AN EMPIRICAL STUDY ON THEMATIC KNOWLEDGE ACQUISITION BASED ON SYNTACTIC CLUES AND HEURISTICS

Rey-Long Liu* and Von-Wun Soo**

Department of Computer Science
National Tsing-Hua University
HsinChu, Taiwan, R.O.C.
Email: dr798303@cs.nthu.edu.tw* and soo@cs.nthu.edu.tw**

Abstract

Thematic knowledge is a basis of semantic interpretation. In this paper, we propose an acquisition method to acquire thematic knowledge by exploiting syntactic clues from training sentences. The syntactic clues, which may be easily collected by most existing syntactic processors, reduce the hypothesis space of the thematic roles. The ambiguities may be further resolved by the evidences either from a trainer or from a large corpus. A set of heuristics based on linguistic constraints is employed to guide the ambiguity resolution process. When a trainer is available, the system generates new sentences whose thematic validities can be justified by the trainer. When a large corpus is available, the thematic validity may be justified by observing the sentences in the corpus. Using this way, a syntactic processor may become a thematic recognizer by simply deriving its thematic knowledge from its own syntactic knowledge.

Keywords: Thematic Knowledge Acquisition, Syntactic Clues, Heuristics-guided Ambiguity Resolution, Corpus-based Acquisition, Interactive Acquisition

1. INTRODUCTION

Natural language processing (NLP) systems need various knowledge including syntactic, semantic, discourse, and pragmatic knowledge in different applications. Perhaps due to the relatively well-established syntactic theories and formalisms, there were many syntactic processing systems either manually constructed or automatically extended by various acquisition methods (Asker92, Berwick85, Brent91, Liu92b, Lytinen90, Samuelsson91, Simmons91 Sanfilippo92, Smadja91 and Sekine92). However, the satisfactory representation and acquisition methods of domain-independent semantic, discourse, and pragmatic knowledge are not yet developed or computationally implemented. NLP systems often suffer the dilemma of semantic representation. Sophisticated representation of semantics has better expressive power but imposes difficulties on acquisition in practice. On the other hand, the poor adequacy of naive semantic representation may deteriorate the performance of NLP systems. Therefore, for plausible acquisition and processing, domain-dependent semantic bias was often employed in many previous acquisition systems (Grishman92b, Lang88, Lu89, and Velardi91).

In this paper, we present an implemented system that acquires domain-independent thematic knowledge using available syntactic resources (e.g. syntactic processing systems and syntactically processed corpora). Thematic knowledge can represent semantic or conceptual entities. For correct and efficient parsing, thematic expectation serves as a basis for conflict resolution (Taraban88). For natural language understanding and other applications (e.g. machine translation), thematic role recognition is a major step. Thematic relations may serve as the vocabulary shared by the parser, the discourse model, and the world knowledge (Tanenhaus89). More importantly, since thematic structures are perhaps most closely linked to syntactic structures (Jackendoff72), thematic knowledge acquisition may be more feasible when only syntactic resources are available. The consideration of the availability of the resources from which thematic knowledge may be derived promotes the practical feasibility of the acquisition method.

In general, lexical knowledge of a lexical head should (at least) include 1) the number of arguments of the lexical head, 2) syntactic properties of the arguments, and 3) thematic roles of the arguments (the argument structure). The former two components may be either already constructed in available syntactic processors or acquired by many syntactic acquisition systems. However, the acquisition of the thematic roles of the arguments deserves more exploration. A constituent may have different thematic roles for different verbs in different uses. For example, "John" has different thematic roles in (1.1) - (1.4).

(1.1) [Agent John] turned on the light.
(1.2) [Goal John] inherited a million dollars.
(1.3) The magic wand turned [Theme John] into a frog.

Table 1. Syntactic clues for hypothesizing thematic roles

Theta role	Constituent	Animate	Subject	Object	Preposition in PP
Agent(Ag)	NP	y	y	n	by
Goal(Go)	NP	?	y(animate)	y	till,untill,to,into,down
Source(So)	NP	?	y(animate)	y	from
Instrument(In)	NP	n	y(no Ag)	n	with,by
Theme(Th)	NP	?	y	y	of,about
Beneficiary(Be)	NP	y	n	y	for
Location(Lo)	NP,ADJP	n	y	n	at,in,on,under
Time(Ti)	NP(Ti)	n	y	n	at,in,before,after,about,by,on,during
Quantity(Qu)	NP(Qu)	n	-	-	for
Proposition(Po)	Proposition	n	n	n	none
Manner(Ma)	ADVP,PP	n	n	n	in,with
Cause(Ca)	NP	n	y	n	by,for,because of
Result(Re)	NP	n	n	y	in,into

(1.4) The letter reached [Goal John] yesterday.

