Generating Instruction Automatically for the Reading Strategy of Self-Questioning

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

Abstract. Self-questioning is an important reading comprehension strategy, so it would be useful for an intelligent tutor to help students apply it to any given text. Our goal is to help children generate questions that make them think about the text in ways that improve their comprehension and retention. However, teaching and scaffolding self-questioning involve analyzing both the text and the students' responses. This requirement poses a tricky challenge to generating such instruction automatically, especially for children too young to respond by typing. This paper describes how to generate self-questioning instruction for an automated reading tutor. Following expert pedagogy, we decompose strategy instruction into *describing*, *modeling*, *scaffolding*, and *prompting* the strategy. We present a working example to illustrate how we generate each of these four phases of instruction for a given text. We identify some relevant criteria and use them to evaluate the generated instruction on a corpus of 513 children's stories.

Keywords. Self-questioning, reading comprehension strategy, reading tutor

Introduction

Reading is fundamental. Teaching effective reading comprehension strategies can help students understand text better [1-3]. Some strategy instruction has been automated, generally for adolescent or college-age students [e.g. 4]. We focus more on children.

The National Reading Panel [1] found *self-questioning* the most effective strategy to teach, based on effect size for comprehension gains. In this strategy, also known as question generation [5], reader pose questions to themselves about the text. Good questions help the reader infer [6] and retain [7] the meaning of the text.

Wong et al. [6] identified three theoretical frameworks underlying instructional research in self-questioning. First, schema theory suggests that self-questioning activates the reader's background knowledge. Second, by bolstering metacognitive awareness, self-questioning helps students monitor their own comprehension. Third, by making readers process text actively, self-questioning invokes higher-order comprehension processes, such as inferring answers from text already read, or priming the student to notice them in later text, and improves retention.

Our goal is to automate instruction of the self-questioning strategy for a given text. Some earlier work [8] implemented and user-tested automated instruction for other

---

1 Corresponding author. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B070458. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute and the U.S. Department of Education. We also thank the educators, students, and LISTENers who helped generate and analyze our data, and Nell Duke and Christina Trotochaud for their expertise, comments, and advice.

comprehension strategies, such as verbal summarization and visualization, but the instruction was hand-scripted for each text. In contrast, this paper focuses on how to generate self-questioning instruction *automatically* for a given text.

Automatically generated instruction offers flexibility in teaching with *any* text, not just specific texts with particular instruction built in by hand. Automatically generated scaffolding for reading comprehension would apply to the broad class of software that displays text to read, and the large number of readers who need such scaffolding. Thus such a capability could have widespread impact.

1. A Working Example

This section presents a working example of automatically generated instruction in self-questioning; later sections explain how it is generated. The story text, in **boldface**, is a modern adaptation of an Aesop's fable from the collection of stories in Project LISTEN's Reading Tutor [9]. The instruction, in *italics*, consists of spoken tutorial interjections at suitable points in the text to *describe* a comprehension strategy, *model* its use, *scaffold* its application, and *prompt* the child to use it. This sequence is based on an established apprenticeship model that Duke and Pearson [2] instantiated for comprehension strategy instruction. It progressively transfers responsibility for applying the strategy from tutor to student. Student responses are in [brackets].

Text: **Once upon a time a town mouse, on a trip to the country, met a country mouse. They soon became friends. The country mouse took his new friend into the meadows and forest. To thank his friend for the lovely time, he invited the country mouse to visit him in the town. And when the country mouse saw the cheese, cake, honey, jam and other goodies at the house, he was pleasantly surprised.**

1.1. Describe the strategy

Teaching a strategy begins by explaining it. Our explanation is canned and generic. However, the decision of when to insert it is text-specific: at the first opportunity to apply the strategy, so that the tutor can illustrate the strategy right after describing it.

Tutor: *Good readers often ask themselves questions before, during, and after they read, to help them deepen their understanding of a text.*

1.2. Model the strategy

The tutor illustrates the strategy by posing a question about the sentence just read, combining hand-wordsmithed text-independent boilerplate and a <u>text-specific question</u>:

Tutor: *Right now the question I'm thinking about is, <u>why was the country mouse surprised?</u>*

1.3. Scaffold the strategy

Text: **…Suddenly there came the sound of heavy footsteps. The two mice ran. The man of the house had come to get a snack. He saw that mice had gotten some honey. So he decided to send the cat.**

At the next opportunity to apply the strategy, the tutor helps the child construct a question by choosing from 4 characters, 3 question types, and 3 question completers.

