

# STARC (Structured Annotations for Reading Comprehension)

## Guidelines for OneStopQA

Yevgeni Berzak, Jonathan Malamaud, Roger Levy

April 2019

### 1. Questions

Compose three questions for each paragraph.

- The questions should be answerable based on any of the three text difficulty versions.
- The questions should be challenging. Avoid simple string matching.
- The questions should be local to the paragraph: i.e. answerable based only on the text in the given paragraph, without knowledge of the content in preceding paragraphs.
- Correctly answering the question should rely on critical information from a textual region(s) called the "A span" (see further information in section 3).

#### Question types to avoid

- Questions that require numerical reasoning (this is not a math test).
- Questions that require substantial world knowledge.
- Questions that require the reader to speculate (e.g. guessing the likely position of person X on issue Y when it is not explicitly stated in the text).
- Questions that don't refer to specifics of the paragraph. For example, we shouldn't have questions like "What is true of the paragraph", or "What is true according to Dr. X" when most of the paragraph comprises citations of Dr. X.

#### Phrasing

- Question phrasing should not cater to an individual difficulty level of the paragraph. If the wording is the same in both you can use it, otherwise paraphrase or mix (see similar instructions regarding answers phrasing in section 2).
- Avoid using "According to the paragraph/article" in the question.
- Questions ending in "?" are preferred over statements ending with "..."

#### Format

- In questions ending with "...", the "..." should be space separated from the last word. Make sure google docs doesn't change the three dots into one symbol.

## 2. Answers

Answer	Span
(a) correct	A (Critical span)
(b) incorrect, reflects a misunderstanding of the A region.	A (Critical span)
(c) incorrect, pertains to another region of the paragraph	D (Distractor span)
(d) incorrect, and has no support in the paragraph	-

Table 1: Correspondence between answer types and spans

### General Instructions

- The correct answer should not be guessable without reading the paragraph.
- It should also not be possible to eliminate incorrect answers without reading the paragraph (e.g. based on general world knowledge or logical inference).
- While (c) and (d) should all be unambiguously incorrect, they should also make sense and be plausible answers to the question.
- (c) and in some cases (b) could be correct (in part or fully) as a fact but not correct as an answer to the question. It's also fine for (c) to be an incorrect fact as long as it has textual support in the paragraph.
- (d) may refer to general world knowledge.
- (a) should not have substantial portions that are verbatim from the paragraph to avoid correct answers via simple string matching.
- (b) & (c) can often be strengthened by using verbatim words / phrases from the paragraph.

### Answer length

- Strive to have the answers of a given question to be of comparable length. Overall, you may have a tendency to make (b) and (c) longer than (a) and (d); try balancing this out whenever possible by making longer (a) & (d) and shorter (b) and (c).

### Wording patterns

Acceptable patterns for the four possible answers to a given question, in order of preference:

1. All answers share similar wording.
2. Each answer has a distinct wording.
3. Two pairs of similarly worded answers.

4. A pair of similarly worded answers, and distinct wording for the other two answers.
5. Three answers with similar wording and a fourth outstanding answer.

For patterns 3-5, try balancing the answer types within the pattern across questions (e.g. which answer type is the outstanding answer in pattern 5).

Avoid patterns that may help the reader guessing the correct answer:

- (b) should not be the negation / opposite of (a).

### **Answer Phrasing**

- Similarly to question phrasing, answer phrasing should not cater to an individual difficulty level of the paragraph. If the wording is the same in both you can use it, otherwise paraphrase or mix.
- Answers should start with a capital letter. The only exception for this is when the question ends with "...". Answers should not have a period at the end.

## **3. Span Annotations**

- See Table 1 in section 2 for an overview on the relation between answers and spans.
- The A span should contain the essential information for answering the question. Without this information, it should not be possible to answer the question (although in some cases, some relevant information for the question may reside outside the A span).
- A spans do not have to conform to sentence boundaries (i.e. can be/include only a part of a sentence) and can be non continuous (i.e. we can have several A sub-spans for the same question).
- The D span should contain information based on which (c) is constructed.
- The A spans for the first two questions should not overlap.
- The third question should have the same A span as one (and only one) of the first two questions while being substantially different from the question that shares its A span.

### **Format**

- Mark each contiguous part of the A span as <A[Question number]></A[Question number]>. Typically this will be one span, but in some cases the span may be split into several parts.
- Mark the D region as <D[Question number]></D[Question number]>
- A and D region spans should not include the preceding nor following white spaces, but do include following punctuation e.g. "...bla. <A1>bla bla,</A1> bla...".
- The third question should start with "Q1:" or "Q2:" depending on which question's A span it reuses. Mark the A span for the third question as A3 and the (c) distractor span as D3.

### Constraints on A spans

- A1 should come before A2 in the paragraph.
- A3 should repeat A1 and A2 a roughly equal amount of times across the article.
- Whenever possible, the first sentence in the paragraph should not be part of an A region, unless it also includes later text regions. If the first sentence is split into two or more sentences in the elementary version, the second, third etc. sentences in the split can count as later text regions.

### Constraints on D spans

- The D region of a question can be anywhere outside the A region. There is NO requirement on questions with the same / overlapping A regions to have the same/overlapping D regions.

### Text Changes

- The minimum paragraph length is 3 sentences. If any of the three paragraph versions are shorter than 3 sentences, merge it with either the preceding or following paragraph (the shorter between the two).
- British spelling should be converted to American spelling.

## 4. Reviewing Checklist (first review round)

### 1) Check constraints on question, answers, and spans:

- Is the correct answer guessable without reading the text?
- Would you pick the correct answer based on the paragraph?
- Does the A span indeed capture the critical region for answering the question correctly?
- Does (b) represent a misunderstanding of the A span (and only that span)?
- Does (c) have support outside the A span, and is it marked correctly as the D span?

### 2) Check distractor quality:

- Do (b) and (c) answer the questions, and are they plausible answers?
- Is (d) plausible a-priori, but not supported in the text?

### 3) Check the wording of the question and each of the answers:

- Is it tailored only to advanced or elementary?

### 4) Review the paragraph formatting and proofread everything:

- Check for typos in the paragraph (in particular issues with quotation marks, apostrophes, and extraneous spacing).
- All words in British English spelling should be changed to American English spelling.
- All spans should be there and formatted correctly.
- Proofread the question and the answers.

