COMPUTATIONAL MODELS OF
HUMAN INTELLIGENCE

CMHI Report Number 4 1 April 2019

Automated Question Generation System for Genesis

Sayeri Lala

Automatic Question Generation systems automatically generate questions from input such as text. This study implements an Automated Question Generation system for Genesis, a program that analyzes text. The Automated Question Generation system for Genesis outputs a ranked list of questions over content Genesis does not understand. It does this using a Question Generation Module and Question Ranking module. The Question Generation Module determines what content Genesis does not understand and generates questions using rules. The Question Ranking Module ranks the questions by relevance. This Automated Question Generation system was evaluated on a story read by Genesis. The average question relevance among the top 10 generated questions was 2.41 on a scale of 1-3, with 3 being most relevant. 53.8Ranking Module. The results suggest that the Automated Question Generation system produces an optimally ranked list of relevant questions for Genesis.

Keywords: computational models of human intelligence, cognitive AI, story understanding, automated question generation, question ranking .

Automated Question Generation System for Genesis

Sayeri Lala

***Abstract*—Automatic Question Generation systems automatically generate questions from input such as text. This study implements an Automated Question Generation system for Genesis, a program that analyzes text. The Automated Question Generation system for Genesis outputs a ranked list of questions over content Genesis does not understand. It does this using a Question Generation Module and Question Ranking module. The Question Generation Module determines what content Genesis does not understand and generates questions using rules. The Question Ranking Module ranks the questions by relevance. This Automated Question Generation system was evaluated on a story read by Genesis. The average question relevance among the top 10 generated questions was 2.41 on a scale of 1-3, with 3 being most relevant. 53.8% of subjects ranked questions in the same order as the Question Ranking Module. The results suggest that the Automated Question Generation system produces an optimally ranked list of relevant questions for Genesis.**

***Index Terms*—Automated Question Generation, Question Ranking**

I. Introduction

QUESTION answering systems like Apple's Siri have grown increasingly popular as consumers rely more on personal voice assistants. Despite their widespread usage, question answering systems might still face problems when they lack sufficient information to intelligently answer the user's question. For example, if the user asks Siri, "What are good places to eat around here?", Siri returns with a list of top-rated nearby restaurants. Siri could return better results by clarifying what type of restaurant e.g., Chinese fast food, the user is looking for. How can machines be designed to generate questions?

Previous research on Automatic Question Generation (AQG) systems explores their applications in guiding reading and writing. The AQG system for Project LISTENS produces questions that guide elementary students as they read stories [8]. The AQG system for the academic writing support tool forms questions about technical concepts in the paper.

However, the questions generated by these AQGs could already be answered in the text and would not be useful for clarifying ambiguous information. To produce questions over ambiguous content, AQGs should be built on top of natural language processing systems. The natural language processing systems can be used to identify vague information and then generate appropriate questions.

In this study, we built an AQG system on top of Genesis, a text-based natural language processing system. The goal of this AQG system is to generate questions that resolve content not understood by Genesis.

II. Previous Work

The goal of this study is to implement an AQG system which produces a ranked list of questions over text read by Genesis. The questions should identify causal connections in the story that Genesis does not understand. This section reviews literature on current algorithms for AQG and question ranking systems.

A. AQG Systems

AQG systems first identify elements to question and then produce questions.

1) Identifying Source Sentences: Identifying source sentences is finding sentences to question [1]. Studies on reading comprehension strategies show that questions relating prior knowledge to new material are more effective than questions about a single sentence [5]. Questions that relate information across sentences are also more effective than questions about a single sentence [6]. Therefore, sources sentences used by the AQG system for Genesis link information to Genesis knowledge and link information across sentences in the story.

Some AQGs rely on heuristics to discover source sentences. For example, the RoboCHAIR system uses pattern-based selection to identify source sentences [1]. This solution defines a list of linguistic anchors such as key pronoun/verb pairs. One AQG system detects source sentences by identifying key phrases by the Lingo algorithm [3]. The Lingo algorithm identifies concepts within text documents and clusters phrases/words associated with the concept. The AQG system for Genesis uses similar heuristics for source detection.

2) Question Formulation: AQGs use question formulation rules to construct questions from source sentences. The RoboCHAIR system uses syntactic trees and rules to produce grammatically correct questions [1]. One AQG system avoids the complexity of syntactic trees by designing rules that only need to be filled in with the needed informatio [3]n. The needed information is stored in a conceptual graph structure. The node of the conceptual graph is the concept and the edge relations are parameters that the question generation rules are functions of. Another AQG system also generates questions for structured data (i.e., organized in tables) using only rules [7]. The AQG system for Genesis uses the latter approach.

B. Question Ranking

Several methods have been explored to rank the questions generated by AQG systems. Two solutions build machine learning based ranking models. The models are trained on question rankings collected from humans. The logistic regression model ranks questions by acceptability [2,4]. Acceptable questions are grammatically correct and clear. The RoboCHAIR system uses a decision-tree based question ranking model [1]. It ranks questions by acceptability and relevance. Relevance uses a 1-5 scale (1=completely irrelevant, 5=very relevant) and measures how important the question is in relation to the topic.

