# Acquisition of Unbounded Dependency Using Explanation-Based Learning

Rey-Long Liu and Von-Wun Soo
Department of Computer Science
National Tsing-Hua University
HsinChu, Taiwan, R.O.C.

## Abstract

A natural language acquisition model using Explanation-Based Learning (EBL) had been proposed to acquire parsing-related knowledge which includes Context-Free grammar rules and syntactic and thematic features of lexicons. The domain theory that is assumed to be innate to the model includes the theta-theory and the universal feature instantiation principles in Generalized Phrase Structure Grammar (GPSG). In this paper, we show in particular how unbounded dependency may be acquired in the natural language acquisition model. The acquisition problem of unbounded dependency may be further divided into two sub-problems: detecting whether there are moved constituents and finding the places to which the constituents are moved. For these problems, the universal innate domain theory facilitates and constrains the acquisition process which is otherwise intractable.

**Keywords:** Natural Language Acquisition, Explanation-Based Learning, Theta Theory, Universal Feature Instantiation Principles, Knowledge Assimilation.

## 1. Introduction

Parsing involves searching for a set of applicable knowledge pieces to transform a sentence into its corresponding syntactic and/or semantic structure (e.g. the parse tree). This problem solving process needs a knowledge base which is often enhanced, maintained, and tested periodically, especially when the system is applied to different domains. Since natural language is ever evolutionary in nature, extensibility of a natural language processing (NLP) system becomes one of the most critical concerns in real applications.

The Universal Grammar (UG, Chomsky[19]), which is claimed to be innate and universal among various natural languages, is believed to reflect children natural language acquisition phenomena. From this point of view, natural language acquisition may be approached by setting the parameters embedded in UG and learning the particular linguistic requirements (called Periphery Grammar) of the target language. Thus, the introduction of UG not only reduces the hypotheses space and hence makes learning more tractable, but also promotes the portability of of the system, since it not only facilitates adaptive acquisition in various application domains with the same target language (Lehman[10]), but also makes acquisition across different natural languages more possible.

Therefore, a natural language acquisition model (Liu[13]) had been proposed to automatically assimilate and maintain parsing-related knowledge, including Context-Free grammar rules and syntactic and thematic requirements of lexicons. In the model, the knowledge bases of the model consist of two parts: the static part and the dynamic part. The static part contains the universal linguistic principles, including the theta-theory and the universal feature instantiation principles in the Generalized Phrase Structure Grammar formalism (GPSG, Gazdar[4]). They are innate and invariant in learning. The dynamic part contains current parsing-related knowledge of the system (periphery grammar). Through learning, the periphery grammar in the dynamic part is enhanced by following the principles in the static part.

In this paper, we focus on the acquisition of unbounded dependency in the developed explanation-based natural acquisition model. Typically, an unbounded dependency occurs in a construction in which there is an unexpected constituent outside a clause, while within that clause a constituent is correspondingly missing (Chomsky[19]). Wh-questions, relative clauses, and topicalizations, which all involve movement, are the representative examples of unbounded dependencies we consider in this paper.

In fact, the task of unbounded dependency acquisition involves two steps: detecting whether there are moved constituents, and then finding the place to which the constituents are moved. For example, in the sentence "The boy I see is a student", it is necessary for the learning system to determine whether the VP (Verb Phrase) "see" has a missing theme or not. If a theme is missing, the system learns that an NP (Noun Phrase) may be constructed by an NP followed by an S (Sentence) with a theme missing. On the other hand, if no themes are missing, the S cannot have a missing theme. For these problems, the universal innate linguistic principles facilitate and constrain the acquisition process which is otherwise intractable.

In the next section, we describe the framework of the explanation-based natural acquisition model. More detailed elaboration may be found in Liu[13]. In section 3, we show why and how the universal linguistic principles are employed to acquire unbounded dependency. In section 4, experimental results are shown to investigate the performance of the model. The model is also related to previous works and evaluated from various perspectives. In section 5, we conclude the article.

## 2. Explanation-based natural language acquisition

Explanation-Based Learning (EBL, Mitchell[17], Keller[8]) had been widely applied to learning domains in which intensive domain theory may be constructed before learning. Major components of EBL may include Goal Concept, Operationality, Training Example, Domain Theory, and Problem Solver. In learning, the problem solver uses the predefined domain theory to prove (or explain) the given positive training examples to be an instance of the goal concept. The sufficient conditions of the explanation are thus extracted and expressed in terms of the operationality criteria. In later problem solving, when the extracted conditions may be directly applied to the current problem, no further explanation processes are needed. Therefore, through learning, the domain theory is "compiled" into a more efficient version.

A new explanation-based natural language acquisition model had been proposed to learn parsing-related knowledge for the parser (Liu[13]). The relationship between the traditional EBL and the language acquisition model can be summarized as follows:

- Goal concept: Grammatical sentence.
- Operationality: Recognizability of linguistic features of constituents.
- Training examples: Input sentences and their parse trees.
- Domain theory: Universal linguistic principles (static) + Current parsing knowledge (dynamic).
- Problem solver: The parser.
- Explanation tree: Parse tree annotated with sufficient constraints (features).

