In-depth Exploitation of Noun and Verb Semantics to Identify Causation in Verb-Noun Pairs

Mehwish Riaz and Roxana Girju
Department of Computer Science and Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
{mriaz2,girju}@illinois.edu

Abstract

Recognition of causality is important to achieve natural language discourse understanding. Previous approaches rely on shallow linguistic features. In this work, we propose to identify causality in verb-noun pairs by exploiting deeper semantics of nouns and verbs. Particularly, we acquire and employ three novel types of knowledge: (1) semantic classes of nouns with a high and low tendency to encode causality along with information regarding metonymies, (2) data-driven semantic classes of verbal events with the least tendency to encode causality, and (3) tendencies of verb frames to encode causality. Using these knowledge sources, we achieve around 15% improvement in F-score over a supervised classifier trained using linguistic features.

1 Introduction

The identification of cause-effect relations is critical to achieve natural language discourse understanding. Causal relations are encoded in text using various linguistic constructions e.g., between two verbs, a verb and a noun, two discourse segments, etc. In this research, we focus on identifying causality encoded between a verb and a noun (or noun_phrase). For example, consider the following example:

1. At least 1,833 people **died** in **the hurricane**.

In example (1), the verb-noun_phrase pair "died"-"the hurricane" encodes causality where event "died" is the effect of "hurricane" event.

Previously several approaches have been proposed to identify causality between two verbs (Bethard and Martin, 2008; Riaz and Girju, 2010; Do et al., 2011; Riaz and Girju, 2013) and discourse segments (Sporleder and Lascarides, 2008; Pitler and Nenkova, 2009; Pitler et al., 2009). However, the problem of identifying causality in verb-noun pairs has not received a considerable attention. For example, Do et al. (2011) have studied this task but they worked only with a list of predefined nouns representing events. In this work, we focus on the linguistic construction of verb-noun (or noun_phrase) pairs where noun can be of any semantic type.

Traditional approaches for identifying causality mainly employ linguistic features (e.g., lexical items, part-of-speech tags of words, etc.) in the framework of supervised learning (Girju, 2003; Sporleder and Lascarides, 2008; Bethard and Martin, 2008; Pitler and Nenkova, 2009; Pitler et al., 2009) and do not involve deeper semantics of language. Analysis of such approaches by Sporleder and Lascarides (2008) have revealed that the linguistic features are not always sufficient to achieve a good performance on the task of identifying semantic relations including causality. In this work, we propose a model that deeply processes and acquires the specific semantic information about the participants of a verb-noun_phrase (v-np) pair (i.e., noun and verb semantics) to identify causality with a better performance over the baseline model depending merely on shallow linguistic features.

The work in this paper builds on our recent work reported in Riaz and Girju (2014). In that previous model, we identified the semantic classes of nouns and verbs with a high and low tendency to encode causation. For example, a named entity such as LOCATION may have the least tendency to encode causation. We leveraged such information about nouns to filter false positives. Similarly, we utilized the TimeBank's (Pustejovsky et al., 2006) classification of verbal events (i.e., Occurrence, Perception, Aspectual, State, I_State, I_Action and Reporting) and their definitions to claim that the reporting events (e.g., say, tell, etc.)

Proceedings of the SIGDIAL 2014 Conference, pages 161–170,
Philadelphia, U.S.A., 18-20 June 2014.

just describe and narrate other events instead of encoding causality with them. We proposed an Integer Linear Programming (ILP) model (Roth and Yih, 2004; Do et al., 2011) to combine noun and verb semantics with the decisions of a supervised classifier which only relies on linguistic features.

In this paper, we extend our previous model by acquiring and exploiting the following three novel types of knowledge:

1. We learn the information about tendencies of various verb frames to encode causation. For example, our model identifies if the subject of verb "destroy" ("occur") has a high (low) tendency to encode causation. Such information helps gain performance by exploiting causal semantics of each verb frame separately. We also learn and incorporate information about the verb frames in general e.g., how likely it is for the subject of any verb to encode causation with its verb.
2. In Riaz and Girju (2014), we utilized the TimeBank's definition of reporting events to argue that such events have the least tendency to encode causation. Instead of relying on human judgment we now introduce a data intensive approach to identify the TimeBank's classes of events with the least tendency to encode causation.
3. Although, information about the nouns with the least tendency to encode causation helps to filter false positives it can lead to false negatives when metonymic readings are associated with such nouns. Therefore, we introduce a metonymy resolver on top of our current model to avoid false negatives.

We provide details of our previously proposed model in section 3. We introduce new model and discuss its performance in sections 4 and 5. Section 6 concludes the current research.

2 Relevant Work

In Natural Language Processing (NLP), researchers are showing lots of interest in the task of identifying causality due to its various applications e.g., question answering (Girju, 2003), summarization (Chklovski and Pantel, 2004), future prediction (Radinsky and Horvitz, 2013), etc.

Several approaches have been proposed to identify causality in pairs of verbal events (Bethard and Martin, 2008; Riaz and Girju, 2010; Do et al., 2011; Riaz and Girju, 2013) and discourse segments (Sporleder and Lascarides, 2008; Pitler and Nenkova, 2009; Pitler et al., 2009). However causality a pervasive relation of language can be encoded via various linguistic constructions. For example, verbs and nouns are the key components of language to represent events. Therefore in this work we focus on identifying causality in verb-noun pairs.

Previously researchers have followed the path of utilizing linguistic features in the framework of supervised learning (Girju, 2003; Bethard and Martin, 2008; Sporleder and Lascarides, 2008; Pitler and Nenkova, 2009; Pitler et al., 2009). Though linguistic features are important but other sources of knowledge are also critically required to achieve progress on the current task.

In recent years, researchers have proposed unsupervised metrics to identify causality between events (Riaz and Girju, 2010; Do et al., 2011). For example, Riaz and Girju (2010) and Do et al. (2011) introduced unsupervised metrics to learn causal dependencies between events. These metrics mainly depend on probabilities of co-occurrences of events and do not distinguish well causality from any other types of correlation (Riaz and Girju, 2013). In order to overcome this problem Riaz and Girju (2013) proposed some advanced metrics which combine probabilities of co-occurrences of events with the supervised estimates of cause and non-cause relations.