Tutor: *Let's make a question about ___ (the town mouse; the country mouse; the man of the house; the cat).*

Student: [The student chooses *the country mouse* from the on-screen menu of 4 characters.]

Next the tutor invites the student to choose what type of question to make up:

Tutor: *Let's ask a ___ (what; why; how) question.*

Student: [The student chooses *why*.]

The tutor praises the choice:

Tutor: *Great! 'Why' questions can help you understand by making you think!*

The tutor lets the student choose how to complete the question:

Tutor: *Let's complete your question: Why did the country mouse___ (decide to send the cat; try to taste everything before his tummy was full; run)?*

Student: [The student chooses *decide to send the cat*.]

The text may answer a question (1) earlier, (2) never, (3) later, or (4) contradict it, e.g.:

(1) How did the town mouse meet a country mouse?
(2) Why did the man of the house decide to send the cat?
(3) What did the cat forget?
(4) Why did the country mouse decide to send the cat?

The tutor gives feedback on counterfactual questions by inviting the child to try again:

Tutor: *I'm not sure the country mouse decided to send the cat. Would you like to try again?*

Otherwise the tutor chooses randomly from generic non-committal positive feedback:

Tutor: *You made a question!*

1.4. Prompt the strategy

Text: **The two mice, full of terror, hid away.… A large cat was staring around the room. The country mouse and the town mouse went on tiptoes back to their hole. Soon the cat found a snack of his own and forgot about the mice.**

At the next opportunity, the tutor prompts the student to use the strategy without help:

Tutor: *What is a good question to ask about what you just read?*

Student: [The student says (or types, if able to do so), "How are the mice now?"]

The tutor encourages the student to look out for the answer:

Tutor: *OK, keep that in mind and let's see what you find out!*

Text: **… As soon as the danger was gone, the country mouse shook hands with the town mouse. He said, "Thanks so much for everything! But I must rush home now! I can't stand all these shocks! I would rather have a boring meal in the country in peace and safety, than have a splendid meal but eat it in fear."**

After the story, the tutor prompts the student to relate questions, answers, and text, so as to encourage looking in future text for answers to self-generated questions:

Tutor: *We've reached the end of the story. Did you find the answers to your questions?*

Playing back (or displaying) students' questions could remind them of what they asked.

2. Approach

As Sections 1.1-1.4 illustrated, our instruction for self-questioning instantiates a modified version of Duke and Pearson's five-step model [2] for human instruction of a comprehension strategy: 1. Describe the strategy; 2. Model the strategy; 3. Scaffold the strategy; 4. Prompt the strategy; 5. The student practices the strategy independently.

Steps 1 (description) and 4 (prompting) are independent of the particular text, so we implement them using manually scripted generic instructions. Evaluating these steps means user-testing them to see if students understand the description and obey the prompts. We generate no instruction for step 5 (practice), since the student does it independently. Alternatively, before fading out prompting altogether, we could insert a generic prompt before the story to remember to use the strategy while reading the story. To evaluate its effectiveness, we could insert occasional think-aloud prompts, record students' spoken responses, and check them for signs of the prompted strategy.

Sections 3 and 4 address the technological challenge of how to generate instruction for steps 2 (modeling) and 3 (scaffolding) from a given text. Both steps rely on the same underlying capability to generate questions appropriate to teaching the strategy.

3. Modeling the Strategy: Automatic Question Generation

Before generating questions, we must first decide where to insert them. Although automated question generation is emerging as an active area of research [10], relatively little of this work apparently addresses this decision. Some previous work inserted questions to assess [11] or assist [12] comprehension during reading, but after randomly chosen sentences in the text. Its only conclusion about when to insert a question was "not too soon after the prior one," which often provoked hasty guesses.

In contrast to this randomized approach, we insert questions after statements about characters' mental states, such as belief, intention, supposition, emotion, and so on. Such statements typically draw inferential connections between key story elements [13, p. 448], thereby providing opportunities to pose good *why* questions. In our example in section 1, every intervention comes after a mental state expressed in the story. For example, the verb **surprise** or **decide** indicates a change in a character's beliefs or intentions, which young children may not easily recognize [14]. Thus it is natural to ask why the character was surprised or made a particular decision. This model-driven approach contrasts with shallower methods, such as turning a sentence into a cloze question by deleting a word [11], or asking *Which will come next?* with the next 3 sentences in the text (in scrambled order) as the choices [12].