In the model, the problem solver is the parser which uses its parsing knowledge to parse an input sentence. If the highest level goal S-maj (a major sentence) can be achieved, the sentence is proven to be grammatical (the sentence can be successfully parsed). The condition parts of the rules in the knowledge base are expressed in terms of linguistic features such as VERB, NOUN, AGENT, OBJECT, PERSON, ... etc. These features are operational or "efficiently recognizable" (Keller[8]) in the system.

In real world problem domains (e.g. natural language processing), although a preliminary domain theory can be constructed (such as simple grammar rules), it is quite difficult to have a complete and correct domain theory (Hall[6]). The domain theory can be incomplete. It is separated into two major parts: a static part and a dynamic part. The static part includes universal linguistic principles which are invariant and innate to the system, while the dynamic part is augmented through learning.

When an input sentence cannot be proven to be grammatical (i.e. it cannot be successfully parsed), learning is triggered to enhance the dynamic part of the domain theory. The parsing knowledge in the dynamic part includes the argument structures of verbs (e.g. the verb "see" needs an EXPERIENCER argument and a THEME argument), thematic features of nouns (e.g. AGENT, OBJECT, ... etc.), general grammar rules (e.g. S --> NP VP), and some special phrase patterns (e.g. "Although S, S"). Initially, syntactic and thematic features of some verbs and nouns are provided to the dynamic part as the bootstrapping parsing knowledge.

### 2.1 The learning algorithm

As the dynamic part is inadequate to provide actions, learning is triggered. The system can first deduce a correct solution path from the given parse tree (Liu[11], Liu[13]). After executing each action in the solution path, an annotated parse tree can still be constructed as a sufficient condition to explain the input sentence as a grammatical sentence. The new parsing knowledge can be extracted from the annotated parse tree and then assimilated into the dynamic part of the domain theory. The algorithm of the learning module can be thus formalized as follows:

(1) Get the parse tree of the new sentence from the trainer;
(2) Iteratively invoke the parser to annotate all constituents in the parse tree (i.e. apply the current parsing knowledge and the universal linguistic principles to the parse tree);
(3) Extract new rules from the annotated parse tree.
(4) If the first subgoal of the extracted rule is a phrase, assimilate the new rule into the grammar rule base;
Else begin
(5) Try to generalize the rules in the lexicon entry (empirical generalization);
(6) Assimilate the rule into the lexicon entry;
end

In the following sections, we further elaborate the extra parse tree input (step 1), the use of universal linguistic principles (step 2), and the way of knowledge assimilation (step 4 and step 6). Finally, an example is shown to illustrate the learning algorithm.

### 2.2 The parse tree as external guidance

When there is missing knowledge in the domain theory, new knowledge might become too ambiguous to acquire, even though the learning system has exploited all its current knowledge to the largest extent. For example, consider the sentence "Taking exercises is good for your health". The target knowledge is the rule "NP[NUM=-plu,PER=3] --> VP[VF=prp]" which means that a singular (NUM=-plu) third-person (PER=3) Noun Phrase (NP) can be constructed by a Verb Phrase (VP) with present participle verb form (VF=prp). However, if no other information is provided, the learning module cannot segment the sentence into phrases. In that case, there are too many possible kinds of new knowledge. For example, the system can hypothesize that "taking" can be an NP, an S-maj can be implemented by the pattern "taking NP VP", an S-maj can be expanded as "taking exercises VP", ... etc.

However, the given parse tree cannot be a correct "explanation tree" in which the system can find sufficient conditions for the sentence to be grammatical. For example, in a parse tree, the system can deduce a rule "S --> VP" (since S is the mother of VP in the parse tree) which is too general in the sense that the sentence "Eats the hotdog" will also be accepted. To find a sufficient condition, the parse tree should be annotated with critical features by the help of the static part of the domain theory.

### 2.3 The static part -- universal linguistic principles

In the model, the static (and predefined) part of the domain theory contains the "abstract" and universal linguistic principles which guide the acquisition of "operational" knowledge (parsing knowledge) in the dynamic part. It contains the minimal linguistic knowledge which is assumed to be innate to the system and is invariant during learning. It includes the theta-theory and the universal feature instantiation principles. These principles promote the portability of the system and make learning more tractable by reducing the hypothesis space in learning. The universal innate principles in the model are thus defined as follows:

- The theta-theory (Chomsky[19]) proposes a theta criterion which requires that, in the argument structure of a lexical head, each argument must bear one and only one theta-role. For example, in the sentence "John kissed Mary", the head "kissed" assigns the NP "John" the "AGENT" theta-role, and the NP "Mary" the "THEME" theta-role. No arguments can be assigned more than one theta-roles.

