Generating Questions Automatically from Informational Text

Wei CHEN[1], Gregory AIST and Jack MOSTOW
Project LISTEN, School of Computer Science, Carnegie Mellon University

Abstract. Good readers ask themselves questions during reading. Our goal is to scaffold this self-questioning strategy automatically to help children in grades 1-3 understand informational text. In previous work, we showed that instruction for self-questioning can be generated for narrative text. This paper tests the generality of that approach by applying it to informational text. We describe the modifications required, and evaluate the approach on informational texts from Project LISTEN's Reading Tutor.

Keywords. Question generation, informational text, self-questioning, reading tutor

Introduction

Good readers ask themselves questions during reading. The National Reading Panel [1] reported that self-questioning was the most effective reading comprehension strategy, based on comprehension gains. So it would be useful for an intelligent tutor to automatically generate instruction for the self-questioning strategy to help students understand text. Our ultimate goal is to generate self-questioning instruction automatically from any given text, focusing on children's text.

Previous work [2] proposed a two-step approach to generate instruction for modeling and scaffolding the self-questioning strategy: first generate questions from the text, and then augment the questions into strategy instruction. It showed how to generate questions automatically from narrative text; here we test the generality of that approach by extending it to another important genre: informational text. The example below shows a paragraph from the informational text "Life under the Sea Part 1 – the Meaning of Life" and a question generated from the text.

Text: What does it mean to be "alive?" What is the difference, say, between an elephant and a boulder? This seems to be an obvious question but one that may be difficult to answer. All living things are not exactly alike. For example, not all living things breathe air, or have blood, or grow hair, like we do. Likewise, we can't live under water like fish do.

Question: Why can't we live under water like fish do?

Informational text is an important source of knowledge. Reading experts suggest that even young children can benefit from it, if taught the right strategy (e.g. [3, 4]).

[1] Corresponding Author.

Compared to narrative fiction, informational texts have different text structure and serve different reading goals [5]. For example, sentences (1) and (2) came from narrative and informational text, respectively.

(1) Peter thought it best to go away without speaking to the white cat.
(2) Rainbows are seen after it rains and the sun is out.

As exemplified by sentence (1), narrative text involves characters, their behavior, and mental states that drive it. In contrast, informational text does not require characters. In addition, it has more emphasis on descriptions and explanations, which are often used to introduce objective phenomena, as in (2).

The rest of this paper is organized as follows. Section 1 summarizes our previous work on question generation instruction from narrative text and describes how to extend the approach to informational text. Section 2 presents our evaluation criteria the results. Section 3 discusses the generality and importance of our approach based on those results. Section 4 summarizes the contribution, lists the limitations of our approach and points out future work.

1. Question generation for narratives

Previous work [2] described how to generate questions from narrative text and convert questions into modeling and scaffolding instruction. Given a piece of text, the question generation system transforms it into a situation model. A mental state expression (e.g., "think", "regret") indicates an opportunity to prompt strategy instruction. To model the self-questioning strategy for the student, the system could transform the sentence into a question, e.g., "Why did Peter think it best to go away without speaking to the white cat?" To scaffold the strategy, the system led the student to construct a question by choosing a character, a question type, and a completer. For example,

Tutor: *Let's make a question about ___ (Peter; Mr. McGregor; the old mouse; the white cat).*
Student: [The student chooses <u>*Mr. McGregor*</u> from the on-screen menu of 4 characters.]
Tutor: *Let's ask a ___ (what; why; how) question.*
Student: [The student chooses *why*.]
Tutor: *Great! 'Why' questions can help you understand by making you think!*
Tutor: *Let's complete your question: Why did Mr. McGregor ___ (try to put his foot upon Peter; try to find his way straight across the garden; think it best to go away without speaking to the white cat)?*
Student: [The student chooses *try to find his way straight across the garden*.]
Tutor: *I'm not sure Mr. McGregor tried to find his way straight across the garden. Would you like to try again?*

The character and completer choices are all automatically extracted from the story. Constructible questions include the system-generated question and other combinations such as "How did Mr. McGregor try to put his foot upon Peter?" and "What did the white cat think?" ("what"-questions will trigger different multiple choice completers, which is not shown in the example).

What is the scalability of our question generation mechanism? We tested it by adapting the approach to informational text. We cannot use exactly the same approach for informational text because of its different text structure and vocabulary [6]. Therefore, we have to add knowledge to our question generation system to deal with two specific issues in informational text: the "where" of decisions about what sentences to use to generate questions; and the "how" of inference rules used for building the situation model, and question templates to map the text into questions.

