

# Annotation Guidelines

Attached you shall find the sentences of code-mixed data in two formats

1. data.txt
2. data.csv

Use appropriate files as mentioned in the tool overhead.

## Google Drive

[https://drive.google.com/drive/folders/15j1aZZj8YzQBH0bYw\\_uDZtz3aPIUQT7H?usp=sharing](https://drive.google.com/drive/folders/15j1aZZj8YzQBH0bYw_uDZtz3aPIUQT7H?usp=sharing)

Steps to follow common to all the tools

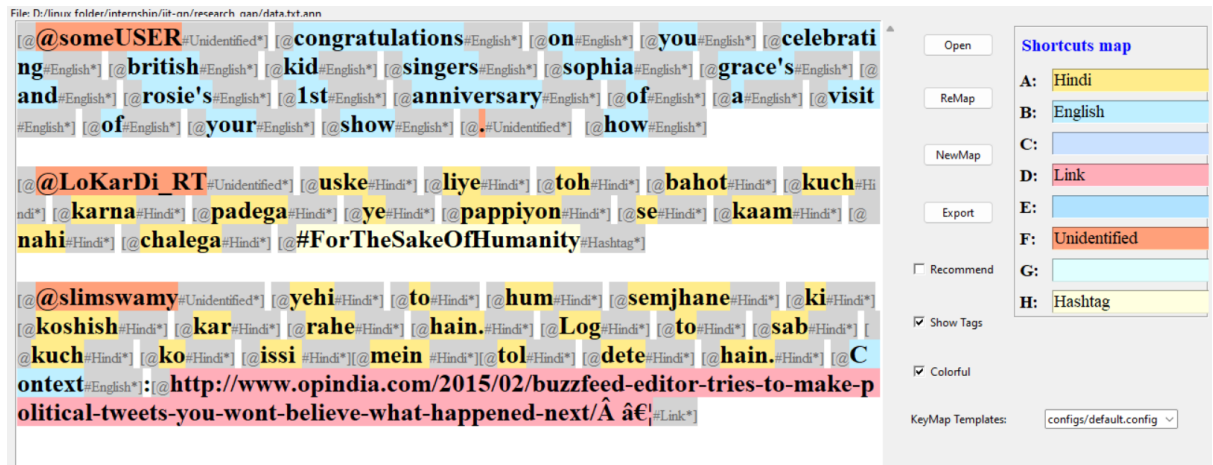
1. Start a timer at the beginning of the annotation of every sentence and record the time taken for every annotated sentence(in seconds). There should be 10 time entries for the 10 corresponding sentences.
2. Store and update all timings in the attached timesheet  
<https://docs.google.com/spreadsheets/d/1e2w1d0Qig1R9J5dC4eIN9uAwQbuBkLgYS-dRxxiH6X0/edit?usp=sharing>
3. End timer on completion of annotation and submit the annotation time taken by you in the google form attached (this form shall be circulated at a later stage, for now note down the annotation timings in a self-created text file).
4. We shall be performing **word-level** annotation. For example, in a sentence “mene khana khaya”, even though the sequence of three words is Hindi you will have to annotate the three words individually, like [mene=Hindi], [khana=Hindi], [khaya=Hindi]. Do not cluster continuous similar annotations. Follow this rule strictly.
5. The sentence annotation time needs to be in **seconds (eg 87s)**, same for the installation and configuration time needs to be in a **second** format (eg 235s).

## YEDDA

<https://github.com/jiesutd/YEDDA>

1. Open the YEDDA folder and type command “py YEDDA.py” for python version 3.9 or above and “python YEDDA.py” for python version below 3.8 in the terminal to start the application. Hit Enter. (**Note:** At the start of the this step, start the timer and end the timer when you finish the installation step and note down the timing under header, “Installation Time”)
2. Click on open and load data.txt to the YEDDA tool (Available in the Data Folder)(**Note:** At the start of the this step, start the timer and end the timer when you finish the configuration step to step 4 and note down the timing under header, “Configuration Time”)
3. Update the shortcut maps using the new default.config file available in the attached drive link at the start of the document ( drive>YEDDA>default.config ). Upload this config file in the location YEDDA>config in the local file system.
4. Tick on Colourful to aid you in visual representations.

- Perform annotations of the loaded sentences on word level, meaning annotate every word with respect to the language they pertain to. Identify them as unidentified (null) when you encounter anything similar to period (.), exclamation (!), question mark (?) etc. Select a word and then press a key to annotate it  
For example: select “congratulations” then hit key B to annotate it as English.

