

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 another language. 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 three 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).

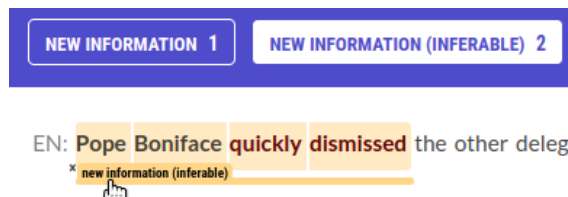
- This could be content that is **added** (e.g., "Charles Dickens was born on 7 February 1812 in Portsea Island" vs "Charles Dickens was born on 7 February 1812") or **changed** (e.g., "Saint Patrick's Day is a religious and cultural holiday" vs "Saint Patrick's Day is a religious and cultural festival". In this example, "festival" is labeled as "new information" because not every holiday is a festival.



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** (e.g., knowing the Michael Jackson is often called The King of Pop) or **commonsense reasoning** (e.g., "70% of the students passed the exam" vs "30% of the students failed the exam", or "Emma burst into tears" vs "Emma cried").

3. Connotation Difference - The words or expressions express the same thing exactly, but have different connotations (i.e., different associations or attitudes expressed even though the literal meaning is the same, such as "the slender man" vs "the scrawny man").



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. This can be useful if it is impossible to even know what the paragraphs say (e.g., if they are sentence fragments or snippets of math equations)
- “**Undo**” () can be used if you need to go back to edit the previously annotated paragraph pair.

Finally, every paragraph pair also has a box with “Optional comments”. You do not need to fill this, but it can be used to give feedback on difficult scenarios.

Special Cases

Some paragraph pairs may have some superficial overlap, 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.

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 Biden”.
- **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 says “he”, while the other refers to a specific person).**
 - A: If it is clear that these refer to the same entity, then do not label it.