- The Head Feature Convention (HFC, Gazdar[4]) says that a mother's HEAD features should be identical to the HEAD features of its head daughter. For example, the verb "eating" is the HEAD of the verb phrase "eating the apple" (the verb phrase is the mother of the verb in a parse tree). Since verb form (VF) is a HEAD feature defined in GPSG, the verb phrase should share the feature "VF=prp" with the verb (via unification).

• The Foot Feature Principle (FFP, Gazdar[4]) allows FOOT features to propagate from any daughter to its mother in the parse tree. For example, the SLASH feature is a FOOT feature in GPSG. If a constituent has a SLASH feature with value NP, there is an NP missing in it. Consider the NP "the boy I like". There is an object NP missing in the verb phrase "like". By following FFP, this SLASH feature will be propagated to the clause "I like".

• The Control Agreement Principle (CAP, Gazdar[4]) says that controllees (such as VPs) agree with their controllers (such as NPs) by showing the features that are essentially properties of the controllers. The AGR feature in GPSG formalism needs to follow this principle. For example, for the verb "likes", an AGR feature with value "NP[NUM=-plu,PER=3]" (Subject-Verb agreement) is encoded. According to the feature, CAP will inform the parser to climb the parse tree upward to check whether there is a singular third-person NP. CAP can deal with semantic processing when the value of an AGR feature includes thematic properties AGENT, THEME, EXPERIENCER, ... etc.) of controllers.

For more detailed description, the reader should refer to Chomsky[19] and Gazdar[4]. The critical roles of these principles on the acquisition of parsing knowledge can be further illustrated by the following examples:

• Suppose the system attempts to learn from an English command sentence "Eat the hotdog", and it has the rules for parsing the NP "the hotdog" and the subcategorization information of "Eat" (e.g. "Eat" needs an NP as object) as the currently available parsing knowledge. If HFC is not employed, even though a parse tree is given, the system might induce the rule "S-maj --> VP" (it comes from the parse tree). The rule is too general in the sense that the sentence "Eats the hotdog" will also be accepted. On the other hand, by following HFC, the VP can be appropriately annotated by critical features which are the basis of the generality of the new rule. In this case, a better rule "S-maj --> VP[VF=bse]" (a VP with base verb form can be a major sentence) can be constructed to enrich the current parsing knowledge bases (see Fig. 1).

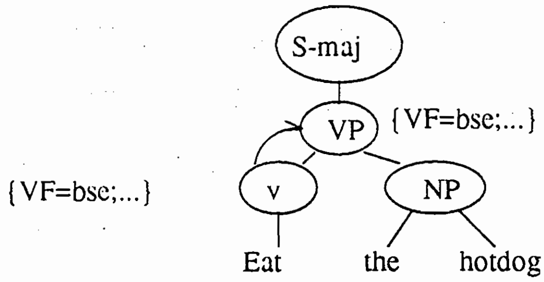


Fig. 1. Head Feature Convention

• Consider the sentence "Taking exercises is good for your health". Suppose the parsing module does not have a rule to construct an NP from a VP with present participle verb form. From the given parse tree and HFC, a VP[VF=prp] can be constructed by the parser. Therefore, the rule "NP --> VP[VF=prp]" can be induced. However, this rule is too general in the sense that the sentence "Taking exercises are good for your health" will also be accepted. On the other hand, if the VP "is good for your health" is parsed, by following CAP, it will restrict the number and person features of the NP to be singular and third-person. Therefore, the target rule "NP[NUM=-plu,PER=3] --> VP[VF=prp]" can be acquired (see Fig. 2).

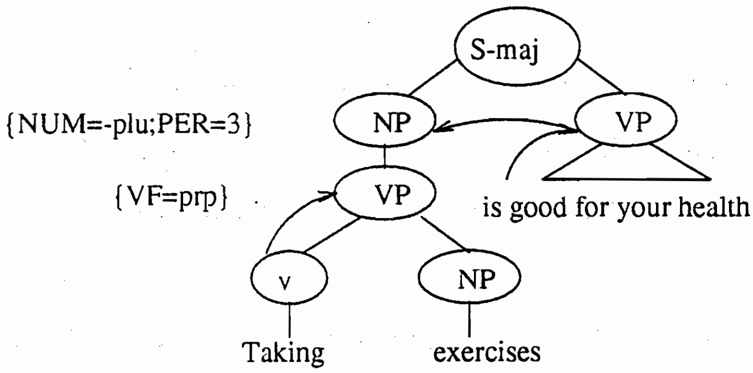


Fig. 2. Head Feature Convention & Foot Feature Principle.

## 2.4 Blame assignment and knowledge assimilation

In this paper, we focus on the problem of incomplete domain theory. Enhancing the dynamic domain theory is simply adding and then properly generalizing new knowledge pieces. From this point of view, the problem of blame assignment is reduced to the problem of finding which knowledge pieces are the missing knowledge. When invoking the parser to parse the sentence (step 2 in the learning algorithm), the learning system keeps track of the activation of rules. When no rules can issue the current action in the solution path, there is a missing rule at this point. After the whole parse tree is annotated, the missing rule may be extracted and assimilated into the dynamic domain theory.