To acquire thematic lexical knowledge, precise thematic roles of arguments in the sentences needs to be determined.

In the next section, the thematic roles considered in this paper are listed. The syntactic properties of the thematic roles are also summarized. The syntactic properties serve as a preliminary filter to reduce the hypothesis space of possible thematic roles of arguments in training sentences. To further resolve the ambiguities, heuristics based on various linguistic phenomena and constraints are introduced in section 3. The heuristics serve as a general guidance for the system to collect valuable information to discriminate thematic roles. Current status of the experiment is reported in section 4. In section 5, the method is evaluated and related to previous methodologies. We conclude, in section 6, that by properly collecting discrimination information from available sources, thematic knowledge acquisition may be more feasible in practice.

2. THEMATIC ROLES AND SYNTACTIC CLUES

The thematic roles considered in this paper and the syntactic clues for identifying them are presented in Table 1. The syntactic clues include 1) the possible syntactic constituents of the arguments, 2) whether animate or inanimate arguments, 3) grammatical functions (subject or object) of the arguments when they are Noun Phrases (NPs), and 4) prepositions of the prepositional phrase in which the arguments may occur. The syntactic constituents include NP, Proposition (Po), Adverbial Phrase (ADVP), Adjective Phrase (ADJP), and Prepositional phrase (PP). In

addition to common animate nouns (e.g. he, she, and I), proper nouns are treated as animate NPs as well. In Table 1, "y", "n", "?", and "-" denote "yes", "no", "don't care", and "seldom" respectively. For example, an Agent should be an animate NP which may be at the subject (but not object) position, and if it is in a PP, the preposition of the PP should be "by" (e.g. "John" in "the light is turned on by John").

We consider the thematic roles to be well-known and referred, although slight differences might be found in various works. The intrinsic properties of the thematic roles had been discussed from various perspectives in previous literatures (Jackendoff72 and Gruber76). Grimshaw88 and Levin86 discussed the problems of thematic role marking in so-called light verbs and adjectival passives. More detailed description of the thematic roles may be found in the literatures. To illustrate the thematic roles, consider (2.1)-(2.9).

(2.1) [Ag The robber] robbed [So the bank] of [Th the money].

(2.2) [Th The rock] rolled down [Go the hill].

(2.3) [In The key] can open [Th the door].

(2.4) [Go Will] inherited [Qu a million dollars].

(2.5) [Th The letter] finally reached [Go John].

(2.6) [Lo The restaurant] can dine [Th fifty people].

(2.7) [Ca A fire] burned down [Th the house].

(2.8) [Ag John] bought [Be Mary] [Th a coat] [Ma reluctantly].

(2.9) [Ag John] promised [Go Mary] [Po to marry her].

When a training sentence is entered, arguments of lexical verbs in the sentence need to be extracted before learning. This can be achieved by invoking a syntactic processor.

244

Table 2. Heuristics for discriminating thematic roles

• **Volition Heuristic (VH):** Purposive constructions (e.g. in order to) and purposive adverbials (e.g. deliberately and intentionally) may occur in sentences with Agent arguments (Gruber76).

• **Imperative Heuristic (IH):** Imperatives are permissible only for Agent subjects (Gruber76).

• **Thematic Hierarchy Heuristic (THH):** Given a thematic hierarchy (from higher to lower) "Agent > Location, Source, Goal > Theme", the passive by-phrases must reside at a higher level than the derived subjects in the hierarchy (i.e. the Thematic Hierarchy Condition in Jackendoff72). In this paper, we set up the hierarchy: Agent > Location, Source, Goal, Instrument, Cause > Theme, Beneficiary, Time, Quantity, Proposition, Manner, Result. Subjects and objects cannot reside at the same level.

• **Preposition Heuristic (PH):** The prepositions of the PPs in which the arguments occur often convey good discrimination information for resolving thematic roles ambiguities (see the "Preposition in PP" column in Table 1).

• **One-Theme Heuristic (OTH):** An argument is preferred to be Theme if it is the only possible Theme in the argument structure.

