

# Answering Questions Using Advanced Semantics and Probabilistic Inference

**Srini Narayanan**

International Computer Science Institute  
1947 Center St., Suite 600,  
Berkeley, CA 94704  
snarayan@icsi.berkeley.edu

**Sanda Harabagiu**

Human Language Technology Research  
Institute - University of Texas at Dallas  
2601 N. Floyd Rd. EC 31, ECSS 3.908A  
Richardson, TX 75083  
sanda@hlt.utdallas.edu

## Abstract

In this paper we argue that access to rich semantic structures derived from questions and answers will enable both the retrieval of more accurate answers to simple questions and enable inference processes that explain the validity and contextual coverage of answers to complex questions. Processing complex questions involves the identifications of several forms of complex semantic structures. Answer Extraction is performed by recognizing event inter-relationships and by inferring over multiple sentences and texts, using background knowledge.

## 1 Introduction

Current Question Answering systems extract answers from large text collections by (1) classifying questions by the answer type they expect; (2) using question keywords or patterns associated with questions to identify candidate answer passages and (3) ranking the candidate answers to decide which passage contains the exact answer. A few systems also justify the answer, by performing abduction in first-order predicate logic [Moldovan et al., 2003]. This paradigm is limited by the assumption that the answer can be found because it uses the question words. Although this may happen sometimes, this assumption does not cover the many cases where an informative answer is missed because its identification requires more sophisticated semantic processing than named entity recognition and the identification

of an answer type. Therefore access to rich semantic structures derived from questions and answers will enable the retrieval of more accurate answers as well as inference processes that explain the validity and contextual coverage of answers.

Several stages of deeper semantic processing may be considered for processing complex questions. A first step in this direction is the incorporation of “semantic parsers” or identifiers of predicate argument structures in the processing of both questions and documents. Processing complex questions consists of: (1) a syntactic parse of the question and of the document collection; (2) Named Entity recognition that along with the syntactic parse enable (3) the identification of predicate-argument structures; and (4) identification of the answer types, which no longer consist of simple concepts, but rather complex conceptual structures, and (5) keywords extraction that allows candidate answers to be identified. Document processing is performed by indexing and retrieval that uses three forms of semantic information: (1) Classes of named entities; (2) Predicate-argument structures and (3) Ontologies of possible answer types. Additionally, as more complex semantic structures are evoked by the question and recognized in documents, indexing and retrieval models are enhanced by taking into account conceptual schemas and topic models. Answer processing is concerned with the recognition of the answer structure, which is a natural extension of recognizing exact answers when they are represented as single concepts. Since many times the answer is merged from several sources, enhanced answer processing also requires a set of special operators for answer fusion.

The rest of the paper is organized as follows. Section 2 presents question processing that uses deeper semantic resources. Section 3 details the methods for answer ex-

traction whereas Section 4 describes methods for representing and reasoning with rich semantic structures. Section 5 summarizes the conclusions.

## 2 Question Processing that uses a variety of semantic resources

Given the size of today’s very large document repositories, one can expect that any complex topic will be covered from multiple points of view. This feature is exploited by the question decomposition techniques, which generate a set of multiple questions in order to cover all of the possible interpretations of a complex topic. However, a set of decomposed questions may end up producing a disparate (and potentially contradictory) set of answers. In order for Q/A systems to use these collections of answers to their advantage, answer fusion must be performed in order to identify a single, unique, and coherent answer.

We view answer fusion as a three-step process. First, an open-domain, template-based answer formalization is constructed based on predicate-argument frames. Second, a probabilistic model is trained to detect relations between the extracted templates. Finally, a set of template merging operators are introduced to construct the merged answer. The block architecture for answer fusion is illustrated in Figure 2. The system functionality is demonstrated with the example illustrated in Figure 3.

Our method first converts the extracted answers into a series of open-domain templates, which are based on predicate-argument frames (Surdeanu et al, 2003). The next component detects generic inter-template relations. Typical “greedy” approaches in Information Extraction (Hobbs et al, 1997; Surdeanu and Harabagiu, 2002) use heuristics that favor proximity for template merging. The example in Figure 3 proves that this is not always the best decision, even for templates that share the same predicate and have compatible slots.