The acquired rules may be assimilated into either the lexicon entries (step 6 in the learning algorithm) or the general grammar rule base (step 4 in the learning algorithm). The way of assimilating knowledge is closely related to the way of retrieving knowledge to use. In the model, indexing is employed for fast assimilation and utilization of knowledge. If the first subgoal of the acquired rule is a phrase, the rule is placed into the grammar rule base. If the first subgoal is a word, the rule is assimilated into the lexicon entry of the word in the dictionary.

### 2.5 An example

When the parser fails to parse the input sentence, learning is triggered, and the user is asked to input a parse tree (Step 1). For example, for the above sentence "Taking exercises is good for your health", the parse tree might be:

```
(S (NP (VP (v taking)
           (NP (n exercises))))
   (VP (v is)
       (adj good)
       (PP (prep for)
           (NP (pos your)
               (n health))))).
```

As described in section 2.2, the system should have the ability to derive the critical features of constituents rather than directly extracts the rules from the parse tree. Therefore, after given a parse tree, the parser is invoked to separately parse the constituents in the sentence (Step 2). After that, the critical features (including syntactic and thematic features) of each parsed constituent are derived. In this case, we assume the subcat pattern "take NP[THM=OBJECT]" has already been in the lexicon entry of "take". Therefore, the first VP "taking exercises" can be successfully parsed. Its feature "VF=prp" is also derived (since "VF" is a HEAD feature). At this time, the parsing module finds that it has a missing rule which allows it to construct an NP from the VP[VF=prp]. Therefore, now the possible new rule is "NP --> VP[VF=prp]".

After parsing the main VP "is good for your health", its feature "AGR=NP[NUM=-plu,PER=3]" is computed, where AGR is also a HEAD feature whose propagation in the parse tree must obey the Head Feature Convention. By following the Control Agreement Principle, the VP needs an NP which must be singular and the third person. This feature specification indicates that the NP constructed from the VP[VF=prp] should have the features "NUM=-plu" and "PER=3". Therefore, the final rule "NP[NUM=-plu,PER=3] --> VP[VF=prp]" can be successfully extracted from the annotated parse tree (Step 3). Since the first subgoal of the rule is a phrase, this rule is considered to be a general phrase structure rule which should be assimilated into the grammar rule base (Step 4).

It should be noted that, this way of computing critical features of constituents is a conservative way of acquiring new knowledge. That is, the computed features might be too specific. For example, consider the sentence "We live in an abundant life". Since the NP "We" has the thematic feature "THM=PERSON", after the sentence is processed, the system will restrict the AGENT of "live" to be an NP with the feature "THM=PERSON". Thus, when other input sentences involving "live" are entered (e.g. the sentence "The dog lives with us"), the learning module will try to generalize the rules which had already been acquired and stored in the lexicon entry of "live" (Step 5). The generalized rule is then assimilated into the lexicon entry of "live" (Step 6). Also note that, in some cases the system needs to generalize knowledge pieces among different lexicon entries.

## 3 Acquisition of unbounded dependency

Since constructions of unbounded dependency frequently occur in natural languages, its acquisition becomes an important task for natural language acquisition. As described above, portability is one of the major concerns of the learning model. Therefore, we need to introduce a minimal and universal innate domain theory to constrain the hypothesis space, and simultaneously, maintain the portability of the system in the sense that it may be applied to various learning situations (e.g. Chinese).

Typically, unbounded dependency occurs in a construction in which there is an unexpected constituent outside a clause, while within that clause its corresponding constituent is missing. We consider in this paper such typical unbounded dependencies as in relative clauses, wh-movements, and topicalizations. To acquire them, the learning system needs to determine whether there are missing constituents and, if so, to which places the missing constituents are moved.

Berwick[2] employs the Subjacency Principle to locate the moved constituents in the sentence. The location process is simply triggered when the syntactic requirements (e.g. the subcategorization frames of verbs) are not satisfied (e.g. an NP is expected but does not appear at its corresponding place). However, when the syntactic requirements have not been completely acquired, the location process might be miss-triggered. For example, many verbs may be both transitive and intransitive. As a verb's transitive subcategorization frame has already been acquired, but its intransitive version has not, a new sentence with no NPs occurring at the object position of the verb causes two possibilities: either the verb may have a intransitive version or there is a missing NP that can be found in other places in the sentence (unbounded dependency). If the ambiguity cannot be resolved, erroneous knowledge, which is not only useless but also harmful to the learning system, may be acquired.

In this paper, FFP and the theta-theory work together to facilitate the acquisition of unbounded dependency. They are consulted as the learning system attempts to acquire the unbounded dependency.

