Belebele Answer Span: English Annotation Guidelines

You'll be provided with a context paragraph and a question related to that paragraph. Your task is to find the *answer span* from the given paragraph where either the exact answer or the evidence of the answer can be found. The span should be as **concise as possible.**

The original <u>Belebele</u> multi-choice answer has been hidden in your file. However, the examples provided below include the multi-choice answer simply to illustrate the differences in the nature of the original answers and the "answer span" - which may help explain some annotation guideline decisions.

Please follow the annotation guidelines below carefully.

- 1. Identify the span of text in which the answer to the question or evidence of the answer can be found (Answer Span column)
- 2. Mark the start of the span with \$! and the end of the span with !\$
- 3. The answer span needs to be as short as possible.
- 4. Finding a relevant span is more important than finding an exact answer.
- 5. Don't include non-relevant punctuation.
- 6. Include the articles "the", "a" and "an".
- 7. Include the preposition if the question is a When/ Where/ How question.
- 8. In cases of multiple instances of the answer string, choose the appropriate one only.
- 9. Mark problematic questions with BB or X.

Examples expanding on guidelines 3-9 are presented next.

3. The answer span needs to be as short as possible.

3 (a) Example of correct annotation:

Paragraph	Question	Multi-choice Answer	Answer Span
The American plan relied on launching coordinated attacks from three different directions. General John Cadwalder would launch a diversionary attack against the British garrison at Bordentown, in order to block off any reinforcements. General James Ewing would take 700 militia across the river at Trenton Ferry, seize the bridge over the Assunpink Creek and prevent any enemy troops from escaping.	Where was there a British garrison located?	Bordentown	The American plan relied on launching coordinated attacks from three different directions. General John Cadwalder would launch a diversionary attack against the British garrison \$!at Bordentown!\$, in order to block off any reinforcements. General James Ewing would take 700 militia across the river at Trenton Ferry, seize the bridge over the Assunpink Creek and prevent any enemy troops from escaping.

3 (b) Example of incorrect annotation: (Note: You don't need to add more context to the span)

Paragraph	Question	Multi-choice Answer	Answer Span
Twentieth century research has shown that there are two pools of genetic variation: hidden and expressed. Mutation adds new genetic variation, and selection removes it from the pool of expressed variation. Segregation and recombination shuffle variation back and forth between the two pools with each generation.	Which process is responsible for adding genetic variation?	Mutation	Twentieth century research has shown that there are two pools of genetic variation: hidden and expressed. \$!Mutation adds new genetic variation!\$, and selection removes it from the pool of expressed variation. Segregation and recombination shuffle variation back and forth between the two pools with each generation.

3 (b) In some cases, the answer can't be found in a concise form in the paragraph, thus requiring a longer span. Find the answer span where one can find the answer to the question, including the additional information that it may entail. Example of correct annotation where longer span is necessary:

Paragraph	Question	Multi-choice Answer	Answer Span
This is called a chemical's pH. You can make an indicator using red cabbage juice. The cabbage juice changes color depending on how acidic or basic (alkaline) the chemical is. The pH level is indicated by the amount of Hydrogen (the H in pH) ions in the tested chemical.	How is the pH level of a chemical measured?	The amount of Hydrogen ions in the chemical	This is called a chemical's pH. You can make an indicator using red cabbage juice. The cabbage juice changes color depending on how acidic or basic (alkaline) the chemical is. The pH level is indicated \$!by the amount of Hydrogen (the H in pH) ions in the tested chemical!\$.

4. Finding a relevant span is more important than finding an exact answer.

4 (a) In some cases, there is no exact answer in the paragraph, and therefore reasoning or deduction would be required to give a fully comprehensive answer. Yet keep in mind that the task is finding the span which points to *evidence for the answer*. In these cases, simply identify the span of text where an answer can be found. In the example below, the referent "this" can later be resolved by the reader by reading the earlier text in the passage (movement of men and materials).

