In: V. Rus and A. Graesser (eds.), online Proceedings of Workshop on the Question Generation Shared Task and Evaluation Challenge, September 25-26, 2008, NSF, Arlington, VA. [http://www.questiongeneration.org/]

Generating Questions: An Inclusive Characterization and a Dialogue-based Application

Paul Piwek
Centre for Research
in Computing
The Open University, UK

Helmut Prendinger
National Institute
of Informatics
Tokyo, Japan

Hugo Hernault
National Institute
of Informatics
Tokyo, Japan

Mitsuru Ishizuka
University of Tokyo
Tokyo, Japan

Abstract

In this paper, we make a case for the intrinsic richness of the Question Generation (QG) task. We examine QG in terms of its input, output and the relation between the two. We then describe a specific approach to QG which depends on λ-abstraction to formalize the relation between questions and answers. This allows us to make fine-grained distinctions between different types of questions. We have applied this approach to the generation of dialogue from monologue, exploiting the coherence relations underlying the monologue.

1 Introduction

This paper argues for an inclusive characterization of Question Generation (henceforth QG), embracing a wide variety of approaches. In our view, such an open-minded approach is most conducive for a new and hopefully soon burgeoning area of research. We strongly support concrete tasks and resources to develop the field, but without losing sight of the wider context of Question Generation as a general problem. As a research topic and emerging community, the chance of making a real impact can, in our view, be maximized by bringing people from a variety of backgrounds together, exploiting cross-fertilization between different approaches, and aiming for a critical mass that can be build up by appealing to a broad audience.

2 Characterizing Question Generation

As a starting point, let us characterize *Question Generation* as *the task of automatically generating questions*. This characterization, though very inclusive, focuses on QG from a *computational* point of view. Consequently, work on QG is seen as primarily belonging to the fields of Artificial Intelligence and Computational Linguistics, with disciplines such Psychology and Formal Logic in a supporting role.

Instead of narrowing the characterization further, we would like to illustrate the richness of the resulting topic of research by asking the following three questions:

1. What is the *input* to the QG task?
2. What is the *output* of the QG task?
3. What is the *relation between* input and output?

The answers to these questions result in a 'map' of the potential ground covered by QG.

The output The second question, *What is the output?*, may at first sight appear straightforward. It does, however, immediately lead to the issue whether we define questions purely *syntactically* (as interrogative sentences) or *semantically* as conveying informational gaps, see, e.g., Piwek (1998). In the latter case, the output may include, for example: 'declarative questions' (Beun, 1990), questions expressed by imperative sentences ('Tell me what time it is'), and also embedded questions ('I know who lives here'). Moreover, even those questions that do have the interrogative form are often inseparably linked to declarative sentences. Take, for example, the question 'What is the reason for this?' in 'For more than a century and a half, [...] has been known as the finest watch in the world. [...] What is the reason for this?'. Van Kuppevelt (1996) calls declarative sentences that set up the context for an interrogative sentence 'feeders'

In: V. Rus and A. Graesser (eds.), online Proceedings of Workshop on the Question Generation Shared Task and Evaluation Challenge, September 25-26, 2008, NSF, Arlington, VA. [http://www.questiongeneration.org/]

Further issues emerge when we consider not only *written* but also *spoken* questions: e.g., the questions (where italics indicate stress) 'Did you see *Mary*?' and 'Did you *see* Mary?' are subtly different (Pierrehumbert and Hirschberg, 1990). And all this presupposes that questions are always expressed verbally; what if *A* and *B* are working on a car, and *A* tells *B* to undo a screw whilst pointing at it. *B* might touch the screw that he thinks *A* indicated, raise his eyebrows and look at *A* for confirmation – now, did *B* ask a question? It seems that non-verbal actions, diagrams, pictures, formal languages, or a combination of any of these can be used to pose questions.

The input Many of the issues that arise when characterizing the output of QG also arise when characterizing its input. Though we might at first think of declarative sentences as the input, there is no principled reason for excluding interrogatives and imperatives. Also, the restriction to purely verbal input (whether written or spoken) seems arbitrary; why not also allow pictures, diagrams, gestures, or a (multimodal) combination of these?

The relation between in- and output At an abstract level, questions can be viewed as expressing informational gaps and answers as supplying the declarative information that fills such gaps. This is true regardless of the modality (written language, speech, gestures, etc.) of the questions or answers. At this level, we can distinguish between different relations that may hold between the input and output in QG. Here we would like to discern three basic cases:

A) *Question reformulation:* We proceed from a question (the input) to another question (the output). In question reformulation the content of the input and output question is identical or sufficiently similar, what changes is the way the question is expressed, i.e., its form. The purpose could, for example, be to translate queries that a system can answer to Natural Language questions that a user can understand (Marciniak, 2008).

B) *Input answers the output question:* For example, given the input *John walks*, the output question could be *Who walks?*, a question that is answered by the input.

C) *Input raises the output question:* In the example below, we have a question and statement which together give rise to another question. Note that the input statement is an indirect answer to the input question; see, e.g., Piwek (1998) .

