Belebele Answer Span: Arabic Annotation Guidelines

You'll be provided with a context paragraph and a question related to that paragraph. Your task is review and validate the pre-annotated *answer span* in the given paragraph. You should edit the span where necessary. The span indicates either the exact answer or where 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. The span should identify the part of the text in which the answer to the question or evidence of the answer can be found
- 2. Where necessary, edit the start of the span with \$! and the end of the span with !\$. (Answer Span column V)
- 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 preposition if the question is a When/ Where/ How question.
- 7. In cases of multiple instances of the answer string, choose the appropriate one only.
- 8. Mark problematic questions with BB or X.

Examples expanding on guidelines 3-8 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
تنتج النباتات الأكسجين الذي يتنفسه البشر، ويمتصون ثاني أكسيد الكربون الذي يخلفه البشر (أي الزفير). تصنع النباتات طعامها من الشمس من خلال التمثيل الضوئي. كما يوفرون الظل. نصنع بيوتنا وملابسنا من النباتات. معظم الطعام الذي نتناوله من النباتات كذلك. بدون النباتات، لا يمكن للحيوانات البقاء.	ما الذي يساعد النباتات في عملية التمثيل الضوئي؟	الشمس	تنتج النباتات الأكسجين الذي يتنفسه البشر، ويمتصون ثاني أكسيد الكربون الذي يخلفه البشر (أي الزفير). تصنع النباتات طعامها من الأالشمس! من علال التمثيل الضوئي. كما يوفرون الظل. نصنع بيوتنا وملابسنا من النباتات . معظم الطعام الذي نتناوله من النباتات كذلك. بدون النباتات، لا يمكن للحيوانات البقاء.

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
العضو الرئيسي للجهاز الدوري هو القلب الذي يضخ الدم. يتحرك الدم بعيداً عن القلب في أنابيب تسمى الأوردة. الشرايين ويعود إلى القلب في أنابيب تسمى الأوردة. أصغر الأنابيب تسمى الشعيرات الدموية.	ما هي الأنابيب الأصغر في جهاز الدورة الدموية؟	الشعيرات الدموية	العضو الرئيسي للجهاز الدوري هو القلب الذي يضخ الدم. يتحرك الدم بعيداً عن القلب في أنابيب تسمى الشرابين ويعود إلى القلب في أنابيب تسمى الأوردة. أصغر الأنابيب تسمى الشعيرات الدموية.!\$

3 (c) 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
ضرب زلز ال متوسط القوة مونتانا الغربية الساعة مداع. وم الاثنين. ولم ترد تقارير فورية عن وقوع أضرار من قبل هيئة المسح الجيولوجي الأمريكي (USGS) ومركز معلومات الزلازل الوطني التابع لها. وكان مركز الزلزال على بعد حوالي 20 كم (15 ميلاً) بين شمال وشمال شرق ديلون، وحوالي 65 كم (40 ميلاً) جنوب بوتي.	أين تمركز الزلزال بالنسبة لبوتي؟	40میلا جنوبا	ضرب زلز ال متوسط القوة مونتانا الغربية الساعة ١٠:١ مساءً. يوم الاثنين. ولم ترد تقارير فورية عن وقوع أضرار من قبل هيئة المسح الجيولوجي الأمريكي (USGS) ومركز معلومات الزلازل الوطني التابع لها. وكان مركز الزلزال على بعد حوالي 20 كم (15 ميلاً) بين شمال وشمال شرق ديلون، \$!وحوالي 65 كم (40 ميلاً) جنوب بوتي!

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
قد يكون للطفرات مجموعة متنوعة من الأثار المختلفة، على حسب نوع الطفرة، وأهميّة قطعة المادّة الوراثيّة المتأثرة، وما إذا كانت الخلايا المتأثرة من خلايا الخط الإنتاشي. يمكن فقط نقل الطفرات في خلايا الخط الجرثومي إلى الأطفال، بينما يمكن أن تسبب الطفرات في أماكن أخرى موت الخلايا أو السرطان.	على ماذا تعتمد قدرة الطفرة في الانتقال إلى الذرية؟	إذا كانت الخلايا من خلايا الخط الإنتاشي	قد يكون للطفرات مجموعة متنوعة من الآثار المختلفة، على حسب نوع الطفرة، وأهميّة قطعة المادّة الوراثيّة المتأثرة، وما إذا كانت الخلايا المتأثرة من خلايا الخط الإنتاشي. \$إلى يمكن فقط نقل الطفرات في خلايا الخط الجرثومي إلى الأطفال. بينما يمكن أن تسبب الطفرات في أماكن أخرى موت الخلايا أو السرطان.

