Papers from the AAAI 2004 Spring Symposium on Language Learning: An Interdisciplinary Perspective, pp. 39-44, Stanford, CA, March 2004.

Learning Semantic Parsers: An Important but Under-Studied Problem

Raymond J. Mooney
Department of Computer Sciences
University of Texas at Austin
1 University Station C0500
Austin, TX 78712-0233
mooney@cs.utexas.edu

Abstract

Computational systems that learn to transform natural-language sentences into semantic representations have important practical applications in building natural-language interfaces. They can also provide insight into important issues in human language acquisition. However, within AI, computational linguistics, and machine learning, there has been relatively little research on developing systems that learn such semantic parsers. This paper briefly reviews our own work in this area and presents semantic-parser acquistion as an important challenge problem for AI.

Introduction

Over the past ten to fifteen years, research in computational linguistics has undergone a dramatic "paradigm shift." Statistical learning methods that automatically acquire knowledge for language processing from empirical data have largely supplanted systems based on human knowledge engineering. The original success of statistical methods in speech recognition (Jelinek 1998) has been particularly influential in motivating the application of similar methods to almost all areas of natural language processing. Statistical methods are now the dominant approaches in syntactic analysis, word sense disambiguation, information extraction, and machine translation (Manning & Schütze 1999).

Nevertheless, there is precious little research in computational linguistics on learning for "deeper" semantic analysis. We use the term *semantic parsing* to refer to the task of mapping a natural-language sentence into a detailed semantic representation or *logical form*. Uncharacteristic of current research, most of our own recent investigations in language learning has focused on learning to parse natural-language into semantic logical form, specifically mapping natural-language database queries into executable Prolog queries (Zelle & Mooney 1993; 1996; Zelle 1995; Tang & Mooney 2000; Tang 2003; Thompson & Mooney 1999; 2003; Thompson 1998). There is a long tradition of representing the meaning of natural language statements and queries in first-order logic (Allen 1995; Dowty, Wall, & Peters 1981; Woods 1978). However, we know of no other recent research specifically on learning to map language into logical form. Nevertheless, we believe semantic parsing is an extremely important, challenging problem, and that broader investigation of this problem would lead to significant advances in computational linguistics as well as potentially deepen our understanding of issues in human language acquisition.

This paper first presents a brief historical view of the shifting emphasis of research on various tasks in natural language processing. Next, it briefly reviews our own work on learning for semantic parsing. Finally, it argues for a broader focus on semantic parsing as a key problem in language learning and for the development of additional, larger corpora of sentences annotated with detailed semantic logical form.

A Brief Historical Review of NLP Research

From the very early days of natural-language processing (NLP) research, answering natural-language questions in a particular domain was a key task (Green *et al.* 1963; Simmons 1965; 1970). Although syntactic analysis was a major component of this task, the production of a semantic interpretation that could be used to retrieve answers was also very important. The semantic analysis of language was a particular focus of NLP research in the 1970's, with researchers exploring tasks ranging from responding to commands and answering questions in a micro-world (Winograd 1972) to answering database queries (Woods 1977; Waltz 1978; Hendrix *et al.* 1978) and understanding short stories (Charniak 1972; Schank 1975; Charniak & Wilks 1976; Schank & Abelson 1977; Schank & Riesbeck 1981).

Research in this era attempted to address complex issues in semantic interpretation, knowledge representation, and inference. The systems that were developed could perform interesting semantic interpretation and inference when understanding particular sentences or stories; however, they tended to require tedious amounts of application-specific knowledge-engineering and were therefore quite brittle and not easily extended to new texts or new applications and domains. The result was systems that could perform fairly in-depth understanding of narrative text; but, were restricted to comprehending three or four specific stories (Dyer 1983).

