Generating Questions and Multiple-Choice Answers using Semantic Analysis of Texts

Jun Araki, Dheeraj Rajagopal, Sreecharan Sankaranarayanan, Susan Holm, Yukari Yamakawa, and Teruko Mitamura
Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Contributions

- This is the first work to automatically generate questions from **multiple sentences**, involving specific **inference steps** such as coreference resolution and paraphrase detection.
- We present another approach which generates questions based on patterns extracted from relationships between events and entities.
- Our system also generates phrase-level distractors to challenge comprehension by using event-event relation annotations.

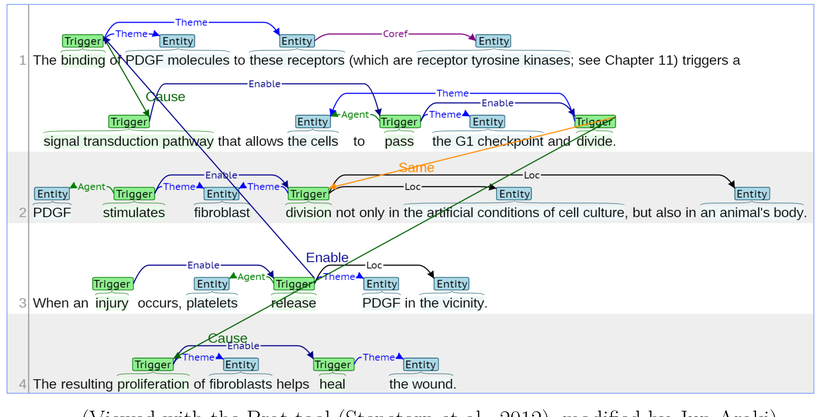
Introduction

- Goal: Automatic **question generation** (QG) for language learners.
- Our research can be useful to the task of creating exam questions and answers since automatic QG makes the task more efficient.
- **Higher-level questions** have more educational benefits for reading comprehension, as compared to simple factoid questions, e.g., (Anderson and Biddle, 1975; Andre 1979; Hamaker 1986).

Research Problems

- Existing QG systems generate questions from a single sentence, relying heavily on syntax and emphasizing grammaticality.
- The majority of questions generated from single sentences tend to be too specific and low-level, ending up essentially assessing the learners' ability to compare sentences.

ProcessBank corpus (Berant et al., 2014)
- 200 paragraphs from a biology textbook, with questions by biologists.
- We utilize human annotations of entities, events and relations.

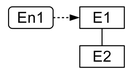
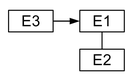
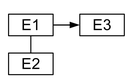
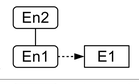
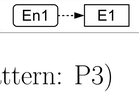


(Viewed with the Brat tool (Stenetorp et al., 2012), modified by Jun Araki)

Question Generation

Question Generation System 1 (QG1)

- Generate questions from multiple sentences using three semantic relations: event coreference, entity coreference, and paraphrases.
- Find answers using templates, and then construct questions.

Semantic relation	Question patterns	Answer	Question templates
Event coreference	P1. 	En1	T1. What [*verbal trigger* + *subsequent arguments*]?
	P2. 	E3	T2. What causes [*nominal trigger* + *subsequent arguments*]? T3. What makes it happen to [*verbal trigger* + *subsequent arguments*]? T4. What makes it happen that [*event clause*]?
	P3. 	E3	T5. What is a result of [*nominal trigger* + *subsequent arguments*]? T6. What happens when [*event clause*]?
Entity coreference	P4. 	En2	T1. What [*verbal event trigger* + *subsequent arguments*]?
Paraphrase	P5. 	En1	

Examples (Pattern: P3)
- Question: "What is a result of the fibroblast division not only in the artificial conditions of cell culture, but also in an animal's body?"
- Answer: "Proliferation of fibroblasts"

Question Generation System 2 (QG2)

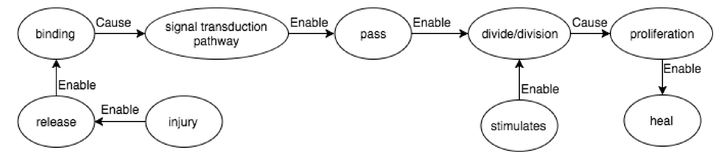
1. Extract generic patterns based on the relations between events and entities from a set of passages.
2. Apply the patterns to unseen passages and generate questions.

Question pattern	Question template
Entity *Result* Trigger	What event should occur before the *Trigger* of *Entity*?
Trigger_1 *Super* Trigger_2	What would happen without *Trigger_1* in *Trigger_2*?
Entity *Theme* Trigger	What would happen without the *Trigger* of *Entity*?
Entity *Location* Trigger	Where was *Trigger* in *Entity*?
Trigger_1 *Cause* Trigger_2	What is caused by *Trigger_1*?

- *Enable* relations tend to produce questions that have answers
- *Theme* relations tend to produce questions that may or may not have answers due to their ambiguity

Examples
- Question: "What event enables division?", Answer: "PDGF."
- Question: "What happens because of PDFG release?", (No answer)

Distractor Generation



Examples (against the example question by QG1)
- Distractor 1: "binding of PDGF molecules to receptor tyrosine kinases"
- Distractor 2: "PDGF stimulates fibroblast"
- Distractor 3: "platelets release PDGF"

Results and Future Work

Evaluation metrics
- Grammatical correctness: 1 (best): no grammatical errors, 2: 1 or 2 grammatical errors, and 3 (worst): 3 or more grammatical errors.
- Answer existence: 1 (yes) and 2 (no)
- Answer correctness: 1 (correct), 2 (partially correct), and 3 (incorrect)
- Inference steps: the number of semantic relations (event or event coreferences, paraphrases, or negations) humans need to understand
- Distractor quality: 1 (worst): confusing due to the overlaps with the correct answer, 2: easily identified, and 3 (best): viable.

System	Grammatical correctness			Answer existence			Inference steps		
	Ann 1	Ann 2	Total	Ann 1	Ann 2	Total	Ann 1	Ann 2	Total
QG1	1.52	1.48	1.50	1.17	1.26	1.21	0.80	0.71	**0.76**
QG2	2.13	2.07	2.10	1.58	1.75	1.67	0.31	0.20	0.27
MH	1.42	1.25	**1.34**	1.20	1.14	**1.17**	0.13	0.19	0.16

Question generation

System	Ann 1	Ann 2	Total
QG1	1.35	1.57	1.46
MH	1.08	1.13	1.11

(a) **Answer correctness**

System	Ann 1	Ann 2	Total
QG1	1.98	1.90	1.94
MH	1.93	1.88	1.91

(b) **Distractor quality**

Observations
- QG1 often fails to earn inference steps because the answer could exist in the same sentence as the question event with implicit relations.
- Distractors are often labeled as 2 ("easily eliminated") because they come from events closely preceding or following the question event.

Future work
- Explore more domain-adaptable question generation strategies.
- Develop an end-to-end system which takes raw text as input.
- Devise more intelligent ways of generating distractors.
- A real user test with non-native English readers.
- Automatically carry out one or more of our evaluation processes.

Acknowledgements

This publication was partly made possible by grant NPRP-08-1337-1-243 from the Qatar National Research Fund (a member of the Qatar Foundation).