### 3.1 Acquisition of unbounded dependencies in relative clauses

Movement in relative clauses may be characterized as A-Bar-movement in GB theory (Chomsky[19]). A constituent is moved from a position that is assigned both a theta-role and

Case to an A-Bar position. Consider the sentence "The boy I see is a student". The parsing module needs to acquire the target rule "NP --> SN S[/=NP]", where "SN" is a nonterminal for simple NPs (without embedding clauses), and "/" is the SLASH feature in GPSG terminology ("/=NP" means "missing an NP"). Similarly, without using FFP to propagate the SLASH feature of the VP "see" to the S "I see", the rule "NP --> SN S", which is too general, will be acquired. However, if the verb "see" is intransitive, the slash feature may not exist. How can the system determine whether the VP "see" has the "/=NP" feature? By following the theta-theory, the SN "the boy" must bear a theta-role. In the sentence, only the verb "see" may assign the "THEME" theta-role to it (the verb "is" only assigns a theta-role to the whole NP "The boy I see"). Therefore, the VP "see" is missing an NP.

The second step in the acquisition of unbounded dependency involves the locating of the moved constituent. In GPSG formalism, FFP propagates the slash feature upward *until* there is a rule which admits the subtree and mentions the corresponding slash feature in its LHS. However, this rule is just the target rule the system needs to acquire (e.g. "NP --> SN S[/=NP]"). In the model, the theta-theory and FFP need to work together to locate the moved constituents. The locating process propagates the slash feature upward, and as the *first* constituent with no theta-roles assigned is encountered in a subtree, the constituent is treated as the moved constituent, and the locating process then terminates.

Similarly, in the case of reduced relative clauses, such as "The boy running in the park" and "The boy seen in the room", FFP and the theta-theory may facilitate the acquisition of unbounded dependencies. In the former example, "running" is allowed to assign an "AGENT" theta-role to "the boy". Therefore, there are no missing NPs in the VP "running". In the latter sentence, since the verb "seen" with the passive participle form cannot assign theta-role to "the boy", an NP must be missing in the VP "seen".

### 3.2 Acquisition of unbounded dependencies in topicalizations

The way of acquiring topicalization constructions is quite similar to the way of acquiring relative clauses. Since there is an "extra" constituent (e.g. NP, AP, or PP) not been assigned any theta roles, there must be some corresponding constituent missing in the structures after the extra constituent. Therefore, the acquisition of unbounded dependency may always be triggered, and locating process may also be succeeded in finding the extra constituent.

However, there might be multiple places from which the extra constituent is moved (Gazdar[4]). For example, consider the sentence "Sandy we want to succeed". It may be interpreted as "We want Sandy to succeed" or "We want to succeed Sandy". In the model, the acquisition of unbounded dependency is triggered whenever necessary. Therefore, the learning system will adopt the first interpretation. Fortunately, no matter which interpretation is adopted, from the acquisition point of view, the target rule "S --> NP S[/=NP]" may be learned.

### 3.3 Acquisition of unbounded dependencies in wh-movement

Wh-movement is also characterized as A-Bar-movement. Therefore, unbounded dependency in wh-movement may be learned in a similar way to acquiring relative clauses. The major difference is that, additional transformation is needed (e.g. in English, the auxiliary-verb inversion). Auxiliary-verb inversion can be treated as special phrase patterns which may be learned in the way discussed in section 2.3. For example, consider the wh-questions "What do you want?". The target rule is "S-maj --> what do S[/=NP]". It requires that, after matching "what" and "do", an S with a missing NP is expected for constructing an S-maj.

Generality of the acquired rules deserves further elaboration here. The reader may question why a better rule "S-maj --> wh aux S[/=NP]", where "wh" denotes a category covering wh-words and "aux" denotes a category covering auxiliary verbs, is not acquired. Unfortunately, universal linguistic principles give no help in this case, since it is the "Peripheral Grammar" that needs to take the responsibility of this kind of generalization. However, peripheral grammar is what the system tries to learn. Without any prior knowledge about the peripheral grammar of a particular natural language, over-generalization might be committed due to either the categories that are not well pre-classified or some special phrase patterns (e.g. not only S but also S) that cannot be generalized in this way. Therefore, the more specific version is preferred by the model. The specific rule "S-maj --> what do S[/=NP]" may be further generalized as more empirical evidences are available (step 5 in the learning algorithm).

