Appendix I: More details on mechanical turk interface for the various steps of the question and answer-options creation process

Instructions You need to read a passage and construct questions about the information presented in the passage. For each question, you also need to provide the correct answer. The question should be answerable from the passage. Most importantly, only those questions are allowed whose answer cannot be determined from a single sentence Below are a few example passages and sample questions (and answers). Paragraph: Sent 1: A Republican bid to keep the government open past Friday includes no new money for the construction of a border wall along the U.S.-Mexico border, according to several congressional aides familiar with ongoing talks. Sent 2: GOP leaders submitted the new offer Tuesday afternoon in an effort to appease Democrats, whose votes are needed to avert a shutdown of federal agencies, several House and Senate aides said. Sent 3: However, Republicans also insisted on increases in border security and defense spending, including money to repair existing fencing and new surveillance technology to patrol the nearly 2,000-mile border. Sent 4: Democrats have indicated that they would support such appropriations so long as no money goes toward an actual wall. Here are a few example questions with our criteria: Question: Who has to be appeased to keep the government open? Correct Answer: "Democrats" Sentences needed to answer this question: ☑ 1 ☑ 2 □ 3 □ 4 Explanation: This is multi-sentence question because, "appease Democrats" is mentioned in Sent 2, while "keep the government open" is part of Sent 1. Question: When was the republican bid submitted? Correct Answer: "Tuesday afternoon" Sentences needed to answer this question: ☑ 1 ☑ 2 □ 3 □ 4 Explanation: Republican bid is mentioned in Sent 1 while Tuesday afternoon occurs in Sent 2. Question: Who does not want a border wall? Correct Answer: "Democrats" Sentences needed to answer this question: ☑ 1 ☑ 2 □ 3 □ 4 Explanation: Democrats will be appeased by the new offer (Sent 2). The new offer includes no money for the wall (Sent 1). This implies that Democrats don't want the wall. Here is another example paragraph: Paragraph: Sent 1: Obama was born on August 4, 1961, at Kapi'olani Maternity & Gynecological Hospital in Honolulu, Hawaii. Sent 2: He is the only President to have been born in Hawaii. Sent 3: He was born to a white mother and a black father. Sent 4: His mother, Ann Dunham (1942-1995), was born in Wichita, Kansas, of mostly English descent, with some German, Irish, Scottish, Swiss, and Welsh ancestry. Question: How old was Obama's mother when he was born? Correct Answer: "19" Sentences needed to answer this question: ☑ 1 🗌 2 🗌 3 ☑ 4 Explanation: Obama was born in 1961. His mother was born in 1942. This question reasoning about time across Sentences 1 and 4 (1961-1942=19) Question: Where was Ann living in August, 1961? Correct Answer: "Hawaii" Sentences needed to answer this question: Explanation: Ann is Obama's mother (Sent 4). She must have been there when Obama was born (Sent 1). Obama was born in Hawaii (Sent 2). Question: Was one of Ann's children a future president? Correct Answer: "Yes' □ 1 🗹 2 🗔 3 🔽 4 Explanation: Ann is Obama's mother (Sent 4). Obama was a future president (Sent 2). Exmple of a bad question: Paragraph: Sent 1: Safti succumbs to Edwina 's charms and falls hopelessly in love with her. Sent 2: Lord Esketh becomes aware of this but Safti saves him from a man-eating tiger during a safari. Question: What is the name of the man who succumbs to Edwina's charms and saves her husband from a man-eating tiger during a safari? Correct Answer: "Safti" Explanation: This is a BAD question because even though it is constructed using phrases from more than one sentence ("succumbs to Edwina's charms" and "a maneating tiger during a safari") but the answer can be guessed by reading ONLY ONE of the two sentences. • For the paragraph below, create at least 3 questions that require multiple sentences (examples above). If you create more than 3 questions, you'd get bonus rewards

· For each generated questions you write the answer in the space followed by the question, as well as denoting what sentences are required for answering it. · Items with * are required.