To detect mental states, we look for 10 categories of modal verbs (including constructions like *think of*) enumerated by following WordNet synset links as detailed elsewhere [15]. We manually removed dozens of words with no modal sense, and added 42 modal verbs we noticed in stories, yielding a total of 239 modal verbs. Since we evaluated the approach on the same set of stories, this may also lead to overfitting. As Table 1 shows, each category is named for a modal verb, such as *decide*, and also contains related words at a semantic distance of 1, 2, or 3 synset links away from it.

Table 1. Modal verb categories with semantic distance to each member, or M if added manually.

Confess: 1. concede, fink, profess, squeal, …; 2. cede, yield, …; 3. give in, succumb, …; M. admit
Decide: 1. determine, settle, …; 2. ascertain, find out, …; 3. finalize, nail down, …; M. choose, prefer
Frustrate: 1. baffle, foil, spoil, torment, …; 2. dun, fail, torture, …; 3. annoy, bother, …; M. hurt
Pretend: 1. dissemble, feign, make believe…; 2. cloak, mask, …; 3. belie, disguise, … M. claim, lie
Regret: 1. repent, rue; 2. atone; 3. aby, abye, expiate
Satisfy: 1. fulfil, fulfill, gratify, live up to, …; 2. accomplish, carry out, …; 3. achieve, attain, reach, …
Surprise: 1. storm; M. affect, impress, move, strike
Think: 1. believe, conceive, guess, imagine, …; 2. envisage, infer, judge, trust, …; 3. understand, …
Want: 1. desire, need, require, …; 2. call for, demand, hope, …; 3. go for, …; M. attempt , intend, try, …
Worry: 1. care, concern, interest, occupy, vex; 2. bewilder, disturb, …; 3. bemuse, …; M. dread, fear

To generate questions about mental states, we adapted a mental state understanding system that inputs a text and transforms the information in the text into a situation model in SCONE [16]. This system is reported in more detail elsewhere [15], so we only summarize it here. The mental state understanding system consists of a parser that generates a semantic representation of the input text, and a mental state inference engine that uses 29 inference rules about mental states to model higher-order cognition in reading by expanding the semantic representation of the text into a situation model of the story. One such inference rule states that a person surprised by something believes it afterwards but not beforehand. Right now the situation model is built only on knowledge about modal verbs, without other world knowledge.

We generate three types of questions, namely *what*, *why*, and *how*. They are specific to the text, unlike previously studied generic multiple-choice *wh-* questions with text-independent choices, e.g., *When does this take place? in the present; in the future; in the past; I can't tell* [12].

We use question templates plus a morphology generator to transform the inferred information in the situation model into questions. A question template consists of a question word and required information for that question type. The question templates are *What did <character> <verb>?, Why/How did <character> <verb> <complement>?,* and *Why was/were <character> <past-participle>?* Characters, verbs, participles, and complements come from the situation model.

3.1. Evaluation of Automated Question Generation

How acceptable are the automatically generated questions? Our criteria are that the question must be grammatically well-formed and appropriate to the story context based on (our) human judgment. Human teachers tend to ask shallow questions [17]. So we did not choose to compare our questions with human-made ones. We evaluated question generation on a corpus of 513 children's stories from Project LISTEN's Reading Tutor, 401 of which contained one or more modal verbs. From the 14,727 sentences in these stories, our question generation tools extracted 1,594 mental state expressions (as questioning points). In a random sample of 100 mental state expressions, 83 are covered by our question templates. To fill in the question templates, we need a complete set of roles of a modal verb, namely the agent and sometimes the direct object. Only 61.8% of the modal verbs were labeled with the agent and if necessary the direct object. They generated 1,537 questions, of which we randomly selected half to evaluate. Of these 769 questions, we rated 548 (71.3%) as acceptable. We know of no published baseline to compare to. This quality is too low to use as is, but is plausibly high enough to be cost-effective in combination with human vetting.