## 4. Experiment and evaluation

For efficiency, the system is implemented in C language on PC-386 computers. There are about five thousand lines of code in the program. The system can acquire thematic features of unknown nouns, argument structures of verbs, general phrase structure rules, and special patterns (such as "Not only S, but also S") which are all essential for a practical parser. About thirty general grammar rules and thousands of lexicon entries are currently in the dynamic part of the system. They are either initially given (for bootstrapping) or acquired by the system. The initially given knowledge includes the syntactic and thematic features of some nouns and verbs and a general set of phrase structure rules such as "S --> NP VP" that can be easily constructed (recall that the agreement in number between the NP and the VP is licensed by the Control Agreement Principle). The features of the words in sentences that trigger the acquisition of new rules should be available. Otherwise, no features can be propagated by the direction of the universal linguistic principles, and in turn, the acquired rules will be erroneous (recall sec. 2.2). On the other hand, when the system tries to acquire features of unknown words, all rules (e.g. argument structures of verbs) for parsing the sentence should be available. Rule acquisition and lexicon feature acquisition depend on each other in learning.

### 4.1 Efficiency of parsing

To show the parsing efficiency after learning, we show some interesting data concerning the effects of the introduced problem solving strategies. We had employed the strategies of common work sharing, dynamic conflict resolution, and knowledge indexing in the parsing module (Liu[13]). Common work sharing keeps a record of both succeeded and failed goals to eliminate redundant exploration. Dynamic conflict resolution resolves ambiguities in parsing by dynamically scanning the history of parsing and current input. Therefore, a set of parsing (diagnostic) rules in traditional Marcus' parsing (Liu[12]), which is quite difficult to maintain and acquire, may be avoided. Indexing adopts the concept of lexicon-driven NLP to assimilate and retrieve relevant knowledge pieces.

In the experiment, we use 77 sentences to test the performance of the problem solver (parser) after learning. Most of the sentences come from a testing corpus originally collected from Chinese students' articles for grammar and style checking. The result is shown in Table 1. Since knowledge indexing maintains knowledge retrieval efficiency after learning, we focus on, under indexing, the effects of common work sharing and dynamic conflict resolution. As the result shows, common work sharing has significant contribution to efficiency. When it is incorporated, dynamic conflict resolution further improves the efficiency. Otherwise, the performance cannot be acceptable. It is interesting to note that, when common works are not shared among alternatives, the overhead caused by redundant invocation of conflict resolution even slows down the global efficiency.

Table 1. Accumulated run time (in second).

| Strategies | Run Time |
|---|---|
| Indexing+Sharing+Resolution | 62.44 |
| Indexing+Sharing+Non-resolution | 97.19 |
| Indexing+Non-sharing+Resolution | 3659.76 |
| Indexing+Non-sharing+Non-resolution | 3006.39 |

The result also shows that, if different learning approaches (Holder[7]), operationality criteria (Keller[8]), or intelligent knowledge selection methods (Minton[16]) are introduced without improving problem solving strategies, many "useful" or "good" knowledge pieces will be discarded because of the poor problem solving performance. As a result, the effective power of EBL may be limited, and even worse, the incomplete domain theory cannot be enhanced.

### 4.2 Minimal domain theory

As described above, the static part consists of the universal linguistic principles which are assumed to be invariant and innate to the system. To design an effective explanation-based natural language acquisition model, the application of universal linguistic knowledge is valuable. As the model is applied to other languages (e.g. Chinese), whether the static part is adequate or not becomes an interesting problem (Huang[30]). We believe that a more concrete and "universal" model can be expected only after analyzing various learning and processing requirements of different languages. This analysis can help us to define the minimal static domain knowledge which is the core of EBL.

In fact, more predefined domain knowledge also introduces more domain constraints which might turn to be obstacles in different learning situations (e.g. different target languages). According to the GPSG formalism, there are still five components that are responsible for licensing natural language sentences but not included as innate domain theory in our model. They are Feature Co-occurrence Restriction (FCR), Feature Specification Default (FSD), Lexical Immediate Dominance Rules (LIDs), Non-Lexical Immediate Dominance Rules (NLIDs), Metarules, and Linear Precedence Statements (LPS). These principles are either the target knowledge to be acquired (e.g. LIDs, NLIDs, LPS) or the principles that need fine-tuning (e.g. FCR, FSD, Metarules) among different natural languages. Although the introduction of FCR, FSD and Metarules makes knowledge representation more compact by reducing redundancies in knowledge bases, to acquire them needs a huge amount of empirical generalization which may be intractable, especially when empirical generalization is expensive in learning. Fortunately, they have no effects on the learnability of various parsing knowledge. In fact, by fast knowledge indexing, enumerating knowledge pieces (possibly redundant from the point of view of FCR, FSD and Metarules) in the general grammar rule base and the lexicon does not deteriorate parsing efficiency.

### 4.3 The validity and availability of the given parse trees

The kinds of input given to a learning system is essential and can vary from different learning methodologies and systems. The learning system utilizes the input to derive (or infer) new knowledge (such as a consistently generalized version of knowledge). In natural language acquisition, additional input is indispensable (the semantic bootstrapping hypothesis, Pinker[20]). In practice, the form and availability of the extra input have a strong effect on the plausibility (including portability and convergence quality) of the model.