Processing complex questions involves the identification of several forms of complex semantic structures. Namely we need to first recognize the answer type that is expected, which is a rich semantic structure, in the case of complex question or a mere concept in the case of a factual question. At least three forms of information are needed for detecting the answer type: (1) question classes and named entity classes; (2) syntactic dependency information, enabling the recognition of (3) predicate-argument structures. Each of the following three

Q-Ex1: When was India’s Prithvi missile launched for the first time?  
 Q-Ex2: From which country did North Korea import its missile launch pad metals?  
 Q-Ex3: What stimulated India’s missile programs?

questions illustrates the significance of the three forms of semantic information in question processing:

For question *Q-Ex1*, the question stem “when” indicates that the answer type is a temporal unit, eventually expressed as a date. To find candidate answers, the recognition of *India* and other related named entities, e.g. *Indian*, as well as the name of the *Prithvi missile* or of its related program are important. Named entity recognition is also important for processing question *Q-Ex2*, because not only “North Korea” needs to be recognized as a country, but names of other countries need to be identified in the candidate answer paragraph. Processing question *Q-Ex2* involves syntactic information as well, e.g. the identification of the complex nominal “missile launch pad metals”. To better process question *Q-Ex2*, additional semantic information in the form of predicate-arguments structures enables the recognition of the answer type more precisely. Instead of looking only for country names when processing the documents, a search for countries that export missile launch pad metals or of counties from which North Korea imports such commodities refines the search space. This is made possible by the transformation of question *Q-Ex2* in the structure illustrated in Figure 2.

Predicate : import  
 Arg 0 : (role = importer) : North Korea  
 Arg 1 : (role = commodity) : missile launch pad metals  
 Arg 2 : (role = exporter) : ANSWER

Figure 2: Predicate-argument structure of question Q-Ex2

The role set for the arguments of predicate “import” was used as it is currently defined in the PropBank project. Predicate-argument structures are also essential to the processing of question *Q-Ex3*, because the question is too ambiguous. The stem “what” and the named entity “India” may relate to a large range of events and entities.

Predicate: Stimulate  
 Arg0 (role = agent): ANSWER (part 1)  
 Arg1 (role = thing increasing): India’s missile program  
 Arg2 (role = instrument): ANSWER (part 2)

Figure 3: Predicate-argument structure for question Q-Ex3

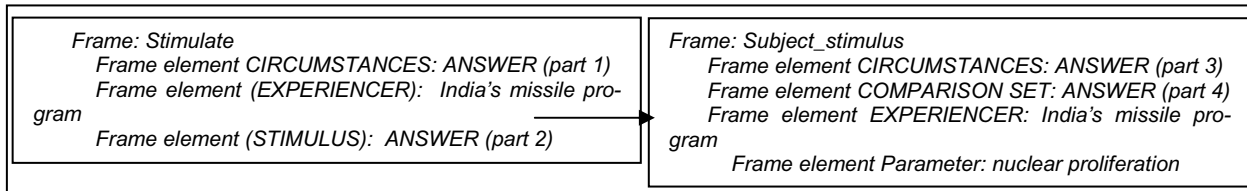


Figure 1: FrameNet structuring of question Q-Ex3

**Q-Sem:** How can a biological weapons program be detected?

**Answer (Q-Sem)**  
*In recent months, Milton Leitenberg, an expert on biological weapons, has been looking at this murkiest and most dangerous corner of Saddam Hussein's armory. He says a series of reports add up to indications that Iraq may be trying to develop a new viral agent, possibly in underground laboratories at a military complex near Baghdad where Iraqis first chased away inspectors six years ago. A new assessment by the United Nations suggests Iraq still has chemical and biological weapons - as well as the rockets to deliver them to targets in other countries. The UN document says Iraq may have hidden a number of Scud missiles, as well as launchers and stocks of fuel. US intelligence believes Iraq still has stockpiles of chemical and biological weapons and guided missiles, which it hid from the UN inspectors.*