• **Uniqueness Heuristic (UH):** No two arguments may receive the same thematic role (exclusive of conjunctions and anaphora which co-relate two constituents assigned with the same thematic role).

If the sentence is selected from a syntactically processed corpus (such as the PENN treebank) the arguments may be directly extracted from the corpus. To identify the thematic roles of the arguments, Table 1 is consulted.

For example, consider (2.1) as the training sentence. Since "the robber" is an animate NP with the subject grammatical function, it can only qualify for Ag, Go, So, and Th. Similarly, since "the bank" is an inanimate NP with the object grammatical function, it can only satisfy the requirements of Go, So, Th, and Re. Because of the preposition "of", "the money" can only be Th. As a result, after consulting the constraints in Table 1, "the robber", "the bank", and "the money" can only be {Ag, Go, So, Th}, {Go, So, Th, Re}, and {Th} respectively. Therefore, although the clues in Table 1 may serve as a filter, lots of thematic role ambiguities still call for other discrimination information and resolution mechanisms.

3. FINDING EXTRA INFORMATION FOR RESOLVING THETA ROLE AMBIGUITIES

The remaining thematic role ambiguities should be resolved by the evidences from other sources. Trainers and corpora are the two most commonly available sources of the extra information. Interactive acquisition had been applied in various systems in which the oracle from the trainer may reduce most

ambiguities (e.g. Lang88, Liu93, Lu89, and Velardi91). Corpus-based acquisition systems may also converge to a satisfactory performance by collecting evidences from a large corpus (e.g. Brent91, Sekine92, Smadja91, and Zernik89). We are concerned with the kinds of information the available sources may contribute to thematic knowledge acquisition.

The heuristics to discriminate thematic roles are proposed in Table 2. The heuristics suggest the system the ways of collecting useful information for resolving ambiguities. Volition Heuristic and Imperative Heuristic are for confirming the Agent role, One-Theme Heuristic is for Theme, while Thematic Hierarchy Heuristic, Preposition Heuristic and Uniqueness Heuristic may be used in a general way.

It should be noted that, for the purposes of efficient acquisition, not all of the heuristics were identical to the corresponding original linguistic postulations. For example, Thematic Hierarchy Heuristic was motivated by the Thematic Hierarchy Condition (Jackendoff72) but embedded with more constraints to filter out more hypotheses. One-Theme Heuristic was a relaxed version of the statement "every sentence has a theme" which might be too strong in many cases (Jackendoff87).

Because of the space limit, we only use an example to illustrate the idea. Consider (2.1) "The robber robbed the bank of the money" again. As

mentioned above, after applying the preliminary syntactic clues, "the robber", "the bank", and "the money" may be {Ag, Go, So, Th}, {Go, So, Th, Re}, and {Th} respectively. By applying Uniqueness Heuristic to the Theme role, the argument structure of "rob" in the sentence can only be

(AS1) "{Ag, Go, So}, {Go, So, Re}, {Th}",

which means that, the external argument is {Ag, Go, So} and the internal arguments are {Go, So, Re} and {Th}. Based on the intermediate result, Volition Heuristic, Imperative Heuristic, Thematic Hierarchy Heuristic, and Preposition Heuristic could be invoked to further resolve ambiguities.

Volition Heuristic and Imperative Heuristic ask the learner to verify the validities of the sentences such as "John intentionally robbed the bank" ("John" and "the robber" matches because they have the same properties considered in Table 1 and Table 2). If the sentence is "accepted", an Agent is needed for "rob". Therefore, the argument structure becomes

(AS2) "{Ag}, {Go, So, Re}, {Th}".

Thematic Hierarchy Heuristic guides the learner to test the validity of the passive form of (2.1). Similarly, since sentences like "The bank is robbed by Mary" could be valid, "The robber" is higher than "the bank" in the Thematic Hierarchy. Therefore, the learner may conclude that either AS3 or AS4 may be the argument structure of "rob":

(AS3) "{Ag}, {Go, So, Re}, {Th}"
(AS4) "{Go, So}, {Re}, {Th}".

Preposition Heuristic suggests the learner to to resolve ambiguities based on the prepositions of PPs. For example, it may suggest the system to confirm: The money is from the bank? If so, "the bank" is recognized as Source. The argument structure becomes

(AS5) "{Ag, Go}, {So}, {Th}".