Considering the importance of employing rich sources of knowledge other than linguistic features for the current task, we have recently proposed a model that incorporates semantic classes of nouns and verbs with a high and low tendency to encode causation (Riaz and Girju, 2014). In this work, we exploit information about verb frames, data-driven verb semantics and metonymies to achieve more progress on our recent work.

3 Model for Recognizing Causality

In this section we provide an overview of our previous model (Riaz and Girju, 2014) for identifying causality in v-np pairs where v (np) stands for verb (noun_phrase). This model works in the following two stages: (1) A supervised classifier is used to make binary predictions (i.e., the label cause (C) or non-cause ($\neg$C)) employing linguistic features, and (2) noun and verb semantics are then combined with the predictions of supervised classifier in the ILP framework to identify causality.

3.1 Supervised Classifier

To the best of our knowledge, there is no data set of v-np pairs with the labels C and ¬C available to us. For the current task we employ some heuristics to extract a training corpus of v-np pairs using FrameNet (Baker et al., 1998). FrameNet provides frame elements for the verbs and hand annotated examples (aka annotations) of these frame elements. Consider the following annotation from FrameNet "They **died** [$_{Cause}$from **shotgun wounds**]" where the frame element "Cause" is given for the verb "died". We remove the preposition "from" from the above annotation of frame element to acquire an instance of v-np (i.e., diedshotgun wounds) pair. We extract all annotations for verbs from FrameNet in which a frame element must contain at least one noun and no verb in it. We found such annotations for 729 distinct frame elements. We manually assigned the labels C and ¬C to these frame elements. Cause, Purpose, Reason, Result, Explanation are some examples of the frame elements to which we assigned the label C. Using the above mentioned assignments of labels C and ¬C to frame elements, we have acquired a training corpus of $4,141$ $(77,119)$ C (¬C) instances from FrameNet. In order to avoid class imbalance while training we employ an equal number of instances of both labels.

Due to space constraints, we refer the reader to Appendix A for the details of linguistic features to build the supervised classifier. We employ both Naive Bayes (NB) and Maximum Entropy (MaxEnt) algorithms to acquire predictions and probabilities of assignments of labels. We set up the following ILP using these probabilities:

$$Z_1 = \max \sum_{\text{v-np} \in I} \sum_{l \in L_1} x_1(\text{v-np}, l) P(\text{v-np}, l) \quad (1)$$

$$\sum_{l \in L_1} x_1(\text{v-np}, l) = 1 \quad \forall \text{v-np} \in I \quad (2)$$

$$x_1(\text{v-np}, l) \in \{0, 1\} \quad \forall \text{v-np} \in I \quad \forall l \in L_1 \quad (3)$$

Here, $L_1 = \{\text{C}, \neg\text{C}\}$, I is the set of all v-np pairs. $x_1(\text{v-np}, l)$ is a binary decision variable set to 1 only if the label $l \in L_1$ is assigned to a v-np pair and only one label out of $|L_1|$ choices can be assigned to a v-np pair (see constraints 2 and 3). In particular, we maximize the objective function Z_1 (1) assigning the labels $l \in \{L_1\}$ to v-np pairs depending on the probabilities of assignments (i.e., $P(\text{v-np}, l)$) obtained through the supervised classifier.

3.2 Noun and Verb Semantics

We automatically acquire and employ semantic classes of nouns and verbs with a high and low tendency to encode causation. Such information helps to reduce errors in predictions of the supervised classifier.

We derive two semantic classes of nouns for our purpose i.e., C_{np} and $\neg\text{C}_{\text{np}}$ where the class C_{np} ($\neg\text{C}_{\text{np}}$) represents the noun_phrases with a high (low) tendency to encode causation. For example, a noun_phrase expression for a location has the least tendency to encode causation unless a metonymic reading is associated with it. In order to acquire these classes, we extract annotations of 936 distinct frame elements from FrameNet in which a frame element must contain at least one noun and no verb in it. These annotations of frame elements roughly represent instances of noun_phrases (np). We manually assigned the labels C_{np} and $\neg\text{C}_{\text{np}}$ to the frame elements. For example, we assign the label $\neg\text{C}_{\text{np}}$ to the frame element "Place" which represents a location (see Appendix B for some examples of the frame elements with labels C_{np} and $\neg\text{C}_{\text{np}}$). We also follow the approach similar to Girju and Moldovan (2002) to employ WordNet senses of nouns to acquire more instances of the classes C_{np} and $\neg\text{C}_{\text{np}}$ (see Appendix B for the details). We have acquired a total of $280,212$ instances of np (50% for each of the two classes i.e., C_{np} and $\neg\text{C}_{\text{np}}$) using both FrameNet and WordNet. Using these instances, we build a supervised classifier to identify the semantic class of np (see Appendix B for the details of features to build the classifier). We incorporate the knowledge of semantic classes of nouns by making the following additions to ILP:

$$Z_2 = Z_1 + \sum_{\text{v-np} \in I - M} \sum_{l \in L_2} x_2(\text{f}_{\text{np}}(\text{v-np}), l) P(\text{f}_{\text{np}}(\text{v-np}), l) \quad (4)$$

$$\sum_{l \in L_2} x_2(\text{f}_{\text{np}}(\text{v-np}), l) = 1 \quad \forall \text{v-np} \in I - M \quad (5)$$

$$x_2(\text{f}_{\text{np}}(\text{v-np}), l) \in \{0, 1\} \quad \forall \text{v-np} \in I - M \quad \forall l \in L_2 \quad (6)$$

$$x_1(\text{v-np}, \neg\text{C}) - x_2(\text{f}_{\text{np}}(\text{v-np}), \neg\text{C}_{\text{np}}) \geq 0 \quad \forall \text{v-np} \in I - M \quad (7)$$