Disenchantment with the knowledge-engineering requirements and brittleness of such systems grew, and research on in-depth semantic interpretation began to wane in the early to mid 1980's. The author's own thesis research in the mid 1980's focused on attempting to relieve the knowledge-engineering bottleneck by using *explanation-based learning* (EBL) to automatically acquire the larger knowledge structures (*scripts* or *schemas*) needed for narrative understanding (DeJong 1981; Mooney & DeJong 1985; DeJong & Mooney 1986). However, this approach still required a large amount of existing knowledge that could be used to construct detailed explanations for simpler stories.

In order to avoid the difficult problems of detailed semantic analysis, NLP research began to focus on building robust systems for simpler tasks. With the advent of statistical learning methods that could successfully acquire knowledge from large corpora for more tractable problems such as speech recognition, part-of-speech tagging, and syntactic parsing, significant progress has been made on these tasks over the past decade (Jelinek 1998; Manning & Schütze 1999). Also, much current NLP research is driven by applications to arbitrary documents on the Internet and World Wide Web (Mahesh 1997), and therefore cannot exploit domain-specific knowledge. Consequently, much current NLP research has more the flavor of traditional *information retrieval* (Sparck Jones & Willett 1997; Baeza-Yates & Ribeiro-Neto 1999), rather than AI research on language *understanding*. This overall trend is succinctly captured by the clever phrase "scaling up by dumbing down."

Unfortunately, there is relatively little research on using learning methods to acquire knowledge for detailed semantic interpretation. Research on corpus-based word-sense disambiguation addresses semantic issues (Ng & Zelle 1997; Ide & Jeanéronis 1998); however, only at the level of interpreting individual words rather than constructing representations for complete sentences. Research on learning for information extraction also touches on semantic interpretation; however, existing methods learn fairly low-level syntactic patterns for extracting specific target phrases (Cardie 1997; Freitag 1998; Bikel, Schwartz, & Weischedel 1999; Califf & Mooney 1999). Research on empirical semantic-role analysis (Fillmore *et al.* 2000; Gildea & Jurafsky 2002) comes close to studying learning for semantic parsing; however, this task only involves marking natural-language phrases as filling specific semantic roles in a given verb frame and does not provide detailed semantic analyses of complete sentences. Unfortunately, there has been only very limited research on learning to map complete sentences into detailed semantic representations, such as parsing database queries into a formal query language (Zelle & Mooney 1996; Miller *et al.* 1996; Kuhn & De Mori 1995).

CHILL: ILP for Semantic Parsing

Our own research on learning semantic parsers has involved the development of a system called CHILL (Zelle 1995) which uses Inductive Logic Programming (ILP) (Muggleton 1992; Lavrac & Dzeroski 1994) to learn a deterministic shift-reduce parser written in Prolog. The input to CHILL is a corpus of sentences paired with their logical forms. The parser learned from this data is able to transform these training sentences into their correct representations, as well as generalizing to correctly interpret many novel sentences.

CHILL is currently able to handle two kinds of semantic representations: a case-role form based on *conceptual dependency* (Schank 1975) and a Prolog-based logical query language. As examples of the latter, consider two sample queries for a database on U.S. geography, paired with their corresponding logical form:

What is the capital of the state with the highest population?

```
answer(C, (capital(S,C), largest(P,
   (state(S), population(S,P))))).
```

What state is Texarkana located in?

```
answer(S, (state(S),
   eq(C,cityid(texarkana,_)),
   loc(C,S))).
```

CHILL treats parser induction as a problem of learning rules to control the actions of a shift-reduce parser. During parsing, the current context is maintained in a stack of previously interpreted constituents and a buffer containing the remaining input. When parsing is complete, the buffer is empty and the stack contains the final representation of the input. There are three types of operators used to construct logical queries. First is the introduction onto the stack of a predicate needed in the sentence representation due to the appearance of a word or phrase at the front of the input buffer. A second type of operator unifies two variables appearing in the current items in the stack. Finally, a stack item may be embedded as an argument of another stack item.