In: *If John's car is in the garage, he is at home. Is John at home?*
Out: *Is John's car in the garage?*

3 Generating Questions in Dialogue

Let us start by briefly motivating our application of QG since, whereas it is obvious why one would like to generate answers, it is not immediately clear why there is a need for the generation of questions. An excellent example of QG in interactive tutoring is provided by Rus et al. (2007) who automatically generate questions corresponding to prompts and hints for a student, with as input the system's expectations regarding a correct answer. Similar to this application, our work also focuses on generating questions whose answer is provided by the input. However, rather than generate individual questions, our system – T2D; Piwek et al. (2007) – generates a complete *dialogue* from an input text that is in monologue form. The resulting dialogue is meant to faithfully convey the information in the monologue, but now as a conversation between an expert and a layman.

The rationale for generating dialogue automatically is that a number of studies suggest that dialogue can be more effective than monologue both for educational and persuasive purposes. For example, Craig et al. (2000) found that dialogue stimulates students to write more in a free recall test and ask twice as many deep-level reasoning questions in a subsequent tutor-guided task on a different topic. Additionally, presenting information in the form of a dialogue is a popular means for engaging and entertaining an audience, as witnessed by the widespread use of dialogue in commercials, news bulletins (between presenters), educational entertainment, and games. Automatically generated dialogue allows emulation of such engaging presentation forms by rendering dialogue as synthesized speech or through Life-like Computer-animated Characters; see, e.g., Prendinger and Ishizuka (2004).

The T2D system generates dialogue from monologue that has been analysed in terms of its underlying coherence relations, in particular, the relations found in Rhetorical Structure Theory (Mann and Thompson, 1988). For each relation (CONDITION, DEFINITION, CONTRAST, etc.), there are one or more transformations that can turn that relation into a Question-Answer pair. Questions are obtained by abstraction (formalized in the λ-calculus) over one or more elements in the input information. When abstracting over the arguments of a rhetorical relation or the relation itself, we can discern at least four theoretical possibilities. For the input 'If it rains, the tiles get wet' (with the underlying CONDITION relation) we have:[1]

- abstract over first argument: *Under what circumstances will the tiles get wet?*
- abstract over second argument: *What if it rains?*
- abstract over both arguments: *Under what conditions will what happen?*
- abstract over relation: *What is the relation between it raining and the tiles getting wet?*

Similar possibilities exist when abstracting over predicates and their arguments (e.g., 'John sees Mary' leading to 'Who sees Mary?', 'John sees who?', 'Who sees who?', and 'What is the relation between John and Mary?') and (generalized) quantifiers and their arguments (e.g., 'Five men walk' leading to 'Five men do what?', 'Five what walk?', 'Five what do what?', 'How many men walk?'). It would be interesting to investigate whether our formalization of questions using abstraction and the relations of Rhetorical Structure Theory can be used in the further refinement of question taxonomies from psychology and education (Graesser et al., 2008; Nielsen et al., 2008).

4 Conclusion

In this paper, we have argued for the intrinsic richness of the Question Generation task by describing the wide variety of possible inputs, outputs and relations between input and output. In this light, we would like to suggest for the QG Shared Task and Evaluation Challenge (STEC) to include, alongside one or more highly focused tasks, an open track for implemented and empirically evaluated work on QG. The second half of this paper briefly described the T2D system (Piwek et al., 2007) and the theoretical framework that underpins it for relating declarative information to questions. T2D exploits coherence relations in the input text, including those that cross sentence boundaries. Dealing with the full problem of monologue-to-dialogue transformation is a task that we hope the QG STEC could perhaps adopt after having explored the task of generating of single questions, a task which can be viewed as a first step towards generating complete dialogues.

[1]Note that Yes/No–questions are dealt with in this framework by abstracting over the modal operator *it is true* that or *it is false that*.

References

R.J. Beun. 1990. The recognition of Dutch Declarative Questions. *Journal of Pragmatics*, 14:39–56.

S. Craig, B. Gholson, M. Ventura, A. Graesser, and the Tutoring Research Group. 2000. Overhearing dialogues and monologues in virtual tutoring sessions. *Intl Journal of Artificial Intelligence in Education*, 11:242–253.

A. Graesser, V. Rus, and Z. Cai. 2008. Question classification schemes. In *This Workshop*.

W. Mann and S. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

T. Marciniak. 2008. Language generation in the context of yahoo! answers. In *This Workshop*.

R. Nielsen, J. Buckingham, G. Knoll, B. Marsh, and L. Palen. 2008. A Taxonomy of Questions for Question Generation. In *This Workshop*.

J. Pierrehumbert and J. Hirschberg. 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. In *Intentions in Communication*, pages 271–311. MIT Press, Cambridge MA.

P. Piwek, H. Hernault, H. Prendinger, and M. Ishizuka. 2007. T2D: Generating Dialogues between Virtual Agents Automatically from Text. In *Procs of IVA07*, LNAI 4722, pages 161–174, Berlin. Springer Verlag.

P. Piwek. 1998. *Logic, Information and Conversation*. Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands.

H. Prendinger and M. Ishizuka, editors. 2004. *Life-Like Characters: Tools, Affective Functions, and Applications*. Cognitive Technologies Series. Springer, Berlin.

V. Rus, Z. Cai, and A.C. Graesser. 2007. Experiments on Generating Questions About Facts. In *Procs. of CICLing 2007*, volume 4394 of *LNCS*, pages 444–455, Berlin. Springer Verlag.

J. van Kuppevelt. 1996. Directionality in discourse: Prominence differences in subordination relations. *Journal of Semantics*, 13:363–395.