Here $L_2 = \{\text{C}_{\text{np}}, \neg\text{C}_{\text{np}}\}$. $\text{f}_{\text{np}}(\text{v-np})$ is a function which returns np of a v-np pair. M is the set of v-np pairs with metonymic readings associated with np. Currently, this set is empty and in section 4.3 we introduce a metonymy resolver to pop-

ulate this set. $x_2(\mathrm{f_{np}}(\text{v-np}), l)$ is a binary decision variable set to 1 only if the label $l \in L_2$ is assigned to np and only one label out of $|L_2|$ choices can be assigned to np (see constraints 5 and 6). Constraint 7 enforces that if an np belongs to the class $\neg\mathrm{C_{np}}$ then its corresponding v-np pair is assigned the label $\neg$C. In particular, we maximize the objective function Z_2 (4) subject to the constraints introduced till now. For each v-np pair, we predict the semantic class of np using our supervised classifier for the labels $l \in L_2$ and set the probabilities – i.e., $P(\mathrm{f_{np}}(\text{v-np}), l) = 1, P(\mathrm{f_{np}}(\text{v-np}), \{L_2\} - \{l\}) = 0$ if the label $l \in L_2$ is assigned to np. Also before running our supervised classifier, we run a named entity recognizer (Finkel et al., 2005) and assign the label $\neg\mathrm{C_{np}}$ to all noun_phrases identified as named entities. We also determine association of metonymies with the noun_phrases identified as named entities.

For the current task we also acquire two semantic classes of verbs i.e., $\mathrm{C_{e_v}}$ and $\neg\mathrm{C_{e_v}}$ where the class $\mathrm{C_{e_v}}$ ($\neg\mathrm{C_{e_v}}$) contains the verbal events with a high (low) tendency to encode causation. In order to derive these two classes we exploit the TimeBank corpus (Pustejovsky et al., 2003) which provides seven semantic classes of verbal events – i.e., Occurrence, Perception, Aspectual, State, I_State, I_Action and Reporting. According to the definitions of these classes, we claim that the reporting events (e.g., say, tell, etc.) just describe and narrate other events instead of encoding causality with them. Using this claim, we consider that all instances of reporting verbal events of TimeBank belong to the class $\neg\mathrm{C_{e_v}}$ and the rest of instances of verbal events lie in the class $\mathrm{C_{e_v}}$. After acquiring instances of the classes $\mathrm{C_{e_v}}$ and $\neg\mathrm{C_{e_v}}$, we build a supervised classifier for these two classes. We use the features introduced by Bethard and Martin (2006) to build this classifier (see Bethard and Martin (2006) for the details). Employing predictions and probabilities of assignments of the labels $\mathrm{C_{e_v}}$ and $\neg\mathrm{C_{e_v}}$ we add the following two constraints to ILP: (1) if the event represented by v belongs to $\neg\mathrm{C_{e_v}}$ then the corresponding v-np pair must be labeled with $\neg$C and (2) if a v-np pair is a causal pair then the event represented by v must be labeled with $\mathrm{C_{e_v}}$.

4 Enriched Verb and Noun Semantics

This section describes the novel contributions of this work i.e., identification of semantics of verb frames, semantic classes of verbal events via a data intensive approach and association of metonymic readings with noun_phrases to identify causality with a better performance.

4.1 Verb Frames

We introduce a method to acquire tendencies of various verb frames to encode causation. Consider the following two examples to understand the tendencies of verb frames of form $\{\text{v}, \text{gr}\}$ to encode causation where v is the verb and gr is the grammatical relation of np with the verb v.

1. **The Great Storm of October 1987** almost totally **destroyed** the eighty year old pinetum at Nymans Garden in Sussex. (Cause (C))
2. **The explosion occurred** in the city's main business area. (Non-Cause ($\neg$C))

In above two examples the nps "The Great Storm of October 1987" and "The explosion" have the grammatical relations of subject with the verbs "destroyed" and "died". In examples (1) and (2) the verb frames {destroy, subject} and {occur, subject} encode cause and non-cause relations. These examples reveal that each verb frame has its own tendency to encode causation. This type of knowledge helps gain performance by exploiting the semantics of each verb frame separately.

We leverage FrameNet annotations to acquire such type of knowledge. We collect all annotations of verbs from FrameNet and assign the labels C and $\neg$C to the frame elements as discussed in section 3.1. In FrameNet, example (1) is given as follows:

3. [$_{Cause}$ The Great Storm of October 1987] [$_{Degree}$ almost totally] **destroyed** [$_{Undergoer}$ the eighty year old pinetum at Nymans Garden in Sussex].

According to our assignments of labels C and $\neg$C to the frame elements, example (1) is given as "[$_{\mathrm{C}}$ The Great Storm of October 1987] [$_{\neg\mathrm{C}}$ almost totally] **destroyed** [$_{\neg\mathrm{C}}$ the eighty year old pinetum at Nymans Garden in Sussex].". After acquiring instances of the labels C and $\neg$C from example (1), we populate the fields of a knowledge base of verb frames (see Table 1). Fields of this knowledge base are $\{\text{v}, \text{gr}\}$, count($\{\text{v}, \text{gr}\}$, C) and count($\{\text{v}, \text{gr}\}, \neg\text{C}$). gr is the dependency relation of the frame element with the verb v. We use Stanford's dependency parser (Marneffe et al., 2006) to collect dependency relations. count($\{\text{v}, \text{gr}\}$, C) (count($\{\text{v}, \text{gr}\}, \neg\text{C}$)) is the count of the label C ($\neg$C) of the frame {v, gr}. As shown in Table 1,

for the frame element "The Great Storm of October 1987", the word "Storm" has the dependency relation of "nsubj" with the verb "destroy". If there exists more than one dependency relations between the frame element and its verb then we choose the very first relation in the text order. According to the counts given in Table 1, {destroy, nsubj} has more tendency to encode a cause relation than the non-cause one. We have acquired 7,156 and 114,898 instances of the labels C and ¬C from FrameNet for populating the knowledge base of verb frames. We compute tendencies of verb frames to encode causality using the following scores:

$$S(\{\text{v, gr}\}, l) = S_1(\{\text{v, gr}\}, l) \times S_2(\{*, \text{gr}\}, l) \quad (8)$$

$$S_1(\{\text{v, gr}\}, l) = \frac{count(\{\text{v,gr}\}, l)}{count(\{\text{v,gr}\}, l) + count(\{\text{v,gr}\}, L_1 - \{l\})}$$

$$S_2(\{*, \text{gr}\}, l) = \frac{count(\{*,\text{gr}\}, l)}{count(\{*,\text{gr}\}, l) + count(\{*,\text{gr}\}, L_1 - \{l\})}$$