Paragraph	Question	Multi- choice Answer	Answer Span
Using ships to transport goods is by far the most efficient way to move large amounts of people and goods across oceans. The job of navies has traditionally been to ensure that your country maintains the ability to move your people and goods, while at the same time, interfering with your enemy's ability to move his people and goods. One of the most noteworthy recent examples of this was the North Atlantic campaign of WWII. The Americans were trying to move men and materials across the Atlantic Ocean to help Britain. At the same time, the German navy, using mainly U-boats, was trying to stop this traffic. Had the Allies failed, Germany probably would have been able to conquer Britain as it had the rest of Europe.	What was the German navy trying to accomplish during WWII?	Preventing Britain from receiving people and goods	Using ships to transport goods is by far the most efficient way to move large amounts of people and goods across oceans. The job of navies has traditionally been to ensure that your country maintains the ability to move your people and goods, while at the same time, interfering with your enemy's ability to move his people and goods. One of the most noteworthy recent examples of this was the North Atlantic campaign of WWII. The Americans were trying to move men and materials across the Atlantic Ocean to help Britain. At the same time, the German navy, using mainly U-boats, was trying to \$!stop this traffic!\$. Had the Allies failed, Germany probably would have been able to conquer Britain as it had the rest of Europe.

5. Don't include non-relevant punctuation

5 (a) Final punctuation is usually not considered relevant

Paragraph	Question	Multi-choice Answer	Answer Span
Things did not go well for the Italians in North Africa almost from the start. Within a week of Italy's declaration of war on June 10, 1940, the British 11th Hussars had seized Fort Capuzzo in Libya. In an ambush east of Bardia, the British captured the Italian Tenth Army's Engineer-in-Chief, General Lastucci. On June 28, Marshal Italo Balbo, the Governor-General of Libya and apparent heir to Mussolini, was killed by friendly fire while landing in Tobruk.	Where was Italo Balbo killed?	Tobruk	Things did not go well for the Italians in North Africa almost from the start. Within a week of Italy's declaration of war on June 10, 1940, the British 11th Hussars had seized Fort Capuzzo in Libya. In an ambush east of Bardia, the British captured the Italian Tenth Army's Engineer-in-Chief, General Lastucci. On June 28, Marshal Italo Balbo, the Governor-General of Libya and apparent heir to Mussolini, was killed by friendly fire while landing \$\frac{\\$!}{\!!} \text{In Tobruk!}\$.

5 (b) Closing punctuation is more likely to be relevant. () " " [] The example below includes the closing bracket in the span.

Paragraph	Question	Multi-choice Answer	Answer Span
MRI is based on a physics phenomenon called nuclear magnetic resonance (NMR), which was discovered in the 1930s by Felix Bloch (working at Stanford University) and Edward Purcell (from Harvard University). In this resonance, magnetic field and radio waves cause atoms to give off tiny radio signals. In the year 1970, Raymond Damadian, a medical doctor and research scientist, discovered the basis for using magnetic resonance imaging as a tool for medical diagnosis. Four years later a patent was granted, which was the world's first patent issued in the field of MRI. In 1977, Dr. Damadian completed the construction of the first "whole-body" MRI scanner, which he called the "Indomitable".	Who discovered nuclear magnetic resonance?	Bloch and Purcell	MRI is based on a physics phenomenon called nuclear magnetic resonance (NMR), which was discovered in the 1930s by \$!Felix Bloch (working at Stanford University) and Edward Purcell (from Harvard University)!\$. In this resonance, magnetic field and radio waves cause atoms to give off tiny radio signals. In the year 1970, Raymond Damadian, a medical doctor and research scientist, discovered the basis for using magnetic resonance imaging as a tool for medical diagnosis. Four years later a patent was granted, which was the world's first patent issued in the field of MRI. In 1977, Dr. Damadian completed the construction of the first "whole-body" MRI scanner, which he called the "Indomitable".

6. In order to maintain consistency across annotators, the span will include the articles "the", "a" and "an".

6 (a) An example of the inclusion of the article "the" in a span:

Paragraph	Question	Multi-choice Answer	Answer Span
Golf is a game in which players use clubs to hit balls into holes. Eighteen holes are played during a regular round, with players usually starting on the first hole on the course and finishing on the eighteenth. The player who takes the fewest strokes, or swings of the club, to complete the course wins. The game is played on grass, and the grass around the hole is mown shorter and called the green.	On a golf course, where is the grass cut shorter?	On the green	Golf is a game in which players use clubs to hit balls into holes. Eighteen holes are played during a regular round, with players usually starting on the first hole on the course and finishing on the eighteenth. The player who takes the fewest strokes, or swings of the club, to complete the course wins. The game is played on grass, and the grass around the hole is mown shorter and called \$1the green!\$.