Figure 5: Complex question and its answer derived from the CNS collection

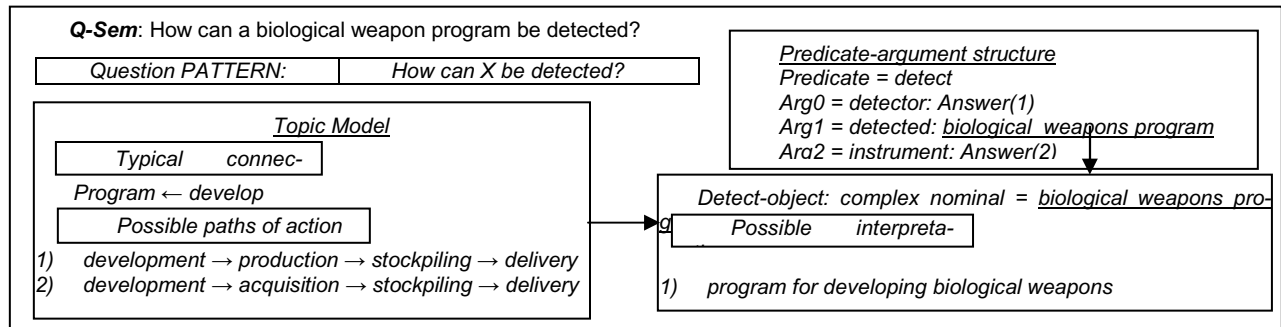


Figure 4: Question processing based on topic semantics

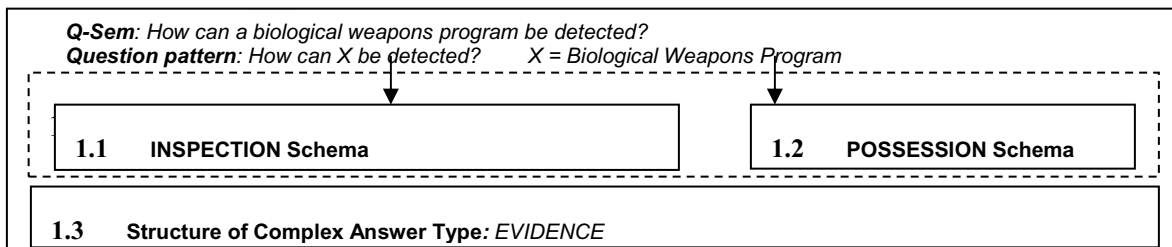
The predicate-argument structure illustrated in Figure 3 indicates that the answer may have the role of the *agent* or even the role of the *instrument*. When semantic information from FrameNet is also used, Figure 4 shows that the answer may have in fact four other semantic roles.

To illustrate the semantic knowledge that needs to be recognized and the inference process that they enable, we shall use one of the questions employed in the AQUAINT Pilot 2 for Dialog Processing of CNS Scenarios, illustrated in Figure 5.

Processing *Q-Sem* cannot be done by simply using the question stem “*how*” to identify manners of detection or even by employing the predicate-argument structure illustrated in Figure 6. The answer contains a single troponym of the verb “*detect*”, namely “*look at*”, and the agent is “*Milton Leitenberg, an expert on biological weapons*”. However returning the name of Milton Leitenberg as the answer is not informative enough.

Instead of relying only on the question stem and the predicate-argument structure, question processing takes advantage of a more complex semantic structure made available by the enhanced architecture: **the topic model**. The topic model contributes to the interpretation of the only argument fully specified in the predicate-argument structure illustrated in Figure 6, namely Arg1 representing the “detected” role, expressed as “the biological weapons program”. The interpretation of this complex nominal is made possible by two semantic representations: (1) the typical connections in the topic model; and (2) the possible paths of action characterizing the topical model as represented in Figure 6.

Figure 6 lists only two of the semantic representation typical of the scenario defined in Figure 8, namely *typical connections* between events and entities or between events. A special kind of relations between events is represented by the possible paths of action. The two paths of actions that are listed in Figure 6 enable the two interpretations of the detected object. It is to be noted that such semantic knowledge as the one represented in the topic model is not available from WordNet or FrameNet at present, and thus need to be encoded and made accessible to the Q/A system. For structuring the complex answer type expected by question *Q-Sem*, a set



of conceptual schemas need also to be recognized. Figure 7 shows some of the schemas instantiated by the question processing. The inspection schema is evoked by the question verb “*detect*”; the possession schema is evoked by the complex nominal “*biological weapons program*”.

Along with the answer structure, the enhanced question processing module generates the structure of the intentions uncovered by question *Q-Sem*. The general intention of finding evidence that there is a biological weapons program in Iraq is structured in four different representations illustrated in Figure 8. Intention structures are also dependent on the topic model.