Counts of first component (S_1) can be taken from the knowledge base of verb frames of form $\{\text{v, gr}\}$. The second component (S_2) with counts $count(\{*, \text{gr}\}, l)$ and $count(\{*, \text{gr}\}, L_1 - \{l\})$ captures tendencies of verb frames in general. For example, what is the tendency of any subject to encode causality with its verb i.e., the score $S_2(\{*, nsubj\}, \text{C})$. We populate the knowledge base of Table 1 with equal number of C and ¬C instances to calculate counts for S_2. We make the following additions to ILP to incorporate information about verb frames:

$$Z_3 = Z_2 + \sum_{\substack{\text{v-np} \in I \wedge \\ g(\text{v-np}) \in KB \wedge \\ \text{f}_{\text{np}}(\text{v-np}) \in \text{C}_{\text{np}}}} \sum_{l \in L_1} x_3(g(\text{v-np}), l) S(g(\text{v-np}), l) \quad (9)$$

$$\sum_{l \in L_1} x_3(g(\text{v-np}), l) = 1 \quad \forall \begin{array}{c} \text{v-np} \in I \wedge \\ g(\text{v-np}) \in KB \wedge \\ \text{f}_{\text{np}}(\text{v-np}) \in \text{C}_{\text{np}} \end{array} \quad (10)$$

$$x_3(g(\text{v-np}), l) \in \{0, 1\} \quad \forall l \in L_1, \forall \begin{array}{c} \text{v-np} \in I \wedge \\ g(\text{v-np}) \in KB \wedge \\ \text{f}_{\text{np}}(\text{v-np}) \in \text{C}_{\text{np}} \end{array} \quad (11)$$

$$x_3(g(\text{v-np}), l) \leq x_1(\text{v-np}, l) \quad \forall l \in L_1, \quad \forall \begin{array}{c} \text{v-np} \in I \\ \wedge g(\text{v-np}) \in KB \\ \wedge \text{f}_{\text{np}}(\text{v-np}) \in \text{C}_{\text{np}} \end{array} \quad (12)$$

$$x_1(\text{v-np}, l) \leq x_3(g(\text{v-np}), l) \quad \forall l \in L_1, \quad \forall \begin{array}{c} \text{v-np} \in I \wedge \\ g(\text{v-np}) \in KB \wedge \\ \text{f}_{\text{np}}(\text{v-np}) \in \text{C}_{\text{np}} \end{array} \quad (13)$$

Here, KB is the knowledge base of verb frames and $g(\text{v-np})$ is the function which returns the verb frame i.e., $\{\text{v, gr}\}$. This function returns NULL value if there is no grammatical relation between v and np in an instance. The above changes in ILP are only applicable for the v-np pairs with

{v, gr}	count($\{\text{v, gr}\}$, C)	count($\{\text{v, gr}\}$, ¬C)
{destroy,nsubj}	1	0
{destroy,advmod}	0	1
{destroy,dobj}	0	1
[$_{\text{C}}$ The Great Storm of October 1987] [$_{\neg\text{C}}$ almost totally] **destroyed** [$_{\neg\text{C}}$ the eighty year old pinetum at Nymans Garden in Sussex].		

Table 1: A knowledge base of verb frames. This knowledge base is populated using the instances of C and ¬C labels given in this table.

$g(\text{v-np}) \in KB$ and np identified as of class C_{np} because we have already filtered the cases of $np \in \neg\text{C}_{\text{np}}$ in section 3.2. $x_3(g(\text{v-np}), l)$ is a binary decision variable set to 1 only if the label $l \in L_1$ is assigned to $g(\text{v-np})$ and only one label out of $|L_1|$ choices can be assigned to $g(\text{v-np})$ (see constraints 10 and 11). We add information about verb frames using constraints 12 and 13. These constraints enforce the predictions of the supervised classifier of causality (section 3.1) to be consistent with the predictions using tendencies of verb frames (i.e., score $S(\{\text{v, gr}\}, l)$). We maximize objective function (9) subject to the above constraints. We remove those $\{\text{v, gr}\}$ from KB which have $count(\{\text{v, gr}\}, \text{C}) + count(\{\text{v, gr}\}, \neg\text{C}) < 5$ to avoid wrong predictions based on the small counts of verb frames.

4.2 Data-driven Verb Semantics

In section 3.2 we considered that reporting events belong to the class $\neg\text{C}_{\text{e}_\text{v}}$ with the least tendency to encode causation using the definition of these events in the TimeBank corpus. Instead of relying on definitions of events we now introduce a data intensive approach to automatically identify the class $\neg\text{C}_{\text{e}_\text{v}}$ of verbal events. In order to identify this class we extract training instances of verbal events encoding C and ¬C relations. Verbal events encode cause-effect relations using verb-verb (e.g., Five shoppers were **killed** when a car **blew up**.) and verb-noun linguistic constructions. Therefore for the current purpose we use the following two types of training instances: (A) a training corpus of 240K instances of verb-verb (v_i-v_j) pairs encoding C and ¬C relations (named as Training$_{\text{v}_\text{i}\text{-}\text{v}_\text{j}}$) (we refer the reader to Riaz and Girju (2013) for the details of this training corpus) and (B) the training corpus v-np instances introduced in section 3.1 (named as Training$_{\text{v-np}}$).

Following is the procedure to derive $\text{V}_{\neg\text{C}} \subseteq \text{V}$ where V={Occurrence, Perception, Aspectual, State, I_State, I_Action, Reporting} and the set $\text{V}_{\neg\text{C}}$ contains the TimeBank's semantic classes

with the least tendency to encode a cause relation.

