

Discourse Structure for Context Question Answering

Joyce Y. Chai

Rong Jin

Department of Computer Science and Engineering

Michigan State University

East Lansing, MI 48864

jchai@cse.msu.edu, rongjin@cse.msu.edu

Abstract

In a real-world setting, questions are not asked in isolation, but rather in a cohesive manner that involves a sequence of related questions to meet user's information needs. The capability to interpret and answer questions based on context is important. In this paper, we discuss the role of discourse modeling in context question answering. In particular, we motivate a semantic-rich discourse representation and discuss the impact of refined discourse structure on question answering.

1 Introduction

In a real-world setting, questions are not asked in isolation, but rather in a cohesive manner that involves a sequence of related questions to meet user's information needs. The capability to interpret and answer questions based on context is important. For example, Figure 1 shows an example of a series of context questions. In this example, the interpretation of Q2 and Q4 depends on the resolution of "it" and "this" from the context respectively. Although neither Q3 nor Q6 requires any anaphora resolution, the interpretation of Q3 depends on Q2 while the interpretation of Q6 depends solely on itself. Furthermore, in Q5, there are no explicit references. Its interpretation depends on a preceding question (e.g., Q4), however, in a different manner.

This example indicates that interpreting each of these questions and extracting answers needs to be situated in a particular context as the QA session proceeds. There are situations where a question is "complete" enough and its interpretation does not depend on the previous questions (Q6). There are also situations where the interpretation of a question depends on preceding questions no matter whether it

requires anaphora or ellipsis resolution. Based on these observations, a natural question to ask is what makes the use of discourse differently in different situations? What is the role of discourse in context question answering?

To address these questions, a key issue, in our mind, is that every question and its answer have a discourse status with respect to an entire QA session. This discourse status includes two aspects. The first aspect relates to discourse roles of entities in a question and the corresponding answer. Entities (such as noun phrase, verb phrase, preposition phrase, etc) in a question carry distinctive roles that indicate what is the topic or focus of a question in terms of the overall information seeking discourse. Topic relates to the "aboutness" of a question and focus relates to a specific perspective of the topic. The second aspect of discourse status relates to discourse transitions that indicate how discourse roles are changed from one question to another as the interaction proceeds and how such changes reflect the progress of user information needs. Both discourse roles and discourse transitions determine whether the context is useful, and if so, how to use the context to interpret a question.

This paper takes an initial attempt to investigate the discourse status for context question answering. In particular, it motivates a semantic-rich discourse representation that captures both discourse roles of a question and discourse transitions between questions. Through examples, this paper further discusses the potential impact of this refined discourse structure on context question answering.

Q1: What is the name of the volcano that destroyed the ancient city of Pompeii?

Q2: When did it happen?

Q3: how many people were killed?

Q4: how tall was this volcano?

Q5: Any pictures?

Q6: Where is Mount Rainier?

Figure 1: An example of context questions

2 Semantic-rich Discourse Modeling

For processing single questions, an earlier study shows that an impressive improvement can be achieved when more knowledge-intensive NLP techniques are applied at both question and answer processing level (Harabagiu et al., 2000). For context questions, a parallel question would be whether rich contextual knowledge will help interpret subsequent questions and extracting answers. To address this question, we propose a semantic-rich discourse modeling that captures both discourse roles of questions and discourse transitions between questions, and investigate its usefulness in context question answering.

2.1 Discourse Roles

In context question answering, each question is situated in a context. In addition to the semantic information carried by important syntactic entities (such as noun phrase, verb phrase, preposition phrase, etc), each question also carries distinctive discourse roles with respect to the whole question answering discourse. Specifically, the discourse roles can be categorized based on both the informational and intentional perspectives of discourse (Hobbs, 1996), as well as the presentation aspect of both questions and answers.

The intentional perspective relates to the purpose of a question. In a fully interactive question answering environment, instead of asking questions, a user may need to reply to a clarification question prompted by the system or may need to simply ask for a confirmation. Therefore, it is important to capture the intention from the user (Grosz and Sidner 1986). The informational perspective relates to the information content of a question, in particular, the topic and the focus based on the semantics of the content. In addition to the intentional and informational aspects, there is also a presentational aspect of discourse that relates to both the input modality (i.e., questions) and the output modality (i.e., answers). For example, a user may explicitly ask for images or pictures of a person or event. The presentation aspect is particularly important to facilitate multimodal multimedia question answering. Therefore, for a given question, three types of discourse roles: Intent, Content, and Media can be captured to reflect the intentional, informational, and presentational perspectives of discourse respectively.

