PhraseNet: Towards Context Sensitive Lexical Semantics*

Xin Li†, Dan Roth†, Yuancheng Tu‡
Dept. of Computer Science†
Dept. of Linguistics‡
University of Illinois at Urbana-Champaign
{xli1,danr,ytu}@uiuc.edu

Abstract

This paper introduces PhraseNet, a context-sensitive lexical semantic knowledge base system. Based on the supposition that semantic proximity is not simply a relation between two words in isolation, but rather a relation between them in their context, English nouns and verbs, along with contexts they appear in, are organized in PhraseNet into *Consets*; Consets capture the underlying lexical concept, and are connected with several semantic relations that respect contextually sensitive lexical information. PhraseNet makes use of WordNet as an important knowledge source. It enhances a WordNet synset with its contextual information and refines its relational structure by maintaining only those relations that respect contextual constraints. The contextual information allows for supporting more functionalities compared with those of WordNet. Natural language researchers as well as linguists and language learners can gain from accessing PhraseNet with a word token and its context, to retrieve relevant semantic information.

We describe the design and construction of PhraseNet and give preliminary experimental evidence to its usefulness for NLP researches.

1 Introduction

Progress in natural language understanding research necessitates significant progress in lexical semantics and the development of lexical semantics resources. In a broad range of natural language applications, from prepositional phrase attachment (Pantel and Lin, 2000; Stetina and Nagao, 1997), co-reference resolution (Ng and Cardie, 2002) to text summarization (Saggion and Lapalme, 2002), semantic information is a necessary component in the inference, by providing a level of abstraction that is necessary for robust decisions.

Inducing that the prepositional phrase in "`They ate a cake with a fork`" has the same grammatical function as that in "`They ate a cake with a spoon`", for example, depends on the knowledge that "cutlery" and "tableware" are the hypernyms of both "fork" and "spoon". However, the noun "fork" has five senses listed in WordNet and each of them has several different hypernyms. Choosing the correct one is a context sensitive decision.

WordNet (Fellbaum, 1998), a manually constructed lexical reference system provides a lexical database along with semantic relations among the lexemes of English and is widely used in NLP tasks today. However, WordNet is organized at the *word* level, and at this level, English suffers ambiguities. Stand-alone words may have several meanings and take on relations (e.g., hypernyms, hyponyms) that depend on their meanings. Consequently, there are very few success stories of automatically using WordNet in natural language applications. In many cases, reported (and unreported) problems are due to the fact that WordNet enumerates all the senses of polysemous words; attempts to use this resource automatically often result in noisy and non-uniform information (Brill and Resnik, 1994; Krymolowski and Roth, 1998).

PhraseNet is designed based on the assumption that, by and large, semantic ambiguity in English disappears when local context of words is taken into account. It makes use of WordNet as an important knowledge source and is generated *automatically* using WordNet and machine learning based processing of large English corpora. It enhances a WordNet synset with its contextual information and refines its relational structure, including relations

*Research supported by NSF grants IIS-99-84168, ITR-IIS-00-85836 and an ONR MURI award. Names of authors are listed alphabetically.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2003		**00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
PhraseNet: Towards Context Sensitive Lexical Semantics	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Illinois,Department of Computer Science ,Urbana,IL,61801**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**8**	

such as hypernym, hyponym, antonym and synonym, by maintaining only those links that respect contextual constraints. However, PhraseNet is not just a functional extension of WordNet. It is an independent lexical semantic system allied with proper user interfaces and access functions that will allow researchers and practitioners to use it in applications.

As stated before, PhraseNet, is built on the assumption that linguistic context is an indispensable factor affecting the perception of a semantic proximity between words. In its current design, PhraseNet defines "context" hierarchically with three abstraction levels: abstract syntactic skeletons, such as

$$[(S) - (V) - (DO) - (IO) - (P) - (N)]$$

which stands for Subject, Verb, Direct Object, Indirect Object, Preposition and Noun(Object) of the Preposition, respectively; syntactic skeletons whose components are enhanced by semantic abstraction, such as $[\boldsymbol{Peop} - send - \boldsymbol{Peop} - gift - on - \boldsymbol{Day}]$ and finally concrete syntactic skeletons from real sentences as $[they - send - mom - gift - on - Christmas]$.