2. Locating question opportunities in informational text

2.1. Mental states in informational text

Our previous work relied on mental states to locate possible questioning points. However, mental states are not as central in informational texts as in narrative texts, in terms of their role in understanding the main idea of the text. Using the same set of mental state expressions, we found that mental states occurred 1382 times in 162 narrative texts (8.5 modal terms per text) from Project LISTEN's Reading Tutor [7] and 727 times in 105 informational texts which we used as our training data (6.9 modal terms per text). This shows that words and phrases indicating mental states occur more frequently in narrative text than in informational text, although the gap does not seem very big. On the other hand, mental states may vary by text genre in terms of what relationships between clauses they represent (e.g., causal vs. coincidental vs. mood-setting). In narrative text, mental states are normally attached to a character in the story, as in "Peter thought". Moreover, mental states of a character tend to reveal the motivation of the character and thus are likely to indicate causal relationships to events involving the character. In contrast, informational text does not always contain any character in the same sense as in narrative stories. Instead, informational text often contains mental states of the reader and author, as in sentence (3) (*italics* highlights mental state expressions), scientists' belief and common belief, as in sentence (4), and even mental states of research subjects, as in sentence (5). The mental state expressed in sentence (5) has similar meaning to those in narrative text (i.e., mental states of a character that are the motive force or result of some events or phenomena), whereas the two cases exemplified in sentences (3) and (4) are different.

(3) If you're an American citizen 18 years of age or older, you probably *think* you have the right to vote for presidential candidates in the national election.

(4) It is *thought* that they use this structure to detect prey, perhaps being able to distinguish the weak electrical signals given off by injured animals.

(5) Fish have "noses" (called nares) that don't look anything like our own, yet their *purpose* is to smell chemicals in the water.

2.2. Extension to other categories of question indicators in informational text

Based on the observations, we assume that merely using mental states as indicators of good questions is not enough for informational text. Our criteria for selecting questioning indicators are that the indicator has to signal key information about the text and it should be feasible for an intelligent tutor to recognize and generate good questions. According to expert pedagogy, teaching text structure is important in comprehending informational text [e.g. 8, 9]. However, automatically figuring out global text structure is still an active research problem which has not been solved completely, so we started with discourse markers that indicate modality (i.e., possibility and necessity) and some causal relationships (i.e., conditions and temporal context).

2.2.1. Causality: conditional and temporal contexts

Conditional and temporal context are very important in informational text. Compared to narratives, conditional context and temporal context in informational text are more likely to describe causation. For example, in sentence (2), the temporal expression “after it rains and the sun is out” describes a causal condition of the formation of a rainbow; (6) below shows another example of conditional context.

(6) *If humans removed all the kelp from the sea* soon all the other sea life would start to suffer as well.

To search for linguistic expressions that indicate conditional contexts, we enumerated 4 words and constructions we noticed in the training data as questioning points, namely “if”, “even if”, “only if”, and “as long as”. In our training data, if-constructions occurred 37 times. To find temporal expressions, we used the ASSERT semantic role labeler [10] to annotate the corpus automatically. Then the system looks for expressions marked by the ARGM-TMP tag [11] for “temporal expression”. The system found 763 such temporal expressions in the training data. We noticed four kinds of temporal expressions in our training data: a general condition that express a general condition such as “after it rains and the sun is out”, a date or time such as “in 1999”, a duration of time such as “for several hours”, and companionship such as “while she was reading”. Right now we focus only on the first type of temporal expression, which tends to indicate causality. To filter out the other three types of temporal expressions, we used regular expressions to detect dates, duration (i.e., started with the word “for”), and expressions that indicate companionship (i.e., started with the word “while”). We also noticed that some words about frequency such as “usually” and “sometimes” can lead to trivial “when”-questions, and they are often tagged individually with ARGM-TMP as in “[ARGM-TMP usually]”, which is not as informative for our purpose of finding causality. To filter them out, we used a heuristic which is to ignore temporal expressions that contain only one word. This heuristic filtered out 35.8% (273) temporal expressions, yielding 490 questioning points for questions about temporal context.

2.2.2. Linguistic modality: possibility and necessity

Linguistic modality such as possibility and necessity are also important in informational text. Linguistic modality is often expressed by auxiliary verbs. The most frequent auxiliary verbs can be hypothetical (e.g. “would”), predictive (e.g. “will”), or prescriptive (e.g. “should”, “ought to”, “must”). As sentence (7) shows, the word “should” expresses goats’ need for covered shelters. Thus reasonable question coming out of this sentence is “Why should goats have covered shelters?”

(7) All goats <u>should</u> have covered shelters where they can escape the weather.

We identified 8 auxiliary verbs and constructions from the training data to extract modality patterns, including “would”, “will”, “should”, “shall”, “could”, “ought to”, “must” and “may”. These words and phrases appeared in total 179 times in the 105 training texts.