Since the Question Generation module for Genesis uses question formulation rules that do not depend on grammar, the questions are generally grammatically correct. Thus the Question Ranking module orders questions only by relevance.

III. System Architecture

This section describes the components involved in building an AQG system for Genesis. As shown in Fig. 1, after Genesis reads a story, it outputs an elaboration graph. The AQG System uses the elaboration graph to produce a ranked list of questions. The questions aim to help Genesis understand the causal relations in the story.

A. Genesis

The Question Generating (QG) and Ranking (QR) Modules are built on top of the Genesis (Fig. 2), an artificial intelligence program which models human comprehension. Genesis reads English text and analyzes it using common sense if-then rules and concepts. It then outputs an elaboration graph (Fig. 3), depicting its understanding of how events within the story are causally related.

B. Question Generation (QG) Module

This module produces questions in two stages: the first is source detection and the second is question formulation.

1) Source Detection: Source detection is identifying story elements to question. Sources are chosen so that Genesis can learn causal relations.

Source types:

Disconnected: A disconnected source is a story element with no connections in the elaboration graph (Fig. 3). Genesis is unable to find any causal connection between this and remaining story elements.
The module detects these elements by searching for nodes in the elaboration graph that lack parent and children nodes (Fig. 3).

Ambiguous: An ambiguous source is a story element with signal words *entails*. This signal word identifies vague causal relationships.

Surprising: A surprising contrast source is a pair of contrasting story elements *A* and *B*, whose contrast is unexplained to Genesis.
The module first obtains the contrasting pairs from Genesis. Genesis identifies contrasting pairs in the story via a list of contrasts specified in its set of concepts. The module then determines if the contrast is explained. The contrast is explained if *A* and *B* have different parents in the elaboration graph; otherwise, the contrast is unexplained.

Explicit, unknown causal relations: An explicit causal relation is a story element with the signal word because. An unknown, explicit causal relation source is an explicit causal relation that does not completely match a Genesis rule (Fig. 4).

Currently, this source type is partitioned into two cases:

- Case 1: **Unexpected Consequences**
  In this case, only the if-clause (Fig. 4) of the unknown explicit causal relation *X* matches the if-clause of some Genesis rule *Y*.

  The then-clause (Fig. 4) of source X is an unknown consequence of the if-clause of both source X and Genesis rule Y. This source type suggest edits or additions to Genesis rule *Y*.

- Case 2: **Completely Unknown Rule**
  In this case, the if clause does not match to any if-clause among the Genesis rules. This source type suggests a new Genesis rule.
  Note: this case was not supported during the evaluation process for the AQG system.

  The module identifies explicit unknown causal relations by first identifying explicit causal relations and then using Genesis matcher functions to determine if there is a complete match between the explicit causal relation and an existing Genesis rule.

C. Question Formulation

Questions are generated by the question formulation rules (Table 1). The rules are a function of the source type.

Source type (for some identified source X):

Disconnected: The QG module identifies A, the set of story elements that could be causally connected to X. A is found using a proximity search heuristic. These story elements immediately precede X, and have a common topic with X. If A is empty, the module asks how X impacts the story. Otherwise, for every element *a* in A it asks if X occurs because of element *a*.

Ambiguous: In this source type, it is not clear why one event A causes another event B. The generated question asks how A causes B.

Surprising Contrast: Since the contrast between story elements A and B is unexplained to Genesis, the generated question asks how both A and B can occur. The question is more specific if A and B have parents in the elaboration graph.

Explicit, unknown causal relations:

- **Unexpected consequences**
  Since the then-clause of X does not match the then-clause of Y, the generated question asks about the causal relationship between the then clauses. This question clarifies rule Y by asking whether the then clause of X causes the then clause of Y (or vice versa).

- **Completely Unknown Rule**
  Since the if-clause does not match the if-clause of existing Genesis rules, the QG module hypothesizes a new rule Z for Genesis by generalizing the content in the unknown explicit causal relation X via Genesis helper functions. The generated question asks if the hypothesized rule Z is correct.

D. Question Ranking (QR) Module

This module orders questions by relevance. Relevance measures how essential the question is in order for Genesis to understand causal relations in the story.

Question relevance is determined by the source type of element X, with 1 being least relevant to 4 being most relevant.

Relevance of Source type:

1) Disconnected
2) Unknown, explicit causal relations (i.e., Unexpected Consequences)
3) Ambiguous
4) Surprising Contrasts

Surprising contrasts indicate what content Genesis perceives as conflicts in logic in the story. Since Genesis does not understand how both elements in a contrasting pair can occur, questions about the surprising contrast are most useful for resolving these conflicts in logic.

Ambiguous sources indicate ambiguous causal relations. Genesis does not understand how A causes B though it knows that A leads to B. Questions clarifying these ambiguous causal relations are useful to Genesis. However, since ambiguous sources do not indicate conflicting logic, they are less relevant than surprising contrast sources.