Intuitively, while "candle" and "cigarette" would score poorly on semantic similarity without any contextual information, their occurrence in sentences such as "`John tried to light a candle/cigarette`" may highlight their connection with the process of burning. PhraseNet captures such constraints from the contextual structures extracted automatically from natural language corpora and enumerates word lists with their hierarchical contextual information. Several abstractions are made in the process of extracting the context in order to prevent superfluous information and support generalization.

The basic unit in PhraseNet is a *conset*, a word in its context, together with all relations associated with it. In the lexical database, consets are chained together via their similar or hierarchical contexts. By listing every context extracted from large corpora and all the generalized contexts based on those attested sentences, PhraseNet will have much more consets than synsets in WordNet. However, the organization of PhraseNet respects the syntactic structure together with the distinction of senses of each word in its corresponding contexts.

For example, rather than linking all hypernyms of a polysemous word to a single word token, PhraseNet connects the hypernym of each sense to the target word in every context that instantiates that sense. While in WordNet every word has an average of 5.4 hypernyms, in PhraseNet, the average number of hypernyms of a word in a conset is 1.5[1].

In addition to querying WordNet semantic relations to disambiguate consets, PhraseNet also maintains frequency records of each word in its context to help differentiate consets and makes use of defined similarity between contexts in this process [2].

Several access functions are built into PhraseNet that allow retrieving information relevant to a word and its context. When accessed with words and their contextual information, the system tends to output more *relevant* semantic information due to the constraint set by their syntactic contexts.

While still in preliminary stages of development and experimentation and with a lot of functionalities still missing, we believe that PhraseNet is an important effort towards building a contextually sensitive lexical semantic resource, that will be of much value to NLP researchers as well as linguists and language learners.

The rest of this paper is organized as follows. Sec. 2 presents the design principles of PhraseNet. Sec. 3 describes the construction of PhraseNet and the current stage of the implementation. An application that provides a preliminary experimental evaluation is described in Sec. 4. Sec. 5 discuses some related work on lexical semantics resources and Sec. 6 discusses future directions within PhraseNet.

2 The Design of PhraseNet

Context is one important notion in PhraseNet. While the context may mean different things in natural language, many previous work in statistically natural language processing defined "context" as an n-word window around the target word (Gale et al., 1992; Brown et al., 1991; Roth, 1998). In PhraseNet, "context" has a more precise definition that depends on the grammatical structure of a sentence rather than simply counting surrounding words. We define "context" to be the syntactic structure of the sentence in which the word of interest occurs. Specifically, we define this notion at three abstraction levels. The highest level is the abstract syntactic skeleton of the sentence. That is, it is in the form of the different combinations of six syntactic components. Some components may be missing as long as the structure is from a legitimate English sentence. The most complete form of the abstract syntactic skeleton is:

$$[(S) - (V) - (DO) - (IO) - (P) - (N)] \quad (1)$$

which captures all of the six syntactic components such as Subject, Verb, Direct Object, Indirect Object, Preposition and Noun(Object) of Preposition, respectively, in the sentence. And all components are assumed to be arranged to obey the word order in English. The lowest level of contexts is the concrete instantiation of the stated syntactic skeleton, such as $[Mary(S) - give(V) - John(DO) - gift(IO) - on(P) - birthday(N)]$ and

[1] The statistics is taken over $200,000$ words from a mixed corpus of American English.

[2] See details in Sec. 3

$[I(S) - eat(V) - bread(DO) - with(P) - hand(N)]$ which are extracted directly from corpora with grammatical lemmatization done during the process. Therefore, all word tokens are in their lemma format. The middle layer(s) consists of generalized formats of the syntactic skeleton. For example, the first example given above can be generalized as $[\boldsymbol{Peop}(S) - give(V) - \boldsymbol{Peop}(DO) - \boldsymbol{Possession}(IO) - on(P) - \boldsymbol{Day}(N)]$ by replacing some of its components with more abstract semantic concepts.

PhraseNet organizes nouns and verbs into "consets" and a "conset" is defined as a context with all its corresponding pointers (edges) to other consets. The context that forms a conset can be either directly extracted from the corpus, or at a certain level of abstraction. For example, both $[Mary(S) - eat(V) - cake(DO) - on(P) - birthday(N), \{p_1, p_2, \ldots, p_n\}]$ and $[\boldsymbol{Peop}(S) - eat(V) - Food(DO) - on(P) - Day(N), \{p_1, p_2, \ldots, p_n\}]$ are consets.