### 3 Answer Extraction based on semantic processing

In the baseline architecture, the answer type as well as the predicate-argument structure determine the answer extraction. Complex questions like *Q-Sem* are provided with an answer by filling the semantic information of their complex answer types.

Figure 9 illustrates the answer extracted for question *Q-Sem* in the form of: (1) the text where the answer is found (2) the semantic structure of the answer with information derived from the text; and (3) pointers linking the fillers of the semantic structure of the answer with the text source. Such pointers may be supplemented with traces of the inferential processes. The answer type, labeled “evidence-combined” has several semantic classes that are in turn filled with semantic representations for (1) the content of evidence; (2) the source of the evidence; (3) the quality of evidence and (4) the judgment of evidence. The content structure lists both predicate-argument-like structures as well as such attributes as: (a) the justification, accounting for the conceptual schema that identified the content; (b) the status of the event/state recognized by the schema; (c) the likelihood of the eventuality of the event/state and (d) intentions and abilities from the past, present or future. The source representation is also structured as (a) author, (b) type and (c) reliability. The quality of the inferred answer is measured by (a) the judges; (b) the judge types; (c) judgment manner and (d) judgment stage. Finally, a

qualitative assessment of the reliability of the answer is given, to complement the reliability score computed through probabilistic inference.

## 4 Representing and Reasoning with Rich Semantic Structures for Advanced QA

The ICSI work on embodied cognition is based on cognitive linguistics. Research in this field has shown that many concepts and inferences are based on a relatively small number of /image schemas/, which are deeply embodied and are apparently universal across languages. For example, the ideas of container, goal and oppositional force occur prominently in language and thought.

### 4.1 Cross-Linguistic Conceptual Schemas and Inference

Much of narrative text relies on a relatively constrained set of conceptual schemas. For instance, the example above uses some of the most potent general schemas: POSSESSION, EVASION, SPATIAL RELATION, EVENT STRUCTURE, and SOURCE-PATH-GOAL which involves an agent trying to obtain a particular *goal* (finding WMD) by moving along a *path* of actions. These are all basic embodied schemas whose inferential structure is common cross-linguistically. Furthermore, these schemas are often sources of metaphor (PHYSICAL POSSESSION maps to INFORMATION POSSESSION, SPATIAL LOCATIONS MAP TO INFORMATION STATES (*murky and dangerous corner*), PHYSICAL ACTIONS (*look*) MAP to ABSTRACT ACTIONS (*scrutinize information*) [35]). It appears that only a few dozen such general schemas suffice to describe a very wide range of scenarios. These have been extensively studied for many languages by linguists, but only recently formalized (as part of our AQUAINT Phase 1 effort). Now that we have the formalism in hand, we believe and hope to demonstrate in Phase II that the combination of embodied schemas with metaphorical mappings to other domains can yield a significant improvement in inferential power over traditional approaches.

### 4.2 Reasoning about Event Structure

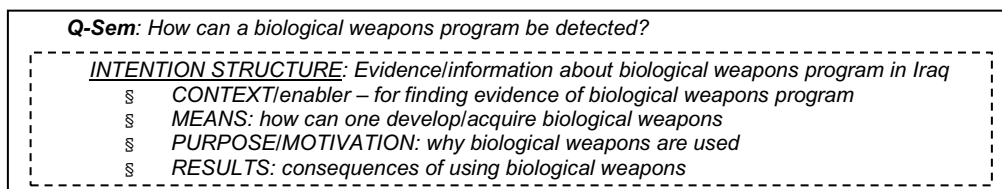


Figure 6: Intentional Structure

Performing QA with complex scenarios requires sophisticated reasoning about actions and events. For instance, in the example above knowing the *stage* of a process (interrupted inspection due to a *chase away*

markers of *evidentials*) and often 3) *abductive* (Iraq's goal of chasing away the inspectors was probably to be able to hide the WMD). In all complex scenarios event descriptions are 1) *dynamic* (*has been looking, still has,*