1. **Input:** Training corpus, V
2. **Output:** Set $\mathrm{V}_{\neg\mathrm{C}}$
3. For each training instance k employ the supervised classifier of Bethard and Martin (2006) to do the following:
   (a) if $k \in \mathrm{Training}_{\mathrm{v_i\text{-}v_j}}$ then identify the semantic class (sc) of both events represented by both verbs $\mathrm{v_i}$ and $\mathrm{v_j}$ and add this information to a set i.e., T = T $\cup$ ($k_{\mathrm{v_i}}$, $\mathrm{sc_{v_i}}$, l) $\cup$ ($k_{\mathrm{v_j}}$, $\mathrm{sc_{v_j}}$, l) where $\mathrm{sc_{v_i}}$ is the semantic class of event of the verb $\mathrm{v_i}$ of instance k and $l \in \{\mathrm{C}, \neg\mathrm{C}\}$.
   (b) Else if $k \in \mathrm{Training}_{\mathrm{v\text{-}np}}$ then identify the semantic class (sc) of event represented by the verb v and set T = T $\cup$ (k_{v}, $\mathrm{sc_v}$, l).
4. Using results of step 3, calculate tendency of each semantic class sc $\in$ V to encode non-causality (i.e., $score(\mathrm{sc}, \neg\mathrm{C})$) as follows:

$$score(\mathrm{sc}, \neg\mathrm{C}) = score_1(\mathrm{sc}, \neg\mathrm{C}) \times score_2(\mathrm{sc}, \neg\mathrm{C})$$
$$score_1(\mathrm{sc}, \neg\mathrm{C}) = \left(\frac{count(\mathrm{sc}, \neg\mathrm{C})}{count(\mathrm{sc})} - \frac{count(\mathrm{sc}, \mathrm{C})}{count(\mathrm{sc})}\right)$$
$$score_2(\mathrm{sc}, \neg\mathrm{C}) = \left(\frac{count(\mathrm{sc}, \neg\mathrm{C})}{count(\neg\mathrm{C})} - \frac{count(\mathrm{sc}, \mathrm{C})}{count(\mathrm{C})}\right)$$

   where count(m, n) is the number of instances of verbal events with the labels m and n and count(m) is the number of instances of verbal events with the label m.
5. Acquire a ranked list of semantic classes $\mathrm{list_{sc}}$ = [$\mathrm{sc_1}$, $\mathrm{sc_2}$, ... sc_m] s.t. $score(\mathrm{sc}_i, \neg\mathrm{C}) \geq score(\mathrm{sc}_{i+1}, \neg\mathrm{C})$. From this list we remove the class sc_i if either $score_1(\mathrm{sc}_i, \neg c) < 0$ or $score_2(\mathrm{sc}_i, \neg c) < 0$.

   ▷ The following steps are used to determine the cutoff class $\mathrm{sc}_i \in \mathrm{list_{sc}}$ s.t. the semantic classes {$\mathrm{sc_1}$, $\mathrm{sc_2}$, ..., $\mathrm{sc_{i\text{-}1}}$} have the least tendency to encode causation.
6. $\mathrm{result_{sc_{-1}}} = 0$ and $\mathrm{result_{sc_0}} = 0$.
7. Remove sc_i from the front of $\mathrm{list_{sc}}$ and do the following:
   (c) Predict the label (l) $\neg$C for all tuples of form $(m, \mathrm{sc}, l) \in$ T if sc $\in$ {$\mathrm{sc_1}$, $\mathrm{sc_2}$, ..., sc_i} and predict C for the rest of the tuples.
   (d) Using predictions from step (c), calculate the $\mathrm{result_{sc_i}}$ = F1-score $\times$ accuracy for the label $l \in \{\mathrm{C}, \neg\mathrm{C}\}$.
   (e) If $\mathrm{result_{sc_i}} - \mathrm{result_{sc_{i\text{-}1}}} < \mathrm{result_{sc_{i\text{-}1}}} - \mathrm{result_{sc_{i\text{-}2}}}$ then output {$\mathrm{sc_1}$, $\mathrm{sc_2}$, ..., sc_{i-1}}
   (f) Else go to step 7.

Using the above procedure, we obtain the sets {Aspectual} and {Reporting, I_State} with $\mathrm{Training_{v_i\text{-}v_j}}$ and $\mathrm{Training_{v\text{-}np}}$ corpora. We consider that the Aspectual, Reporting and I_State events of the TimeBank corpus belong to the class $\neg\mathrm{C_{e_v}}$ and rest of the events lie in $\mathrm{C_{e_v}}$. Using these semantic classes we apply the constraints introduced in section 3.2.

4.3 Metonymy Resolution:

Metonymy resolution is the task to determine if a literal or non-literal reading is associated with a

{**v, gr**}	**count**({v, gr}, $\mathrm{C_{np}}$)	**count**({v, gr}, $\neg\mathrm{C_{np}}$)
{kill,nsubj}	1	0
{kill,dobj}	0	1
[$\mathrm{C_{np}}$Pissed off Angelus] just kills [$\neg\mathrm{C_{np}}$me]		

Table 2: A knowledge base of verb frames. This knowledge base is populated using the instances of $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$ labels given in this table.

natural language expression (Markert and Nissim, 2009). Consider the following example:

4. **The United States** has **killed** Osama bin Laden and has custody of his body. (Cause (C))

In example (4) "The United States" refers to a non-literal reading i.e., the event of "raid in Abbottabad on May 2, 2011 by the United States" rather than merely referring to a literal sense i.e., a country. The association of non-literal reading with "The United States" results in killing event. Previously, researchers have worked with hand-annotated selectional restrictions violation for this task (Markert and Nissim, 2009). In the example (4) a country cannot "kill" someone and thus a metonymic reading is associated with it. In this work we identify association of metonymies with noun_phrases via verb frames and prepositions as explained below in this section.

In the first part of our approach we employ violations of tendencies of verb frames to identify if a non-literal reading is associated with a noun_phrases. Particularly, we build a knowledge base of verb frames using $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$ classes as discussed in section 4.1. Consider the knowledge base given in Table 2 populated using the following FrameNet annotations "[$_{Stimulus}$Pissed off Angelus] just kills [$_{Experiencer}$me]." with assignments of labels $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$ to the frame elements. We populate the knowledge base using only those FrameNet annotations in which a frame element does not contain a verb.

Now we introduce our method to identify the association of non-literal reading with the "The United States" in example (4). The supervised classifier predicts the class $\neg\mathrm{C_{np}}$ for the np "The United States". However, in the current state of knowledge base (Table 2) P({destroy, nsubj}, $\mathrm{C_{np}}$) > P({destroy, nsubj}, $\neg\mathrm{C_{np}}$) where P is the probability. The prediction of $\neg\mathrm{C_{np}}$ for "The United States" violates the above probabilities. Considering this violation, we predict the association of metonymy with np.