Two types of relational pointers are defined currently in PhraseNet: Equal and Hyper. Both of these two relations are based on the context of each conset. *Equal* is defined among consets with same number of components and same syntactic ordering, i.e, some contexts under the same abstract syntactic structure (the highest level of context as defined in this paper). It is defined that the *Equal* relation exists among consets whose contexts are with same abstract syntactic skeleton, if there is only one component at the same position that is different. For example, $[Mary(S) - give(V) - John(DO) - gift(IO) - on(P) - birthday(N), \{p_1, p_2, \ldots, p_n\}]$ and $[Mary(S) - send(V) - John(DO) - gift(IO) - on(P) - birthday(N), \{p_1, p_2, \ldots, p_k\}]$ are equal because the syntactic skeleton each of them has is the same, i.e., $[(S) - (V) - (DO) - (IO) - (P) - (N)]$ and except one word in the verb position that is different, i.e., "give" and "send", all other five components at the corresponding same position are the same. The *Equal* relation is transitive only with regard to a specific component in the same position. For example, to be transitive to the above two example consets, the *Equal* conset should be also different from them only by its verb. The *Hyper* relation is also defined for consets with same abstract syntactic structure. For conset A and conset B, if they have the same syntactic structure, and if there is at least one component of the context in A that is the hypernym of the component in that of B at the corresponding same position, and all other components are the same respectively, A is the *Hyper* conset of B. For example, both $[Molly(S) - hit(V) - Body(DO), \{p_1, p_2, \ldots, p_j\}]$ and $[Peop(S) - hit(V) - Body(DO), \{p_1, p_2, \ldots, p_n\}]$ are *Hyper* consets of $[Molly(S) - hit(V) - nose(DO), \{p_1, p_2, \ldots, p_k\}]$. The intuition behind these two relations is that the *Equal* relation can cluster a list of words which occur in exactly the same contextual structure and if the extreme case occurs, namely when the same context in all these equal consets with regard to a specific syntactic component groups virtually any nouns or verbs, the *Hyper* relation can be used here for further disambiguation.

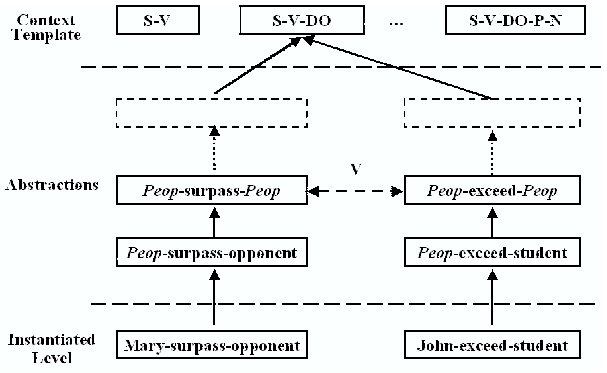

Figure 1: **The basic organization of PhraseNet:** The upward arrow denotes the *Hyper* relation and the dotted two-way arrow with a V above denotes the *Equal* relation that is transitive with regard to the V component.

To summarize, PhraseNet can be thought of as a graph on consets. Each node is a context and edges between nodes are relations defined by the context of each node. They are either *Equal* or *Hyper*. *Equal* relation can be derived by matching consets and it is easy to implement while building the *Hyper* relation requires the assistance of WordNet and the defined *Equal* relation. Semantic relations among words can be generated using the two types of defined edges. For example, it is likely that the target words in all equal consets with transitivity have similar meaning. If this is not true at the lowest lower of contexts, it is more likely to be true at higher, i.e., more generalized level. Figure 1 shows a simple example reflecting the preliminary design of PhraseNet.

After we get the similar meaning lists based on their contexts, we can build interaction from this word list to WordNet and inherit other semantic relations from WordNet. However, each member of a word list can help to disambiguate other members in this list. Therefore, it is expected that with the pruning assisted by list members, i.e., the disambiguation by truncating semantic relations associated with each synset in WordNet, the extract meaning in the context together with all other semantic relations such as hypernyms, holonyms, troponyms, antonyms can be derived from WordNet.