5. Don't include non-relevant punctuation

5 (a) Final punctuation is usually not considered relevant

Paragraph	Question	Multi-choice Answer	Answer Span
العضو الرئيسي للجهاز الدوري هو القلب الذي يضخ الدم. يتحرك الدم بعيداً عن القلب في أنابيب تسمى الشرايين ويعود إلى القلب في أنابيب تسمى الأوردة. أصغر الأنابيب تسمى الشعيرات الدموية.	أي جزء من الدورة الدموية يجلب الدم نحو القلب؟	الأوردة	العضو الرئيسي للجهاز الدوري هو القلب الذي يضخ الدم. يتحرك الدم بعيداً عن القلب في أنابيب تسمى الشرابين ويعود إلى القلب في أنابيب تسمى الأوردة! أصغر الأنابيب تسمى الشعيرات الدموية.

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
"يعتمد التصوير بالرنين المغناطيسي على ظاهرة فيزيائية تسمى الرنين المغناطيسي النووي (NMR) والتي اكتشفت في ثلاثينيات القرن الماضي عن طريق فيليكس بلوخ (من جامعة ستانفورد) وإدوارد بورسيل فيليكس بلوخ (من جامعة ستانفورد) وإدوارد بورسيل المغناطيسية والموجات الراديوية الذرات تبعث باشارات راديو دقيقة. في عام 1970، اكتشف رايموند داماديان، وهو طبيب وعالم أبحاث، الأساس الذي بني عليه استخدام التصوير بالرنين المغناطيسي كأداة لتشخيص الطبي. بعد أربع سنوات، تم منح براءة اختراع، والتي كانت أول براءة اختراع والتي كانت أول براءة اختراع في العالم يتم محبال التصوير بالرنين المغناطيسي. في عام 1977 استكمل الدكتور داماديان إنشاء أول ماسح ضوني للتصوير بالرئين المغناطيسي. في ضوني للتصوير بالرئين المغناطيسي. في اضوني للتصوير بالرئين المغناطيسي "كامل الجسم"،	من اكتشف الرنين المغناطيسي النووي؟	بلوخ وبورسيل	"يعتمد التصوير بالرنين المغناطيسي على ظاهرة فيزيانية تسمى الرنين المغناطيسي النووي ، (NMR) والتي اكتشفت في ثلاثينيات القرن الماضي عن طريق \$ إفيليكس بلوخ (من جامعة ستانفورد) وإدوارد بورسيل (من جامعة ملوفارد)! في هذا الرنين، تجعل المجالات المغناطيسية والموجات الراديوية الذرات تبعث إشارات راديو دقيقة. في المحالات الأساس الذي بنبي عليه استخدام التصوير بالربين أبحاث، الأساس الذي بنبي عليه استخدام التصوير بالرنين منح باداة التشخيص الطبي. بعد أربع مسوات، تم العالم يتم منحها في مجال التصوير بالربين المغناطيسي. في عام 1977 استكمل الدكتور داماديان إنشاء أول ماسح ضوئي للتصوير بالرئين المغناطيسي "كامل الجسم"، في عام 1977 استكمل الدكتور داماديان إنشاء أول ماسح والذي يسمى اليوم ب"الذي لا يقهر"."

- 6. If the question is a When/ How question then include the preposition or adverb that would normally form part of an appropriate answer. Note that sometimes the preposition is already part of the word in the form of a morpheme.
- 6 (a) An example of a "how-question" where the preposition should be included in the span.

