Discovering verb relations in corpora: distributional versus non-distributional approaches

Maria Teresa Pazienza[1], Marco Pennacchiotti[1], and Fabio Massimo Zanzotto[2]

[1] University of Roma Tor Vergata, Via del Politecnico 1, Roma, Italy,
{pazienza, pennacchiotti}@info.uniroma2.it,
[2] DISCo, University of Milano Bicocca, Via B. Arcimboldi 8, Milano, Italy,
zanzotto@disco.unimib.it

Abstract. Verbs represent a way in which ontological relationships between concepts and instances are expressed in natural language utterances. Moreover, an organized network of semantically related verbs can play a crucial role in applications. For example, if a Question-Answering system could exploit the direction of the entailment relation win → play, it may expand the question "Who played against Liverpool?" with "X won against Liverpool" and it may avoid the expansion of "Who won against Liverpool?" in "X played against Liverpool" that would be wrong. In this paper, we present a survey of the methods proposed to extract verb relations in corpora. These methods can be divided in two classes: those using the Harris distributional hypothesis and those based on point-wise assertions. These methods are analysed and compared.

1 Introduction

Learning domain ontologies (or semantic models) from texts is a largely debated problem that undergoes to different names according to the research area: terminology extraction [1], named entity recognition [2], lexical acquisition, and information extraction [3].

Models for terminology extraction and structuring [1] have always had the aim to extract terms from domain corpora and structuring them in *is-a* hierarchies [4] or different conceptual networks. Terms are strictly related to ontologies as they are the linguistic counterpart of domain concepts.

Named entity extraction as well as information extraction give the possibility of populating ontologies with concept and relationship instances. Traditionally, these techniques are based on a pre-existing semantic model (or ontology) of the domain defining relevant concepts called named entity categories (e.g., `location`, `team`, `person`) and relevant relations between concepts called templates (e.g., `Match(Home,Visitors,Result,Score,Date)`). Given a text, the resulting systems aim to extract named entities, that are instances of the categories, and fillers for the template slots. For example, given the sentence "*Marcus Camby won for NY at the Yankees Stadium*", these systems have to extract:

– 3 named entities: "*Marcus Camby*" as instance of `person`, "*NY*" as instance of `team`, and "*Yankees Stadium*" as instance of `location`;
– a partial instance of the `Match` template:

`Match(New_York_Yankees,-,HomeWinner,-,-)`

These activities are clearly very difficult. What is written in the sample sentence is distant with respest to its interpretation according to the template. However, some systems really achieve good performances in some Information Extraction tasks [2].

Ontology learning from texts is not then proposing a new challenge to natural language processing (NLP). It is pushing the scalability of these methods. These latter give the basis over which more scalable models can be produced.

In particular, verb and verb phrases, one of the linguistic counterpart of ontological relationships between concepts and instances, have been fairly studied. A clear model of verb semantics could be helpful in determining a clear model of the relationships between concepts. Relations between verbs play then a very important role. This will be clarified with an example. Consider the question "*What college did Marcus Camby play for?*". A system could find the answer in the snippet "*Marcus Camby won for Massachusetts*" only if the question verb *play* is related to the verb *win* according to some ontological resource even if *play* and *win* have a different meaning.

Understanding verb relations and discovering such a relations between verbs in corpora is a very interesting research field. This can give some insights to model better ontological resources. Moreover, some results and methods to induce such relations between verbs from corpora are already available. These are mainly based on two hypotheses:

– the Harris' Distributional Hypothesis [5]
– the pointwise assertion hypothesis [6]

These two hypotheses originated very different methods for discovering relations between words in general and between verbs in particular.

The rest of the paper is organised as follows. Sec. 2 shortly revises the possible verb relations. This give an important view on how relationships between concepts should be modelled. Sec. 3 and Sec. 4 describe the automatic models to extract these verb relations from corpora. Sec. 3 and Sec. 4 respectively describes the methods based on the distributional hypothesis and the pointwise assertion methods.

2 Verb relations: a classification

WordNet [7] is a very large lexcial knowledge base. Its organization can give hints to better understand which relations hold between verbs. Verbs are basically 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 [7] 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 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 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* $Y \rightarrow Y$ *have*), as 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.