In the next two sections we describe our current implementation of these operations and preliminary experiments we have done with them.

2.1 Accessing PhraseNet

Retrieval of information from PhraseNet is done via several access functions that we describe below. PhraseNet is designed to be accessed via multiple functions with flexible input modes set by the user. These functions may allow users to exploit several different functionalities of PhraseNet, depending on their goal and amount of resources they have.

An access function in PhraseNet has two components. The first component is the input, which can vary from a single word token to a word with its complete context. The second component is the functionality, which ranges over simple retrieval and several relational functions, modelled after WordNet relations.

The most basic and simplest way to query PhraseNet is with a single word. In this case, the system outputs all contexts the word can occur in, and its related words in each context.

PhraseNet can also be accessed with input that consists of a single word token along with its context information. Context information refers to any of the elements in the syntactic skeleton defined in Eq. 1, namely, Subject(S), Verb(V), Direct Object(DO), Indirect Object(IO), Preposition(P) and Noun(Object) of the Preposition(N). The contextual roles S, V, DO, IO, P or N or any subset of them, can be specified by the user or derived by an application making use of a shallow or full parser. The more information the user provides, the more specific the retrieved information is.

To ease the requirements from the user, say, in case no information of this form is available to the user, PhraseNet will, in the future, have functions that allow a user to supply a word token and some context, where the functionality of the word in the context is not specified. See Sec. 6 for a discussion.

Function Name	Input Variables	Output
PN_WL	Word [, Context]	Word List
PN_RL	Word [, Context]	WordNet relations
PN_SN	Word [, Context]	Sense
PN_ST	Context	Sentence

Table 1: **PhraseNet Access Functions:** PhraseNet access functions along with their input and output. $[i]$ denotes optional input. PN_RL is a family of functions, modelled after WordNet relations.

Table 1 lists the functionality of the access functions in PhraseNet. If the user only input a word token without any context, all those designed functions will return each context the input word occurs together with the wordlist in these contexts. Otherwise, the output is constrained by the input context. The functions are described below:

PN_WL takes the optional contextual skeleton and one specified word in that context as inputs and returns the corresponding wordlist occurring in that context or a higher level of context. A parameter to this function specifies if we want to get the complete wordlist or those words in the list that satisfy a specific pruning criterion. (This is the function used in the experiments in Sec. 4.)

PN_RL is modelled after the WordNet access functions. It will return all words in those contexts that are linked in PhraseNet by their *Equal* or *Hyper* relation. Those words can help to access WordNet to derive all lexical relations stored there.

PN_SN is modelled after the semantic concordance in (Landes et al., 1998). It takes a word token and an optional context as input, and returns the sense of the word in that context. Similarly to PN_RL this function is implemented by appealing to WordNet senses and pruning the possible sense based on the wordlist determined for the given context.

PN_ST is not implemented at this point, but is designed to output a sentence that has same structure as the input context, but use different words. It is inspired by the work on reformulation, e.g., (Barzilay and McKeown, 2001).

We can envision many ways users of PhraseNet can make use of the retrieved information. At this point in the life of PhraseNet we focus mostly on using PhraseNet as a way to acquire semantic features to aid learning based natural language applications. This determines our priorities in the implementation that we describe next.

3 Constructing PhraseNet

Constructing PhraseNet involves three main stages: (1) extracting syntactic skeletons from corpora, (2) constructing the core element in PhraseNet: consets, and (3) developing access functions.

The first stage makes use of fully parsed data. In constructing the current version of PhraseNet we used two corpora. The first, relatively small corpus of the 1.1 million-word Penn-State Treebank which consists of American English news articles (WSJ), and is fully parsed. The second corpus has about 5 million sentences of the TREC-11 (Voorhees, 2002), also containing mostly American English news articles (NYT, 1998) and parsed with Dekang Lin's minipar parser (Lin, 1998a).

In the near future we are planning to construct a much larger version of PhraseNet, using Trec-10 and Trec-11 data sets, which cover about 8 GB of text. We believe that the size is very important here, and will add significant robustness to our results.