Combining (AS5) with (AS3) or (AS5) with (AS2), the learner may conclude that the argument structure of "rob" is "{Ag}, {So}, {Th}".

In summary, as the arguments of lexical heads are entered to the acquisition system, the clues in Table 1 are consulted first to reduce the hypothesis space. The heuristics in Table 2 are then invoked to further resolve the ambiguities by collecting useful information from other sources. The information that the heuristics suggest the system to collect is the thematic validities of the sentences that may help to confirm the target thematic roles.

The confirmation information required by Volition Heuristic, Imperative Heuristic, and Thematic

Hierarchy Heuristic may come from corpora (and of course trainers as well), while Preposition Heuristic sometimes needs the information only available from trainers. This is because the derivation of new PPs might generate ungrammatical sentences not available in general corpora. For example, (3.1) from (2.3) "The key can open the door" is grammatical, while (3.2) from (2.5) "The letter finally reached John" is ungrammatical.

(3.1) The door is opened *by* the key.
(3.2) *The letter finally reached *to* John.

Therefore, simple queries as above are preferred in the method.

It should also be noted that since these heuristics only serve as the guidelines for finding discrimination information, the sequence of their applications does not have significant effects on the *result* of learning. However, the number of queries may be minimized by applying the heuristics in the order: Volition Heuristic and Imperative Heuristic -> Thematic Hierarchy Heuristic -> Preposition Heuristic. One-Theme Heuristic and Uniqueness Heuristic are invoked each time current hypotheses of thematic roles are changed by the application of the clues, Volition Heuristic, Imperative Heuristic, Thematic Hierarchy Heuristic, or Preposition Heuristic. This is because One-Theme Heuristic and Uniqueness Heuristic are constraint-based. Given a hypothesis of thematic roles, they may be employed to filter out impossible combinations of thematic roles without using any queries. Therefore, as a query is issued by other heuristics and answered by the trainer or the corpus, the two heuristics may be used to "extend" the result by further reducing the hypothesis space.

4. EXPERIMENT

As described above, the proposed acquisition method requires syntactic information of arguments as input (recall Table 1). We believe that the syntactic information is one of the most commonly available resources. It may be collected from a syntactic processor or a syntactically processed corpus. To test the method with a public corpus as in Grishman92a, the PENN TreeBank was used as a syntactically processed corpus for learning. Argument packets (including VP packets and NP packets) were extracted from ATIS corpus (including JUN90, SRI_TB, and TI_TB tree files), MARI corpus (including AMBIG and WBUR tree files), MUC1 corpus, and MUC2 corpus of the treebank. VP packets and NP packets recorded syntactic properties of the arguments of verbs and nouns respectively.

246

Table 3. Argument extraction from TreeBank

Corpus	Sentences	Words	VP packets	Verbs	NP packets	Nouns
ATIS	1373	15286	1716	138	959	188
MARI	543	9897	1067	509	425	288
MUC1	1026	22662	1916	732	907	490
MUC2	3341	73548	6410	1556	3313	1177

Since not all constructions involving movement were tagged with trace information in the corpus, to derive the arguments, the procedure needs to consider the constructions of passivization, interjection, and unbounded dependency (e.g. in relative clauses and wh-questions). That is, it needs to determine whether a constituent is an argument of a verb (or noun), whether an argument is moved, and if so, which constituent is the moved argument. Basically, Case Theory, Theta Theory (Chomsky81), and Foot Feature Principle (Gazdar85) were employed to locate the arguments (Liu92a, Liu92b).

Table 3 summarizes the results of the argument extraction. About 96% of the trees were extracted. Parse trees with too many words (60) or nodes (i.e. 50 subgoals of parsing) were discarded. All VP packets in the parse trees were derived, but only the NP packets having PPs as modifiers were extracted. These PPs could help the system to hypothesize argument structures of nouns. The extracted packets were assimilated into an acquisition system (called EBNLA, Liu92a) as syntactic subcategorization frames. Different morphologies of lexicons were not counted as different verbs and nouns.

As an example of the extracted argument packets, consider the following sentence from MUC1:

"..., at la linea, ..., where a FARC front ambushed an 11th brigade army patrol".

The extraction procedure derived the following VP packet for "ambushed":

ambushed (NP: a FARC front) (WHADVP: where) (NP: an 11th brigade army patrol)

The first NP was the external argument of the verb. Other constituents were internal arguments of the verb. The procedure could not determine whether an argument was optional or not.