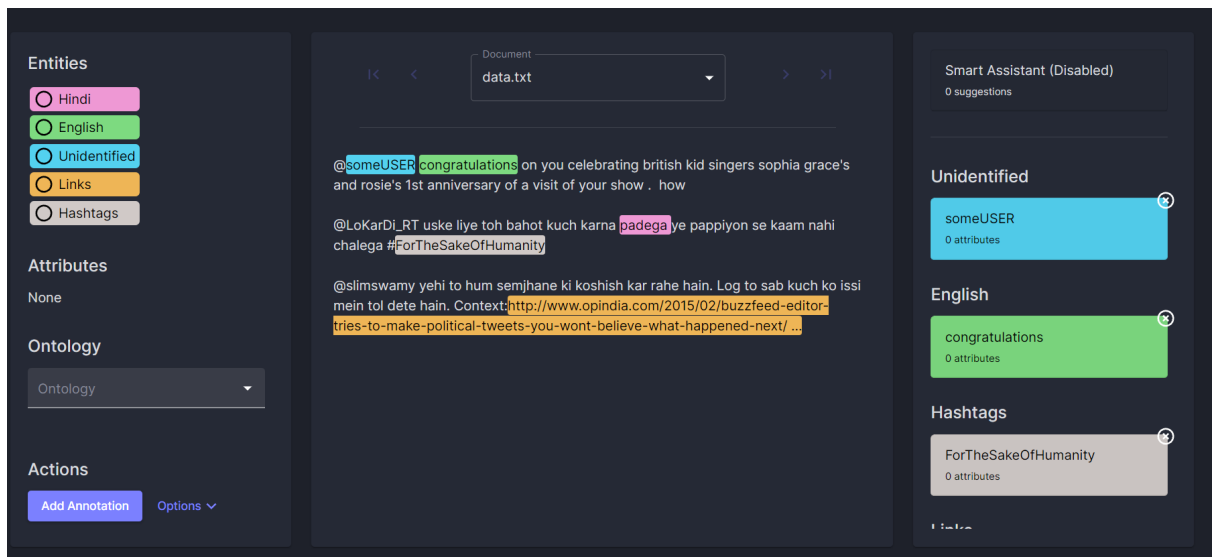


- Click on export and then ok.
- An ‘ann’ and a ‘bmes’ file shall be downloaded, and submit the files to the admin for further analysis.

## Markup

<https://getmarkup.com/dashboard>

- Create a new account on Markup. (**Note:** At the start of this step, start the timer and end the timer when you finish the configuration step to step 5 and note down the timing under header, “Configuration Time”)(You can enter the “Installation Time” = 0, for this tool)
- Create a new session
- Upload a configuration file present in the drive>Markup directory.
- Upload data.txt available in drive.
- Click on start session.
- Select individual words, select an entity (on the left) and then click on add annotation. Continue this process for every word in the sentence. Here is a small example for the same.

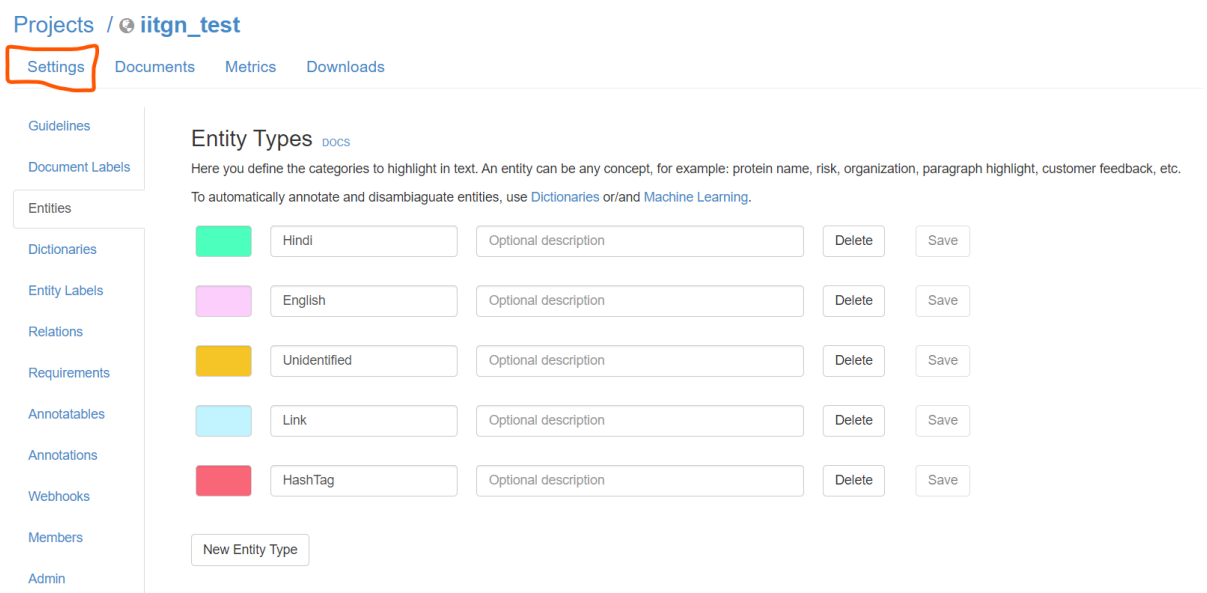


7. Click on options and select data.txt (ann) and export the file. Submit the annotation file to the admin.
8. Exit session.