In our model, giving a parse tree of an unrecognized sentence to the system seems to be a strong assumption. From the parse tree, we can have not only categories of words but also phrase structures of the input sentence. However, there are still many things remaining to be learned. No parsers can completely parse sentences using general phrase structure rules only. The information in the parse tree is properly generalized according to the linguistic principles and current parsing knowledge. The system can thus derive practically essential knowledge (syntactic and thematic knowledge) based on the informative initial input knowledge.

In fact, the extra input can range from syntactic association to semantic association (or both) to the current sentence. The critical point is what kind of information the input provides. Giving syntactic information (Zernik[28], Lytinen[15], Liu[11], Liu[13]) to the system allows the acquisition of more syntactic (and perhaps semantic) information, while entering semantic information (Berwick[2], Siskind[24], Pinker[20], Zernik[27]) facilitates the acquisition of more semantic information.

Another aspect of providing extra input is the availability of the input. In practice, providing complicated semantic association is a very heavy burden for a naive user. In language acquisition, we can also rely on a large "pre-processed" corpus. However, to what extent the corpus should be pre-processed? As pointed out in section 2.2 (and in Zernik[28] also), a minimally pre-processed corpus allowing only co-occurrence acquisition contributes little in phrase structure and lexicon acquisition. Two constituents that are conceptually related (e.g. a verb and its argument) may not be co-located because they are distant from each other, while two constituents that are conceptually unrelated may still be co-located due to inadequate information in the minimally pre-processed corpus (Basili[1], Smadja[25]). Furthermore, co-location acquisition needs a large corpus and a large memory. To reduce these difficulties, a partial parser (Basili[1], Sekine[22], Smadja[25]), a tagger (Zernik[29]), and/or a set of predefined syntactic and semantic categories (Basili[1], Smadja[25]) need to be constructed before learning. However, the limitations (e.g. the incorrect analysis on the text and incomplete set of categories) coming from these preprocessing may also be introduced.

Machine-readable dictionaries were also the available sources of the training input in recent years (Sanfilippo[21]). To acquire knowledge from them, a pre-processor (e.g. a parser) is needed for processing the description text part and the example part in lexical entries. When the system tries to learn from multiple dictionaries or multiple lexical entries, filtering and combining information from different sources are needed. These processing modules are the basic requirements, and hence the limitations, of the learning model.

Interactive acquisition (Lang[9], Liu[11], Lu[14], Simmons[23], Velard[26]) shows another alternative for giving additional information to the system. The confirmation information is available only if there is a well-trained trainer monitoring the learning behavior of the system. In addition, the number of questions needed for justifying the generated hypotheses may become a critical bottleneck (Liu[11]).

The parse trees assumed in the model can come from the trainer, the existing incomplete parsers, and the parse tree bank constructed for research evaluation (Grishman[5]). Currently, we are trying to transform an on-line parse tree corpus (PENN tree bank in the CD-ROM from Association of Computational Linguistics Data Collection Initiative) into the form suitable in the model. By exploiting the large available parse tree bank, the system can converge to a more complete parser without relying on the parse trees given by users.

### 4.4 Future work in the acquisition of unbounded dependency

The acquisition of unbounded dependency in "missing-object" constructions has not yet been well-developed in the model. For example, in the sentence "Kim is easy to please", there is an NP missing in the VP "please". However, for the sentence "Kim is eager to please", the VP "please" does not have any NP missing (Gazdar[4]). GPSG deals with the problem by using lexical immediate dominance (lexical ID) rules of "easy" and "eager". However, from the acquisition point of view, the incorporated universal linguistic principles have no help to the discrimination of the two sentence structures.

## 5. Conclusion

In this paper, we consider the effects of incorporating universal linguistic principles from the viewpoint of computational natural language acquisition. Portability and learnability are the major concerns of the explanation-based natural language acquisition model. Currently, we find the theta-theory and the universal feature instantiation principles may play the critical role as the domain theory in EBL. The acquired knowledge can be properly generalized (without causing over-generalization) by following the guidance of these principles. In the acquisition of unbounded dependency, these principles facilitate not only the triggering of the chaining process, but also the locating of the moved constituents. The acquired operational knowledge, including Context-Free grammar rules and syntactic and thematic requirements of lexicons, becomes new domain theory for later parsing and learning.

## Acknowledgement

This research is supported in part by NSC under the grant NSC81-0408-E-007-02.

## References

[1] Basili R., Pazienza M. T., and Velardi P., *Computational Lexicons: the Neat Examples and the Odd Exemplars,* Proc. of the 3rd Conference on Applied NLP, pp. 96-103, 1992.

[2] Berwick R. C., *The Acquisition of Syntactic Knowledge,* The MIT Press, Cambridge, Massachusetts, London, England, 1985.

[3] Flann N. S. and Dietterich T. G., *A Study of Explanation-Based Methods for Inductive Learning,* Machine Learning, 4, 187-226, 1989.

[4] Gazdar G., Klein E., Pullum G. K., and Sag I. A., *Generalized Phrase Structure Grammar,* Harvard University Press, Cambridge, Massachusetts, 1985.