Unknown, explicit causal relations indicate rules that are unknown to Genesis. Since Genesis understands explicit causal relations, questions about these sources are less relevant for understanding the story. They are useful for helping Genesis clarify existing causal connections in the story and learn new rules.

Disconnected elements indicate story elements that Genesis perceives as having no effect on the story. Since these elements do not contain conflicting logic or ambiguous information, questions about these sources are less relevant for understanding the story. Since questions about unknown, explicit causal relations allow Genesis to clarify existing causal relations in the story, questions on disconnected elements are less relevant compared to questions on unknown, explicit causal relations.

Currently, ranking questions over source elements of the same type is arbitrary.

IV. Evaluation

The goal of the AQG system is to generate questions that allow Genesis to understand the causal relationships in the story. The AQG system was evaluated according to two criteria:

1) Relevance of generated questions
2) Ranking of generated questions

Relevance is defined as how useful the question is for understanding the causal relations in the story.

The AQG system should produce questions that are highly relevant for understanding the story. It should also rank questions so that questions that are more relevant are ranked higher.

A. Materials

The AQG system was evaluated on a rendering of Macbeth read by Genesis. The story contains all source types: disconnected, ambiguous, surprising contrasts, and unexpected consequences. Genesis had a set of rules and a list of contrasts that represent common sense knowledge that human readers have.

B. Procedure

Human participants evaluated the relevance and ranking of the questions produced by the AQG system for the story read by Genesis. The subjects were students in college who can read, speak, and write in English with proficiency.

The subjects independently read the same Macbeth rendering that Genesis read. The subjects were not familiar with the plot of Macbeth beforehand to ensure that their understanding of the story was based only on the Macbeth rendering.

Afterwards, subjects filled out a survey asking them to:

- evaluate the relevance of the top 10 questions produced by the AQG system
- rank subsets of questions relative to each other

Relevance was evaluated using a 1-3 Likert scale, with 1=irrelevant, 2= somewhat relevant, and 3=very relevant. Questions were ranked such that top ranked questions were more relevant than other questions. The questions were ordered randomly to reduce potential bias in assessing question relevance and ranking.

The relevance scores were averaged across the questions to evaluate the general relevance of the generated questions. The relevance scores were also averaged across questions of each source type to evaluate the general relevance of each source type.

The rankings collected on each subsets of questions was averaged. The average ranking on each subset was compared against the corresponding rankings produced by the AQG.

V. Results and Discussion

The AQG system produced 64 questions for the Macbeth rendering, with the distribution of question type displayed in Fig. 5.

Question Relevance

The average relevance scores across questions for each source type are displayed in Fig. 6. The average relevance score across the top 10 questions was 2.41, with standard deviation 0.73.

The top 10 questions produced by Genesis contained questions for the following source types: 3 ambiguous, 3 surprising contrasts, 2 unexpected consequences, and 2 disconnected. The most relevant questions were for the unexpected consequence and surprising contrast source types (i.e., relevance scores of 2.7 and 2.6 respectively). Questions for ambiguous sources had an average relevance score of 2.4. The least relevant questions were for the disconnected source type, with an average relevance of 2.2.

Since the lowest relevance score across the top 10 questions was 2.2, and the average relevance score was 2.41, this suggests that the questions produced by the AQG system are relevant.

Question Ranking

Let A >B indicate that A is more relevant than B.

Subset 1 contained 3 questions of different source types: Ambiguous, Surprising Contrast, and Unexpected Consequence. The distribution in the human rankings across questions in subset 1 is displayed in Fig 7.

53.8% of subjects ranked Surprising contrast >Ambiguous >Unexpected Consequence. This ranking matches the current ranking algorithm implemented in the Question Ranking module.

The result suggests that:

- Questions resolving conflicting information (e.g., surprising contrasts) are most important to understanding a story. Conflicting information in the story indicates either gaps in logic in the story or in the readers knowledge. In either case, the gap in logic must be resolved for the reader to understand the causal relations in a story.

- Questions resolving ambiguous causal relations are less important than questions resolving conflicting information.

  Since this question type clarifies vague causal relations, they are important but not as important as questions resolving conflicting information.

- Questions over unexpected consequences are less important than the above question types.

  These questions clarify true but imprecise causal relations. The causal relations are not as vague as the causal relations for ambiguous sources however.

  For example, one such question produced by the AQG system was:

  Does Lady Macbeth persuade Macbeth to want to become king because Lady Macbeth wants to become queen?

  The story explicitly states, “Lady Macbeth persuades Macbeth to want to become king because Lady Macbeth is greedy, which is a true but less precise causal relation compared to the one hypothesized by the question. Since the causal relation is true, the reader does not learn as much information from this question type as they would from the above question types.

Subset 2 contained 2 questions of different source types: Ambiguous, and Unexpected Consequence. The distribution in the human rankings across questions in subset 2 is displayed in Fig 8.

85.7% of subjects ranked Ambiguous >Unexpected Consequence. This result matches the current ranking algorithm implemented in the Question Ranking Module. The result indicates that questions clarifying vague causal relations (i.e., ambiguous sources) are more useful than questions clarifying less vague causal relations (i.e., unexpected consequence sources).