2.3. Question generation process for informational text

Our questions are generated from the situation model, which is built from inference rules. The question generation system calls for one inference rule for each target conditional, temporal or modality expressions. Based on semantic categories of the target expressions, we defined 6 inference rules, which build various sub-contexts and store elements of statements in a situation model. For example, one inference rule for modeling temporal context can be paraphrased as “create a temporal context to store the when-statement; re-order existing temporal contexts based on time order.”

We added 4 question templates to transform the information retrieved from situation models to questions. The question template for conditional context is “What would happen if <x>?” For temporal context, we used two templates: “When would <x>?” and “What would happen (when) <x>?” For linguistic modality, we used “Why <aux-verb> <x>?” Since we aimed at questions about general conditions, which do not concern about tense, we included auxiliary verbs such as “would” in the question templates. Therefore, we do not need morphology generation for verbs, as we did for narrative text questions. Table 2 shows questions made from sentences (2), (6) and (7).

Table 2. Questions generated from temporal, conditional, and modality expressions.

Sentence number	Resulting question
(2)	a. When would rainbows be seen? b. What would happen after it rains and the sun is out?
(6)	What would happen if humans removed all the kelp from the sea?
(7)	Why should all goats have covered shelters?

3. Results

We evaluated the quality of the generated questions by the same criteria we used for mental state questions, i.e., the question has to be grammatically correct and it has to make sense in the context of the text. To evaluate our approach, we used a separate set of 26 informational texts from the Reading Tutor as our test data, which does not overlap with the training data. From the total 444 sentences in the test corpus, the system extracted 15 if-constructions, 44 temporal expressions and 33 auxiliary verbs. From these expressions, our question generation system generated 180 questions, including 15 questions about conditional contexts, 88 questions about temporal information, and 77 questions about modality (one auxiliary verb can participate in multiple semantic role tagging). We hand evaluated the questions in each of the three categories.

For questions about conditional contexts, three if-clauses turned out to be direct objects such as in “Scientists wondered *if meat-eating Tyrannosaurus rex had ever eaten Triceratops.*” Others were adverbs such as in “*If humans removed all the kelp from the sea* soon all the other sea life would start to suffer as well.” We rated 86.7% (13) of them as acceptable. The errors were due to unresolved coreference and ambiguity of “if” under different contexts. For example, the sentence “If so, then you have eaten kelp” resulted in an unacceptable question “What would happen if so?” by failing to resolve what “so” refers to. Also, some phrases like “as if” changed the meaning of “if” which in our case was defined to set a conditional context. The sentence “Sit beside a quiet pool of water and you'll soon see water striders skating as *if on ice*” resulted in an out-of-context question “What would happen if on ice?”

Of the 88 questions about temporal information, we scored 65.9% (58) as acceptable. 20 bad questions were due to parsing errors. For example, in the parsing result “If the pressure changes over a large area it can cause [ARG1 winds] to [TARGET blow] [ARGM-TMP in a huge circle]”, the tagger erroneously tagged “in a huge circle” as a temporal expression, leading to the unacceptable question “What would happen when in a huge circle?” The other 10 error questions came from undetected constructions that do not belong to the first type of temporal expressions, such as “at present” and “some day”. For example, from the sentence “At present totem poles are sold to people who collect them and to museums”, a question was “When would totem poles be sold to people who collect them and to museums?”, which is not asking something that the sentence is to convey.

Of modality questions, we rated 87.0% (71) as acceptable. All the observed errors were caused by parsing errors (including conference and negation errors). We used semantic roles as parameters to build the situation model, but sometimes the semantic roles are only partially tagged. For example, in “[ARG0 Skin cells] [ARGM-MOD must] [ARGM-DIS also] [TARGET make] [ARG1 sure] to keep harmful things out of the body while letting helpful things in”, the incomplete semantic role labeling yielded a question “Why must skin cells make sure?”

4. Discussion

The goal of this paper is to extend our question generation approach for narrative fiction to handle informational text. This problem involves two issues: a) how well does the approach work on informational text? b) how much additional work does it take to extend question generation from narrative to informational text? Section 3 has reported the quality of questions generated by the system. During the evaluation, we have noticed that some generated questions may not have explicit answers in the text, such as if-clauses as the direct object of a verb (e.g. “What would happen *if meat-eating Tyrannosaurus rex had ever eaten Triceratops*?”). This makes the question itself very interesting, because the question drives the student to think about a possible result that would be led by the condition, and it may not be obvious what the answers are from the text. Similar to the case in narratives, the inference rules we used for informational text can be used for extracting answers and detecting counter-factual questions, which is helpful for providing feedback to students in a complete instruction scenario. To adapt our approach to informational text, we kept the question generation process and same language technology tools, and we added three types of knowledge. Table 2 compares the knowledge we used for both genres.