These discourse roles can be further characterized by a set of features. For example, Intent can be represented by Act and Motivator, where Act indicates

whether the user is requesting information from the system or replying to a system question. Motivator corresponds to the information goal as to what type of action is expected from the system, for example, whether information retrieval or confirmation (Chai et al, 2003). We will not elaborate Intent here since it has been widely modeled for most dialog systems.

Content can be characterized as Target, Topic and Focus. Target indicates the expected answer type such as whether it should be a proposition (e.g., for why and how questions), or a specific type of entity (e.g., TIME and PLACE).

Topic indicates the “aboutness” or the scope related to a question. Focus indicates the current focus of attention given a particular topic. Focus always refers to a particular aspect of Topic. Since the informational perspective of discourse should capture the semantics of what has been conveyed, Topic and Focus are linked with the semantic information of a question, for example, semantic roles as described in (Gildea and Jurafsky 2002). Semantic roles concern with the roles of constituents in a question in terms of its predicate-argument structure. The discourse roles link the semantic roles of individual questions together with respect to the discourse progress through Topic and Focus.

For example, Topic can be of type Activity or Entity. Activity can be further categorized by ActType, Participant, and Peripheral. ActType indicates the type of the activity; Participant indicates entities that are participating in the activity with different semantic roles. Peripheral captures auxiliary information such

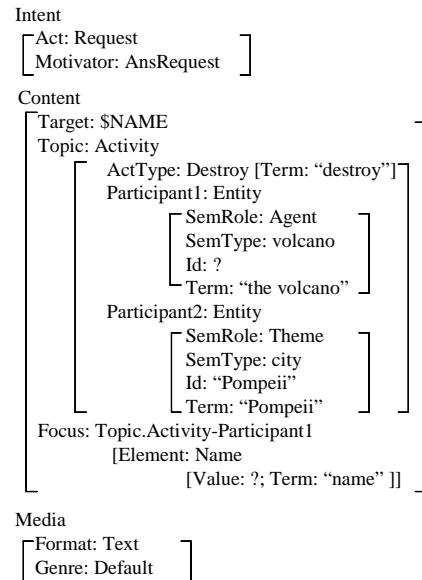


Figure 2: Discourse roles for Q1.

as the time, place, purpose, and reason for such an activity. Entity can be categorized by SemRole, SemType, Id, Element, and Constraint. SemRole indicates the semantic role of the entity in a particular activity (if any). SemType represents the semantic type of the entity. Element indicates the specific features associated with the entity. Constraint specifies the constraints need to be satisfied to identify the entity, and Id specifies the particular identifier of the entity that particularly corresponds to pronouns, demonstratives, and definite noun phrases.

Media indicates the desired information media, which can be further characterized as Format and Genre as shown in Figure 2. Format indicates whether it is an image, a table, or text, etc. Genre specifies the answer needs such as summary or list. If it is a list, how many should be in the list as in the question “number ten largest cities in the world.”

Figure 2 shows the representation of discourse roles of Q1 using typed feature structures (Carpenter 1992), where Intent indicates that the user is requesting for the system to retrieve an answer. Topic indicates the topic of Q1 is a Destroy Activity, which has two participants. The first participant is some kind of unknown volcano that takes the role of Agent in the activity (i.e., the destroyer). The second participant is the city of Pompeii that takes the role of Theme indicating the thing destroyed. The Focus of Q1 is about the name (i.e., Element) of the entity in the first participant (i.e., Participant1) in the Topic representation.

The granularity of discourse roles can be varied. The finer the granularity, the better is the use of context for inference (as discussed later). However, the finer granularity also implies deeper semantic processing. This semantic rich representation can be used to generate other representations such as queries based on weighted terms for information retrieval or logical forms for deduction and inference (Waldinger et al., 2003). Furthermore, this representation is able to keep all QA sessions in a structured way to support inference, summarization, and collaborative fusion as described later.