3 Distributional Semantics

3.1 The Harris' Distributional Hypothesis

The Distributional Hypothesis [5] has been fairly explored and exploited in learning similarity between words and between more structured expressions. The hypothesis is the following:

> *Words that tend to occur in the same contexts tend to have similar meanings.*

This seems to be a simple statement but it has given very interesting results. It is a very strong assertion as word semantics is modelled using only contextual information. Target words themselves are only used to select contexts from a corpus.

Trying to formalize, given a set of words W and a set of contexts Ctx drawn from a corpus C, we can define a function that associate the contexts to each subset of words:

$$\mathcal{C} : 2^W \rightarrow 2^{Ctx} \tag{1}$$

The Distributional Hypothesis suggests that the similarity sim_w between two sets of words, W_1 and W_2, is computable as the simililarity sim_{ctx} between the related sets of contexts:

$$sim_w(W_1, W_2) \approx sim_{ctx}(\mathcal{C}(W_1), \mathcal{C}(W_2)) \tag{2}$$

The problem is then how to compute the similarity between contexts. Contexts are generally represented on a feature space $F = F_1 \times ... \times F_n$. As any vector space model, this space hosts also the intensional representation of sets of contexts $\mathcal{I}(Ctx')$ (e.g., their centroids). The function that computes the intesional representations is:

$$\mathcal{I} : 2^{Ctx} \rightarrow F \tag{3}$$

The similarity between two sets of contexts (or between their intensional represetations) is then computed over the space F. In the following we will use $w \in W$ and $c \in Ctx$ in the similarities to indicate, respectively, the singletons $\{w\}$ and $\{c\}$.

As any other Machine Learning problem, the methods applying this principle differ according to:

– how contexts are represented in features of the space $\mathcal{F}$ (using bag-of-word models, syntactic representations, or semantic models);
– how, evetually, intensional representations of sets of contexts are computed;
– how the similarity measure over contexts is defined.

Before starting the analysis of the method we need to give another classification. This latter depends on how the similarity is used in an algorithm that generates a lexicon of similar words starting from a corpus C. We can distinguish to two ways: a **direct** approach exploit directly the similarity and an **indirect** approach that, working in the context space, select words sharing similar contexts. In the following, we rewiew some of these learning algorithm with respect to this classification.

3.2 Discovering similarity using the *direct* approach

The **direct** approach is fairly studied and it is applicable when the corpus C is a-priori known. Given a set W of relevant words or linguistic structures, the corpus C is scanned for each element $w \in W$. Contexts for each w are gathered and represented in the feature space $\mathcal{F}$. Pairs (w', w'') in $W \times W$ are ranked according to $sim(w', w'')$ and these are retained as good pairs if their similarity is greater than a threshold α, i.e. $sim(w', w'') > \alpha$. Sometimes, to improve selectiveness this threshold depends on one of the two elements, i.e. $\alpha_{w'}$. The applicability to known corpora depends on two factors. The first is that the set W has to be defined a-priori. This is generally done analysing the frequency, or similar indicators, of the words w in the corpus. The second applicability limitation depends on the fact that the contexts of each w have to be represented in the $\mathcal{F}$ space. This means that the corpus has to be completely scanned. Only after the computation of the similarity between different words can be done.

The general algorithm DH_direct(C), that directly applies the Distributional Hypothesis, is described in the followings. It takes in input a corpus C and return a set of equivalence classes W_i containing words or linguistic sturctures:

DH_direct(C) returns $(W_1,...,W_N)$

Given a corpus C and the releted function $\mathcal{C}$:

1. Let W be the most important elements in C
2. For each $w \in W$:
   (a) retreive all the contexts $\mathcal{C}(w)$ and map them in the feature space F
   (b) eventually compute $\mathcal{I}(\mathcal{C}(w))$
3. For each $(w_i, w_j) \in W \times W$:
   (a) comupute $sim_{ctx}(\mathcal{C}(w_i), \mathcal{C}(w_j))$
   (b) if $sim_{ctx}(\mathcal{C}(w_i), \mathcal{C}(w_j)) > \alpha_{w_i}$ then put w_j in W_i

The above algorithm has been largely employed. As we already discussed in the previous sections, the methods differ for the target of the analysis and the feature space in which the similarity is computed.