Common errors were due to parsing errors and failure to detect certain grammar phenomena, causing bugs in the situation model. Parsing errors caused incorrect assignment of verb arguments. In **Mom likes flowers, they thought**, misidentifying the object of **thought** as **flowers** rather than **Mom likes flowers** led to the ill-formed question *Why did they think flowers?* The overall parsing accuracy on the fully parsed sentences was 85.62%. Undetected grammar phenomena can cause counterfactual questions. For example, in **He put his hands before his eyes as if he were trying to remember something**, the parser correctly identified **trying** as a modal verb and **he** as its agent, but ignored the phrase **as if**, leading to the counterfactual question *Why did he try to remember something?*

4. Scaffolding the Strategy: Helping Students Construct Questions

Scaffolding helps students apply the strategy. As the example in section 1.3 illustrated, the tutor follows a menu-driven scaffolding scenario, guiding the student to construct a question by making a series of multiple-choice decisions. The tutor then gives feedback on the constructed question. Unconstrained spoken or typed questions would be more natural, but giving intelligent feedback on them would be technologically problematic. The scaffolding dialogue constrains the process and result of generating a question, but it lets students participate in making questions. The student's choice at each step affects the subsequent decisions and the reading tutor's feedback.

This scenario is triggered by detecting a mental state expression in the text. For example, the word **decide** signals an opportunity to initiate a scaffolding dialogue. The scenario contains three student decisions: character, question type, and completion.

First, the tutor prompts the student to choose which character to ask about, e.g., *the country mouse*. The tutor prompts for a character first because it is more concrete than question type. The four choices are randomly chosen from characters already mentioned in the story, augmented if necessary by drawing from a list of generic characters – *a girl*, *a boy*, *a woman*, *a man*, *an animal*, *a group of people or animals*. Our example text mentions four characters prior to the scaffolding point: **the country mouse**, **the town mouse**, **the man of the house**, and **the cat.** Candidate characters, in order of preference, are agents of modal verbs, then proper noun phrases, and lastly objects of modal verbs. This heuristic is intended to prefer sentient entities to inanimate objects.

Second, the student chooses the question type. The choices of question types are always the same – *what*, *why*, or *how*. The tutor responds with praise when the student chooses to make a *why* or *how* question, because they typically stimulate deeper comprehension than shallow questions whose answers are "right there in the text."

Third, the student chooses a completion of the question – a proposition or main verb, depending on the question type. For *why* and *how* questions, completers are story verbs (preferably modal ones) with complements, if required, e.g., *decide to send the cat*. So the tutor might prompt, *Why did the country mouse___ (decide to send the cat; try to taste everything before his tummy was full; run)?*

For *what* questions, candidate completions come from modal verbs in the story, augmented if necessary by distinct categories selected at random. Each menu item for a *what* question belongs to a different category, at least one of which contains modal verbs that appeared in the story. Thus the tutor might prompt, *What did the cat ____? (decide; try; fear)*. The resulting *what* question asks for an element of the situation model.

Finally, the tutor provides feedback on the completed question. Feedback on constructed questions is constrained by the situation model, which captures mental states but not the physical world. We therefore assess only whether the completed *why* or *how* question is counterfactual, i.e., not supported by the situation model. The feedback is intended to motivate students to think up good questions. To respond to counterfactual questions, we combine the <u>unsupported proposition</u> with a generic response, as in *<u>I'm not sure the country mouse decided to send the cat.</u> Would you like to try again?.* Otherwise, the response is generic, e.g. *You made a question!* Other responses, e.g. trying to answer the question, might engage students but disrupt reading.

4.1. Evaluation of Scaffolding

We used the same story corpus to evaluate three aspects of scaffolding: (1) Are the menu choices acceptable, that is, grammatical, appropriate, and semantically distinct? (2) Do one or more choices lead to acceptable questions? (3) Is the feedback accurate? To generate scaffolding instruction, we augmented questions generated automatically in Section 3, by adding multiple choices and feedback. The various combinations generated 3,108 distinct questions. We randomly selected half of them to evaluate. We scored only 35.6% as acceptable, due to counterfactual and ungrammatical combinations. However, based on a random sample of 100 3-step construction sequences, 48% of them led to completer menus with at least one acceptable question.

We classified 84.4% of the character choices and 80.9% of the question completer choices as acceptable. Buggy character choices were mainly due to incorrect or missing coreference links. For example, the menu choices *the two mice* and *the country mouse and the town mouse* are semantically equivalent because they refer to the same two entities. Buggy question completer choices were due to parsing errors that made question completers either too short, as in **he believed what** (omitting **she said**), or too long, as in **They thought that the Hen must contain a great lump of gold in its inside <u>and in order to get the gold they killed it</u>** (wrongly including the underlined text).