2.2 Discourse Transitions

Transitions from one question to another also determine how context will be used in interpreting questions and retrieving answers. In this section, we use query formation as an example to illustrate the role of different types of discourse transitions.

Discourse transitions also correspond to the intentional, informational, and presentational perspectives of discourse. Intentional transitions are

closely related to Grosz and Sidner’s “dominance” and “satisfaction precedence” relations, which are more relevant to plan-based discourse (Grosz and Sidner, 1986). Here we focus on informational transitions and presentational transitions that are more relevant to QA systems since they are targeted for information exchange.

Informational transitions are mainly centered around Topics of questions. In context question answering, how questions are related to each other depends on how “topics” of those questions evolve. Currently, we categorize information transitions into three types: *Topic Extension*, *Topic Exploration*, and *Topic Shift*.

Topic Extension

A question concerns a similar topic as that of a previous question, but with different participants, peripheral, or constraints. It has the following subcategories:

Constraint Refinement

A question is about a similar topic with additional or revised constraints. For example:

Q7: What’s the crime rate in Maryland and Virginia?

Q8: What is it ten years ago?

For another example:

Q9: What’s the crime rate in Maryland and Virginia?

Q10: What was it in Alabama and Florida?

In both examples, both questions share the topic of “crime rate”, but concerning different crime rates with different constraints. Interpreting the second question requires not only identifying constraints, but also the relations between constraints. In the first example, the constraints from Q7 need to be used to form a query for Q8. However, constraints from Q9 should not be used for Q10.

Participant Shift

A question is about a similar topic with different participants.

For example:

Q11: In what country did the game of croquet originate?

Q12: What about soccer?

In this example, both questions are about the origination of a certain sport. The Content structure for both questions are the same except for the Participant role, which in Q11 is “croquet” and in Q12 is “soccer”. Therefore, the query created for Q12 would be {country, soccer, originate}, the keyword “croquet” should not appear in the query list.

Topic Exploration

Two questions are concerning the same topic, but with different focus (i.e., asking about different aspects of the topic). For example,

Q13: What is the name of the volcano that destroyed the ancient city of Pompeii?

Q14: When did this happen?

In this example, “this” in Q14 refers to the same activity topic in Q13, but focus on the TIME peripheral information about the activity.

In the following example,

Q15: Where is Mount Rainier?

Q16: How tall is it?

Q15 asks about the location of Mount Rainier (which is an entity topic) and Q16 asks about a different aspect (i.e., the height) of the same entity topic. In both examples, significant terms representing the Topic from the preceding question can be merged with the significant terms in the current question to form a query.

Topic Shift

Two consecutive questions could ask about two different topics. Different topic shifts indicate different semantic relations between two questions.

Activity Topic shifts to another Activity Topic

In the following example,

Q17: What is the name of the volcano that destroyed the ancient city of Pompeii?

Q18: How many people were killed?

The topic of both questions concerns about certain activities. This activity shift indicates that “kill” activity is a *consequence* of “destroy” activity (i.e., Q18 is a consequence of Q17).

Other relations can also be entailed from such a transition such as “effect-cause” relation as in the following example (Harabagiu et al. 2001):

Q19: Which museum in Florence was damaged by a major bomb explosion in 1993?

Q20: How much explosive was used?

Activity Topic shifts to Entity Topic

In the example:

Q21: What is the name of the volcano that destroyed the ancient city of Pompeii?

Q22: How tall is this volcano?

The topic of Q21 is an activity of “destroying” and the focus is the agent of the activity “the volcano”. This focus becomes the topic of Q22. This transition indicates a further probing of a particular participant in an activity that can be independent of the activity itself. Therefore, the terms in Q21 will not be helpful in setting up the stage for processing Q22. Q21 should be used only to resolve reference to the definite noun phrase “this volcano”.

Related to the presentational perspective of a QA discourse, we currently only identify: *Media Shift*. This relation indicates that two questions are about the same information content, but with different preference of media presentation.

For example,

Q25: how tall is Mount Vesuvius?

Q26: Any pictures?

Q26 is asking for the images of the Mount Vesuvius. This indicates that the backend should perform image retrieval rather than text retrieval.