In [8], the DH_direct has been used to extract similarity between verbs from a single corpus. The elements represented in the feature space F were then intensional representations of verb contexts. The actual features were the determined by the pair $(VerbArg, ArgFiller)$ where $VerbArg$ is one of the possible verb arguments (e.g., the $subject, object, modifier-from, modifier-in, ...$) and $ArgFiller$ is a word sequence. The value of this feature was related to the frequency and the inverse document frequency.

In [9], the algoritm has been used to discover equivalence relations between verbal linguistic expressions connecting two arguments X and Y, e.g. *X solved Y* $\approx$ *X found a solution to Y*. Each one of these verbal linguistic expressions is called pattern p. The idea there was to represent in a feature space the possible filler of the slots X and Y. The feature space represents intensionally a set of contexts of each pattern. The features were (s, w) where s is the slot X or Y and w is a possible word filling the slot. Given a set on contexts where the pattern *pat* has been found (the set will be called *pat* as the pattern), the feature values of the $\mathcal{I}(pat)$ are computed as follows:

$$\mathcal{I}_{(s,w)}(pat) = mi(pat, s, w) \quad (4)$$

where $mi(pat, s, w)$ is the point-wise mutual information [10]:

$$mi(pat, s, w) = log\frac{p(pat, s, w)}{p(pat, s)p(s, w)} \quad (5)$$

Probabilities are estimated with the maximun likelihood principle over the corpus. We define the vectors I_s (where s is the slot X or Y) as the vectors having 1 in the feature (s, w) and 0 otherwise. The similarity $sim(p_1, p_2, s)$ between p_1 and p_2 according to the slot s is defined as follows :

$$sim(p_1, p_2, s) = \frac{(\mathcal{I}(p_1) + \mathcal{I}(p_2))I(p_1, p_2)I_s^T}{\mathcal{I}(p_1)I_s^T + \mathcal{I}(p_2)I_s^T} \quad (6)$$

where $I(p_1, p_2)$ is a matrix whose elements on the diagonal are:

$$I_{(s,w),(s,w)}(p_1, p_2) = \begin{cases} 1 \texttt{ if } \mathcal{I}_{(s,w)}(p_1) > 0 \texttt{ and } \mathcal{I}_{(s,w)}(p_2) > 0 \\ 0 \texttt{ otherwise} \end{cases}$$

and elements out of the diagonal are 0. The similarity $sim_p(p_1, p_2)$ between two patterns is then computed as follows:

$$sim_p(p_1, p_2) = \sqrt{sim(p_1, p_2, X) \times sim(p_1, p_2, Y)} \quad (7)$$

3.3 Anchor-based algorithms: *indirect* approaches

The **indirect** approaches (e.g. [11–13]) have been proposed to apply the Distributional Hypothesis also when the corpus C is a-priori not known (e.g. the Web that is potentially infinite). The problem there is that the set W can not be computed in advance. If W were somehow given, the more similar words to a word w can be found only when $sim_w(w, w')$ has been computed for each $w' \in W$. This is generally unfeasible due to the large size of the corpus. The pursued idea is the following. Given a seeding set of words or linguistic structures W_S considered similar or realising a unique semantic relation (e.g., the *is-a* relation such as in [11]), a set C_S of prototypical contexts are extracted. Each element c in the set C_S that has some important property is called *anchor* [**?**]. An anchor highly characterises the contexts where the set of words W_S appears. For each of the element c, it is then possible to derive the set of words:

$$W_c = \mathcal{C}^{-1}(\{c' \in CTX | sim_{ctx}(c, c') > \alpha\}) \quad (8)$$

These sets can be used to enrich the original set of word W with similar words. The similarity is always estimated using the similarity between contexts.

DH_indirect(C,(W'_1,...,W'_N)) returns (W_1,...,W_N)

Given a corpus C and the releted function $\mathcal{C}$:

1. For each W'_i:
   (a) $W_i = W'_i$:
   (b) For each relevant set $W' \subseteq W'_i$:
      i. compute $I_{W'} = \mathcal{I}(\mathcal{C}(W'))$
      ii. extract $W'' =$ $\mathcal{C}^{-1}(\{c' \in CTX | sim_{ctx}(c, I_{W'}) > \alpha\})$
      iii. select the most relevant $W''' \subseteq W''$
      iv. $W_i = W_i \cup W'''$

Hearst [11] by first mined a wide collection of texts to identify, in a set of frequently occurring lexico-syntactic patterns, lexical relations of interest. Main focus was in discovering patterns highlighting hyponymyc lexical relationship between two or more noun phrase in naturally-occurring text. Basic assumption was that the structure of a language can indicate the meanings of lexical items.

