Task Description

Thank you for participating in this task!

This annotation task is to identify pieces of content that differ in paragraph pairs across different languages. You will be given two paragraphs: one in English, and one in Chinese. These are not necessarily translations of each other. For each pair of paragraphs, please do the following:

- Read the first paragraph carefully.
- Read the second paragraph, and select spans (contiguous word sequences) in the second paragraph that differ in meaning with respect to the first paragraph shown. Spans may indicate differences falling into one of the following four categories (with examples given in English for simplicity):

1. New Information - Content in one paragraph which is not given in the other one and which cannot be inferred (using reasoning or background knowledge).

- This could be content that is **added** (e.g., "Charles Dickens was born on 7 February 1812 _in Portsea Island_" vs "Charles Dickens was born on 7 February 1812") or **changed** (e.g., "Saint Patrick's Day is a religious and cultural _holiday_" vs "Saint Patrick's Day is a religious and cultural _festival_". In this example, "festival" is labeled as "new information" because not every holiday is a festival. **Note that we are particularly interested in fine-grained meaning changes like this.**

2. New Information (Inferable) -New content in one paragraph that is not present in the other, but which can reasonably be inferred from it. Inferences can make use of information in the paragraph, background knowledge or commonsense reasoning.

- **An example of background knowledge:**
  - PARAGRAPH 1: "Michael Jackson was heavily influenced by funk, disco and gospel."
  - PARAGRAPH 2: "_The King of Pop_ was heavily influenced by funk, disco and gospel."

    Here, since it is common knowledge that Michael Jackson is known as The King of Pop, but "The King of Pop" does not appear explicitly in the first paragraph, "The King of Pop" should be labeled as "new information (inferable)"

- **Examples of reasoning:**
    - "70% of the students passed the exam" vs "*30% of the students failed the exam*"
    - "Emma burst into tears" vs "*Emma cried*"

One way to think about the "new information (inferable)" label is: the information given is different, but if I *know* the information given in the first paragraph, then I would be able to infer (without extensive web searches or deep subject matter expert knowledge) the information in the second paragraph.

3. Reference - The span corresponds to a reference **which also interrupts the flow of the text**. For example, consider the following paragraph:
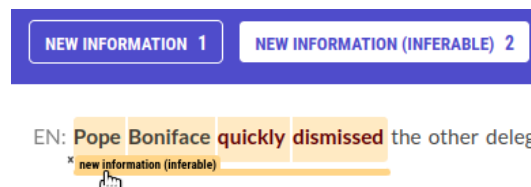
"Though the rebels lacked military training, they displayed skilful use of available local materials and unusual tactics against the disciplined Roman armies. *Frontinus, Stratagems, Book I, 5:20–22 and Book VII:6.* They spent the winter of 73–72 BC training, arming and equipping their new recruits."

It is clear that "Frontinus, Stratagems, Book I, 5:20–22 and Book VII:6." is not part of the main text (it is due to a formatting error) and is actually a reference being cited.

Please no not label references which are correctly formatted and which are clearly part of the original text.

Annotation Instructions

First you will be presented with the English paragraph and asked to annotate the paragraph in the other language. Then the paragraphs will be swapped and you will annotate information in the English paragraph which is not in the other one. To annotate: (1) first select the button for the desired label, and (2) drag the cursor over a span to label it. If some span is labeled by accident, you can remove the label by hovering over it and clicking, as shown below:

After selecting the spans, click the **green check mark (** ✓ **)**button to progress to the next paragraph pair. To save your annotations, please click the **save/floppy disk (** 💾 **)** icon in the top left (or Ctrl+S keys) to save your work.

Notice some words are shown in **red.** These words are more likely (but not guaranteed) to not correspond to any word in the other paragraph; we hope this helps direct your attention to the "most different" parts of the content, and to make it easier to annotate.

Two more buttons are available to you:

- **"Ignore"** (⊘) can be used to move on to the next example without annotating. Please use this sparingly (less than 5% of the time), and only in cases where there is something really wrong with the example. For example, this could happen if the paragraphs are just snippets of math equations that you can't make sense of.
- **"Undo"** (↵) can be used if you need to go back to edit the previously annotated paragraph pair.

Important notes on saving: (1) Please save your work before closing the browser window, (2) Only save once you are certain about your annotations, because it will not let you modify it once you move on. (3) Annotating out of order (skipping ahead and then going back) is not recommended because the autosave may prevent you from doing this.

Finally, every paragraph pair also has a box with **"Optional comments".** Please comment if there was something non-obvious in your annotation (mostly inferable vs same, and inferable vs new). It can be used to give feedback or ask questions if the need arises.

Special Cases

Some paragraph pairs are may have some superficial similarity, but not actually be about the same thing or event. If the paragraphs are not about the same thing, you can click the **red** ✕ button to move on to the next example. **Only use this is** *the entire* **paragraph is completely different.** If even one sentence has some information overlapping, it should be labeled.

Some paragraphs are so close they can be considered perfect translations. In that case nothing needs to be labeled; please do not label anything and just click the **green check mark** button to progress to the next example.

FAQ

- **What if I want to indicate a "deletion" in the paragraph being annotated? (content that occurs in the first paragraph but not in the second)**
  - A: This is equivalent to there being new content in the first paragraph. You can annotate this directly as "new information" in the paragraph once the paragraphs flip.

- **What constitutes "background knowledge"?**
  - A: When deciding if something is "new information (inferable)" you can call upon background knowledge. This is anything that you know and you think is common knowledge, like referring to President Biden as "Joe Bien".

- **I suspect something is not factually correct. What do I do?**
  - A: This task does not concern factual correctness. You do not need to look up whether something is actually true or not, but only how the meanings of the two paragraphs compare to each other.

- **Should I annotate differences in grammar?**
  - A: Only in cases where this would trigger a change in meaning. For example, differences in tense or gender which are clearly typos should not be marked. Similarly, differences in tense may not necessarily indicate differences in when events occur.

- **How should pronouns be handled? (e.g., one paragraph uses "he", while the other refers to a specific person by name).**
  - A: If it is clear that these refer to the same entity, then do not label it.