ent 1: Earth processes	and and abandod another			
	appen now is the same way things h	appened in the past.		
	and mountains slowly wear away.			
	s is at work the same as it was billio	ns of years ago.		
	nt changes, living creatures adapt.			
ent 6: They change over				
	may not be able to adapt.			
ent 8: They become ex				
0	means they die out completely.			
	s study the history of the Earth.			
ent 11: They want to le				
ent 12: They use clues				
	clues to make sense of events.	1		
	ace things in the order they happen to know how long it took for those e			
and 13. They also want	o know now long it took for those e	venta to happen.		
ill out a valid que	tion			
uestion: *				
prrect Answer: *				

Figure 6: Instructions and Interface for the Step 1 of the dataset creation process. In this step, we request turkers to write a question and its correct answer for a given paragraph. We emphasize the requirement for multi-sentence questions, and additionally ask them to indicate which sentence ids would be needed for answering the composed question. We also provide two example paragraphs with 3 multi-sentence questions each and the corresponding answers. We also provide an explanation indicating why these questions fit our criteria. Lastly, we provide an example of bad question before asking them to compose questions. We ask them to write 3-5 questions per paragraph but the figure only shows the form for one question.

will be shown a sentend	e and a question. For each question,			
When saying Yes/No		e provided in the sentence. Only u	the question. If the answer is yes, you have use the information given in the sentence.	e to write the correct answ
	submitted the new offer Tuesday after		crats, whose votes are needed to avert a sh	utdown of federal
Question: Who has to b	appeased to keep the government c	pen?		
Can the above question	be answered using only the information	n provided in the given sentence?		
	rovided in the sentence is enough to			
Answer:	ovided in the sentence is not enough	to answer the question.		
the Democrats				
ancestry.	nn Dunham (1942–1995), was born in Dbama's mother when he was born?	Wichita, Kansas, of mostly English	descent, with some German, Irish, Scottish	, Swiss, and Welsh
Can the above question	be answered using only the information	n provided in the given sentence?		
	rovided in the sentence is enough to ovided in the sentence is not enough	•		
Type the answer here				
•	ation other than that mentioned in the		Obama's mother. It also doesn't say when h r the question. Note that since the response	
nce: , the city's populat	on had grown to an impressive 100,0	00.		
tion: When had Rome's	population grown to 100,000?			
	swered using only the information pr	0		
	ed in the sentence is enough to answ	er the question.		

Figure 7: Instructions and Interface for the Step 2 of the dataset creation process. This step was included to ensure that the questions collected from the previous step are indeed multi-sentence. In this step we show a sentence and a question to the turkers and ask them if they can answer the question using only the information present in the shown sentence. In the instructions, we discourage them from using background knowledge since some of our paragraphs are from non-fictional sources. We also provide examples along with an explanation. The bottom part of the figure shows the form for one sentence-question pair.

Instructions

Writing Answer Options for a Reading Comprehension Task

You are designing a reading comprehension test for your students. Students will read a paragraph; a question based on the paragraph; and a few options for its answer. They have to choose the correct answer from the options. For the test to make sense, you want the correct answer to be part of the options.

However, you don't want the test to be too easy! To make the test difficult for the students, you want to add incorrect options that might be confused with the correct answer to the list of options. A student who tries to answer a question without reading the paragraph carefully might choose the incorrect option! In this task, your goal is to write the list of options that will be shown to your students.

The questions showed to you in this HIT were written by people and might sometimes contain typing mistakes or other kinds of errors. In such cases, you could additionally check the box underneath the question indicating the "question is bad", and also correct it if possible. The answer-options you write would be in response to the corrected questions.

Below we show some good and bad examples of correct and incorrect options. We also show examples of bad questions

Correct Options: A correct option is one that correctly and completely answers the question. Keep in mind that for any question, there can be multiple correct answers. A bad correct option addresses the question only partially or incorrectly. We want to include only good correct options in the list of options. Here are some good and bad examples of correct options:

Paragraph:

Sent 1: Obama was born on August 4, 1961, at Kapi olani Maternity & Gynecological Hospital in Honolulu, Hawaii,

Sent 2: He is the only President to have been born in Hawaii. Sent 3: He was born to a white mother and a black father.

Sent 4: His mother, Ann Dunham (1942-1995), was born in Wichita, Kansas, of mostly English descent, with some German, Irish, Scottish, Swiss, and Welsh ancestry.

Question: How old was Obama's mother when he was born?