To reduce ostensibly different contexts, two important abstractions take place at this stage. (1) Syntactic lemmatization to get the lemma for both nouns and verbs in

the context defined in Eq. 1. For data parsed via Lin's minipar, the lexeme of each word is already included in the parser. (2) Sematic categorization to unify pronouns, proper names of people, locations and organization as well as numbers. This semantic abstraction captures the underlying semantic proximity by categorizing multitudinous surface-form proper names into one representing symbol.

While the first abstraction is simple the second is not. At this point we use an NE tagger we developed ourselves based on the approach to phrase identification developed in (Punyakanok and Roth, 2001). Note that this abstraction handles multiword phrases. While the accuracy of the NE tagger is around 90%, we have yet to experiment with the implication of this additional noise on PhraseNet.

At the end of this stage, each sentence in the original corpora is transformed into a single context either at the lowest level or a more generalized instantiation (with name entity tagged). For example, "`For six years, T. Marshall Hahn Jr. has made corporate acquisitions in the George Bush mode: kind and gentle.`", changes to: $[Peop - make - acquisition - in - mode]$.

The second stage of constructing PhraseNet concentrates on constructing the core element in PhraseNet: consets.

To do that, for each context, we collect wordlists that contain those words that we determine to be admissible in the context(or contexts share the equal relation). The first step in constructing the wordlists in PhraseNet is to follow the most strict definition – include those words that actually occur in the same context in the corpus. This involves all *Equal* consets with the transitive property to a specific syntactic component. We then apply to the wordlists three types of pruning operations that are based on (1) frequency of word occurrences in identical or similar contexts; (2) categorization of words in wordlist based on clustering *all* contexts they occur in, and (3) pruning via the relational structure inherited from WordNet - we prune from the wordlist outliers in terms of this relational structure. Some of these operations are parameterized and determining the optimal setting is an experimental issue.

1. Every word in a conset wordlist has a frequency record associated with it, which records the frequency of the word in its exact context. We prune words with a frequency below k (with the current corpus we choose $k = 3$). A disadvantage of this pruning method is that it might filter out some appropriate words with a low frequency in reality. For example, for the partial context $[strategy - involve - * - * - *]$, we have:

   [**strategy - involve - * - * - ***, $< DO$: advertisement 4, abuse 1, campaign 2, compromise 1, everything 1, fumigation 1, item 1, membership 1, option 3, stock-option 1$>$]

   In this case,"strategy" is the subject and "involve" is the predicate and all words in the list serve as the direct object. The number in the parentheses is the frequency of the token. With $k = 3$ we actually get as a wordlist only: $< advertisment, option >$.

2. There are several ways to prune wordlists based on the different contexts words may occur in. This involves a definition of similar contexts and thresholding based on the number of such contexts a word occurs in. At this point, we implement the construction of PhraseNet using a clustering of contexts, as done in (Pantel and Lin, 2002). An exhaustive PhraseNet list is intersected with word lists generated based on clustered contexts given by (Pantel and Lin, 2002).

3. We prune from the wordlist outliers in terms of the relational structure inherited from WordNet. Currently, this is implemented only using the hypernym relation. The hypernym shared by the highest number of words in the wordlist is kept in the database.

   For example, by searching "option" in WordNet, we get its three senses. Then we collect the hypernyms of 'option' from all the senses as follows:

   05319492(a financial instrument whose value is based on another security)
   04869064(the cognitive process of reaching a decision)
   00026065(something done)

   We do this for every word in the original list and find out the hypernym(s) shared by the highest number of words in the original wordlist. The final pick in this case is the synset 05319492 which is shared by both "option" and "stock option" as their hypernym.

The third stage is to develop the access functions. As mentioned before, while we envision many ways users of PhraseNet can use the retrieved information, at this preliminary stage of PhraseNet we focus mostly on using PhraseNet as a way to supply abstract semantic features that learning based natural language applications can benefit from.

For this purpose, so far we have only used and evaluated the function PN_WL. PN_WL takes as input as specific word and (optionally) its context and returns a lists of words which are semantically related to the target word in the given context. For example,

PN_WL (V= protest, [peop - legislation - * - * - *])= [protest, resist, dissent, veto, blackball, negative, forbid, prohibit, interdict, proscribe, disallow].

This function can be implemented via any of the three pruning methods discussed earlier (see Sec. 4). This wordlists that this function outputs, can be used to augment feature based representations for other, learning based, NLP tasks. Other access functions of PhraseNet can serve in other ways, e.g., expansions in information retrieval, but we have not experimented with it yet.