[5] Grishman R., Macleod C., and Sterling J., *Evaluating Parsing Strategies Using Standardized Parse Files,* Proc. of the Third Applied NLP, pp. 156-161, 1992.

[6] Hall R. J., *Learning by Failing to Explain: Using Partial Explanation to Learn in Incomplete or Intractable Domains,* Machine Learning, 3: 45-77, 1988.

[7] Holder L. B., *The General Utility Problem in Machine Learning,* Proc. of the Seventh

International Machine Learning Conference, pp. 402-410, 1990.

[8] Keller R. M., *Defining Operationality for Explanation-Based Learning,* Artificial Intelligence, 35: 227-241, 1988.

[9] Lang F.-M. and Hirschman L., *Improved Portability and Parsing through Interactive Acquisition of Semantic Information,* Proc. of the 2nd conference on Applied Natural Language Processing, pp. 49-57, 1988.

[10] Lehman J. F., *Adaptive Parsing: A General Method for Learning Idiosyncratic Grammars,* Proc. of the seventh international machine learning conference, pp. 368-376, 1990.

[11] Liu R.-L. and Soo V.-W., *Parsing Driven Generalization for Natural Language Acquisition,* Proc. of ROCLING, R.O.C., pp. 351-376, 1990.

[12] Liu R.-L. and Soo V.-W., *Dealing with Ambiguities in English Conjunctions and Comparatives by A Deterministic Parser,* International Journal of Pattern Recognition and Artificial Intelligence, December, Vol. 4, No. 4, pp. 629-649, 1990.

[13] Liu R.-L. and Soo V.-W., *Augmenting and Efficiently Utilizing Domain Theory in Explanation-Based Natural Language Acquisition,* Proc. of the 9th International Machine Learning Conference, ML92, 1992.

[14] Lu R., Liu Y., and Li X., *Computer-Aided Grammar Acquisition in the Chinese Understanding System CUSAGA,* Proc. of IJCAI, pp. 1550-1555, 1989.

[15] Lytinen S. L. and Moon C. E., *A Comparison of Learning Techniques in Second Language Learning,* Proc. of the 7th Machine Learning conference, pp. 377-383, 1990.

[16] Minton S., *Quantitative Results Concerning the Utility of Explanation-Based Learning,* Proc. of AAAI, pp. 564-569, 1988.

[17] Mitchell T. M., Keller R. M., and Kedar-Cabelli S. T., *Explanation-Based Generalization: A Unifying View,* Machine Learning, 1: 47-80, 1986.

[18] Pazzani M. J., *Detecting and Correcting Errors of Omission After Explanation-based Learning,* Proc. of IJCAI, pp. 713-718, 1989.

[19] Chomsky N., *Lectures on Government and Binding* Foris Publications - Dordrecht, 1981.

[20] Pinker S., *Language Learnability and Language Development,* The Harvard University Press, Cambridge, Massachusetts, London, England, 1984.

[21] Sanfilippo A. and Poznanski V., *The Acquisition of Lexical Knowledge from Combined Machine-Readable Dictionary Sources,* Proc. of the Third Conference on Applied NLP, pp. 80-87, 1992.

[22] Sekine S., Carroll J. J., Ananiadou S., and Tsujii J., *Automatic Learning for Semantic Collocation,* Proc. of the Third Conference on Applied NLP, pp. 104-110, 1992.

[23] Simmons R. F. and Yu Y.-H., *The Acquisition and Application of Context Sensitive Grammar for English,* Proc. of ACL, pp. 122-129, 1991.

[24] Siskind J. M., *Acquiring Core Meanings of Words, Represented as Jackendoff-style Conceptual structures, from Correlated Streams of Linguistic and Non-linguistic Input,* Proc. of the 28th annual meeting of ACL, pp. 143-156, 1990.

[25] Smadja F. A., *From N-Grams to Collocations: An Evaluation of EXTRACT,* Proc. of ACL, pp. 279-284, 1991.

[26] Velard P., Pazienza M. T., and Fasolo M., *How to Encode Semantic Knowledge: A Method for Meaning Representation and Computer-Aided Acquisition,* Computational Linguistic, Vol. 17, No. 2, pp. 153-170, 1991.

[27] Zernik U., *Learning Idioms -- With and Without Explanation,* Proc. of IJCAI, pp. 133-136, 1987.

[28] Zernik U., *Lexicon Acquisition: Learning from Corpus by Capitalizing on Lexical Categories,* Proc. of IJCAI, pp. 1556-1562, 1989.

[29] Zernik U. and Jacobs P., *Tagging for Learning: Collecting Thematic Relation from Corpus,* Proc. of COLING, pp. 34-39, 1990.

[30] Huang C.-R., *Certainty in Functional Uncertainty,* Journal of Chinese Linguistics, Vol. 20, No. 2, pp. 247-287, 1992.