Good Correct options (separated via "//"): 19 // almost twenty

Bad Correct option: 20 Bad Correct option: older than twenty

Explanation: Obama was born in 1961. His mother was born in 1942 and 1961-1942=19. The good options ('19' and 'almost twenty') answer the question correctly.

The bad options give an incorrect answer

Question: Where was Ann living in August, 1961?

Good Correct options (separated via "//"): Honolulu // Hawaii

Bad Correct option: Kansas

Bad Correct option: Scotland

Explanation: Ann is Obama's mother (Sent 4). She must have been there when Obama was born (Sent 1). Obama was born in Hawaii (Sent 2). Hawaii is also in Honolulu. So both Hawaii and Honolulu answer the question correctly. Kansas and Scotland don't answer the question correctly.

Incorrect Options: An incorrect option will be closely related to the content of the paragraph and/or the guestion but will not truthfully answer the guestion. They should not be trivial, in the sense that they should not be easily eliminated as the answer without reading the paragraph. They are presented as options to confuse your students and to make the reading comprehension task challenging for them. Here are some good and bad examples of incorrect options:

Paragraph:

- Sent 1: It was hot that day.
- Sent 2: The temperature on the wall of the backyard was showing something well over 100 F.
- Sent 3: Meanwhile Tom, at home, was trying finish the remainder of carrots from last night, and packing for his trip to Chicago tomorrow. Sent 4: As employees of the Art Museum, Tom and his older cousin often had to travel to Chicago.
- Question: What was the temperature outside, when Tom was eating carrots?
- Good Incorrect options (separated via "//"): Not very hot // Far below 100 F
- Bad Incorrect option: Carrots Bad Incorrect option: Wall

Explanation: The good options appear in the paragraph (at least partially). They are also related to the content of the question (temperature), but do not answer the guestion correctly. Your student will have to carefully read the paragraph to identify that these don't answer the guestion correctly. The bad options appear in the paragraph (e.g. Wall) and sometimes also in the question (e.g. carrots) but they are trivially incorrect. A student can easily determine that they don't answer the question correctly without even reading the paragraph.

Question: What did Tom not visit?

Good Incorrect options (separated via "//"): Tom's older cousin // Art Museum of Chicago

Bad Incorrect option: A flower Bad Incorrect option: Two Chitoz bags

Explanation: The good options are closely related to the content of the paragraph and question, but do not answer the question correctly. Your student will have to carefully read the paragraph to identify that these don't answer the question correctly. The bad options are not in the paragraph at all, and don't challenge the student's comprehension skills.

Bad questions: Sometimes a question could have grammatical errors, typing mistakes, etc. or might not make sense in the context of the paragraph (for instance, it might not be related to the content of the paragraph or not be answerable at all). In such cases, please mark the question as a bad question. If the question contains minor errors please correct it in the boxed space provided underneath the question, otherwise leave the boxed space empty. Here are some examples of bad questions.

Paragraph:

Sent 1: It was hot that day.

Sent 2: The temperature on the wall of the backyard was showing something well over 100 F.

Sent 3: Meanwhile Tom, at home, was trying finish the remainder of carrots from last night, and packing for his trip to Chicago tomorrow. Sent 4: As employees of the Art Museum, Tom and his older cousin often had to travel to Chicago.

Question: What was the temperature outside, when Pom was eating carrots?

Is this a bad question?

Corrected Question, if bad: What was the temperature outside, when Tom was eating carrots? Explanation: It is easy to identify that there is a typing mistake, and Pom should be Tom.

Question: When did Tom walk out of the party?

Is this a bad question?

Corrected Question, if bad

Explanation: The question doesn't make sense for this paragraph. The paragraph does not discuss Tom attending any party. So, this is a bad question and cannot be corrected.