The problem lies in finding surface linguistic expressions that, frequently and reliably, indicate relations of interest. Limited to hyponymy relation, Hearst identified a set of lexico-syntactic patterns:

– occurring frequently and in several text genres
– recognizable with little or no pre-encoded knowledge

By first these patterns were discovered by observation, then, to find new patterns automatically, a list of terms for which the specified relation hold was gathered. The corpus was mined to find expressions in which these terms occurred near one another. Then commonalities among these environments were extracted and assumed to indicate the relation of interest. Each time the specific list of terms is used as an anchor. Hearst declares the method has been successfull for hyponimy relation, while not for meronymy. Such a different result has been ascribed to the "naming" nature of hyponymy relation.

Ravichandran and Hovy [12] explored the power of surface text patterns for open-domain Question Answering systems. They recognized that in several Q/A systems specific types of answer are expressed by using characteristic phrases (that could be described in regular expressions). They described a pattern-learning algorithm and focused on scaling relation extraction to the Web. In fact with their algorithm it is possible to infer surface patterns from a small set of instance seeds (the anchor in this approach) by extracting substrings relating seeds in corpus sentences. The presence of any variant of the answer term causes the same treatement as for the original answer term. Nevertheless the patterns cannot handle long-distance dependencies. The approach has been tested on several relations providing good results for specific relations (such as birthdate) while lower precision revealed for generic and frequent ones (as is-a and part-of). Also this algorithm is in the line of learning then extracting approach.

Szpektor et al. [13] defined a fully unsupervised learning algorithm for web-based extraction of entailment relations. By assuming that the same meaning can be expressed in a text in a huge variety of surface forms, they focused on acquiring paraphrase patterns that represent different forms in which a certain meaning can be expressed. This approach has been applied to acquire entailment relations from the Web. Paraphrase acquisition results in finding linguistic structures (called templates) sharing the same lexical elements describing the context of a sentence. These lexical elements are used as anchors. Templates extracted from different sentences, while connecting the same anchors, are assumed to paraphrase each other. We recognize a structuring model for which we distinguish between syntactic anchors (such as subject, object, verb) and a context anchor (such as prepositional phrase). Main problems relate to both finding matching anchors and identifying template structure. Specific algorithms have been developed for both problems. A broad range of semantic relations varying from synonymy to more complex entailment have been extracted.

4 Non-distributional approaches

Since the Distributional Hypothesis [5] suggests equivalence between words, the related methods can discover only symmetric relations. However, consider again the question "*What college did Marcus Camby play for?*". A 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 were "*What college did Marcus Camby won for?*", the snippet "*Marcus Camby played for Massachusetts*" cannot be used. This is why *winnig* entails *playing* but not vice-versa. The relation between *win* and *play* is asymmetric. These kinds of relation cannot be easily discovered using the distributional hypothesis.

The idea that some *point-wise assertions* carry relevant semantic information (as in [6]). This point-wise assertions can be detected at three levels: at the *sentence* level as done in [14], at the level of relations between sentences [15], and, finally, at the level of document as in [16]. In the following sections we will review these methods.

4.1 Single sentence patterns

In [14] it has been observed that class-level and word-level selectional preferences [17] offer a very interesting place where to search for these asymmetric entailment relations between verbs. Indeed, selectional preferences indicate an entailment relations between verbs 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 $win(x) \rightarrow play(x)$.

To exploit the previous principle for entailment detection, we need to find the specific verb selectional preferences. Our method consists of two steps. Firstly, it is necessary to translate the verb selectional expectation in specific Subject-Verb lexico-syntactic patterns ($\mathcal{P}(v_t, v_h)$). Secondly, we need to define a statistical measure ($\mathcal{S}(v_t, v_h)$) that captures the verb preference. This measure will describe how much relations between target verbs (v_t, v_h) are *stable* and commonly agreed.

