROC values is needed. As Table 2 shows, \mathcal{P}_{nom} has the best AROC value (59.94%) indicating a more interesting behaviour with respect to \mathcal{P}_{hb} and \mathcal{P}_{pe} . It is respectively 2 and 3 absolute percent point higher. Moreover, the combinations *nom + hb + pe* and *nom + pe* that includes the \mathcal{P}_{nom} pattern have a very high performance considering the difficulty of the task, i.e. 66% and 64%. If compared with the combina-

	AROC	best accuracy
<i>hb</i>	56.00	57.11
<i>pe</i>	57.00	55.75
<i>nom</i>	59.94	59.86
<i>nom + pe</i>	64.40	61.33
<i>hb + pe</i>	61.44	58.98
<i>hb + nom + pe</i>	66.44	63.09
\overline{hb}	61.64	62.73
$\overline{hb} + pe$	69.03	64.71
$\overline{hb} + nom + pe$	70.82	66.07

Table 2: Performances in the general case: *ent* vs. *random*

	AROC	best accuracy
<i>hb</i>	43.82	50.11
<i>nom</i>	54.91	54.94
\overline{hb}	56.18	57.16
$\overline{hb} + nom$	49.35	51.73
$\overline{hb} + nom$	57.67	57.22

Table 3: Performances in the *complex* case: *ent* vs. \overline{ent}

tion *hb + pe* that excludes the \mathcal{P}_{nom} pattern (61%), the improvement in the AROC is of 5% and 3%. Moreover, the shape of the *nom + hb + pe* ROC curve in Fig. 1.(a) is above all the other in all the points.

In the second experiment we compared methods in the more *complex* task of dividing the *ent* set from the \overline{ent} set. In this case methods are asked to determine if *win* \rightarrow *play* is a correct entailment and *play* \rightarrow *win* is not. Results of these set of experiments is presented in Tab. 3. The nominalized pattern *nom* preserves its discriminative power. Its AROC is over the chance line even if, as expected, it is worse than the one obtained in the *general* case. Surprisingly, the *happens-before* (*hb*) set of patterns seems to be not correlated the entailment relation. The temporal relation v_h -happens-before- v_t does not seem to be captured by those patterns. But, if this evidence is seen in a positive way, it seems that the patterns are better capturing the entailment when used in the reversed way (\overline{hb}). This is confirmed by its AROC value. If we observe for example one of the implications in the *True Set*, *reach* \rightarrow *go* what is happening may become clearer. Sample sentences respectively for the *hb* case and the \overline{hb} case are “*The group therefore elected to go to Tyso and then reach Anskaven*” and “*striving to reach personal goals and then go beyond them*”. It seems that in the second case *then* assumes an *enabling* role more than only a *temporal* role. After this sur-

prising result, as we expected, in this experiment even the combined approach $\overline{hb} + nom$ behaves better than $hb + nom$ and better than \overline{hb} , respectively around 8% and 1.5% absolute points higher (see Tab. 3).

The above results imposed the running of a third experiment over the *general* case. We need to compare the entailment indicators derived exploiting the new use of hb , i.e. \overline{hb} , with respect to the methods used in the first experiment. Results are reported in Tab. 2 and Fig. 1.(b). As Fig. 1.(b) shows, the \overline{hb} has a very interesting behaviour for small values of $1 - Sp(t)$. In this area it behaves extremely better than the combined method $nom + hb + pe$. This is an advantage and the combined method $nom + \overline{hb} + pe$ exploit it as both the AROC and the shape of the ROC curve demonstrate. Again the method $nom + \overline{hb} + pe$ that includes the \mathcal{P}_{nom} pattern has 1,5% absolute points with respect to the combined method $\overline{hb} + pe$ that does not include this information.

7 Conclusions

In this paper we presented a method to discover asymmetric entailment relations between verbs and we empirically demonstrated interesting improvements when used in combination with similar approaches. The method is promising and there is still some space for improvements. As implicitly experimented in (Chklovski and Pantel, 2004), some beneficial effect can be obtained combining these “non-distributional” methods with the methods based on the Distributional Hypothesis.

References

- Timoty Chklovski and Patrick Pantel. 2004. Verbo-CEAN: Mining the web for fine-grained semantic verb relations. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcellona, Spain.
- Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information and lexicography. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada.
- Oren Glickman and Ido Dagan. 2003. Identifying lexical paraphrases from a single corpus: A case study for verbs. In *Proceedings of the International Conference Recent Advances of Natural Language Processing (RANLP-2003)*, Borovets, Bulgaria.
- Oren Glickman, Ido Dagan, and Moshe Koppel. 2005. Web based probabilistic textual entailment. In *Proceedings of the 1st Pascal Challenge Workshop*, Southampton, UK.
- David M. Green and John A. Swets. 1996. *Signal Detection Theory and Psychophysics*. John Wiley and Sons, New York, USA.
- Zellig Harris. 1964. Distributional structure. In Jerrold J. Katz and Jerry A. Fodor, editors, *The Philosophy of Linguistics*, New York. Oxford University Press.
- Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 15th International Conference on Computational Linguistics (CoLing-92)*, Nantes, France.
- Frank Keller and Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, 29(3), September.
- Dekan Lin and Patrick Pantel. 2001. DIRT-discovery of inference rules from text. In *Proc. of the ACM Conference on Knowledge Discovery and Data Mining (KDD-01)*, San Francisco, CA.
- George A. Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, November.
- Guido Minnen, John Carroll, and Darren Pearce. 2001. Applied morphological processing of english. *Natural Language Engineering*, 7(3):207–223.
- Emmanuel Morin. 1999. *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques*. Ph.D. thesis, Université de Nantes, Faculté des Sciences et de Techniques.
- Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of the 40th ACL Meeting*, Philadelphia, Pennsylvania.
- Philip Resnik and Mona Diab. 2000. Measuring verb similarity. In *Twenty Second Annual Meeting of the Cognitive Science Society (COGSCI2000)*, Philadelphia.
- Philip Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania.
- Harold R. Robison. 1970. Computer-detectable semantic structures. *Information Storage and Retrieval*, 6(3):273–288.
- Peter D. Turney. 2001. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proc. of the 12th European Conference on Machine Learning*, Freiburg, Germany.