Paragraph	Question	Multi-choice Answer	Answer Span
ستطاع اختراع العجلات ذات العوارض أن يجعل العربات الأشورية أخف وزناً وأسرع وأكثر العربات الأشورية أخف وزناً وأسرع وأكثر بمكن أن تخترق السهام المطلقة من الأقواس القاتلة دروع الجنود المتنافسين. قدم الأشوريون أول سلاح فرسان في حوالي عام 1000 قبل الميلاد. سلاح الفرسان هو جيش يقاتل على ظهور الخيل. لم يكن السرج قد اخترع بعد، أذلك قاتل الفرسان الأشوريون على ظهور خيولهم العارية.	كيف كانت تُخاض المعارك التي تستخدم الفرسان؟	على ظهر الخيل	استطاع اختراع العجلات ذات العوارض أن يجعل العربات الأشورية أخف وزناً وأسرع وأكثر استعداداً لتجاؤز الجنود والعربات الأخرى. يمكن أن تخترق السهام الفطلقة من الأقواس القاتلة دروع الجنود المتنافسين. قدم الأشوريون أول سلاح فرسان في حوالي عام 1000 قبل الميلاد. سلاح الفرسان هو جيش يقاتل \$1عنى ظهور الخيل!\$. لم يكن السرج قد اخترع بعد، لذلك قاتل الفرسان الأشوريون على ظهور خيولهم العارية

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

Paragraph	Question	Multi-choice Answer	Answer Span
بلغ الطراز القوطي ذروته في فترة ما بين القرنين العاشر أو الحادي عشر إلى القرن الرابع عشر. في البداية، أثرت الثقافة البيزنطية في الشرق على الملابس تأثيرًا شديدًا. وعلى الرغم من ذلك، فنظرا لبطء قنوات التواصل، قد تتأخر الأنماط في الغرب قدر ما يقارب و 25 إلى 30 سنة. في نهاية العصور الوسطى، بدأت أوروبا الغربية في تطوير أسلوبها الخاص. واحدة من أكبر التطورات في ذلك الوقت نتيجة للحروب الصليبية بدأ الناس في استخدام الأزرار لربط الملابس.	متى توقفت أوروبا الغربية عن الاعتماد الشديد على التأثيرات وبدأت في تطوير أسلوبها الخاص؟	في نهاية العصور الوسطى	بلغ الطراز القوطي ذروته في فترة ما بين القرنين العاشر أو الحادي عشر إلى القرن الرابع عشر في البداية، أثرت الثقافة البيز نطية في الشرق على الملابس تأثيرًا شيدًا. وعلى الرغم من ذلك، فنظرا لبطء قنوات التواصل، قد تتأخر الأنماط في الغرب قدر ما يقارب 25 إلى 30 سنة. وروبا الغربية في تطوير أسلوبها الخاص. أوروبا الغربية في تطوير أسلوبها الخاص. واحدة من أكبر التطورات في ذلك الوقت نتيجة للحروب الصليبية بدأ الناس في استخدام الأزرار لربط الملابس

7. 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
"أفادت تقارير باستمرار أعمال النهب على نطاق واسع حتى صباح اليوم التالي وذلك لغياب ضباط إنفاذ القانون عن شوارع مدينة "بيشكك". وصف أحد المراقبين مدينة بيشكيك بأنها تغرق في حالة من "فقدان السيطرة"، حيث تجولت عصابات من الناس في الطرقات وقاموا بنهب مخازن السلع الاستهلاكية. حمّل العديد من سكان مدينة "بشكيك" المنظاهرين في الجنوب حالة الفوضى."	من ألقى باللوم على المتظاهرين من الجنوب في أعمال النهب؟	سكان بيشكك	"أفادت تقارير باستمرار أعمال النهب على نطاق واسع حتى صباح اليوم التالي وذلك لغياب ضباط إنفاذ حتى صباح اليوم التالي وذلك لغياب ضباط إنفاذ القانون عن شوارع مدينة "بيشكك". وصف أحد المراقبين مدينة بيشكك بأنها تغرق في حالة من "فقدان السيطرة"، حيث تجولت عصابات من الناس في الطرقات وقاموا بنهب مخازن السلع الاستهلاكية. حمّل العديد من \$إسكان مدينة "بشكيك"!\$ المتظاهرين في الجنوب حالة الفوضى."