For each of the following questions, write good correct and incorrect options that would be shown to your students. Additionally, if the question is bad please indicate that and correct it if possible:

 Paragraph: Sent 1: Convicted murderer Seth Baxter awakens chained to a table beneath a pendulum blade . Sent 2: A videotape informs him that crushing his hands between the presses will release him ; he does so , but the blade still swings down and violently cuts him in half , while someone watches through a hole in the wall . Sent 3: The scene cuts to Agent Peter Strahm , who kills Jeff Denlon in self-defense and is sealed in the sickroom . Sent 4: He finds a hidden passage with a tape recorder that warns him to stay in the sickroom , but ignores it . Sent 5: He is attacked by a pig-masked figure in the passage and awakens with his head sealed in a box slowly filling with water , which he survives by performing a tracheotomy using a pen . Sent 6: Outside the plant , Detective Mark Hoffman delivers Corbett Denlon to the police and claims they are the only survivors , and is shocked when Strahm is brought out alive as well . Sent 7: Jill Tuck is met by John Kramer 's attorney , who is administering his will . Sent 9: She opens it with a key hung around her neck and then leaves without disclosing its contents . Sent 9: She opens it with a key hung around her neck and then leaves without disclosing its contents . Sent 11: He is informed of the death of Agent Lindsey Perez while taking Strahm 's phone and goes to the hospital to meet Strahm , who says that Hoffman 's name was Perez's last words .
Question: Whose death was Hoffman informed of? Correct Answer(s) (separated via "//") Incorrect Answer(s) (separated via "//")

Figure 8: Instructions and Interface for the Step 3 of the dataset creation process. This step was included to collect correct and in-correct answer options for the questions in our dataset. In the instructions we define what we mean by correct and incorrect answer options– correct options answer the given question correctly and completely, whereas incorrect options do not answer the question but are non-trivial and cannot be easily eliminated as an incorrect answer without reading the passage. We also show examples of good and bad inputs for both along with explanations. The bottom part of the figure shows the form for entering correct and incorrect answer-options for one paragraph-question pair.