However, the ranking scheme used by the Question Ranking module is not consistent with the average relevance scores for questions of each source type. The ranking algorithm ranks S >A >U >D. Questions for ambiguous sources have lower relevance score (2.4) compared to questions for unexpected consequences (2.7). Also, questions for surprising contrasts and unexpected consequences have nearly the same relevance score. Disconnected sources have the lowest relevance score (2.2) which is consistent with the ranking algorithm. Since only a small set of questions were ranked relative to one another, more results need to be collected to evaluate the potential discrepancy between relevance scores and ranks.

VI. Conclusions and Further Work

The AQG system built for Genesis uses the elaboration graph to inquire about causalities that Genesis does not understand. The system uses rules to generate questions from the source element. It ranks them according to the type of the source element.

The results suggest that the questions produced by the AQG system are relevant. The ranking scheme used by the AQG system seems somewhat consistent with the rankings produced by humans. Further studies will be done to evaluate the ranking scheme.

Several directions could be explored to expand and evaluate the question generation module.

There might be additional types of source elements in the elaboration graph worth inquiring. For example, a story element with an outgoing connection but no incoming connection might constitute a given or unexplained condition in the story. Examples of given information include a character's attributes e.g., "Lady Macbeth is greedy" and goals e.g., "The student wanted to get an A on the test". Characterizing the kinds of given information in a story , such as a character's attributes or goals, might be the first step in formulating relevant questions via the question generation module.

Clarifying the given conditions could help Genesis discover concepts that generalize across various stories. Understanding why Lady Macbeth is greedy could lead Genesis to learn that people lust for power and status. Understanding why a student strives for high grades could help Genesis understand incentives for success in education.

Another possible direction to explore includes generating and ranking new questions in response to answers to previous questions. This would require implementing a teacher interface, via which Genesis presents its questions and acquires answers. It would also require a learning module enabling Genesis to learn rules and concepts from the answers. The AQG system will have to support generating and ranking new types of questions that facilitate Genesis learning.

Other extensions include enabling the QG module to use information Genesis learned across different stories. This requires implementing tools enabling Genesis to efficiently search for information from previously read stories.

Future work for the QR module includes ranking questions derived from sources of the same type. One solution is to use machine learned based ranking models. This requires determining a feature set characterizing each source element and question, and collecting relevance labels from humans.

Revising the evaluation procedure could be helpful to understanding mechanisms for question generation in humans. In this study, humans assessed the relevance of the questions generated by Genesis. However, it could be also be worth investigating the types of questions humans might ask given the same story and base knowledge. Such observations could inspire question generation methods for Genesis that better model how humans raise questions.

Acknowledgment

The author is very grateful to Professor Patrick Winston for all his mentorship on this work and for the sponsorship provided by the MIT SuperUROP program.

Based on a paper submitted under the MIT EECS SuperUROP program, May 2016, revised April 2019.

References

[1] S. Pollak, B. Lesjak, J. Kranjc, V. Podpecan, M. Znidarsic, N. Lavrac. "RoboCHAIR: Creative assistant for question generation and ranking." *2015 IEEE Symposium Series on Computational Intelligence*. (2015): 1468-75. Print.

[2] M. Heilman, N. Smith. "Good Question! Statistical Ranking for Question Generation." *NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*. (2010): 609-617. Print.

[3] M. Liu, R. A. Calvo, A. Aditomo, L.A. Pizzato. “Using Wikipedia and Conceptual Graph Structures to Generate Questions for Academic Writing Support.” *IEEE Transactions on Learning Technologies*. 5.3.(2012): 251-263. Print.

[4] M. Heilman, N. Smith. “Automatic Question Generation System.” *2014 International Conference on Recent Trends in Information Technology, ICRTIT 2014*. (2014): Print.

[5] B. Davey, S. McBride. “Effects of Question-Generation Training on Reading Comprehension.” *Journal of Education Psychology*. 78.4.(1986): 256-262.Print.

[6] B. Davey, S. McBridge. “Generating Self-Questions after Reading: A Comprehension Assist for Elementary Students.” *The Journal of Educational Research*. 80.1.(1986): 43-46. Print.

[7] A. Shirude, S. Totala, S.Nikhar.,Dr V. Attar, J. Ramanand. “Automated Question Generation Tool for Structured Data.” *2014 International Conference on Recent Trends in Information Technology, ICRTIT 2014*. (2014): Print.

[8] J. Mostow, W. Chen. “Generating Instruction Automatically for the Reading Strategy of Self-Questioning.” *14th International Conference on Artificial Intelligence in Education*. (2009): Print.