With the experiments we are doing right now, PhraseNet only takes inputs with the context information in the format of Eq. 1. Semantic categorization and syntactic lemmatization of the context is required in order to get matched in the database. However, PhraseNet will, in the future, have functions that allow a user to supply a word token and more flexible contexts.

4 Evaluation and Application

In this section we provide a first evaluation of PhraseNet. We do that in the context of a learning task.

Learning tasks in NLP are typically modelled as classification tasks, where one seeks a mapping $g : X \rightarrow c_1, ..., c_k$, that maps an instance $x \in X$ (e.g., a sentence) to one of $c_1, ..., c_k$ – representing some properties of the instance (e.g., a part-of-speech tag of a word in the context of the sentence). Typically, the raw representation – sentence or document – are first mapped to some feature based representation, and then a learning algorithm is applied to learn a mapping from this representation to the desired property (Roth, 1998). It is clear that in most cases representing the mapping g in terms of the raw representation of the input instance – words and their order – is very complex. Functionally simple representations of this mapping can only be formed if we augment the information that is readily available in the input instance with additional, more abstract information. For example, it is common to augment sentence representations with syntactic categories – part-of-speech (POS), under the assumption that the sought-after property, for which we seek the classifier, depends on the syntactic role of a word in the sentence rather than the specific word. Similar logic can be applied to semantic categories. In many cases, the property seems not to depend on the specific word used in the sentence – that could be replaced without affecting this property – but rather on its 'meaning'.

In this section we show the benefit of using PhraseNet in doing that in the context of Question Classification.

Question classification (QC) is the task of determining the semantic class of the answer of a given question. For example, given the question: "`What Cuban dictator did Fidel Castro force out of power in 1958?`" we would like to determine that its answer should be a name of a person. Our approach to QC follows that of (Li and Roth, 2002). The question classifier used is a multi-class classifier which can classify a question into one of 50 fine-grained classes.

The baseline classifier makes use of syntactic features like the standard POS information and information extracted by a shallow parser in addition to the words in the sentence. The classifier is then augmented with standard WordNet or with PhraseNet information as follows. In all cases, words in the sentence are augmented with additional words that are supposed to be semantically related to them. The intuition, as described above, is that this provides a level of abstract – we could have potentially seen an equivalent question, where other "equivalent" words occur.

For WordNet, for each word in a question, all its hypernyms are added to its feature based representation (in addition to the syntactic features). For PhraseNet, for each word in a question, all the words in the corresponding conset wordlist are added (where the context is supplied by the question).

Our experiments compare the three pruning operations described above. Training is done on a data set of 21,500 questions. Performance is evaluated by the precision of classifying 1,000 test questions, defined as follows:

$$Precison = \frac{\#\ of\ correct\ predictions}{\#\ of\ predictions} \qquad (2)$$

Table 2 presents the classification precision before and after incorporating WordNet and PhraseNet information into the classifier. By augmenting the question classifier with PhraseNet information, even in this preliminary stage, the error rate of the classifier can be reduced by 12%, while an equivalent use of WordNet information reduces the error by only 5.7%.

Information Used	Precision	Err Reduction
Baseline	84.2%	0%
WordNet	85.1%	**5.7%**
PN: Freq. based Pruning	84.4%	1.3%
PN: Categ. based Pruning	85%	5.1%
PN: Relation based Pruning	86.1%	**12%**

Table 2: **Question Classification with PhraseNet Information** Question classification precision and error rate reduction compared with the baseline error rate(15.8%) by incorporating WordNet and PhraseNet(PN) information. 'Baseline' is the classifier that uses only syntactic features. The classifier is trained over 21,500 questions and tested over 1000 TREC 10 and 11 questions.

5 Related Work

In this section we point to some of the related work on syntax, semantics interaction and lexical semantic resources in computational linguistics and natural language processing. Many current syntactic theories make the common assumption that various aspects of syntactic alternation are predicable via the meaning of the predi-

cate in the sentence (Fillmore, 1968; Jackendoff, 1990; Levin, 1993). With the resurgence of lexical semantics and corpus linguistics during the past two decades, this so-called linking regularity triggers a broad interest of using syntactic representations illustrated in corpora to classify lexical meaning (Baker et al., 1998; Levin, 1993; Dorr and Jones, 1996; Lapata and Brew, 1999; Lin, 1998b; Pantel and Lin, 2002).