In the corpora, most packets were for a small number of verbs (e.g. 296 packets for "show" were found in ATIS). Only 1 to 2 packets could be found for most verbs. Therefore, although the parse trees could provide good quality of argument packets, the information was too sparse to resolve thematic role ambiguities. This is a weakness embedded in most corpus-based acquisition methods, since the learner

might finally fail to collect sufficient information after spending much effort to process the corpus. In that case, the ambiguities need to be temporarily suspended. To seed-up learning and focus on the usage of the proposed method, a trainer was asked to check the thematic validities (yes/no) of the sentences generated by the learner.

Excluding packets of some special verbs to be discussed later and erroneous packets (due to a small amount of inconsistencies and incompleteness of the corpus and the extraction procedure), the packets were fed into the acquisition system (one packet for a verb). The average accuracy rate of the acquired argument structures was 0.86. An argument structure was counted as correct if it was unambiguous and confirmed by the trainer. On average, for resolving ambiguities, 113 queries were generated for every 100 successfully acquired argument structures. The packets from ATIS caused less ambiguities, since in this corpus there were many imperative sentences to which Imperative Heuristic may be applied. Volition Heuristic, Thematic Hierarchy Heuristic, and Preposition Heuristic had almost equal frequencies of application in the experiment.

As an example of how the clues and heuristics could successfully derive argument structures of verbs, consider the sentence from ATIS:

"The flight going to San Francisco ...".

Without issuing any queries, the learner concluded that an argument structure of "go" is "{Th}, {Go}". This was because, according to the clues, "San Francisco" could only be Goal, while according to One-Theme Heuristic, "the flight" was recognized as Theme. Most argument structures were acquired using 1 to 2 queries.

The result showed that, after (manually or automatically) acquiring an argument packet (i.e. a syntactic subcategorization frame plus the syntactic constituent of the external argument) of a verb, the acquisition method could be invoked to upgrade the syntactic knowledge to thematic knowledge by issuing only 113 queries for every 100 argument packets. Since checking the validity of the generated sentences is not a heavy burden for the trainer (answering 'yes'

247

or 'no' only), the method may be attached to various systems for promoting incremental extensibility of thematic knowledge.

The way of counting the accuracy rate of the acquired argument structures deserves notice. Failed cases were mainly due to the clues and heuristics that were too strong or overly committed. For example, the thematic role of "the man" in (4.1) from MARI could not be acquired using the clues and heuristics.

(4.1) Laura ran away with the man.

In the terminology of Gruber76, this is an expression of accompaniment which is not considered in the clues and heuristics. As another example, consider (4.2) also from MARI.

(4.2) The greater Boston area ranked eight among major cities for incidence of AIDS.

The clues and heuristics could not draw any conclusions on the possible thematic roles of "eight".

On the other hand, the cases counted as "failed" did not always lead to "erroneous" argument structures. For example, "Mary" in (2.9) "John promised Mary to marry her" was treated as Theme rather than Goal, because "Mary" is the only possible Theme. Although "Mary" may be Theme in this case as well, treating "Mary" as Goal is more fine-grained.

The clues and heuristics may often lead to acceptable argument structures, even if the argument structures are inherently ambiguous. For example, an NP might function as more than one thematic role within a sentence (Jackendoff87). In (4.3), "John" may be Agent or Source.

(4.3) John sold Mary a coat.

Since Thematic Hierarchy Heuristic assumes that subjects and objects cannot reside at the same level, "John" must not be assigned as Source. Therefore, "John" and "Mary" are assigned as Agent and Goal respectively, and the ambiguity is resolved.

In addition, some thematic roles may cause ambiguities if only syntactic evidences are available. Experiencer, such as "John" in (4.4), and Maleficiary, such as "Mary" in (4.5), are the two examples.

(4.4) Mary surprised John.
(4.5) Mary suffers a headache.

There are difficulties in distinguishing Experiencer, Agent, Maleficiary and Theme. Fortunately, the verbs with Experiencer and Maleficiary may be enumerated before learning. Therefore, the argument structures of these verbs are manually constructed rather than learned by the proposed method.