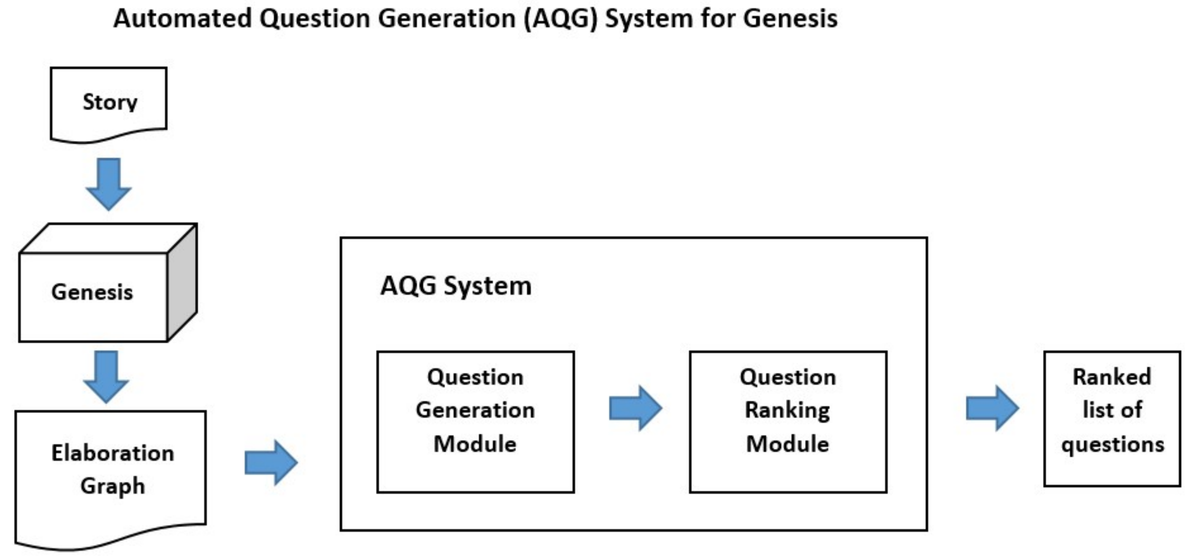


Figure 1: The AQG system produces a ranked list of questions over a story analyzed by Genesis

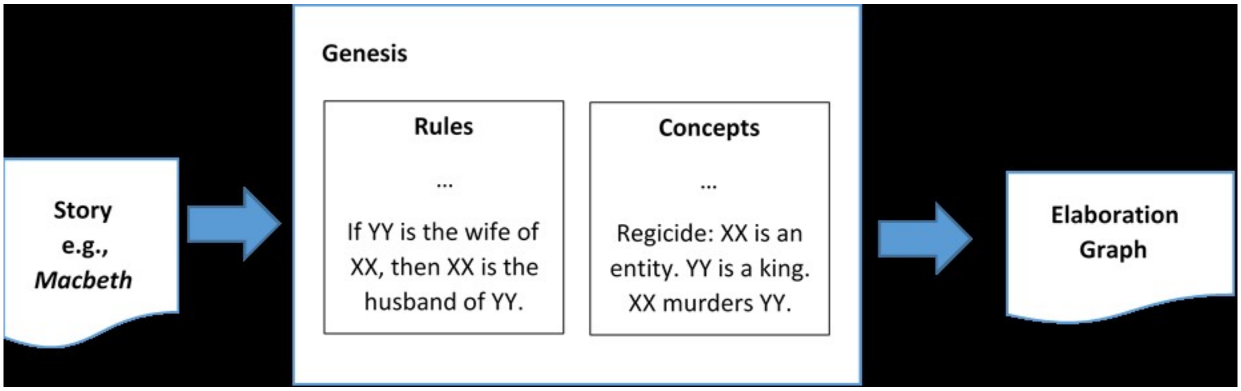

Genesis
Rules
...
If YY is the wife of XX, then XX is the husband of YY.
Concepts
...
Regicide: XX is an entity. YY is a king. XX murders YY.
Story e.g., *Macbeth*
Elaboration Graph

Figure 2: Genesis analyzes a story using rules and concepts and produces an Elaboration Graph.

