**Annotation Software, Guideline and Examples**

Hi Sarah and Yulu, here I will show you how to utilize this software to do annotations.

**Using the software**

1. Unzip the “gold\_align.zip” file by typing **Unzip gold\_align.zip** in command line.

A screenshot of a cell phone

Description automatically generated

1. Change directory to the gold\_align folder.
2. Make sure you have Node.js installed in your computer, if not, download it from the link here <https://nodejs.org/en/download/>
3. Once Node.js is installed, in your terminal, type command **./RUN** to start the server. A successful run of the software will show you the following info in the terminal.



1. Now, you can gain access to the annotation software interface by opening your web browser and visit <http://localhost:2001/>, expected interface is shown below.

A screenshot of a cell phone

Description automatically generated

1. I have created an account for you with User name **Yulu/Sarah** and you should pick the Task **annotation** and Dataset **Update\_corpus**.
2. Click the select button and then the batch\_000.tsv. Then you can click the pen icon, then click “resume editing from the latest” or directly press the middle button showing UNDERWAY. You will now reach the annotation interface.



A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated



**Annotation Interface**

1. Interface Explanation: Below is a real example of the interface.
2. You can click the Contexts button to hide or show the original sentence pair.
3. For each cell in the grid, one-click change the color to black and double click to grey. If you click the cell three times, it will then change back to white.
4. **Detailed annotation guideline** will be given in the next page, once you finished one pair, do not forget to click the “pen icon” in the middle to save the change ( the software might be able to auto-save, but we can make it more reliable through saving by ourselves.

A screenshot of a cell phone

Description automatically generated

**Annotation Guideline**

1. The goal of the annotation is to find all word/phrase alignments, with special interest on the alignments where two sides are not identical. (Majority of the sentence pair should have a large overlapped part (identical ones)
2. Currently, of the provided file, you might find some already contain annotations. Those annotations are automatically produced by a word-aligner model and some may contain multiple errors. Please only treat them as a reference for help you understand the annotation and **do your own annotation.**
3. There are mainly two types of annotation.
   1. **BLACK (**green block) **Sure alignment**: two words are identical or paraphrase of each other.

A screenshot of a cell phone

Description automatically generated



* 1. **GREY** (red block) **Possible Alignments:** long phrase/sequence of word which might not have strict one-one black annotation, but they are having the similar meanings.

In the following example:

**“Popular place to visit” --- “destination for some tourists”** are colored as GREY.

If a cell has been annotated black (**destination – place**), you can leave the color as it is.

A screenshot of text

Description automatically generated



* 1. For the remaining ones, which you cannot find matching in another sentence, leave those cells blank. (Blue block)