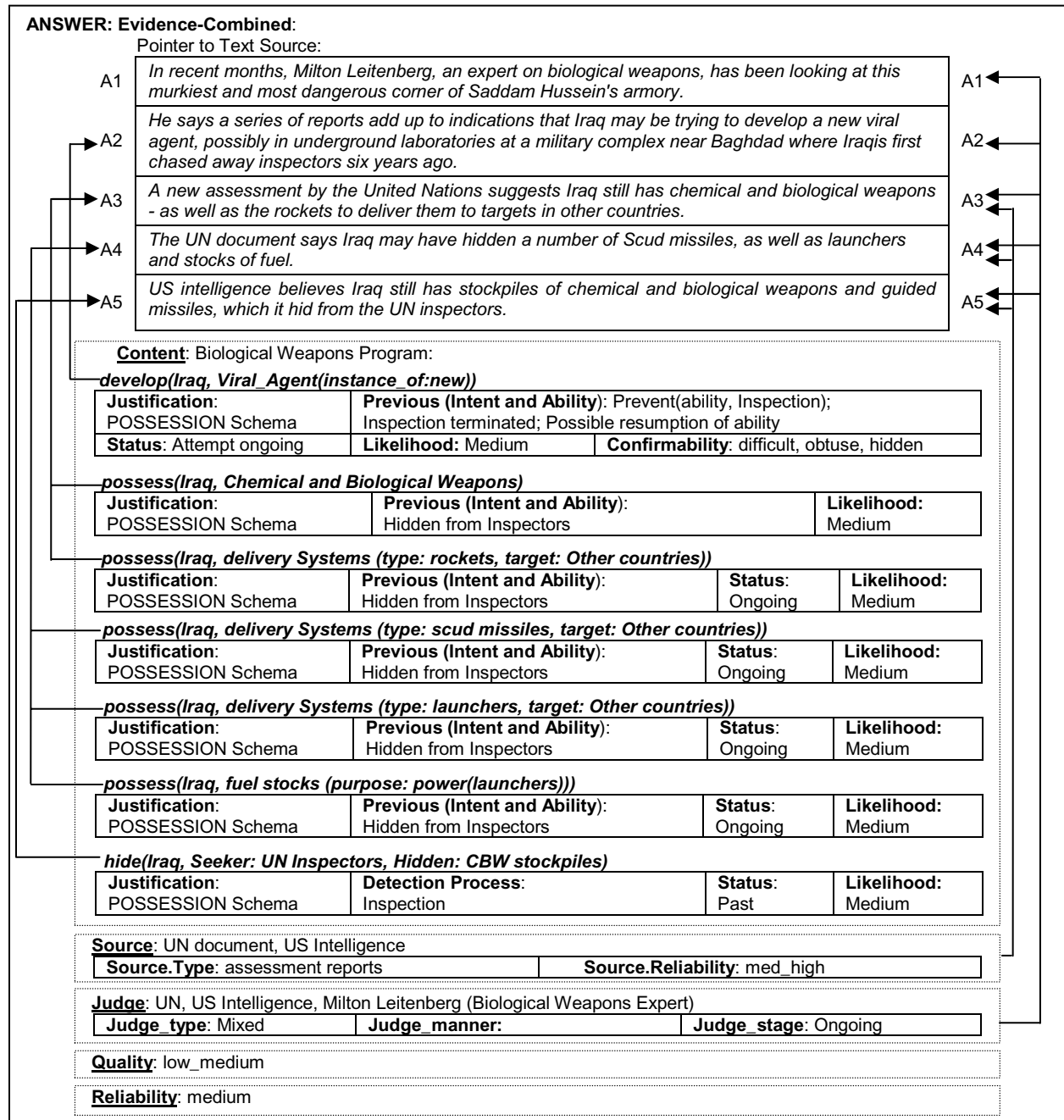


Figure 7: Answer to complex question

event), gives valuable *predictive* information (Iraq may have hidden the WMD) as well as *pre-suppositional* information (Iraq had WMD before the inspections (signaled by the use of *still has* in the example scenario)). Of course, this information is *probabilistic* (Iraq is only *likely* to have WMD (note the use of *indications, may be, suggests, believes,* and other linguistic

*trying to develop,* etc.), 2) *resource specific* (WMD deployment needs launchers, inspections need inspectors), 3) *context sensitive* (all inferences are conditioned on the scenario context which suggests among other things that a) Iraq intends and is probably trying to develop WMD and b) it is likely that Iraq will try to hide these WMD from the Inspectors), often 4) *figurative* (as in