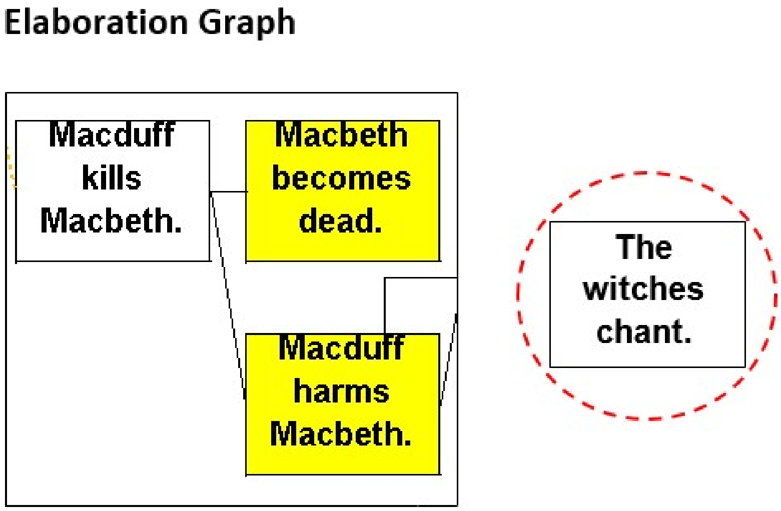
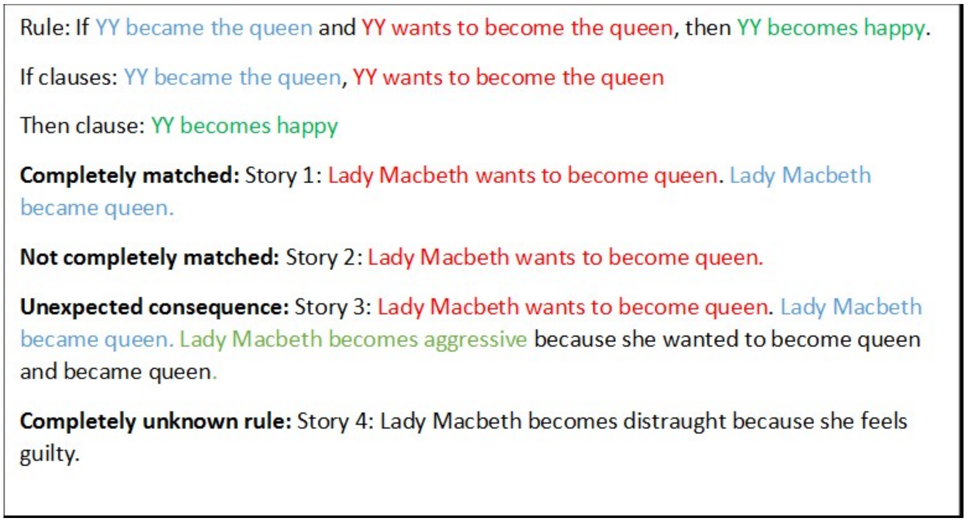


Figure 3: A portion of the Elaboration Graph on Macbeth. The story elements in the black outlined box are connected (each has 1+ parent and/or 1+ child node). The story element in the red outlined circle is disconnected (it has no parent and no children nodes). Yellow elements are Genesis predictions and white elements are explicit in the story.



Rule: If YY became the queen and YY wants to become the queen, then YY becomes happy.

If clauses: YY became the queen, YY wants to become the queen

Then clause: YY becomes happy

Completely matched: Story 1: Lady Macbeth wants to become queen. Lady Macbeth became queen.

Not completely matched: Story 2: Lady Macbeth wants to become queen.

Unexpected consequence: Story 3: Lady Macbeth wants to become queen. Lady Macbeth became queen. Lady Macbeth becomes aggressive because she wanted to become queen and became queen.

Completely unknown rule: Story 4: Lady Macbeth becomes distraught because she feels guilty.

Figure 4: A Genesis rule consists of *if* and *then* clauses. A completely matched rule occurs when the story element matches all the *if* clauses. An Unexpected Consequence source has the same if-clause but a different then-clause compared to a Genesis rule. A Completely Unknown rule is one with an unknown if-clause.

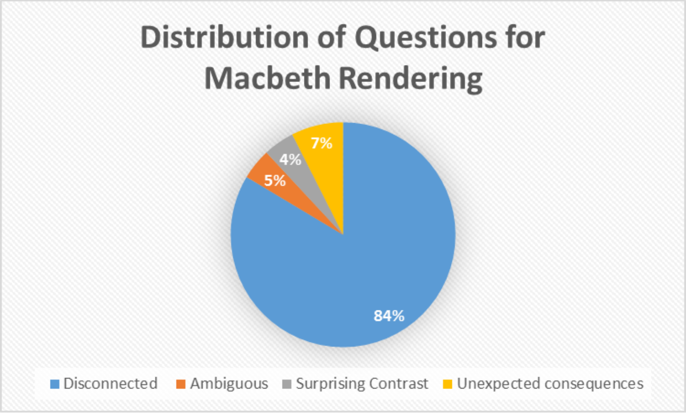


Figure 5: The distribution in source types for questions produced by the AQG System over the a Macbeth rendering.