In summary, given two consecutive questions (Q_i , Q_{i+1}), a certain transition exists from Q_i to Q_{i+1} . These transitions determine how the context, for example, preceding questions and answers can be used in interpreting the following question and identifying the potential answers. Here we only list several examples to show the importance of these transitions, which are by no means complete. We plan to identify a list of salient transitions for processing context questions as well as their implications (e.g., semantic relations) in interpreting context questions.

2.3 Discourse Processing

Given the above discussion, the goal of discourse modeling for context question answering is to automatically identify the discourse roles of a question and discourse relations between questions as the QA session proceeds. This may be a difficult task that requires rich knowledge and deep semantic processing. However, the recent advancement in semantic processing and discourse parsing has provided an excellent foundation for this task.

The discourse roles are higher-level abstracts of the semantic roles as those provided in FrameNet (Baker et al., 1998) and Propbank (Kingsbury and Palmer 2002). Recent corpus-based approaches to identify semantic roles (Roth et al 2002, Gildea and Jurafsky 2002; Gildea and Palmer 2002; Surdeanu et al., 2003) have been successful in identifying domain independent semantic relations with respect to the predicate-argument structure. Furthermore, recent work also provides discourse annotated corpora with rhetorical relations (Carlson, et al., 2003) and techniques for discourse parsing for texts (Soricut and Marcu, 2003). All these recent advances make the semantic-rich discourse modeling possible.

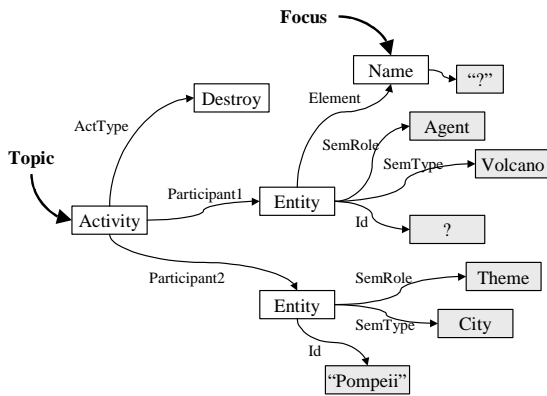
For example, a collection of context questions (and answers) can be annotated in terms of their discourse roles and relations. Specifically, the following information can be either automatically identified or manually annotated:

- Syntactic structures automatically identified from a parser (Collins, 1997);
 - Semantic roles of entities in the question (Gildea and Jurafsky 2002; Gildea and Palmer 2002; Surdeanu et al., 2003);
 - Discourse roles either manually annotated or identified by rules that map directly from semantic roles to discourse roles.
 - Discourse transitions automatically determined once discourse roles are identified for each question.
 - Semantic relations between questions manually annotated.
 - Answers provided by the system.
- Based on this information, important features can be

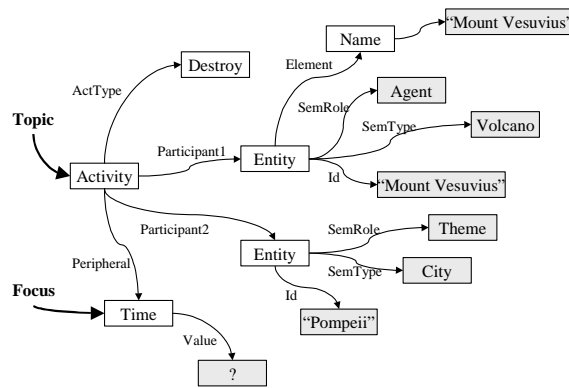
identified. Different learning models such as decision trees or Bayesian classifier can be applied to learn the classifier for discourse roles and relations. Strategies can be built to take into account of discourse roles and relations from preceding questions and answers to process a subsequent question and extract answers. These models can then be applied to process new context questions.

3 Refined Discourse Structure in Context Question Answering

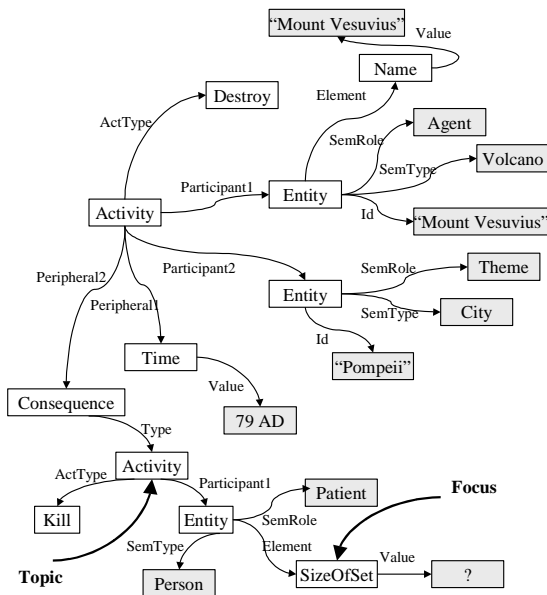
Based on the above discussion, during the question answering process, a discourse structure can be created to capture the discourse roles of each