5. RELATED WORK

To explore the acquisition of domain-independent semantic knowledge, the universal linguistic constraints postulated by many linguistic studies may provide general (and perhaps coarse-grained) hints. The hints may be integrated with domain-specific semantic bias for various applications as well. In the branch of the study, GB theory (Chomsky81) and universal feature instantiation principles (Gazdar85) had been shown to be applicable in syntactic knowledge acquisition (Berwick85, Liu92a, Liu92b). The proposed method is closely related to those methodologies. The major difference is that, various thematic theories are selected and computationalized for thematic knowledge acquisition. The idea of structural patterns in Montemagni92 is similar to Preposition Heuristic in that the patterns suggest general guidance to information extraction.

Extra information resources are needed for thematic knowledge acquisition. From the cognitive point of view, morphological, syntactic, semantic, contextual (Jacobs88), pragmatic, world knowledge, and observations of the environment (Webster89, Siskind90) are all important resources. However, the availability of the resources often deteriorated the feasibility of learning from a practical standpoint. The acquisition often becomes "circular" when relying on semantic information to acquire target semantic information.

Predefined domain linguistic knowledge is another important information for constraining the hypothesis space in learning (or for semantic bootstrapping). From this point of view, lexical categories (Zernik89, Zernik90) and theory of lexical semantics (Pustejovsky87a, Pustejovsky87b) played similar roles as the clues and heuristics employed in this paper. The previous approaches had demonstrated theoretical interest, but their performance on large-scale acquisition was not elaborated. We feel that, requiring the system to use available resources only (i.e. syntactic processors and/or syntactically processed corpora) may make large-scale implementations more feasible. The research investigates the issue as to what extent an acquisition system may acquire thematic knowledge when only the syntactic resources are available.

McClelland86 showed a connectionist model for thematic role assignment. By manually encoding training assignments and semantic microfeatures for a limited number of verbs and nouns, the connectionist network learned how to assign roles. Stochastic approaches (Smadja91, Sekine92) also employed available corpora to acquire collocational data for resolving ambiguities in parsing. However, they acquired numerical values by observing the whole

training corpus (non-incremental learning). Explanation for those numerical values is difficult to derive in those models. As far as the large-scale thematic knowledge acquisition is concerned, the incremental extensibility of the models needs to be further improved.

6. CONCLUSION

Preliminary syntactic analysis could be achieved by many natural language processing systems. Toward semantic interpretation on input sentences, thematic lexical knowledge is needed. Although each lexicon may have its own idiosyncratic thematic requirements on arguments, there exist syntactic clues for hypothesizing the thematic roles of the arguments. Therefore, exploiting the information derived from syntactic analysis to acquire thematic knowledge becomes a plausible way to build an extensible thematic dictionary. In this paper, various syntactic clues are integrated to hypothesize thematic roles of arguments in training sentences. Heuristics-guided ambiguity resolution is invoked to collect extra discrimination information from the trainer or the corpus. As more syntactic resources become available, the method could upgrade the acquired knowledge from syntactic level to thematic level.

Acknowledgement

References

[Asker92] Asker L., Gamback B., Samuelsson C., *EBL2: An Application to Automatic Lexical Acquisition*, Proc. of COLING, pp. 1172-1176, 1992.

[Berwick85] Berwick R. C., *The Acquisition of Syntactic Knowledge*, The MIT Press, Cambridge, Massachusetts, London, England, 1985.

[Brent91] Brent M. R., *Automatic Acquisition of Subcategorization Frames from Untagged Text*, Proc. of the 29th annual meeting of the ACL, pp. 209-214, 1991.

[Chomsky81] Chomsky N., *Lectures on Government and Binding*, Foris Publications - Dordrecht, 1981.

[Gazdar85] Gazdar G., Klein E., Pullum G. K., and Sag I. A., *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge. Massachusetts, 1985.

[Grimshaw88] Grimshaw J. and Mester A., *Light Verbs and Theta-Marking*, Linguistic Inquiry, Vol. 19, No. 2, pp. 205-232, 1988.

[Grishman92a] Grishman R., Macleod C., and Sterling J., *Evaluating Parsing Strategies Using Standardized Parse Files*, Proc. of the Third Applied NLP, pp. 156-161, 1992.

[Grishman92b] Grishman R. and Sterling J., *Acquisition of Selectional Patterns*, Proc. of COLING-92, pp. 658-664, 1992.

[Gruber76] Gruber J. S., *Lexical Structures in Syntax and Semantics*, North-Holland Publishing Company, 1976.

[Jackendoff72] Jackendoff R. S., *Semantic Interpretation in Generative Grammar*, The MIT Press, Cambridge, Massachusetts, 1972.