*murky and dangerous corner, reports add up to indications, UN document says etc.*), may incorporate the specific 7) *perspectives* of the various participants (*Milton Leitenberg, US intelligence, UN inspectors etc.*)

Over the last decade and in Phase I of AQUAINT, we have developed a rich model of event structure that has been shown to be capable of capturing the event structure distinctions in complex text in a variety of languages (Narayanan99a, Narayanan99b, Chang et al, 2002, 2003). The model forms the basis and provides the underlying operational semantics for the DARPA funded DAML-S process model (NM 2003, ISWC 2001) for the semantic web. This model is also being used in the ongoing ARDA video event recognition ontology effort (Hobbs and Nevatia 2003).

#### 4.3 Building Deep Semantic Structure From Text

We now have a set of wide-coverage lexical resources such as FrameNet (FN) and WordNet (WN) that can potentially aid knowledge formation in the rapid development of scenario models for inference in QA. An explicit representation of such semantic information is needed to fully realize use in text interpretation and inference. Previously we have worked out a formalism that unpacks the shorthand of frames into structured event representations. These dynamic representations allow annotated FrameNet data to parameterize event simulations based on the PIN model (Section II.B.7) (Chang et al 2002) in a manner capable of producing the fine-grained, context-sensitive inferences required for language processing. We anticipate that wide-coverage resources will be useful for the focused data AQUAINT Phase II task and we propose to enable developers to access these resources like FrameNet through a Java API. We propose to undertake the following related tasks. 1) Build a common API to WN and FN so we can combine the resources. 2) Produce an OWL-S (<http://www.daml.org/services>) port of FrameNet so that FrameNet information can be combined with other ontologies and in particular with specialized domain ontologies of use to the DoD and intelligence communities. 3) Build PIN (Section II.B.7) models of FrameNet frames that are of use to the CNS scenarios. 4) Evaluate the ease of using FrameNet based models and the amount of human intervention required to instantiate them as Probabilistic Inference Networks. 5) Explore further automation of the mapping from FrameNet frame descriptions to Probabilistic Inference Networks.

#### 4.4 A Construction Grammar based deep semantic analyzer

The ICSI group has developed a powerful semantic grammar formalism – Embodied Construction Grammar (ECG), partially with Phase 1 AQUAINT support. It will not be possible to develop robust, full coverage ECG grammars from this base during phase 2 and efforts in this task will focus on the detailed analysis of complex questions in context. An ECG grammar exploits constructions, starting from individual words and extending through complex linguistic forms, in a similar manner to other unification grammars such as HPSG. Central novel ideas are use of conceptual links, the **evokes** mechanism for activating other concepts, use of **roles** for describing schemas, and a **meaning** section that specifies introduced semantic relations.

Given that an ECG grammar can map linguistic form to deep semantic relations, it remains to build systems that exploit this capability. John Bryant has built such an analyzer as part of the ICSI Phase 1 effort and his Master's thesis. It is basically a unification based chart parser using chunking methods for efficiency and robustness. One major innovation is the use of deep semantic unification in the basic matching step – this improves both efficiency and robustness. The ECG semantic analysis system has been coupled to the ICSI inference engine of task 7 to produce a pilot complete QA system for news stories. For Phase 2, ICSI will extend the existing system in several ways. The semantic unification methodology will be extended to handle linguistic and situational context. This is a natural extension and has the promise of providing much more robust integration over extended discourse. In addition, there will be specific effort aimed at the analysis of queries and the supporting declarations. This is intended to address the fact that analysts ask much more complex questions than Phase 1 systems can understand.

#### 4.5 Probabilistic Inference Networks for combining ontological, statistical and linguistic knowledge for advanced QA

Modern inference systems deal with the ambiguity and uncertainty inherent in any large, real-world domain using probabilistic reasoning. Such models have many advantages, including the ability to deal with missing and uncertain data. Bayesian networks have worked extremely well in moderate sized cases, but do not scale to situations of the size and complexity needed here to model QA with complex scenarios as in the CNS data. To handle such data, we need techniques that combine reasoning about uncertainty with relational knowledge bases and dynamic linguistic knowledge and context. In general, reasoning with linguistic structure, ambiguity, and dynamics requires modeling coordinated temporal processes and complex, structured states. A significant

amount of work has gone into different aspects of overall problem.