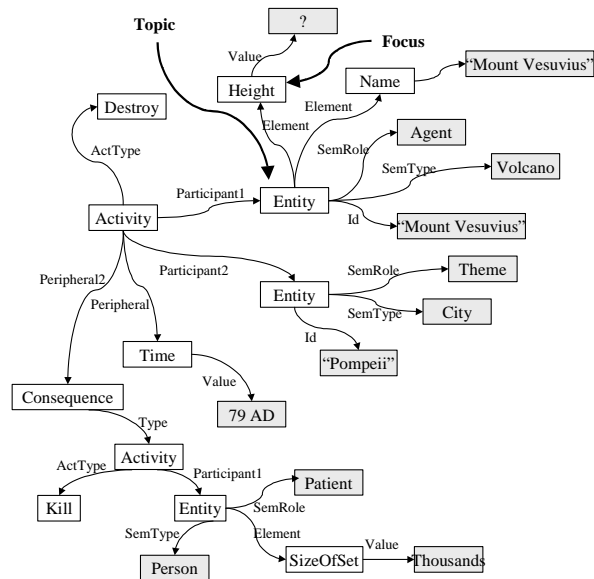
(a) Discourse representation after processing Q1



(b) Discourse representation after processing Q2



(c) Discourse representation after processing Q3



(d) Discourse representation after processing Q4

question and discourse relations between questions. Similar to information extraction for free texts, this refined discourse structure captures the salient information extracted from the question answering process. This discourse provides a structured information space that indicates what type of information has been exchanged and how information obtained at different stages is related. In other words, we can also consider this representation as the “mental map” of user information needs. This mental map will potentially provide a basis to improve question interpretation and answer extraction through inference, summarization, and collaborative question answering.

3.1 Discourse Representation

The typed feature structures can be represented as Directed Acyclic Graph (DAG). Thus the described discourse structure can be represented as semantic networks using DAGs. For example, Figure 3 shows the discourse representation after processing the each of the first four questions in Figure 1. In this network, each node is either a specific value (i.e., leaf nodes) or a typed feature structure itself (i.e., internal node). Each directed link corresponds to a particular feature. Note that because of the space limit, not everything represented in the feature structure in Figure 2 is shown here in the semantic network. For example, the type of an activity (e.g., Destroy) by itself is a feature structure (in Figure 2) that further consists of the specific term used in the question. This term is not shown but is included in the semantic nets.

As context question answering proceeds, the semantic network(s) for discourse grows, with different pointers of Topic and Focus. For example, Figure 3(a) represents Q1, where Topic points an Activity feature structure and Focus points to the Name Element of the Participant1 in the Activity. From Q1 to Q2, there is a transition of *Topic Exploration* which indicates that Q2 is about the same topic, but with a different focus. Therefore, in Figure 3(b), the Topic points to the same activity, but the Focus now points to the peripheral Time information of that activity. Next, Q3 is about a different topic involving activity Kill. However, since there is a consequence relation from Q2 to Q3, the activity asked in Q3 actually fulfills the Peripheral role of Consequence for the previous activity as shown in Figure 3(c). Finally, in Q4, there is a gap between Q3 and Q4, however, there is a transition of *Probing* from Q2 to Q4. Now the Topic becomes the Participant in Q1 as shown in Figure 3(d).

3.2 Potential Impacts

The growth of the semantic networks represents the overall information needs of a user and how such information needs are related. Since this is a structured representation, it can be queried and used to facilitate context question answering, for example, in the following aspects:

- Query expansion and answer retrieval
- Inference and summary for question answering
- Collaborative question answering

To process questions, most systems will first form a query of keywords to represent the current question and to retrieve relevant passages that may contain the potential answers. In context question answering, since the interpretation of a question may depend on preceding questions, some keywords from preceding questions may need to be included in the query for the current question. The fine-grained discourse structure will enhance answer retrieval through more controlled selection of terms from preceding questions and answers. For example, strategies can be developed to select query terms depending on the discourse relations. Different discourse roles and transitions may lead to different weighting schemes for query expansion.