[Jackendoff87] Jackendoff R. S., *The Status of Thematic Relations in Linguistic Theory*, Linguistic Inquiry, Vol. 18, No. 3, pp.369-411, 1987.

[Jacobs88] Jacobs P. and Zernik U., *Acquiring Lexical Knowledge from Text: A Case Study*, Proc. of AAAI, pp. 739-744, 1988.

[Lang88] Lang F.-M. and Hirschman L., *Improved Portability and Parsing through Interactive Acquisition of Semantic Information*, Proc. of the second conference on Applied Natural Language Processing, pp. 49-57, 1988.

[Levin86] Levin B. and Rappaport M., *The Formation of Adjectival Passives*, Linguistic Inquiry, Vol. 17, No. 4, pp. 623-661, 1986.

[Liu92a] Liu R.-L. and Soo V.-W., *Augmenting and Efficiently Utilizing Domain Theory in Explanation-Based Natural Language Acquisition*, Proc. of the Ninth International Machine Learning Conference, ML92, pp. 282-289, 1992.

[Liu92b] Liu R.-L and Soo V.-W., *Acquisition of Unbounded Dependency Using Explanation-Based Learning*, Proc. of ROCLING V, 1992.

[Liu93] Liu R.-L. and Soo V.-W., *Parsing-Driven Generalization for Natural Language Acquisition*, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 7, No. 3, 1993.

[Lu89] Lu R., Liu Y., and Li X., *Computer-Aided Grammar Acquisition in the Chinese Understanding System CUSAGA*, Proc. of IJCAI, pp. 1550-1555, 1989.

[Lytinen90] 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.

[McClelland86] McClelland J. L. and Kawamoto A. H., *Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences*, in Parallel Distributed Processing, Vol. 2, pp. 272-325, 1986.

[Montemagni92] Montemagni S. and Vanderwende L., *Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionary*, Proc. of COLING-92, pp. 546-552, 1992.

[Pustejovsky87a] Pustejovsky J. and Berger S., *The Acquisition of Conceptual Structure for the Lexicon*, Proc. of AAAI, pp. 566-570, 1987.

[Pustejovsky87b] Pustejovsky J, *On the Acquisition of Lexical Entries: The Perceptual Origin of Thematic Relation*, Proc. of the 25th annual meeting of the ACL, pp. 172-178, 1987.

[Samuelsson91] Samuelsson C. and Rayner M., *Quantitative Evaluation of Explanation-Based Learning as an Optimization Tool for a Large-Scale Natural Language System*, Proc. of IJCAI, pp. 609-615, 1991.

[Sanfilippo92] Sanfilippo A. and Pozanski V., *The Acquisition of Lexical Knowledge from Combined Machine-Readable Dictionary Sources*, Proc. of the Third Conference on Applied NLP, pp. 80-87, 1992.

[Sekine92] 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.

[Simmons91] Simmons R. F. and Yu Y.-H., *The Acquisition and Application of Context Sensitive Grammar for English*, Proc. of the 29th annual meeting of the ACL, pp. 122-129, 1991.

[Siskind90] 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 the ACL, pp. 143-156, 1990.

[Smadja91] Smadja F. A., *From N-Grams to Collocations: An Evaluation of EXTRACT*, Proc. of the 29th annual meeting of the ACL, pp. 279-284, 1991.

[Tanenhaus89] Tanenhaus M. K. and Carlson G. N., *Lexical Structure and Language Comprehension*, in Lexical Representation and Process, William Marson-Wilson (ed.), The MIT Press, 1989.

[Taraban88] Taraban R. and McClelland J. L., *Constituent Attachment and Thematic Role Assignment in Sentence Processing: Influences of Content-Based Expectations*, Journal of memory and language, 27, pp. 597-632, 1988.

[Velardi91] Velardi 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.

[Webster89] Webster M. and Marcus M., *Automatic Acquisition of the Lexical Semantics of Verbs from Sentence Frames*, Proc. of the 27th annual meeting of the ACL, pp. 177-184, 1989.

[Zernik89] Zernik U., *Lexicon Acquisition: Learning from Corpus by Capitalizing on Lexical Categories*, Proc. of IJCAI, pp. 1556-1562, 1989.

[Zernik90] Zernik U. and Jacobs P., *Tagging for Learning: Collecting Thematic Relation from Corpus*, Proc. of COLING, pp. 34-39, 1990.