nstructions	
Answering a l	Reading Comprehension Task
	ding comprehension test. You will read a paragraph and some questions, and a few options for its answer. You will have to choose the correct answer(
rom the options.	
lote that sometimes, m	ore than one option can answer the question correctly and completely. In those cases, please select all such options.
lere is an example:	
Deveryon	
Paragraph:	
	August 4, 1961, at Kapi'olani Maternity & Gynecological Hospital in Honolulu, Hawaii.
	ent to have been born in Hawaii. He was born to a white mother and a black father. His mother, Ann Dunham (1942-1995), was born in Wichita, Kansas, scent, with some German, Irish, Scottish, Swiss, and Welsh ancestry.
or mostly English de	sonn, with some connun, non, cootion, cwiss, and moist anotary.
Question: How old	vas Obama's mother when he was born?
teenager	
in his 40s	
mid 20s	
almost twenty	
	was born in 1961. His mother was born in 1942 and 1961-1942=19. The good option (almost twenty) answers the question correctly. The bad answers
give an incorrect and	wer.
Question: Where w	s Ann living in August, 1961?
 Kapi'olani Mat 	
Honolulu, Haw	
Wichita	
England	
•	
Honolulu Explanation: Ann is	Obama's mother (Sent 4). She must have been there when Obama was born (Sent 1). Obama was born in Hawaii (Sent 2). Hawaii is also in Honolulu. So
Honolulu Explanation: Ann is both Hawaii and Ho	olulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Honolulu	
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C	holulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
 Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So 	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w	h, select all the correct answer(s): NN A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight th humanitarian group, w Sent 3: Two bombs lai	h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest ath Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group. Sent 3: Two bombs lat area, the group said in	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported.
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap. Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs laa area, the group said in Sent 4: "It was a mirac	h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest ath Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs lau area, the group said in Sent 4: It was a mirac Sent 5: Graham, who I	h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian nich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured, ' the statement added. tas called on the international community to take out Sudan's air assets and establish a no-fity zone in the region, said in a statement Thursday that he
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs laa area, the group said in Sent 4: "It was a mirac Sent 5: Graham, who i blamed Sudan's air foo Sent 6: At least four cl	bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. b h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured, * the statement added. las called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said.
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs la area, the group said in Sent 4: "It was a mirac Sent 5: Graham, who i blamed Sudan's air fol Sent 6: At least four cl Sent 7: "We are deep!"	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight H humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirat Sent 5: Graham, who lo blamed Sudan's air foi Sent 7: "We are deeply Sent 7: "We are deeply Sent 7: We are deeply Son of the famed Rev.	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. In select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest at Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. Ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injured, " the statement added. las called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Graham.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight 1 humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirara Sent 5: Graham, who I biamed Sudan's air for Sent 6: At least four cl Sent 7: "We are deeply son of the famed Rev. Sent 8: More than 78,(bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. b h, select all the correct answer(s): VN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured, " the statement added. las called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he cort the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Grahan Billy Graham. 00 people have field South Kordofan and Biue Nile states since August of last year after an armed rebelion took root, the United Nations reported.
2 Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, wa Sent 3: Two bombs laa area, the group said in Sent 4: "It was a mirac Sent 5: Graham, who i blamed Sudan's air fo Sent 6: At least four cl Sent 7: "We are deeply son of the famed Rev. Sent 9: The Sudanese	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirad Sent 5: Graham, who blamed Sudan's air fon Sent 6: At least four cl Sent 7: "We are deepl son of the famed Rev. Sent 8: More than 784, Sent 9: The Sudanesse raised concerns over or	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirad Sent 5: Graham, who blamed Sudan's air fon Sent 6: At least four cl Sent 6: At least four cl Sent 7: "We are deepl son of the famed Rev. Sent 8: More than 78,0, Sent 9: The Sudanesse raised concerns over co	Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly.
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirad Sent 5: Graham, who l blamed Sudan's air foo Sent 6: At least for dep Sent 7: "We are deeply son of the famed Rev. Sent 9: The Sudanesee raised concerns over or Sent 10: Decades of co Sent 10: Decades of co	 Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. A christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. Ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured, " the statement added. as called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Grahan 3illy Graham. 00 people have fled South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which hav willian casualties. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005.
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight th humanitarian group, w Sent 3: Two bombs lat area, the group said in Sent 4: "It was a mirat Sent 5: Graham, who blamed Sudan's air fot Sent 6: At least four d Sent 7: "We are deephy son of the famed Rev. Sent 8: More than 78, Sent 8: The Sudaness son of the famed Rev. Sent 8: The Sudaness arised concerns over of Sent 10: Decades of c Question: What did th	 Nolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest ath Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ick supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injured, " the statement added. le that no one was injured," the statement added. as called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said. concorred for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Graham. 00 people have fled South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which have vilian casualties. vili war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005.