6 (b) Example of inclusion of the article "a" in a span:

Paragraph	Question	Multi-choice Answer	Answer Span
The atom can be considered to be one of the fundamental building blocks of all matter. Its a very complex entity which consists, according to a simplified Bohr model, of a central nucleus orbited by electrons, somewhat similar to planets orbiting the sunsee Figure 1.1. The nucleus consists of two particles - neutrons and protons. Protons have a positive electric charge while neutrons have no charge. The electrons have a negative electric charge.	The particles that orbit the nucleus have which type of charge?	Negative charge	The atom can be considered to be one of the fundamental building blocks of all matter. Its a very complex entity which consists, according to a simplified Bohr model, of a central nucleus orbited by electrons, somewhat similar to planets orbiting the sun - see Figure 1.1. The nucleus consists of two particles - neutrons and protons. Protons have a positive electric charge while neutrons have no charge. The electrons have \$!a negative electric charge!\$.

7. If the question is a When/ Where/ How question – then include the preposition or adverb that would normally form part of an appropriate answer.

7 (a) An example of how-question where the preposition should be included in the span.

Paragraph	Question	Multi-choice Answer	Answer Span
Subcultures bring together likeminded individuals who feel neglected by societal standards and allow them to develop a sense of identity. Subcultures can be distinctive because of the age, ethnicity, class, location, and/or gender of the members. The qualities that determine a subculture as distinct may be linguistic, aesthetic, religious, political, sexual, geographical, or a combination of factors. Members of a subculture often signal their membership through a distinctive and symbolic use of style, which includes fashions, mannerisms, and argot.	How do members of a particular subculture often signify their association with the group?	By using style as a form of symbolism	Subcultures bring together like-minded individuals who feel neglected by societal standards and allow them to develop a sense of identity. Subcultures can be distinctive because of the age, ethnicity, class, location, and/or gender of the members. The qualities that determine a subculture as distinct may be linguistic, aesthetic, religious, political, sexual, geographical, or a combination of factors. Members of a subculture often signal their membership \$!through a distinctive and symbolic use of style!\$, which includes fashions, mannerisms, and argot.

7 (b) An example of a when-question where the preposition should be included in the span.

Paragraph	Question	Multi-choice Answer	Answer Span
Gothic style peaked in the period between the 10th - 11th centuries and the 14th century. At the beginning dress was heavily influenced by the Byzantine culture in the east. However, due to the slow communication channels, styles in the west could lag behind by 25 to 30 year. Towards the end of the Middle Ages western Europe began to develop their own style. one of the biggest developments of the time as a result of the crusades people began to use buttons to fasten clothing.	When did western Europe stop relying heavily on influences and start developing its own style?	Around the end of the Middle Ages	Gothic style peaked in the period between the 10th - 11th centuries and the 14th century. At the beginning dress was heavily influenced by the Byzantine culture in the east. However, due to the slow communication channels, styles in the west could lag behind by 25 to 30 year. \$!towards the end of the Middle Ages!\$ western Europe began to develop their own style. one of the biggest developments of the time as a result of the crusades people began to use buttons to fasten clothing.

7 (c) An example of a how-question case where the preposition should be included in the span.

Paragraph	Question	Multi-choice Answer	Answer Span
The invention of spoke wheels made Assyrian chariots lighter, faster, and better prepared to outrun soldiers and other chariots. Arrows from their deadly crossbows could penetrate the armor of rival soldiers. About 1000 B.C., the Assyrians introduced the first cavalry. A cavalry is an army that fights on horseback. The saddle had not yet been invented, so the Assyrian cavalry fought on the bare backs of their horses.	How are battles that utilize a calvary fought?	on horseback	The invention of spoke wheels made Assyrian chariots lighter, faster, and better prepared to outrun soldiers and other chariots. Arrows from their deadly crossbows could penetrate the armor of rival soldiers. About 1000 B.C., the Assyrians introduced the first cavalry. A cavalry is an army that fights \$!on horseback!\$. The saddle had not yet been invented, so the Assyrian cavalry fought on the bare backs of their horses.