Furthermore, the information captured in the discourse structure can help make predication about what the user information need is and therefore provide more intelligent services to help user find answers. For example, semantic and discourse relations between different topics and focuses of a series of questions can help a system infer and predict the overall interest of a user. Although answers to each question may come from different sources, based on the structured discourse (e.g., in semantic network), the system can aggregate information and generate summaries.

Another potential impact of the refined structured discourse is to facilitate collaborative question answering. Very often, various users may have a similar interest about a set of topics. The structured discourse built for one user can be used to help answers questions from another user. A user may have a certain information goal in mind, but does not know what types of questions to ask. Therefore, a user’s question may be very general and vague, such as “what happened to Pompeii?” This question needs to be decomposed into a set of smaller questions. The discourse structure that connects different aspects of topics together can provide some insight on how such decomposition should be made. Furthermore, the discourse structure from a skilled user can enable the

system to intelligently direct a novice user in his information seeking process.

4 Discussion

TREC 10 Question Answering Track initiated a context-task that was designed to investigate the system capability to track context through a series of questions. As described in (Voorhees 2001), there were two unexpected results of this task. First, the ability to identify the correct answer to a question in the later series had no correlation with the capability to identify correct answers to preceding questions. Second, since the first question in a series already restricted answers to a small set of documents, the performance was determined by whether the system could answer a particular type of question, rather than the ability to track context. Because of these unexpected results, the context task has been stopped in the following TREC evaluations (Voorhees 2002).

The reasons that TREC 10 did not achieve the expected results, in our opinion, lie in two aspects. The first aspect relates to the uniqueness of open domain context question answering. In open domain QA, first, there may be many occurrences of correct answers in various part(s) of document(s). Second, there may be multiple paths (e.g., different combination of key query terms) that can lead to one occurrence of the correct answer. Therefore, the correct answer to a previous question may not be critical in finding answers to subsequent questions. This phenomenon may provide an opportunity to find answers without explicitly modeling context (i.e., by keeping track of the discourse objects from answers), but rather identifying and using relevant context.

For example, in the LCC system (that achieved the best result for the context task in TREC 10), the discourse was not explicitly represented (Harabagiu et al 2001). Instead of resolving references using discourse information, the LCC system first identifies the question that contains the potential referents and uses those questions and the current question to identify the target paragraph. Thus, question interpretation does not depend on the answers, but rather depends on the context that is dynamically identified as a list of preceding questions. Now the question is whether the system will achieve even better results (e.g., correctly find answers to the rest of the eight questions) with some context representation?

Another more important question to be asked is whether the design of the context task just happened to provide an opportunity to achieve good results

without modeling the context. As discussed in (Harabagiu et al., 2001), answers to 85% of context questions actually occurred within the same paragraph as the answers to the previous questions. Therefore, just using preceding questions, the system was able to find the target paragraph and the final answer really depended on the capability to identify different types of answers in that paragraph. What if a series of questions were designed differently so that questions are related but answers are scattered in different documents or paragraphs. Will the shallow processing of discourse succeed in finding the answers?

Furthermore, the ultimate goal of QA systems is to be able to access information from different sources (e.g., unstructured text or structured database) and to provide intelligent dialog capability. One important question we need to address is what kind of discourse representation will be sufficient to support these capabilities? For example, to access structured databases, the answer to a previous question usually narrows down the search spaces in the database for subsequent questions. Thus, previous answers usually determine where in the database an answer can be found. Therefore, it is important to keep track of previous questions and answers in some kind of structure for later use.