- 8. 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 QA, and as such it may not be possible to find an exact answer or *any* answer:
- 8 (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.

Example of BB case:

Paragraph	Question	Original Answer	Answer Span	Belebele Problem?
كانت الموازنة في الطاقة نظاماً حاولت من خلاله الدول الأوروبية أن تُحافظ على السيادة الوطنية الخاصة بجميع الدول الأوروبية. نص المفهوم على أن جميع الدول الأوروبية بجب أن تسيع لمنع دولة و احدة من أن تصبيح قوية، وبالتالي عادةً ما غيرت الحكومات الوطنية تحالفاتها للحفاظ على التوازن. كانت حرب الخلافة الإسبانية هي الحرب الأولى التي كانت قضيتها المركزية هي ميزان القوى. كان ذلك بمثابة تغيير مهم، حيث لم يعد لدى القوى الأوروبية ذريعة لكونها حروباً دينية. وهكذا، مستكون حرب الثلاثين عاماً آخر حرب توصف بأنها حرب دينية.	في أي بلد حصلت الحرب الأولى في أوروبا والتي كانت قضيتها المركزية مرتبطة بتوازن القوى بدلاً من أن تكون في سياق ديني؟	إسبانيا	كانت الموازنة في الطاقة نظاماً حاولت من خلاله الدول الأوروبية أن تُحافظ على السيادة الوطنية الخاصة بجميع الدول الأوروبية. نص المطنية الخاصة بجميع الدول الأوروبية بجب أن سعى لمنع دولة واحدة من أن تصبيح قوية، وبالتالي عادةً ما غيرت الحكومات الوطنية تحالفاتها للحفاظ على التوازن. كانت والحرب الأولى التي الخلافة الإسبانية! هي الحرب الأولى التي كانت قضيتها المركزية هي ميزان القوى. كان نلك بمثابة تغيير مهم، حيث لم يعد لدى القوى الأوروبية ذريعة لكونها حروباً دينية. وهكذا، ستكون حرب الثلاثين عاماً آخر حرب توصف بأنها حرب دينية	ВВ

8 (b) Questions for which an answer (or evidence for an answer) could not be found at all have been marked as X (as per the English annotations)

Example of X case:

Paragraph	Question	Original Answer	Answer Span	Belebele Problem?
كانت الموازنة في الطاقة"بعد الهرم الأكبر في محافظة "الجيزة" هو الوحيد من العجانب السبع التي لا تزال قالجيزة" هو الوحيد من العجانب السبع التي لا تزال قانم حتى يومنا هذا. تم بناؤه من قبل المصربين في القرن الثالث من الحقبة الحالية، يعتبر الهرم الأعظم الميث. تحتوي هضبة الجيزة، أو "مدينة الجيزة الجنانزية" بوادي الموتى المصري، على عدة أهر امات الجنانزية" بوادي الموتى المصري، على عدة أهر امات الجوائزية بوادي الهرم الأكبر) ومقابر صغيرة ومعابد وأبو الهول. تم إنشاء الهرم الأكبر لتكريم الفر عون خوفو، وتم بناء العديد من الأهر امات والمقابر والمعابد الأصغر لتكريم زوجات خوفو وأفراد أسرته."	ماذا يمكن العثور عليه في هضبة الجيزة؟	جميع العجائب السبع	كانت الموازنة في الطاقة"يعد الهرم الأكبر في محافظة "الجيزة" هو الوحيد من العجائب السبح التي لا تز ال قائمة حتى يومنا هذا. تم بناؤه من قبل المصريين في القرن الثالث من الحقبة الحالية، يعتبر الهرم الأعظم واحداً من الأهرام الضخمة المبينة على شرف الفرعون المبيت. تحتوي هضبة الجيزة، أو "مدينة الجيزة الجنائزية" بوادي الموتى المصري، على عدة أهر امات (أكبرها الهرم الأكبر) ومقابر صغيرة ومعابد وأبو الهول. تم إنشاء الهرم الأكبر لمتكريم الفرعون خوفو، وتم بناء العديد من الأهرامات والمقابر والمعابد المصيعر لتكريم توواد أسرته."	X