8. In some cases, the exact original answer string can be found multiple times in the paragraph. Choose the appropriate one only.

Paragraph	Question	Multi-choice Answer	Answer Span
Eighteen percent of Venezuelans are unemployed, and most of those who are employed work in the informal economy. Two thirds of Venezuelans who work do so in the service sector, nearly a quarter work in industry and a fifth work in agriculture. An important industry for Venezuelans is oil, where the country is a net exporter, even though only one percent work in the oil industry.	Which line of work mentioned in the passage employs the least number of Venezuelans?	Oil	Eighteen percent of Venezuelans are unemployed, and most of those who are employed work in the informal economy. Two thirds of Venezuelans who work do so in the service sector, nearly a quarter work in industry and a fifth work in agriculture. An important industry for Venezuelans is oil, where the country is a net exporter, even though only one percent work in the \$!oil!\$ industry.

- 9. The Belebele dataset was initially curated as a multiple-choice reading comprehension dataset. As a result, the questions were not crafted in a way that is suitable for extractive QA, and as such it may not be possible to find an exact answer or *any* answer:
- 9 (a) You may come across some answers that cannot be found simply in a span of text i.e. reasoning or deduction is needed through reading the rest of the passage. In these cases, mark the span where evidence for the answer can be found. Then mark it as BB to indicate that it's problematic due to the nature of the Belebele dataset.

In some cases, you may need to unhide the original answer (column H) to help you decide whether or not it should be labelled it as BB.

Example of BB case:

Paragraph	Question	Original Answer	Multi-choice Answer	Belebele Problem?
The balance of power was a system in which European nations sought to maintain the national sovereignty of all European states. The concept was that all European nations had to seek to prevent one nation from becoming powerful, and thus national governments often changed their alliances in order to maintain the balance. The War of Spanish Succession marked the first war whose central issue was the balance of power. This marked an important change, as European powers would no longer have the pretext of being religious wars. Thus, the Thirty Years' War would be the last war to be labeled a religious war.	In which country was the first war in Europe whose central issue was said to relate to the balance of power rather than having a religious context?	Spain	The balance of power was a system in which European nations sought to maintain the national sovereignty of all European states. The concept was that all European nations had to seek to prevent one nation from becoming powerful, and thus national governments often changed their alliances in order to maintain the balance. The War of Spanish Succession marked the first war whose central issue was the balance of power. This marked an important change, as European powers would no longer have the pretext of being religious wars. Thus, the Thirty Years' War would be the last war to be labeled a religious war.	BB

9 (b) You may also be asked a true or false question. These are also not useful for our QA dataset. Remove the passage text and mark with X.

Paragraph	Question	Multi- choice Answer	Answer Span
Virtual teams are held to the same standards of excellence as conventional teams, but there are subtle differences. Virtual team members often function as the point of contact for their immediate physical group. They often have more autonomy than conventional team members as their teams may meet according to varying time zones which may not be understood by their local management. The presence of a true "invisible team" (Larson and LaFasto, 1989, p109) is also a unique component of a virtual team. The "invisible team" is the management team to which each of the members report. The invisible team sets the standards for each member.	Based on the passage, which statement regarding physical and virtual teams is not true?	Convention al teams are usually held to a higher standard	x

9 (c) In some cases, inference or calculation is required to establish the answer to the question. In other words, there is no particular span in which the answer can be found. Remove the passage text and mark as X.

Paragraph	Question	Multi- choice Answer	Answer Span
MRI is based on a physics phenomenon called nuclear magnetic resonance (NMR), which was discovered in the 1930s by Felix Bloch (working at Stanford University) and Edward Purcell (from Harvard University). In this resonance, magnetic field and radio waves cause atoms to give off tiny radio signals. In the year 1970, Raymond Damadian, a medical doctor and research scientist, discovered the basis for using magnetic resonance imaging as a tool for medical diagnosis. Four years later a patent was granted, which was the world's first patent issued in the field of MRI. In 1977, Dr. Damadian completed the construction of the first "whole-body" MRI scanner, which he called the "Indomitable".	In what year was the first patent granted for medical imaging resonance?	1974	х