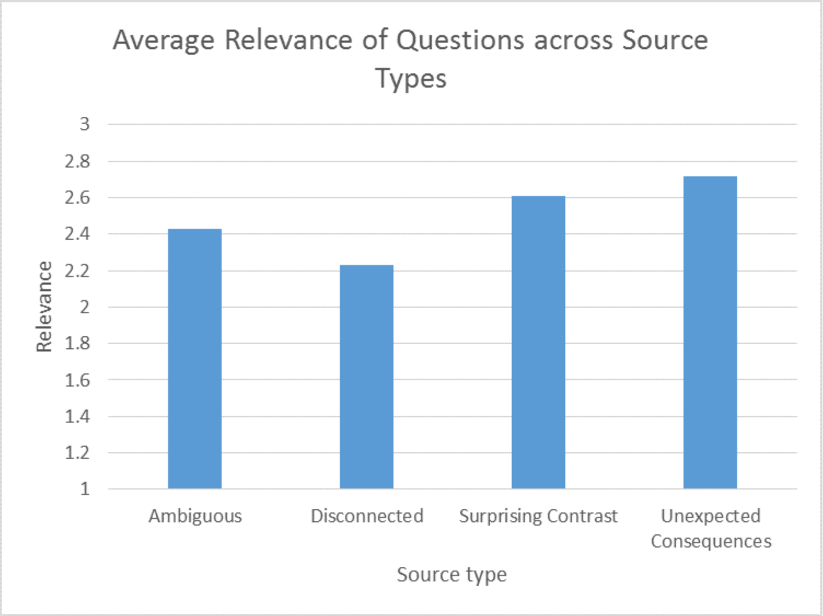


Figure 6: The average relevance scores for questions produced from each source type. Relevance was evaluated using the scale: 1= irrelevant, 2= somewhat relevant, 3=very relevant.

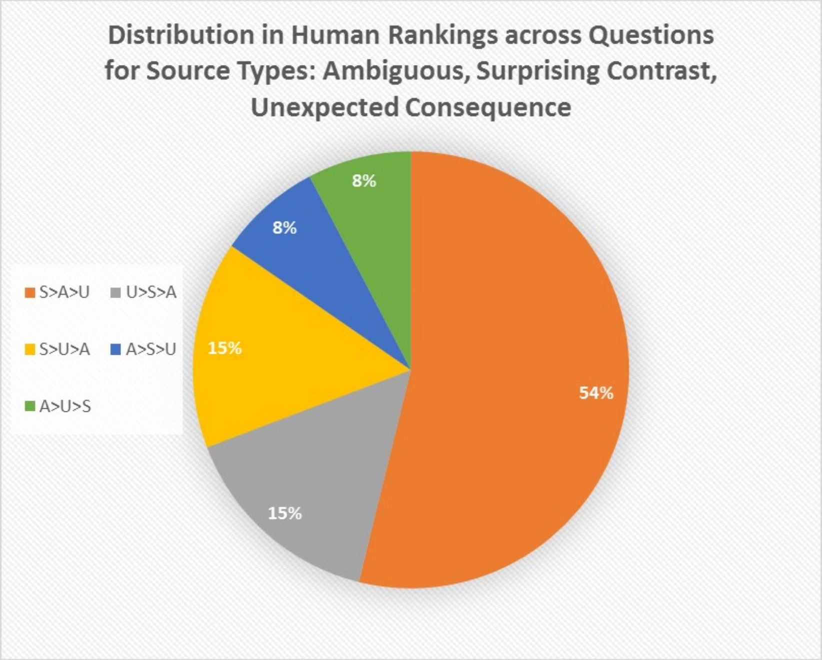


Figure 7: The relative question ranks given by humans to subset 1 of questions. This subset contained 3 questions produced from different sources: ambiguous (A), surprising contrast (S), and unexpected consequence (U).