Generating good questions requires inference, which is a natural language understanding problem. We know that natural language understanding is “AI-complete” because of the inference problem. We do not attempt to solve the entire inference problem, but to identify some inferences that we know how to make. At the knowledge representation level, we built only partial situation models (i.e., about conditional and temporal context and modality). We looked for the types of inferences that are feasible to extract and do not rely on world knowledge beyond the sentence (or story). The only information we needed for capturing important question indicators was knowledge of discourse markers such as if-constructions, temporal expressions, and auxiliary verbs.

Table 2. Comparison of question generation for informational text and narrative text.

Genre	Linguistic patterns	Type of questions	Generation templates
Narrative	mental state expressions	"What", "Why" and "How" questions about mental states	What did <character> <verb>?
			Why/How did <character> <verb> <complement>?
			Why was/were <character> <past-participle>?
Informational text	if-constructions	"What-would-happen-if" question about conditional context	What would happen if <x>?
	temporal expressions	"When-would-x-happen" question about temporal context	When would <x> happen?
		"What-would-happen-when" question about temporal context	What would happen (when) <x>?
	auxiliary verbs	"Why" question about possibility and necessity	Why <auxiliary-verb> x?

5. Conclusion, Limitations and Future Work

In this paper, we tested the generality of our question generation approach by extending it to another genre: informational text. We described an approach to generate questions from informational text, which could then be used to generate modeling and scaffolding instruction for the reading comprehension strategy of self-questioning. We extended the question generation approach to informational text by adding three types of knowledge: a) discourse markers for locating opportunities for questions; b) inference rules for managing information in a situation model; c) question templates for converting information into questions. We proposed three types of questions for informational text: questions about conditional context, questions about temporal information, and questions about possibility and necessity. We also demonstrated how discourse markers, such as conjunctions and certain kinds of verbs, can be used as indicators of good places to ask questions about text.

So far, we covered only three types of questions to generate from informational text. There are many other important features of informational text that can cause difficulty for young children, such as its non-linear text structure and implicit causality. In this paper, we explored discourse markers as causal implication. Future work includes extending the existing approach to interesting inference rules that can automatically discover implicit logical relationships in the text and build global text structures so that other important questions can be derived. As a next step, we will test our working assumption that there is a significant difference between the role of mental state expressions in understanding narrative fictions and informational text. We will also use questions generated by human teachers to evaluate our assumption that temporal context, conditional context and linguistic modality are important for self-questioning. We have already done pilot study for our system that generates self-questioning instruction from narratives. We will integrate the question generation approach for informational text into the intelligent tutoring system, user-test it and improve its performance based on pedagogical gains.

References (Project LISTEN publications are at www.cs.cmu.edu/~listen)

[1] NRP. Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. 2000, http://www.nichd.nih.gov/publications/nrppubskey.cfm: Washington, DC.

[2] Mostow, J. and W. Chen. Generating Instruction Automatically for the Reading Strategy of Self-Questioning. *The 14th International Conference on Artificial Intelligence in Education* 2009. Brighton, UK.

[3] Duke, N.K., V.S. Bennett-Armistead, and E. Roberts, eds. *Incorporating Informational Text in the Primary Grades*. Comprehensive Reading Instruction Across the Grade Levels, ed. C. Roller. 2002, DE: International Reading Association: Newark.

[4] Moss, B. Teaching Expository Text Structures through Information Trade Book Retellings: Teachers Can Help Students Understand Common Expository Text Structures by Having Them Retell Information Trade Books. *The Reading Teacher*, 2004. *57*.

[5] Meyer, B.J.F. Prose analysis: Purposes, procedures, and problems. In C.C.B.M. Pressley, Editor, *Understanding expository text*, 11-64. Erlbaum: Hillsdale, NJ, 1985.

[6] Purcell-Gates, V. and N.K. Duke. Explicit explanation/teaching of informational text genres: A model for research. *Crossing Borders: Connecting Science and Literacy conference* 2001. Baltimore, MD.

[7] Mostow, J. and J. Beck. When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens. *Conceptualizing Scale-Up: Multidisciplinary Perspectives* 2003. Park Hyatt Hotel, Washington, D.C.

[8] Duke, N.K. and V.S. Bennett-Armistead. *Reading & Writing Informational Text in the Primary Grades: Research-Based Practices*. 2003: Teaching Resources.

[9] Anderson, E. and J.T. Guthrie. Motivating children to gain conceptual knowledge from text: The combination of science observation and interesting texts. *The Annual Meeting of the American Educational Research Association* 1999. Montreal, Canada.

[10] Pradhan, S.S., W. Ward, K. Hacioglu, J.H. Martin, and D. Jurafsky. Shallow Semantic Parsing using Support Vector Machines. *the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004)* 2004. Boston, MA.

[11] Palmer, M., D. Gildea, and P. Kingsbury. The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics Journal*, 2005. *31*(1).