FrameNet (Baker et al., 1998) produces a semantic dictionary that documents combinatorial properties of English lexical items in semantic and syntactic terms based on attestations in a very large corpus. In FrameNet, a frame is an intuitive structure that formalizes the links between semantics and syntax in the results of lexical analysis. (Fillmore et al., 2001) However, instead of derived via attested sentences from corpora automatically, each conceptual frame together with all its frame elements has to be constructed via slow and labor-intensive manual work. FrameNet is not constructed automatically based on observed syntactic alternations. Though deep semantic analysis is built for each frame, lack of automatic derivation of the semantic roles from large corpora[3] confines the usage of this network drastically.

Levin's classes (Levin, 1993) of verbs are based on the assumption that the semantics of a verb and its syntactic behavior are predictably related. She defines 191 verb classes by grouping 4183 verbs which pattern together with respect to their diathesis alternations, namely alternations in the expressions of arguments. In Levin's classification, it is the syntactic skeletons (such as np-v-np-pp)to classify verbs directly. Levin's classification is validated via experiments done by (Dorr and Jones, 1996) and some counter-arguments are in (Baker and Ruppenhofer, 2002). Her work provides a a small knowledge source that needs further expansion.

Lin's work (Lin, 1998b; Pantel and Lin, 2002) makes use of distributional syntactic contextual information to define semantic proximity. Dekang Lin's grouping of similar words is a combination of the abstract syntactic skeleton and concrete word tokens. Lin uses syntactic dependencies such as "Subj-people", "Modifier-red", which combine both abstract syntactic notations and their concrete word token representations. He applies this method to classifying not only verbs, but also nouns and adjectives. While no evaluation has ever been done to determine if concrete word tokens are necessary when the syntactic phrase types are already presented, Lin's work indirectly shows that the concrete lexical representation is effective.

WordNet (Fellbaum, 1998) by far is the most widely used semantic database. However, this database does not always work as successfully as researchers have expected (Krymolowski and Roth, 1998; Montemagni and Pirelli, 1998). This seems to be due to lack of topical context (Harabagiu et al., 1999; Agirre et al., 2001) as well as local context (Fellbaum, 1998). By adding contextual information, many researchers, (e.g., (Green et al., 2001; Lapata and Brew, 1999; Landes et al., 1998)), have already made some improvements over it.

The work on the importance of connecting syntax and semantics in developing lexical semantic resources shows the importance of contextual information as a step towards deeper level of processing. With hierarchical sentential local contexts embedded and used to categorize word classes automatically, we believe that PhraseNet provides the right direction for building useful lexical semantic database.

6 Discussion and Further Work

We believe that progress in semantics and in developing lexical resources is a prerequisite to any significant progress in natural language understanding. This work makes a step in this direction by introducing a context-sensitive lexical semantic knowledge base system, PhraseNet. We have argued that while current lexical resources like WordNet are invaluable, we should move towards contextually sensitive resources. PhraseNet is designed to fill this gap, and our preliminary experiments with it are promising.

PhraseNet is an ongoing project and is still in its preliminary stage. There are several key issues that we are currently exploring. First, given that PhraseNet draws part of it power from corpora, we are planning to enlarge the corpus used. We believe that the data size is very important and will add significant robustness to our current results. At the same time, since constructing PhraseNet relies on machine learning techniques, we need to study extensively the effect of tuning these on the reliability of PhraseNet. Second, there are several functionalities and access functions that we are planning to augment PhraseNet with. Among those is the ability of allowing a user to query PhraseNet even without explicitly specifying the role of words in the context. This would reduce the requirement for users and applications using PhraseNet. Finally, current PhraseNet has no lexical information about adjectives and adverbs, which may contain important distributional information about their modified nouns or verbs. We would like to take this information into consideration in the near future.

[3]The attempt to label these semantic roles automatically in (Gildea and Jurafsky, 2002) assumes knowledge of the frame and covers only 20% of them.

References

E. Agirre, O. Ansa, D. Martinez, and E. Hovy. 2001. Enriching wordnet concepts with topic signatures.

C. Baker and J. Ruppenhofer. 2002. Framenet's frames vs. levin's verb classes. In *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*.