In the second part of our approach we identify tendencies of prepositions to encode causation

and use violation of these tendencies to identify metonymies. For this purpose, we use the training corpus of v-np pairs with $4,141$ C and $77,119$ $\neg$C training instances (see section 3.1). We employ only those training instances in which a preposition appears between v and np and there appears no verb between them. From these instances, we acquire a set of prepositions that appear between v and np. Using this set of prepositions (PR) as input to the following procedure, we acquire a set of prepositions ($\mathrm{PR_C}$) with the highest tendency to encode causation:

1. **Input:** Training Corpus of v-np pairs, PR
2. **Output:** $\mathrm{PR_C}$
3. Calculate tendency of each preposition $\mathrm{pr} \in \mathrm{PR}$ to encode causality (i.e., $score(\mathrm{pr}, \mathrm{C})$) as follows:

$$score(\mathrm{pr}, \mathrm{C}) = score_1(\mathrm{pr}, \mathrm{C}) \times score_2(\mathrm{pr}, \mathrm{C})$$
$$score_1(\mathrm{pr}, \mathrm{C}) = (\frac{count(\mathrm{pr}, \mathrm{C})}{count(\mathrm{pr})} - \frac{count(\mathrm{pr}, \neg\mathrm{C})}{count(\mathrm{pr})})$$
$$score_2(\mathrm{pr}, \mathrm{C}) = (\frac{count(\mathrm{pr}, \mathrm{C})}{count(\mathrm{C})} - \frac{count(\mathrm{pr}, \neg\mathrm{C})}{count(\neg\mathrm{C})})$$

4. Acquire a ranked list of prepositions $\mathrm{list_{pr}}$ = [pr_1, pr_2, ... pr_m] s.t. $score(\mathrm{pr}_i, \mathrm{C}) \geq score(\mathrm{pr}_{i+1}, \mathrm{C})$. From this list we remove pr_i if either $score_1(\mathrm{pr}_i, \mathrm{C})$ or $score_2(\mathrm{pr}_i, \mathrm{C}) < 0$.
5. $\mathrm{result_{pr_{-1}}} = 0$, $\mathrm{result_{pr_0}} = 0$
6. Remove pr_i from the front of the $\mathrm{list_{pr}}$ and do the following:
   (a) Predict the label C for all v-np training instances with pr $\in \{\mathrm{pr}_1, \mathrm{pr}_2, \ldots, \mathrm{pr}_i\}$ and assign the label $\neg$C to the rest of the instances.
   (b) Using predictions from step (a) calculate the $\mathrm{result_{pr_i}}$ = F1-score $\times$ accuracy.
   (c) If $\mathrm{result_{pr_i}}$-$\mathrm{result_{pr_{i-1}}} < \mathrm{result_{pr_{i-1}}}$-$\mathrm{result_{pr_{i-2}}}$ then output $\{\mathrm{pr}_1, \mathrm{pr}_2, \ldots, \mathrm{pr}_{i-1}\}$.
   (d) Else go to step 6.

The above procedure outputs the set $\mathrm{PR_C} = \{$for, by$\}$. Now we introduce method to identify association of non-literal reading for the example "All weapon sites in Iraq were **destroyed** by **the United States**" where "the United States" $\in \neg\mathrm{C_{np}}$ as identified by the supervised classifier. However, the preposition "by" has a high tendency to encode causation and thus "the United States" may encode causation. Therefore, there is a possibility that this noun_phrase has a non-literal sense attached to it which results in encoding causality. Using this method, we predict metonymies only for the v-np instances where preposition appears between v and np and there appears no verb between them. If any of two methods of metonymy resolution predicts the association of metonymy with np then we add v-np to the set M used in ILP (see section 3.2).

5 Evaluation and Discussion

In this section we present experiments and discussion on the performance achieved for the current task. In order to evaluate our model, we generated a test set of instances of v-np pairs. For this purpose, we collected three wiki articles on the topics of Hurricane Katrina, Iraq War and Egyptian Revolution of 2011. We apply a part-of-speech tagger and a dependency parser on all sentences of these three articles (Toutanova et al., 2003; Marneffe et al., 2006). We extracted all v-np pairs from each sentence of these articles. For each of the these three articles, we selected first 500 instances of v-np pairs. Two annotators were asked to provide the labels C and $\neg$C to the instances of v-np pairs using the annotation guidelines from Riaz and Girju (2010). We have achieved a 0.64 kappa score for the human inter-annotator agreement on a total of 1,500 v-np instances. This results in a total of 1,365 instances of v-np pairs with 11.86% C pairs.

In this section, we present performance of the following models (see Table 3):

1. **Baseline**: NB and MaxEnt (McCallum, 2002) supervised classifiers using only the shallow linguistic features (see section 3.1).
2. **Basic noun and verb semantics**: ILP with the addition of semantic classes of nouns without metonymy (denoted by $+\mathrm{N_{!\,M}}$) and the addition of semantic classes of verbs where $\neg\mathrm{C_{e_v}}$={(R)eporting events} (denoted by $+\mathrm{N_{!\,M}}+\mathrm{V_{\{R\}}}$). These models represent the work proposed in Riaz and Girju (2014) (section 3).
3. **Noun semantics with metonymies**: ILP with the addition of noun semantics involving metonymies resolved via verb frames (denoted by $+\mathrm{N_{M_1}}$), metonymies resolved via verb frames {v, gr} where gr $\in$ GR = {csubj, csubjpass, nsubj, nsubjpass, xsubj, dobj, iobj, pobj, agent} a set of core dependency relations of subjects and objects (denoted by $+\mathrm{N_{M_{1_{GR}}}}$) and metonymies resolved via both verb frames and prepositions (denoted by $+\mathrm{N_{M_{1_{GR}}+M_2}}$).
4. **Verb frames and data-driven verb semantics**: ILP with the addition of information about verb frames (denoted by $+\mathrm{N_M}$+VF where $\mathrm{M} = \mathrm{M_{1_{GR}}} + \mathrm{M_2}$), data-driven verb semantics i.e., $\neg\mathrm{C_{e_v}}$ ={(A)spectual, (R)eporting, (I)_(S)tate events} (denoted by $+\mathrm{N_M}+\mathrm{V_{\{A,R,IS\}}}$) and both verb frames and data-driven verb semantics (denoted by $+\mathrm{N_M}$+VF+$\mathrm{V_{\{A,R,IS\}}}$)