A generic parsing shell is provided to the system, and the initial parsing operators are produced through an automated analysis of the training data using general templates for each of the operator types described above. During learning, these initial overly-general operators are specialized so that the resulting parser deterministically produces only the correct semantic interpretation of each of the training examples. The introduction operators require a semantic lexicon as background knowledge that provides the possible logical representations of specific words and phrases. CHILL initially required the user to provide this lexicon; however, we have also developed a system called WOLFIE that learns this lexicon automatically from the same training corpus (Thompson & Mooney 1999; 2003).

CHILL has been used successfully to learn natural-language interfaces for three separate databases: 1) a small database on U.S. geography, 2) a database of thousands of restaurants in northern California, and 3) a database of computer jobs automatically extracted from the Usenet newsgroup `austin.jobs` (Califf & Mooney 1999). After training on corpora of a few hundred queries, the system learns parsers that are reasonably accurate at interpreting novel queries for each of these applications. For the geography domain, the system has learned semantic parsers for Spanish, Japanese, and Turkish, as well as English. Below are some of the interesting novel English queries that the geography system can answer although it was never explicitly trained on queries of this complexity:

- What states border states through which the Mississippi runs?

- What states border states that border states that border states that border Texas?
- What states border the state that borders the most states?
- What rivers flow through states that border the state with the largest population?
- What is the longest river that flows through a state that borders Indiana?

We are currently using CHILL and WOLFIE to learn a semantic parser for English instructions for coaching a Robocup soccer team. The Robocup robotic soccer competition recently introduced a coaching competition in which a formal language called CLANG (Coach Language) is used to provide advice to a team of agents in the Robocup simulator league.[1] By automatically learning a parser that maps English statements into CLANG, coaches could more easily provide advice in natural language. This work is part of a new project on advisable learning agents that exploit both natural-language instruction and reinforcement learning. We are also developing new approaches to learning semantic parsers for this task that exploit existing syntactic grammars and parsers for both the desired natural language and the target formal semantic-representation language.

Learning Semantic Parsers as a Challenge Problem

Our research on CHILL demonstrates that machine learning methods can be used to automatically induce semantic parsers given only a corpus of sentences annotated with logical form. It also demonstrates that a lexicon of word meanings can also be learned from such corpora and then used as an important component of semantic-parser acquisition. However, our approach of using ILP to learn a deterministic symbolic parser is only one approach to this important problem. Recently, we have modified our approach to learn non-deterministic probabilistic parsers (Tang & Mooney 2000). Nevertheless, additional research is needed exploring alternative approaches to learning semantic parsers that are potentially more effective from an engineering perspective or more faithful models of human language acquisition.

One of the primary impediments to research in learning for semantic parsing is the lack of large, diverse corpora. Although corpora consisting of tens of thousands of annotated sentences exist for tasks such as part-of-speech tagging and syntactic parsing (Marcus, Santorini, & Marcinkiewicz 1993), very little data exists for semantic analysis. The recent development of FrameNet (Fillmore *et al.* 2000) and PropBank (Kingsbury, Palmer, & Marcus 2002) are interesting and important steps; however, the analyses in these corpora provide only shallow semantic annotation of semantic roles or predicate-argument structure and do not include detailed, semantic, logical form that is language independent and capable of directly supporting inference and question answering.[2] Consequently, the identification of important and representative tasks and the construction of significant corpora of semantically interpreted sentences, is a leading requirement for furthering research in this area. Although developing a good semantic representation and annotating sentences with logical form can be a fairly time-consuming and difficult job, a dedicated effort similar to that already undertaken to produce large treebanks could produce very sizable and useful semantic corpora.