The second reason that TREC 10 did not achieve expected results relates to evaluation methodology. In context QA, good performance depends on two important components: the capability of representing and using the relevant context (both explicitly or implicitly) and the general capability of interpreting questions and extracting answers. The level of sophistication in either component will influence the final performance. Thus, by comparing the final answers to each context question, the evaluation of the context task in TREC 10 was not able to isolate the effect of one component from another. It is not feasible to identify that when an answer is not identified, whether it is because of poor representation and use of the discourse information or it is because of the general limitations of the capability to process certain types of questions. Therefore, to study the role of discourse in context question answering, a more controlled evaluation mechanism is desired. For example, one approach is to keep the general processing capability as a constant and vary the representations of discourse and strategies to use the discourse so that their different impacts on the final answer extraction can be learned.

As a summary, the experience in the TREC 10 context task is very valuable. It does not discount the importance of context modeling for context questions. But rather, it motivates a more in-depth investigation of the role of discourse in context question answering.

5 Conclusion

Questions are not asked in isolation, but rather in a cohesive manner that involves a sequence of related questions to meet user's information needs. It is important to understand the role of discourse to support this cohesive question answering.

By all means, a QA discourse can be represented as coarse as a list of keywords extracted from previous questions or as sophisticated as a fine-grained representation as described in this paper. There is a balance between how much we like to represent the context and how far we can get there. Given recent advances in text-based domain independent semantic processing and discourse parsing, as well as the availability of rich semantic knowledge sources, we believe it is the time to start from the other end of the spectrum to examine the possibility and impact of semantic-rich discourse representation for open-domain question answering.

References

- Baker, C., Fillmore, C., and Lowe, J. 1998. The Berkeley FrameNet project. In Proceedings of COLING/ACL, pp. 86-90, Montreal, Canada
- Carpenter, R. 1992. The Logic of Typed Feature Structures. Cambridge University Press.
- Chai, J., Pan, S., and Zhou, M. 2003. MIND: A Context-based Multimodal Interpretation Framework in Conversational Systems, *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, Eds. O. Bernsen, L. Dybkjaer and J. van Kuppevelt, Kluwer Academic Publishers.
- Carlson, L., Marcu, D., and Okurowski, M. 2003. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In Jan van Kuppevelt and Ronnie Smith, editors, *Current Directions in Discourse and Dialogue*, Kluwer Academic publishers.
- Collins, M. 1997. Three Generative, Lexicalized Models for Statistical Parsing. In Proceedings of the 35th Annual meeting of the Association for Computational Linguistics (ACL 1997): 16-23, Madrid, Spain.
- Fillmore, C. and Atkins, B. 1998. FrameNet and lexicographic relevance, *Proceedings of the First International Conference on Language Resources and Evaluation*, Granada, Spain.
- Grosz, B. J. and Sidner, C. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175-204. 1986.
- Gildea, D. and Jurafsky, D. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245-288.
- Gildea, D. and Palmer, M. 2002. The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL 2002)*: 239-246, Philadelphia, PA.
- Harabagiu, S., Pasca, M., and Maiorano, S. Experiments with Open-domain Textual Question Answering. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, 2000
- Harabagiu, S., et al., Answering Complex, List and Context Questions with LCC's Question-Answering Server. In the *Proceedings of TREC 2001*, 2001.
- Hobbs, J. 1996. On the relations between the informational and Intentional Perspectives on Discourse. In *Burning Issues in Discourse: An inter-Disciplinary Account* (eds. E. Hovy and D. Scott), volume 151 of NATO ASI Series, Series F: Computer and Systems Sciences, pp 139-157. Springer-Verlag, Berlin, Germany
- Kingsbury, P. and Palmer, M. 2002. From Treebank to Propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands, Spain.
- Roth, D., et al. 2002. Question Answering via Enhanced Understanding of Questions. *Proceedings of TREC2002*.
- Surdeanu, M., Harabagiu, S., Williams, J., and Aarseth, P. 2003. Using Predicate-Argument Structures for Information Extraction. In *Proceedings of the 41th Meeting of the Association for Computational Linguistics (ACL 2003)*, Sapporo, Japan.
- Soricut, R. and Marcu, D. 2003. Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *Proceedings of HLT-NAACL*. Edmonton, Canada.
- Waldinger, R., et al., 2003. Deductive Question Answering from Multiple Resources, *New Directions in Question Answering*, AAAI, 2003
- Voorhees, E. 2001. Overview of TREC 2001 Question Answering Track. *Proceedings of TREC2001*.
- Voorhees, E. 2002. Overview of TREC 2002 Question Answering Track. *Proceedings of TREC2002*.