S	B	+$N_{!M}$	+$N_{!M}$+$V_{\{R\}}$	+N_{M_1}	+$N_{M_{1_{GR}}}$	+$N_{M_{1_{GR}}+M_2}$	+N_M+VF	+N_M+$V_{\{A,R,IS\}}$	+N_M+VF +$V_{\{A,R,IS\}}$
A	28.86	71.86	73.40	71.35	71.42	71.64	72.96	75.16	76.19
P	13.52	26.18	27.21	26.29	26.34	27.54	28.39	29.93	30.82
R	92.59	75.30	74.07	78.39	78.39	85.18	83.95	81.48	80.86
F	23.60	38.85	39.80	39.37	39.44	41.62	42.43	43.78	44.63
A	61.46	80.73	81.17	80.65	80.73	81.02	81.39	81.75	82.05
P	19.46	32.02	32.72	32.41	32.52	34.09	34.66	35.25	35.64
R	71.60	55.55	55.55	58.02	58.24	64.19	64.19	64.19	63.58
F	30.60	40.63	41.18	41.59	41.68	44.53	45.02	45.51	45.67

Table 3: Performance of (B)aseline, +$N_{!M}$, +$N_{!M}$+$V_{\{R\}}$, +N_{M_1}, +$N_{M_{1_{GR}}}$, +$N_{M_{1_{GR}}+M_2}$, +N_M+VF, +N_M+$V_{\{A,R,IS\}}$ and +N_M+VF+$V_{\{A,R,IS\}}$ (see text for details) in terms of (S)cores of (A)ccuracy, (P)recision, (R)ecall, (F)-score. The row 1 (2) of this table presents results over NB (MaxEnt) baseline supervised classifier, respectively.

Table 3 shows that MaxEnt gives a very high accuracy and F-score as compared with NB. Model +$N_{!M}$+$V_{\{R\}}$ with basic noun and verb semantics introduced in section 3.2 results in more than 10% improvement in F-score over NB and MaxEnt classifiers relying only on shallow linguistic features. Model +N_M+VF+$V_{\{A,R,IS\}}$ with enriched verb and noun semantics brings more than 4% improvement in F-score over +$N_{!M}$+$V_{\{R\}}$ with MaxEnt as baseline. We perform statistical significance test using bootstrap sampling method given in Berg-Kirkpatrick et al. (2012) (see Berg-Kirkpatrick et al. (2012) for the details). +N_M+VF+$V_{\{A,R,IS\}}$ brings significant improvement in F-score over +$N_{!M}$+$V_{\{R\}}$ with p-value 0.0.

Though +$N_{!M}$ gives significantly better F-score over baseline, it drops recall by more than 16%. Metonymy resolution helps perform quite better by recovering more than 8% recall with +$N_{M_{1_{GR}}+M_2}$ over +$N_{!M}$. +$N_{M_{1_{GR}}+M_2}$ also results in 3.9% improvement in F-score over +$N_{!M}$ with MaxEnt as baseline model (significant improvement with p-value 0.0). Metonymies resolved via verb frames with all and core grammatical relations (i.e., set GR) recover more than 2% recall and slightly improve F-score.

Model with the addition of information of verb frames (i.e.,+N_M+VF) brings 0.49% improvement in F-score over +$N_{M_{1_{GR}}+M_2}$ using MaxEnt as baseline model (significant improvement with p-value 0.027). Model with the addition of data-driven verb semantics (i.e., +N_M+$V_{\{A,R,IS\}}$) results in 0.98% improvement in F-score over +$N_{M_{1_{GR}}+M_2}$ using MaxEnt as baseline model (significant improvement with p-value 0.0021). Overall the model +N_M+VF+V$\{A, R, IS\}$ yields more than 16% (20%) F-score (accuracy) over the baseline models build via NB and MaxEnt.

5.1 Error Analysis

We performed error analysis for the model +N_M+VF+$V_{\{A,R,IS\}}$ by randomly selecting 50 False Positives (FP) and 50 False Negatives (FN).

For 32% FP instances information about verb frames is not available in the knowledge base of verb frames. To avoid this problem researchers should exploit some abstractions e.g., {semantic sense of v, gr} frames. Our model fails to identify the class $\neg C_{np}$ for noun_phrases of 29% FP instances due to the lack of enough training data for the semantic classes of nouns. In 21% FP instances v and np are not even relevant to each other. Our model first needs to determine relevance between v and np before identifying causality. Remaining 18% instances have v and np in temporal only sense, comparison relation or both represent parts of same event. There is need to extract more knowledge sources to better distinguish causality from any other type of relation.

77% FN instances are classified as non-causal due to the lack of enough v-np training data and require more sources of knowledge e.g., background knowledge. On remaining 23% FN instances our model fails to identify C_{np} class due to the lack of enough training data for the semantic classes of nouns.

6 Conclusion

This work has revealed that enriched semantics of nouns and verbs help gain significant improvement in performance over a baseline relying only on shallow linguistic features. Through empirical evaluation and error analysis of our model we have highlighted strengths and weaknesses of our model for the current task. Our work has provided a novel direction to exploit semantics of participants of causal relations to solve the challenge of identifying causality.

References

Collin F. Baker, Charles J. Fillmore and John B. Lowe. 1998. The Berkeley FrameNet project. *In proceedings of COLING-ACL. Montreal, Canada*.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in NLP. *In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)..*

Steven Bethard and James H. Martin. 2006. Identification of Event Mentions and their Semantic Class. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Steven Bethard and James H. Martin. 2008. Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations. *In proceedings of ACL-08: HLT*.

Timothy Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. *In proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-04). Barcelona, Spain.*

Quang X. Do, Yee S. Chen and Dan Roth. 2011. Minimally Supervised Event Causality Identication. *In proceedings of EMNLP-2011.*

Jenny R. Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *In Proceedings of the Association for Computational Linguistics (ACL)*.