## 5 Conclusions

In this paper we show that, while there has been much progress in natural language analysis there is still a large gap in representing knowledge and reasoning with it for advanced QA. We have developed a method for processing complex questions which involves the identification of several forms of complex semantic structures. This involves the development of a powerful semantic grammar formalism - Embodied Construction Grammar (ECG) and applying it to the analysis of complex questions. Answer Extraction will be performed by recognizing event inter-relationships, recognized by novel relation extraction techniques. Question extraction at the level the AQUAINT program seeks requires a mechanism for performing inference over multiple sentences and texts, using background knowledge. We propose to build a software package which we refer to as Probabilistic Inference Networks (PIN) to provide a mechanism for performing context-sensitive inference over multiple sentences and discourse fragments, using encoded knowledge.

## References

- A. L. Berger, S. A. Della Pietra and V. J. Della Pietra. *A maximum entropy approach to natural language processing*. Computational Linguistics, 22(1):39-72, March 1996.
- Jennifer Chu-Carroll and Sandra Carberry. *Generating Information-Sharing Subdialogues in Expert-User Consultation*. In Proceedings of the 14<sup>th</sup> International Joint Conference on Artificial Intelligence (IJCAI-95), pages 1243-1250, Montreal, Canada, 1995.
- Michael Collins. *A New Statistical Parser Based on Bigram Lexical Dependencies*. In Proceedings of the 34<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, ACL-96, pages 184-191, 1996.
- Abdessamad Echihabi and Daniel Marcu. *A Noisy-Channel Approach to Question Answering*. In Proceedings of the 41<sup>st</sup> Annual Meeting of the Association for Computational Linguistics (ACL-2003), Sapporo, Japan, 2003.
- Christiane Fellbaum (Editor). *WordNet – An Electronic Lexical Database*. MIT Press, 1998.
- C.J. Fillmore and C.F. Baker and H. Sato. *The Frame-Net Database and Software Tools*. In M. González-Rodríguez and C.P. Suárez-Araujo, Eds. In Proceedings of the Third International Conference on Language Resources and Evaluation. Las Palmas, Spain, 2002.
- Michael Fleischman, Eduard Hovy and Abdessamad Echihabi. *Offline Strategies for Online Question Answering: Answering Questions Before They Are Asked*. In Proceedings of the 41<sup>st</sup> Annual Meeting of the Association for Computational Linguistics (ACL-2003), pages 1-7, Sapporo, Japan, 2003.
- Sanda Harabagiu, Marius Pa ca and Steven Maiorano. *Experiments with Open-Domain Textual Question Answering*. In Proceedings of the 18<sup>th</sup> International Conference on Computational Linguistics (COLING-2000), pages 292-298, Saarbrücken, Germany, 2000.
- Sanda Harabagiu, Dan Moldovan, Christine Clark, Mitchell Bowden, John Williams and Jeremy Bensley. *Answer Mining by Combining Extraction Techniques with Abductive Reasoning*. In Proceedings of the 12<sup>th</sup> Text Retrieval Conference (TREC 2003).
- Jerry R. Hobbs, Doug E. Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel and Mabry Tyson. *FASTUS: A Cascaded Finite-State Transducer for Extracting Information Natural-Language Text*. In Finite State Language Processing, Edited by Emmanuel Roche and Yves Schabes, MIT Press, 1997.
- E.H. Hovy, U. Hermjakob, C.-Y. Lin, “The Use of External Knowledge in Factoid QA”, TREC-10 conference, November 2001.
- Dan I. Moldovan, Sanda M. Harabagiu, Marius Pa ca, Rada Mihalcea, Roxana Gîrju, Richard Goodrum and Vasile Rus. *The Structure and Performance of an Open-Domain Question Answering System*. In Proceedings of the 38<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-2000), 2000.
- John Prager, Eric Brown, Anni Coden and Dragomir Radev. *Question-answering by predictive annotation*. In Proceedings of the 23<sup>rd</sup> annual international ACM SIGIR conference on Research and development in information retrieval, pages: 184-191, Athens, Greece, 2000.
- Dragomir R. Radev and Kathleen McKeown. *Generating natural languages summaries from multiple online sources*. Computational Linguistics, 24(3):469-500, 1998.