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight th humanitarian group, w Sent 3: Two bombs lat area, the group said in Sent 4: "It was a mirat Sent 5: Graham, who blamed Sudan's air fot Sent 6: At least four d Sent 7: "We are deephy son of the famed Rev. Sent 8: More than 78, Sent 8: The Sudaness son of the famed Rev. Sent 8: The Sudaness arised concerns over of Sent 10: Decades of c Question: What did th	bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): VN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ide insuports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and Igniting grass fires across the a statement No injuries were reported. le that no one was injured," the statement added. uas called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he co for the strik. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Grahan 3illy Graham. 00 people have fled South Kordofan and Bile Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which hav Willian casualties. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group said in sent 3: Two bombs la area, the group said in Sent 3: Two bombs la area, the group said in Sent 4: "It was a mirara. Sent 5: Graham, who I blamed Sudan's air fol Sent 6: At least four cl Sent 7: We are deeply and the famed Rev. Sent 8: More than 78,(Sent 9: The Sudaness raised concerns over o Sent 10: Decades of c C Buestion: What did th school (Bad	bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): VN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured," the statement added. uas called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he co for the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Grahan 3019 Graham. 00 people have fled South Kordofan and Bile Nile states since August of last year after an armed rebellion took root, the United Nations reported. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005. uive as betwoy? (Is this a bad question? e.g. typos, or not making sense?) answer? e.g. typos, or not making sense?
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sont 2: At least eight t humanitarian group, w Sent 3: Two bombs lai area, the group said in Sent 4: "It was a mirat sent 4: "It was a mirat sent 5: Graham, who blamed Sudan's air for Sent 6: At least four d Sent 7: "We are deeph son of the famed Rev. Sent 8: More than 78, Sent 9: The Sudanese on of the famed Rev. Sent 8: More than 78, Sent 9: The Sudanese on of the famed Rev. Sent 10: Decades of c C Question: What did th	bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): VN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured, "the statement added. uas called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he correcter for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities," added Grahan Bily Graham. 00 people have fled South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which hav willian esualties. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005. uil war between? e.g. typos, or not making sense?) (Bad answer? e.g. typos,
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs la area, the group said in Sent 4: "It was a mirara Sent 5: Graham, who Ib biamed Sudan's air for Sent 6: At least four cl Sent 7: "We are deeply biamed Sudan's air for Sent 6: At least four cl Sent 7: Decades of Cuestion: What did th	<pre>bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured,* the statement added. usa called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities,* added Grahan Billy Graham. O0 people have fied South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which hav vilian casualties. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005. bombs destroy? (Is this a bad question? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not ma</pre>
Honolulu Explanation: Ann is both Hawaii and Ho both Hawaii and Ho the following paragrap Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t humanitarian group, w Sent 3: Two bombs lai area, the group said in blamed Sudan's air fo Sent 4: "It was a mirad Sent 5: Graham, who blamed Sudan's air for Sent 7: "We are deepl son of the famed Rev. Sent 8: More than 784," Sent 8: The Sudanese raised concerns over o Sent 10: Decades of c Command Cente Various churches C lairch (Bad people (Bad Atta's personalitii	bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): VN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest ath Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian tich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured," the statement added. as called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he concerned for the welfare and lives of the geople of South Kordofan and we condern the bombing of churches and Christian facilities," added Graham 3011 Graham. 00 people have fied South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. will are asualties. will are between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005. tombs destroy? (Is this a bad question? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?)
Honolulu Explanation: Ann is both Hawaii and Ho the following paragrap: Paragraph: Sent 1: (C bombing raid to hit So Sent 2: At least eight t mumaitarian group, w Sent 3: Two bombs la area, the group said in Sent 4: "It was a miraa Sent 5: Graham, who l blamed Sudan's air for Sent 6: At least four d Sent 7: "We are deeply son of the famed Rev. Sent 8: More than 784, Sent 9: The Sudanese arised concerns over o Sent 10: Decades of c C Guestion: What did th	<pre>bolulu answer the question correctly. Kansas and Scottland don't answer the question correctly. h, select all the correct answer(s): NN) A Christian evangelical group said Thursday that a Bible school backed by American evangelist Franklin Graham was destroyed in the latest th Kordofan, an oil-rich Sudanese province that borders the newly created independent country of South Sudan. ombs were dropped in the area Wednesday during the school's first day of classes, according to a statement by Samaritan's Purse, Graham's Christian ich supports the school. ded inside the compound located in the region's Nuba Mountains destroying two Heiban Bible College buildings and igniting grass fires across the a statement No injuries were reported. le that no one was injured,* the statement added. usa called on the international community to take out Sudan's air assets and establish a no-fly zone in the region, said in a statement Thursday that he ce for the strike. urches have been destroyed since August, the group said. concerned for the welfare and lives of the people of South Kordofan and we condemn the bombing of churches and Christian facilities,* added Grahan Billy Graham. O0 people have fied South Kordofan and Blue Nile states since August of last year after an armed rebellion took root, the United Nations reported. government is thought to have responded to the rebellion by conducting sustained air raids with the use of Russian-made Antonov bombers, which hav vilian casualties. will war between the north and south, costing as many as 2 million lives, formally ended with a U.Sbrokered peace treaty in 2005. bombs destroy? (Is this a bad question? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not making sense?) (Bad answer? e.g. typos, or not ma</pre>