We classified 60.4% of constructable questions as counterfactual. Detection of counterfactual questions was 90.0% accurate. 7.5% of the counterfactual questions were misclassified as factual, for example, *Why did he think that would help him in making amends for his own sins?*, as described in Section 3.1. To prevent praise for an undetected counterfactual question, we praise the choice of question type (*why*) instead of the question itself. 14.0% of the factual questions were misclassified as counterfactual. Detection accuracy depends on the accuracy of the situation model. For example, *Why did the country mouse hide away?* was misclassified as counterfactual because the program did not understand that **the country mouse** was one of **the two mice**.

5. Contributions, Limitations, and Future Work

We have defined the problem of generating comprehension strategy instruction for a given text. We focused in particular on the strategy of self-questioning. Future work may generate instruction for other comprehension strategies currently taught or scripted by humans, such as visualizing and summarizing. We focused on the genre of narrative fiction, in which characters and their mental states are central. We are starting to address other text genres, such as informational text, where characters and mental states may be absent altogether, so question generation must exploit other deep structure in the text, such as temporal and causal relations.

Based on expert human pedagogy, we decomposed the problem of automatically generating instruction into describing, modeling, scaffolding, and prompting the strategy, and illustrated these four phases in detail in a working example. We showed how to extend automated question generation technology to generate such instruction.

We proposed some relevant criteria for evaluating the generated instruction, and used them to measure performance on hundreds of children's stories. We are working to reduce and mitigate its errors, improve its feedback, identify and impose more stringent criteria, incorporate the instruction into the Reading Tutor, user-test it, refine it accordingly, and ultimately assess its impact on children's comprehension of the text.

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] Duke, N.K. and P.D. Pearson. Effective Practices for Developing Reading Comprehension. In S.J. Samuels, Editor, *What Research Has To Say about Reading Instruction*, 205-242. International Reading Association: Newark, DE, 2002.

[3] McNamara., D.S., ed. *Reading comprehension strategies: Theories, interventions, and technologies*. 2007, Erlbaum: Mahwah, NJ.

[4] Johnson-Glenberg, M.C. Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders. *Reading and Writing*, 2005. *18*: p. 755–786.

[5] Rosenshine, B., C. Meister, and S. Chapman. Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 1996. *66*(2): p. 181-221.

[6] Wong, B.Y.L. Self-questioning instructional research: A review. *Review of Educational Research*, 1985. *55*(2): p. 227-268.

[7] Wood, E., G. Miller, S. Symons, T. Canough, and J. Yedlicka. Effects of Elaborative Interrogation on Young Learners' Recall of Facts. *Elementary School Journal*, 1993. *94*(2): p. 245-254.

[8] Zhang, X., J. Mostow, N.K. Duke, C. Trotochaud, J. Valeri, and A. Corbett. Mining Free-form Spoken Responses to Tutor Prompts. *Proceedings of the First International Conference on Educational Data Mining*, 234-241. 2008. Montreal.

[9] Mostow, J. and J. Beck. When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens. In S.-K. McDonald, Editor, *Scale-Up in Education*, 183-200. Rowman & Littlefield Publishers: Lanham, MD, 2007.

[10] Rus, V. and A. Graesser. Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge, http://www.questiongeneration.org/. 2008. NSF, Arlington, VA.

[11] Beck, J.E., J. Mostow, and J. Bey. Can automated questions scaffold children's reading comprehension? *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, 478-490. 2004. Maceio, Brazil.

[12] Mostow, J., J. Beck, J. Bey, A. Cuneo, J. Sison, B. Tobin, and J. Valeri. Using automated questions to assess reading comprehension, vocabulary, and effects of tutorial interventions. *Technology, Instruction, Cognition and Learning*, 2004. *2*: p. 97-134.

[13] Nelson, K. *Language in Cognitive Development: The Emergence of the Mediated Mind*. Cambridge University Press 1998.

[14] Nguyen, L. and D. Frye. Children's Theory of Mind: Understanding of Desire, Belief and Emotion with Social Referents. *Social Development*, 2001. *8*(1): p. 70-92.

[15] Chen, W. Understanding Mental States in Natural Language. *Proceedings of the 8th International Workshop on Computational Semantics*, 61-72. 2009. Tilburg, Netherlands.

[16] Fahlman, S.E. Marker-Passing Inference in the Scone Knowledge-Base System. *First International Conference on Knowledge Science, Engineering and Management (KSEM'06)* 2006. Guilin, China: Springer-Verlag).

[17] Dillon, J.T. *The Practice of Questioning*. 1990: New York: Routledge.