The above idea requires a 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 personified hypothesis verb is the subject of a clause governed by the text verb* v_t. Given the two investigated verbs (v_t, v_h) the assertion is formalized in a set of textual entailment lexico-syntactic patterns, that we call *nominalized patterns* $\mathcal{P}_{nom}(v_t, v_h)$. This set will contain the following textual patterns:

$\mathcal{P}_{agent}(v_t, v_h) =$
$\{$"$agent(v_h)|_{number:sing}\quad v_t|_{person:third,tense:present}$",
"$agent(v_h)|_{number:plur}\quad v_t|_{person:nothird,tense:present}$",
"$agent(v_h)|_{number:sing}\quad v_t|_{tense:past}$",
"$agent(v_h)|_{number:plur}\quad v_t|_{tense:past}$"$\}$

where $agent(v)$ is the noun deriving from the personification of the verb v and elements such as $l|_{f_1,...,f_N}$ are the tokens generated from lemmas l by applying constraints expressed via the feature-value pairs $f_1, ..., f_N$. For example, in the case of the verbs *play* and *win*, the related set of textual entailment expressions derived from the patterns are $\mathcal{P}_{nom}(win, play) = \{$"`player wins`", "`players win`", "`player won`", "`players won`"$\}$.

Given a pair v_t and v_h we may thus define the following *entailment strength indicator* $\mathcal{S}(v_t, v_h)$. Specifically, the measure $\mathcal{S}_{nom}(v_t, v_h)$ we use is derived from point-wise mutual information [10]:

$$\mathcal{S}_{nom}(v_t, v_h) = log\frac{p(v_t, v_h|nom)}{p(v_t)p(v_h|pers)} \tag{9}$$

where *nom* is the event of having a nominalized textual entailment pattern and *pers* is the event of having an agentive nominalization of verbs.

4.2 Multiple sentence patterns

The lexico-syntactic patterns introduced in [15] have been developed to detect six kinds of verb relations: *similarity*, *strength*, *antonymy*, *enablement*, and *happens-before*. Even if, as discussed in [15], these patterns are not specifically defined as entailment detectors, they can be fairly useful for this purpose. In particular, some of these patterns can be use to investigate the *backward-presupposition* entailment. Verb pairs related by backward-presupposition are not completely temporally included one in the other (cf. Sec. 2): the entailed verb v_h precedes the entailing verb v_t. One set of lexical patterns in [15] seem to capture the same idea: the *happens-before* (*hb*) patterns. Indeed, it is used to detect not temporally overlapping verbs, whose relation is semantically very similar to entailment. These patterns in fact show a positive relation with the entailment relation under investigation. The following table reports the *happens-before* lexico-syntactic patterns ($\mathcal{P}_{hb}$) described in [15]:

$\mathcal{P}_{hb}(v_t, v_h) =$
$\{$"$v_h|_{t:inf}$ `and then` $v_t|_{t:pres}$",
"$v_h|_{t:inf}$ `* and then` $v_t|_{t:pres}$",
"$v_h|_{t:past}$ `and then` $v_t|_{t:pres}$",
"$v_h|_{t:past}$ `* and then` $v_t|_{t:pres}$",
"$v_h|_{t:inf}$ `and later` $v_t|_{t:pres}$",
"$v_h|_{t:past}$ `and later` $v_t|_{t:pres}$",
"$v_h|_{t:inf}$ `and subsequently` $v_t|_{t:pres}$",
"$v_h|_{t:past}$ `and subsequently` $v_t|_{t:pres}$",
"$v_h|_{t:inf}$ `and eventually` $v_t|_{t:pres}$",
"$v_h|_{t:past}$ `and eventually` $v_t|_{t:pres}$"$\}$

where `*` is mached with any word. For example, for the classical pair $(v_t, v_h) = (win, play)$, the *happen-before* patterns are realised in $\mathcal{P}_{hb}(win, play) = \{$"`play * and then win`", "`played and then win`", ..., "`play and eventually win`"$\}$. Also in [15], a mutual-information-related measure is used as statistical indicator. The two methods thus seem to be fairly in line.