Figure 9: Instructions and Interface for the Step 4 of the dataset creation process. In this step we ask the annotators to read the questions and answer-options, and indicate if they find any errors in them. We also ask them to select all answer-options that correctly answer the question (illustrated through examples). Any possibly discrepancies or errors were later reviewed and corrected manually. The bottom part of the figure shows a sample form for one question.

Appendix II: Examples from the corpus

Sent 1: Dirk Diggler was born as Steven Samuel Adams on April 15, 1961 outside of Saint Paul, Mini	asoto
Sent 1: Dirk Diggier was born as Steven Samuel Adams on April 15, 1961 outside of Saint Paul, Mini Sent 2: His parents were a construction worker and a boutique shop owner who attended church every	
	Sunday and
believed in God. Sent 3: Looking for a career as a male model, Diggler dropped out of school at age 16 and left home.	
Sent 4: He was discovered at a falafel stand by Jack Horner.	
Sent 5: Diggler met his friend, Reed Rothchild, through Horner in 1979 while working on a film.	
Question: How old was Dirk when he met his friend Reed?	
A)* 18 B) 16	C) 2
Reasoning needed: Temporal and mathematical reasoning	1 (1
One needs to identify that Dirk was born in 1961 (Sent 1) and he met Reed in 1979 (Sent 5). At that ti	me ne must nav
been 18 (=1979-1961) years old.	
Sent 1: The hijackers attacked at 9:28.	
Sent 2: While traveling 35,000 feet above eastern Ohio, United 93 suddenly dropped 700 feet.	
Sent 3: Eleven seconds into the descent, the FAA's air traffic control center in Cleveland received the	irst of two radio
transmissions from the aircraft.	
Sent 4: During the first broadcast, the captain or first officer could be heard declaring "Mayday" amid	the sounds of a
physical struggle in the cockpit.	
Sent 5: The second radio transmission, 35 seconds later, indicated that the fight was continuing.	
Sent 6: The captain or first officer could be heard shouting: "Hey get out of here-get out of here-get out	
Sent 7: On the morning of 9/11, there were only 37 passengers on United 93-33 in addition to the 4 hi	jackers.
Sent 8: This was below the norm for Tuesday mornings during the summer of 2001.	
Sent 9: But there is no evidence that the hijackers manipulated passenger levels or purchased addition	al seats to
facilitate their operation.	
Sent 10: The terrorists who hijacked three other commercial flights on 9/11 operated in five-man team	s.
Sent 11: They initiated their cockpit takeover within 30 minutes of takeoff.	
Sent 12: On Flight 93, however, the takeover took place 46 minutes after takeoff and there were only t	our hijackers.
Question: Which two factors were different between the three other hijacked planes and United 93?	
the day of the takeover	
A) The amount of time that passed before the takeover started	
B)* United 93 took longer and had less hijackers	
C) The airline operating the planes	
D) The weather and fuel used by the airplane	
E) The navigation system used by the planes	
Reasoning needed: Discourse relation (contrast)	
One needs to identify that the discourse marker however in Sent 12 indicates a contrast relation betwee	
the flights mentioned in Sent 10. Also, only in Sent 12 indicates that the number of hijackers were few	er than in the
contrasted other flights.	
Question: What was below average for this particular day?	
A) the number of passengers in the first class.	
B)* the number of passengers on board.	
C) the number of hijackers	
D) the amount of air traffic in the skies	
E) the temperature	
Reasoning needed: Event coreference	
One needs to identify that This in Sent 8 co-refers to (event of) number of passengers in Sent 7. Note the	that Sent 12
contains only four hijackers and understanding that only indicates a smaller number of entities than ex	pected (as in
previous question), might mislead a system into believing that (C) is the correct answer.	

Figure 10: Examples from our MultiRCcorpus. Each example shows relevant excerpts from a paragraph; multisentence question; and corresponding answer-options. The correct answer(s) is indicated by a *. Each example also contains a list of reasoning phenomenon needed to answer that question.

Sent 1: Suddenly, Amy screamed. Sent 2: I whirled around and threw up my arm just in time to knock Dvorov away. Sent 3: His eyes were glowing red, and he had bared a set of fangs that could probably take my hand off. Sent 8: On the floor was Dvorov's head, separated from his body by several feet and a growing puddle of dark blood. Sent 9: Nepthys stood over him with a blue sword in his hand. Sent 10: It flashed in the strobe where it wasn't streaked with Dvorov's blood. Question: Who cut off Dvarov's head? A)* Nepthys B) the wolf Reasoning needed: Implicit semantic role (agent) and causation From sentences 8-10, indicate that Nepthys stood over Dvorov's (dead) body with a sword which was streaked with Dvorov's blood. Therefore, one can infer that the blood on the sword must have been caused by the action of killing of Dvorov and that Nepthys is the implicit agent of this action. Question: What was the weapon used to separate Dvorov's head from his body? B)* a blue sword C)* a sword A) a hook Reasoning needed: Implicit semantic role (instrument) and causation From sentences 8-10, indicate that Nepthys stood over Dvorov's (dead) body with a sword which was streaked with Dvorov's blood. Therefore, one can infer that the blood on the sword must have been caused by the separation of Dvorov's head from his body and that the sword must have been the implicit instrument of this action.

Figure 11: More example questions and corresponding reasoning phenomena from our MultiRCcorpus.