Roxana Girju and Dan Moldovan. 2002. Mining Answers for Causation Questions. *In American Associations of Artificial Intelligence (AAAI), 2002 Symposium.*

Roxana Girju. 2003. Automatic detection of causal relations for Question Answering. *Association for Computational Linguistics ACL, Workshop on Multilingual Summarization and Question Answering Machine Learning and Beyond 2003*.

Katja Markert and Malvina Nissim 2009. Data and models for metonymy resolution. *Language Resources and Evaluation Volume 43 Issue 2, Pages 123−138*.

Marie-Catherine de Marneffe, Bill MacCartney and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. *In Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.

Andrew K. McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu.

Emily Pitler, Annie Louis and Ani Nenkova. 2009. Automatic Sense Prediction for Implicit Discourse Relations in Text. *In proceedings of ACL-IJCNLP, 2009.*

Emily Pitler and Ani Nenkova. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. *In proceedings of ACL-IJCNLP, 2009.*

James Pustejovsky, Patrick Hanks, Roser Saur, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro and Marcia Lazo. 2003. The TIMEBANK Corpus. *In Proceedings of Corpus Linguistics*.

Kira Radinsky and Eric Horvitz. 2013. Mining the Web to Predict Future Events. *In proceedings of sixth ACM international conference on Web search and data mining, WSDM '13.*

Mehwish Riaz and Roxana Girju. 2010. Another Look at Causality: Discovering Scenario-Specific Contingency Relationships with No Supervision. *In proceedings of the IEEE 4th International Conference on Semantic Computing (ICSC)*.

Mehwish Riaz and Roxana Girju 2013. Toward a Better Understanding of Causality between Verbal Events: Extraction and Analysis of the Causal Power of Verb-Verb Associations. *Proceedings of the annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*.

Mehwish Riaz and Roxana Girju 2014. Recognizing Causality in Verb-Noun Pairs via Noun and Verb Semantics. *Proceedings of the Workshop on Computational Approaches to Causality in Language EACL, 2014*.

Dan Roth and Wen-tau Yih 2004. A Linear Programming Formulation for Global Inference in Natural Language Tasks. *In Proceedings of the Annual Conference on Computational Natural Language Learning (CoNLL)*.

Caroline Sporleder and Alex Lascarides. 2008. Using automatically labelled examples to classify rhetorical relations: An assessment. *Journal of Natural Language Engineering Volume 14 Issue 3, July 2008 Pages 369−416*.

Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *In Proceedings of Human Language Technology and North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.

Appendix A. Supervised Classifier

In this appendix, we provide a set of linguistic features taken from Riaz and Girju (2014) to identify causality in v-np pairs employing a supervised classifier (see section 3.1 for the details).

- **Lexical Features:** verb, lemma of verb, noun_phrase, lemmas of all words of noun_phrase, head noun of noun_phrase, lemmas of all words between verb and noun_phrase.
- **Syntatic Features:** part-of-speech tags of verb and head noun of noun_phrase.
- **Semantic Features:** We adopted this feature from Girju (2003) to capture semantics of nouns. The 9 noun hierarchies of WordNet i.e., entity, psychological feature, abstraction, state, event, act, group, possession, phenomenon are used as this feature. Each of these hierarchies is set to 1 if any sense of head noun of noun_phrase lies in that hierarchy, otherwise set to 0.
- **Structural Features:** This feature is applied by considering both subject (i.e., sub_in_np) and object (i.e., obj_in_np) of verb (v). For example, for the pair v-np the variable sub_in_np is set to 1 if the subject $\in$ np, set to 0 if the subject $\notin$ np and set to -1 if the subject is not available in an instance. The subject and object of a verb are its core arguments and may sometime be part of an event represented by a verb. Therefore, these arguments may have high tendency to encode non-causation with their verb.
- **Pairs:** The following pairs (verb, head noun of noun_phrase), (subject$_{verb}$, head noun of noun_phrase) and (object$_{verb}$, head noun of noun_phrase) are used to capture relations.

Appendix B. Noun Semantics

In this appendix, some examples of the frame elements of FrameNet and the WordNet senses belonging to the classes $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$ are given in Tables 4 and 5 (see section 3.2 for the details). We employ training instances acquired using the FrameNet annotations and WordNet senses for building a supervised classifier for the classes $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$. Following is the list of features we use for this supervised classifier:

- **Lexical Features:** All words of noun_phrase, lemmas of all words of noun_phrase, head noun of noun_phrase, first two (three) (four) letters of head noun of noun_phrase, last two, (three) (four) letters of head noun of noun_phrase.
- **Word Class Features:** part-of-speech tags of all words of noun_phrase and head noun of noun_phrase.
- **Semantic Features:** Frequent sense of head noun of noun_phrase.

SC	FrameNet Labels
c_{np}	Event, Goal, Purpose, Cause, Internal cause, External cause, Result, Means, Reason, Phenomena, Coordinated event, Action, Activity, Circumstances, Desired goal, Explanation
$\neg c_{np}$	Artist, Performer, Duration, Time, Place, Distributor, Area, Path, Direction, Sub-region Frequency, Body part, Area, Degree, Angle, Fixed location, Path shape, Addressee, Interval

Table 4: Some examples of the frame elements of FrameNet to which we assign the semantic classes $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$.

SC	WordNet Senses
c_{np}	{act, deed, human action, human activity}, {phenomenon}, {state}, {psychological feature}, {event}, {causal agent, cause, causal agency}
$\neg c_{np}$	{time period, period of time, period}, {measure, quantity, amount}, {group, grouping}, {organization, organisation}, {time unit, unit of time}, {clock time, time}

Table 5: This table shows our selected WordNet senses of nouns belonging to classes $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$. For example, using the information provided in this table we assume that any noun concept whose all senses of WordNet lie in the semantic hierarchy of the sense {time period, period of time, period} is of class $\neg\mathrm{C_{np}}$. We use English Gigaword corpus to collect instances of noun (or noun_phrases) and label them with $\mathrm{C_{np}}$ and $\neg\mathrm{C_{np}}$ according to their senses in WordNet.