Part of the resistance to exploring semantic analysis is that, given the current state of the art, it almost inevitably leads to domain dependence. However, many useful and important applications require NLP systems that can exploit specific knowledge of the domain to interpret and disambiguate queries, commands, or statements. The goal of developing general learning methods for this task is exactly to reduce the burden of developing such systems, in the same way that machine learning is used to overcome the knowledge-acquisition bottleneck in developing expert systems. It is largely the difficulty of engineering specific applications that has prevented natural-language interface technology from becoming a wide-spread method for improving the user-friendliness of computing systems. Learning technology for automating the development of such domain-specific systems could help overcome this barrier, eventually resulting in the wide-spread use of natural-language interfaces.

Studying semantic parser acquisition as a computational problem may also lead to insights into human language acquisition. Sentences paired with semantic analyses are a more cognitively plausible input to human language learning than treebanks or other supervised corpora. By inferring the meaning of sentences uttered in context, children may be able to construct examples of semantically annotated sentences. Of course, unlike the corpora used to train CHILL, in which each sentence is labeled with a unique semantic form, corpora in which each sentence is ambiguous and annotated with a small number of potential alternative semantic forms may be a more realistic model of human learning (Siskind 1996). In this case, the learned parser should be able to construct *one* of the specified alternative semantic representations for each of the sentences in the corpus. Such an induction problem is similar to the *multiple instance problem* in machine learning (Dietterich, Lathrop, & Lozano-Perez 1997) and could potentially be solved using analogous methods.

Using such ambiguously labeled corpora, one could develop and evaluate computational models of semantic lexicon learning that utilize various constraints proposed in the psycholinguistic literature (Gleitman & Landau 1994; Bloom 2000). In particular, one could explore methods that exploit syntactic and morphological cues when acquiring word meanings. For example, an enhancement of the WOLFIE lexicon-acquisition component of our CHILL system should be able to model a form of *syntactic bootstrapping* that seems to aid children's acquisition of verb meanings (Gleitman 1990; Pinker 1994). In one illustrative experiment, young children are shown a movie in which a duck is pushing a rabbit into a squatting position, but at the same time both animals are wheeling their free arms in a

[1] See information on the Robocup coach competition at `http://www.uni-koblenz.de/ fruit/orga/rc03/`

[2] Some of the relatively small corpora we have developed are available from `http://www.cs.utexas.edu/users/ml`.

circle. Half the children are told: "Oh look, the duck is gorping the rabbit," and the other half are told: "The duck and the rabbit are gorping." Given the utterance "The duck is gorping the rabbit," children prefer to associate "gorping" with an action in which a duck is the agent and a rabbit is the patient. By using syntactic and morphological analysis and previously learned word meanings, a simple parser can easily produce the following thematic analysis of this sentence: (Action:"gorp" Agent:Duck Patient:Rabbit). If possible meanings for the sentence inferred from context include both:

(Action:Move Agent:Duck Patient:ArmOf(Duck) Manner:CIRCULAR)
(Action:Move Agent:Duck Patient:Rabbit Manner:DOWN),

a simple pattern matching of the partially analyzed sentence to its possible referents could easily be used to generate and prefer the hypothesis stating that "gorp" means: (Action: Move Agent: X Patient: Y Manner: DOWN). After first learning the meanings of a few words, it can progress to learn syntactic knowledge that allows it to map some sentences to their thematic form, and then utilize information from partially analyzed sentences to acquire additional word meanings. The results of this model can be compared to those of the human experiments referenced above and used to generate predictions on the effect of such bootstrapping methods on the rate of learning.

Conclusion

The problem of learning to map natural-language sentences to detailed, semantic, logical form is an important problem that few researchers in computational linguistics have addressed. Our own research has shown that semantic parsers (and their requisite semantic lexicons) can be automatically induced from a corpus of sentences annotated with logical form. However, many problems remain, including (among others):

- Development of and experimentation with a more diverse set of larger corpora of sentences annotated with detailed semantic representations.
- Development of methods that can learn from corpora in which sentences are annotated with multiple potential alternative meanings inferred from context.
- Development of methods that model known aspects of human language acquisition such as syntactic bootstrapping.