4.3 Document patterns

The other approach we analyse is the "quasi-pattern" used in [16] 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 that the probability that H holds alone, i.e.:

$$p(H|T) > p(H) \tag{10}$$

In [16], words in the two sentences H and T are supposed to be mutually independent: consequently, the previous relation between H and T probabilities holds also for word pairs. Then, a special case can be applied to verb pairs:

$$p(v_h|v_t) > p(v_h) \tag{11}$$

Equation (11) 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 thus possible to formalize the idea in *probabilistic entailment "quasi-patterns"* $\mathcal{P}_{pe}$ reported in the following:

$$\begin{array}{l}
\mathcal{P}_{pe}(v_t, v_h) = \\
\{\text{``}v_h|_{person:third,t:pres}\text{''} \wedge \text{``}v_t|_{person:third,t:pres}\text{''}, \\
\text{``}v_h|_{t:past}\text{''} \wedge \text{``}v_t|_{t:past}\text{''}, \\
\text{``}v_h|_{t:pres\_cont}\text{''} \wedge \text{``}v_t|_{t:pres_cont}\text{''}, \\
\text{``}v_h|_{person:nothird,t:pres}\text{''} \wedge \text{``}v_t|_{person:nothird,t:pres}\text{''}\}
\end{array}$$

Also according to [16] point-wise mutual information is a relevant statistical indicator for entailment, as it is strictly related to eq. (11).

5 Conclusions

In this paper, we presented a survey of the methods proposed to extract verb relations in corpora. We tried to demonstate that these standard methods can be divided in two classes: methods using the Harris distributional hypothesis an methods based on point-wise assertions. As ontology learning from texts is not then proposing a new challenge to natural language processing (NLP) but is pushing the scalability of these methods, we believe that reviewing these traditional method can help in building better models for this task.

References

1. Jacquemin, C.: Spotting and Discovering Terms through Natural Language Processing. Massachusetts Institue of Technology, Cambrige, Massachussetts, USA (2001)
2. MUC-7: Proceedings of the seventh message understanding conference (MUC-7). In: Columbia, MD, Morgan Kaufmann (1997)
3. Pazienza, M.T., ed.: Information Extraction. A Multidisciplinary Approach to an Emerging Information Technology. Number 1299 in LNAI. Springer-Verlag, Heidelberg, Germany (1997)
4. Morin, E.: Extraction de liens sémantiques entre termes à partir de corpus de textes techniques. PhD thesis, Univesité de Nantes, Faculté des Sciences et de Techniques (1999)
5. Harris, Z.: Distributional structure. In Katz, J.J., Fodor, J.A., eds.: The Philosophy of Linguistics, New York, Oxford University Press (1964)
6. Robison, H.R.: Computer-detectable semantic structures. Information Storage and Retrieval **6** (1970) 273–288
7. Miller, G.A.: WordNet: A lexical database for English. Communications of the ACM **38** (1995) 39–41
8. Glickman, O., Dagan, I.: 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 (2003)
9. Lin, D., Pantel, P.: DIRT-discovery of inference rules from text. In: Proc. of the ACM Conference on Knowledge Discovery and Data Mining (KDD-01), San Francisco, CA (2001)
10. Church, K.W., Hanks, P.: Word association norms, mutual information and lexicography. In: Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada (1989)
11. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 15th International Conference on Computational Linguistics (CoLing-92), Nantes, France (1992)
12. Ravichandran, D., Hovy, E.: Learning surface text patterns for a question answering system. In: Proceedings of the 40th ACL Meeting, Philadelphia, Pennsilvania (2002)
13. Szpektor, I., Tanev, H., Dagan, I., Coppola, B.: Scaling web-based acquisition of entailment relations. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcellona, Spain (2004)
14. Zanzotto, F.M., Pazienza, M.T., Pennacchiotti, M.: Discovering entailment relations using textual entailment patterns. In: Proc. of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, Ann Arbor, Michigan, Association for Computational Linguistics (2005) 13–18
15. Chklovski, T., Pantel, P.: VerbOCEAN: Mining the web for fine-grained semantic verb relations. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcellona, Spain (2004)
16. Glickman, O., Dagan, I., Koppel, M.: Web based probabilistic textual entailment. In: Proceedings of the 1st Pascal Challenge Workshop, Southampton, UK (2005)
17. Resnik, P.: Selection and Information: A Class-Based Approach to Lexical Relationships. PhD thesis, Department of Computer and Information Science, University of Pennsylvania (1993)