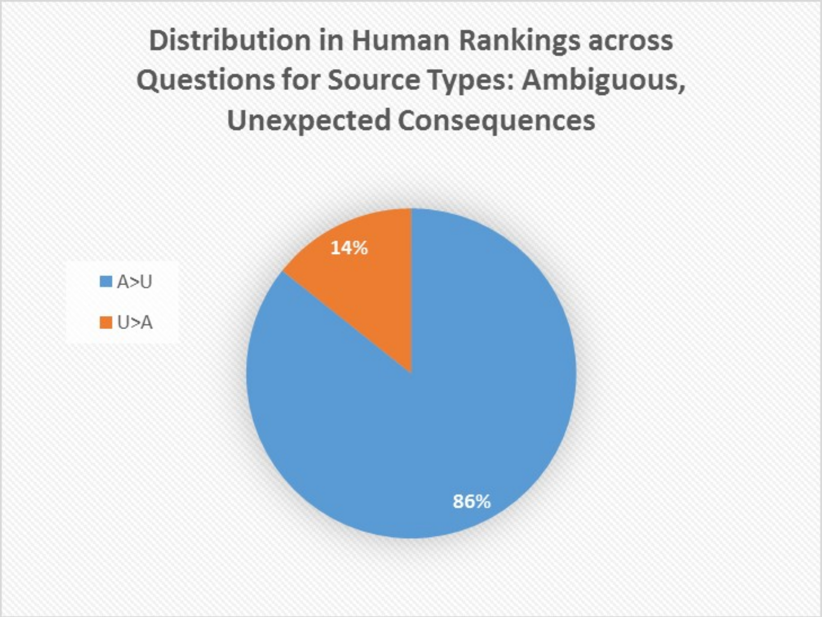
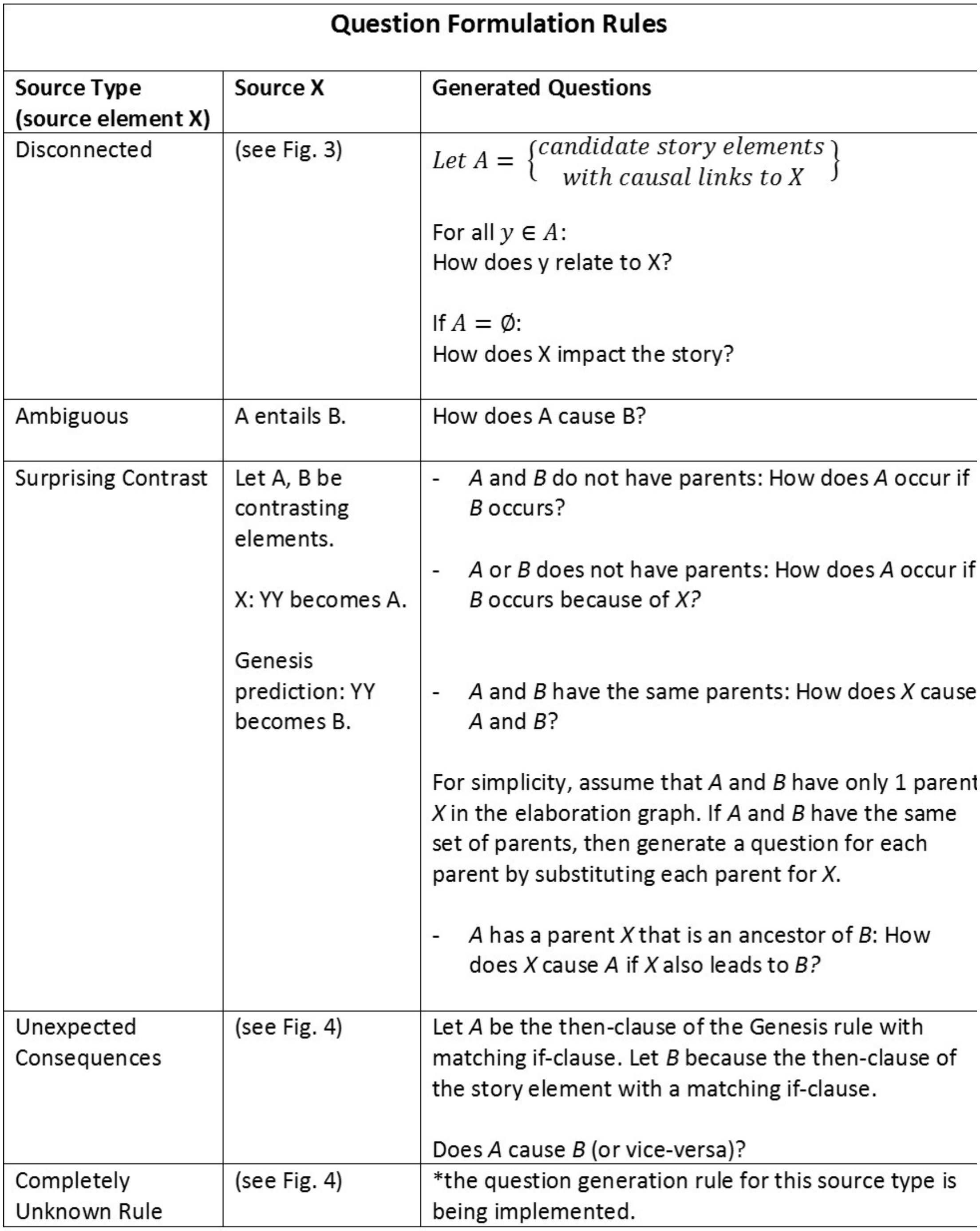


Figure 8: The relative question ranks given by humans to subset 2 of questions. This subset contained 2 questions produced from different sources: ambiguous (A) and unexpected consequence (U).



Question Formulation Rules

Source Type (source element X)	Source X	Generated Questions
Disconnected	(see Fig. 3)	$Let\ A = \left\{\begin{matrix} candidate\ story\ elements \\ with\ causal\ links\ to\ X \end{matrix}\right\}$ For all $y \in A$: How does y relate to X? If $A = \emptyset$: How does X impact the story?
Ambiguous	A entails B.	How does A cause B?
Surprising Contrast	Let A, B be contrasting elements. X: YY becomes A. Genesis prediction: YY becomes B.	- *A* and *B* do not have parents: How does *A* occur if *B* occurs? - *A* or *B* does not have parents: How does *A* occur if *B* occurs because of *X?* - *A* and *B* have the same parents: How does *X* cause *A* and *B*? For simplicity, assume that *A* and *B* have only 1 parent *X* in the elaboration graph. If *A* and *B* have the same set of parents, then generate a question for each parent by substituting each parent for *X*. - *A* has a parent *X* that is an ancestor of *B*: How does *X* cause *A* if *X* also leads to *B?*
Unexpected Consequences	(see Fig. 4)	Let *A* be the then-clause of the Genesis rule with matching if-clause. Let *B* because the then-clause of the story element with a matching if-clause. Does *A* cause *B* (or vice-versa)?
Completely Unknown Rule	(see Fig. 4)	*the question generation rule for this source type is being implemented.

Figure 9: Table 1. The question formulation rules. Generated questions are function of source type.