Consequently, I strongly encourage others to consider investigating learning for semantic parsing. A larger community of researchers investigating this problem is critical to making important and significant progress.

Acknowledgments

This research was supported by the National Science Foundation under grants IRI-9704943 and IIS-0117308 and by the Defense Advanced Research Projects Agency under grant HR0011-04-1-007.

References

Allen, J. F. 1995. *Natural Language Understanding (2nd Ed.)*. Menlo Park, CA: Benjamin/Cummings.

Baeza-Yates, R., and Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. New York: ACM Press.

Bikel, D. M.; Schwartz, R.; and Weischedel, R. M. 1999. An algorithm that learns what's in a name. *Machine Learning* 34:211–232.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Califf, M. E., and Mooney, R. J. 1999. Relational learning of pattern-match rules for information extraction. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, 328–334.

Cardie, C. 1997. Empirical methods in information extraction. *AI Magazine* 18(4):65–79.

Charniak, E., and Wilks, Y., eds. 1976. *Computational Semantics*. Amsterdam: North-Holland.

Charniak, E. 1972. Toward a model of children's story comprehension. Technical Report TR266, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

DeJong, G. F., and Mooney, R. J. 1986. Explanation-based learning: An alternative view. *Machine Learning* 1(2):145–176. Reprinted in *Readings in Machine Learning*, Jude W. Shavlik and Thomas G. Dietterich (eds.), Morgan Kaufman, San Mateo, CA, 1990.

DeJong, G. F. 1981. Generalizations based on explanations. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, 67–70.

Dietterich, T. G.; Lathrop, R. H.; and Lozano-Perez, T. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* 89(1-2):31–71.

Dowty, D. R.; Wall, R. E.; and Peters, S. 1981. *Introduction to Montague Semantics*. Dordrecht, Holland: D. Reidel.

Dyer, M. 1983. *In Depth Understanding*. Cambridge, MA: MIT Press.

Fillmore, C. J.; Baker; F., C.; and Sato, H. 2000. The FrameNet database and software tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, 1157–1160.

Freitag, D. 1998. Toward general-purpose learning for information extraction. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98 (ACL/COLING-98)*, 404–408.

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

Gleitman, L., and Landau, B., eds. 1994. *The Acquisition of the Lexicon*. Cambridge, MA: MIT Press.

Gleitman, L. R. 1990. Structural sources of verb meaning. *Language Acquisition* 1:3–55.

Green, B. F.; Wolf, A. K.; Chomsky, C.; and Laughery, K. 1963. Baseball: An automatic question answerer. In Feigenbaum, E. A., and Feldman, J., eds., *Computers and Thought*. New York: McGraw Hill. 207–216. Reprinted in *Readings in Natural Language Processing*, Barbara J. Grosz, K. Spark Jones, and B. Lynn Webber (eds.), Morgan Kaufman, Los Altos, CA, 1986.

Hendrix, G. G.; Sacerdoti, E.; Sagalowicz, D.; and Slocum, J. 1978. Developing a natural language interface to complex data. *ACM Transactions on Database Systems* 3(2):105–147.

Ide, N. A., and Jeanéronis. 1998. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics* 24(1):1–40.

Jelinek, F. 1998. *Statistical Methods for Speech Recognition*. Cambridge, MA: MIT Press.

Kingsbury, P.; Palmer, M.; and Marcus, M. 2002. Adding semantic annotation to the Penn treebank. In *Proceedings of the Human Language Technology Conference*.

Kuhn, R., and De Mori, R. 1995. The application of semantic classification trees to natural language understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(5):449–460.

Lavrac, N., and Dzeroski, S. 1994. *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood.

Mahesh, K., ed. 1997. *Papers from the AAAI Spring Symposium on Natural Language Processing for the World Wide Web*. Stanford, CA: AAAI Press.

Manning, C. D., and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marcus, M.; Santorini, B.; and Marcinkiewicz, M. A. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics* 19(2):313–330.