C. Baker, C. Fillmore, and J. Lowe. 1998. The Berkeley FrameNet project. In Christian Boitet and Pete Whitelock, editors, *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pages 86–90, San Francisco, California. Association for Computational Linguistics, Morgan Kaufmann Publishers.

R. Barzilay and K. R. McKeown. 2001. Extracting paraphrases from a parallel corpus. In *Proceeding of the 10th Conference of the European Chapter of ACL*.

E. Brill and P. Resnik. 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proc. of COLING*.

P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer. 1991. Word sense disambiguation using statistical methods. In *Proceedings of ACL-1991*.

B. Dorr and D. Jones. 1996. Role of word-sense disambiguation in lexical acquisition.

C. Fellbaum. 1998. In C. Fellbaum, editor, *WordNet: An Electronic Lexical Database*. The MIT Press.

C. J. Fillmore, C. Wooters, and C. F. Baker. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*, HongKong.

C. J. Fillmore. 1968. The case for case. In Bach and Harms, editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart, and Winston, New York.

W. A. Gale, K. W. Church, and D. Jarowsky. 1992. A method for disambiguation word senses in large corpora. *Computers and the Humanities*, 26(5-6):415–439.

D. Gildea and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, September.

R. Green, L. Pearl, B. J. Dorr, and P. Resnik. 2001. Lexical resource integration across the syntax-semantics interface. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, Pittsburg, June.

S. M. Harabagiu, G. A. Miller, and D. I. Moldovan. 1999. Wordnet2 - a morphologically and semantically enhanced resources. In *Proceedings of ACL-SIGLEX99: Standardizing Lexical Resources*, pages 1–8, Maryland.

R. Jackendoff. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

Y. Krymolowski and D. Roth. 1998. Incorporating knowledge in natural language learning: A case study. In *COLING-ACL'98 workshop on the Usage of WordNet in Natural Language Processing Systems*.

S. Landes, C. Leacock, and R. I. Tengi. 1998. Building semantic concordances. In C. Fellbaum, editor, *WordNet: an Electoronic Lexical Database*, pages 199–216. The MIT Press.

M. Lapata and C. Brew. 1999. Using subcategorization to resolve verb class ambiguity. In *Proceedings of EMNLP*, pages 266–274.

B. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, IL.

X. Li and D. Roth. 2002. Learning question classifiers. In *Proceedings of COLING*.

D. Lin. 1998a. Dependency-based evaluation of minipar. In *In Workshop on the Evaluation of Parsing Systems Granada Spain*.

D. Lin. 1998b. An information-theoretic definition of similarity. In *Proc. 15th International Conf. on Machine Learning*, pages 296–304. Morgan Kaufmann, San Francisco, CA.

S. Montemagni and V. Pirelli. 1998. Augmenting WordNet-like lexical resources with distributional evidence. an application-oriented perspective. In S. Harabagiu, editor, *Use of WordNet in Natural Language Processing Systems: Proceedings of the Conference*, pages 87–93. Association for Computational Linguistics.

V. Ng and C. Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of 40th Annual Meeting of the ACL*, TaiPei.

P. Pantel and D. Lin. 2000. An unsupervised approach to prepositional phrase attachment using contextually similar words. In *Proceedings of Association for Computational Linguistics*, Hongkong.

P. Pantel and D. Lin. 2002. Discovering word senses from text. In *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

V. Punyakanok and D. Roth. 2001. The use of classifiers in sequential inference. In *NIPS-13; The 2000 Conference on Advances in Neural Information Processing Systems*, pages 995–1001. MIT Press.

D. Roth. 1998. Learning to resolve natural language ambiguities: A unified approach. In *Proc. National Conference on Artificial Intelligence*, pages 806–813.

H. Saggion and G. Lapalme. 2002. Generating indicative-informative summaries with sumum. *Computational Linguistics*, 28(4):497–526.

J. Stetina and M. Nagao. 1997. Corpus based pp attachment ambiguity rosolution with a semantic dictionary. In *Proceedings of the 5th Workshop on Very Large Corpora*, Beijing and Hongkong.

E. Voorhees. 2002. Overview of the TREC-2002 question answering track. In *The Eleventh TREC Conference*, pages 115–123.