Miller, S.; Stallard, D.; Bobrow, R.; and Schwartz, R. 1996. A fully statistical approach to natural language interfaces. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96)*, 55–61.

Mooney, R. J., and DeJong, G. F. 1985. Learning schemata for natural language processing. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, 681–687.

Muggleton, S. H., ed. 1992. *Inductive Logic Programming*. New York, NY: Academic Press.

Ng, H. T., and Zelle, J. 1997. Corpus-based approaches to semantic interpretation in natural language processing. *AI Magazine* 18(4):45–64.

Pinker, S. 1994. How could a child use verb syntax to learn verb semantics? *Lingua* 92:377–410.

Schank, R. C., and Abelson, R. P. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Schank, R. C., and Riesbeck, C. K. 1981. *Inside Computer Understanding: Five Programs plus Miniatures*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Schank, R. C. 1975. *Conceptual Information Processing*. Oxford: North-Holland.

Simmons, R. F. 1965. Answering English questions by computer: A survey. *Communications of the Association for Computing Machinery* 8:53–70.

Simmons, R. F. 1970. Natural language question-answering systems: 1969. *Communications of the Association for Computing Machinery* 13:15–30.

Siskind, J. M. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61(1):39–91.

Sparck Jones, K., and Willett, P., eds. 1997. *Readings in Information Retrieval*. San Francisco, CA: Morgan Kaufmann.

Tang, L. R., and Mooney, R. J. 2000. Automated construction of database interfaces: Integrating statistical and relational learning for semantic parsing. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora(EMNLP/VLC-2000)*, 133–141.

Tang, L. R. 2003. *Integrating Top-down and Bottom-up Approaches in Inductive Logic Programming: Applications in Natural Language Processing and Relational Data Mining*. Ph.D. Dissertation, Department of Computer Sciences, University of Texas, Austin, TX.

Thompson, C. A., and Mooney, R. J. 1999. Automatic construction of semantic lexicons for learning natural language interfaces. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, 487–493.

Thompson, C. A., and Mooney, R. J. 2003. Acquiring word-meaning mappings for natural language interfaces. *Journal of Artificial Intelligence Research* 18:1–44.

Thompson, C. A. 1998. *Semantic Lexicon Acquisition for Learning Natural Language Interfaces*. Ph.D. Dissertation, Department of Computer Sciences, University of Texas, Austin, TX. Also appears as Artificial Intelligence Laboratory Technical Report AI 99-278 (see http://www.cs.utexas.edu/users/ai-lab).

Waltz, D. L. 1978. An English language question answering system for a large relational database. *Communications of the Association for Computing Machinery* 21(7):526–539.

Winograd, T. 1972. *Understanding Natural Language*. Orlando, FL: Academic Press.

Woods, W. A. 1977. Lunar rocks in natural English: Explorations in natural language question answering. In Zampoli, A., ed., *Linguistic Structures Processing*. New York: Elsevier North-Holland.

Woods, W. A. 1978. Semantics and quantification in natural language question answering. In Yovits, M., ed., *Advances in Computers, vol. 17*. New York: Academic Press. 2–64. Reprinted in *Readings in Natural Language Processing*, Barbar J. Grosz, Karen Spark Jones, and Bonnie Lynn Webber (eds.), Morgan Kaufman, Los Altos, CA, 1986.

Zelle, J. M., and Mooney, R. J. 1993. Learning semantic grammars with constructive inductive logic programming. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, 817–822.

Zelle, J. M., and Mooney, R. J. 1996. Learning to parse database queries using inductive logic programming. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, 1050–1055.

Zelle, J. M. 1995. *Using Inductive Logic Programming to Automate the Construction of Natural Language Parsers*. Ph.D. Dissertation, Department of Computer Sciences, University of Texas, Austin, TX. Also appears as Artificial Intelligence Laboratory Technical Report AI 96-249.