## TagTog

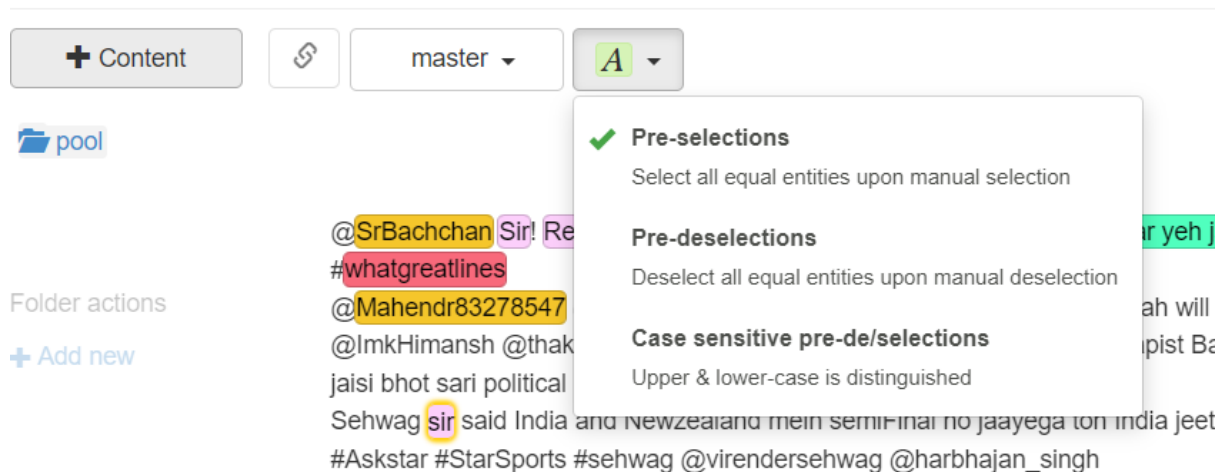
<https://www.tagtog.net/>

1. Create a new account on TagTog. (**Note:** At the start of the this step, start the timer and end the timer when you finish the configuration step to step 7 and note down the timing under header, “Configuration Time”)(You can enter the “Installation Time” = 0, for this tool)
2. Create a new project and add a project name of your choice.
3. In the settings tab add the following entities



4. Click on the documents tab and then click on “+ Content” or add content button.
5. Select input type as file and then upload the data.txt file to the project
6. Click on the freshly added data.txt file.

7. Ensure the “pre-selections” button is selected



8. Make sure you annotate individual words
9. Select individual word to annotate them

@Mahendr83278547 @IndiaToday Teri kimat dokodi ki ho gayi  
 @ImkHimansh @thakur I will say it better .. jb ek Constitutional  
 jaisi bhot sari political party ki dukhti nabj h  
 Sehwag said India and Newzealand mein semiFinal ho jayega

Update them in case they are incorrect by clicking on the selected word.

@Mahendr83278547 @IndiaToday Teri kimat dc  
 @ImkHimansh @thakur I will say it better .. jb ek  
 jaisi bhot sari political party ki dukhti nabj h  
 Sehwag sir said India and Newzealand mein ser  
 #Askstar #StarSports #sehwag @virendersehwag @harbhajan\_singh

10. Select the download button/symbol and choose EntitiesOnlyClasses Tsv option and download the annotation file.

## UbiAI Tool

<https://ubiai.tools/>

1. Create a new account on the UbiAI Tool Website. (**Note:** At the start of the this step, start the timer and end the timer when you finish the configuration step to step 6 and note down the timing under header, “Configuration Time”)(You can enter the “Installation Time” = 0, for this tool)
2. Create a new project and add the project name.
3. Select span based annotation
4. Add the following entities

Step 3/5

## Entities Labels

Add your entities to be able to annotate your documents.

\*\* You can skip this for now.

[+ Add new label set](#)

Add your Label one by one

HINDI 1

ENGLISH 2

UNIDENTIFIED 3

LINK 4

HASHTAG 5

[Back](#)

[Next](#)

Use this box to choose colours, different colours are acceptable, they are just for your visual difference

[+ Add new label set](#)



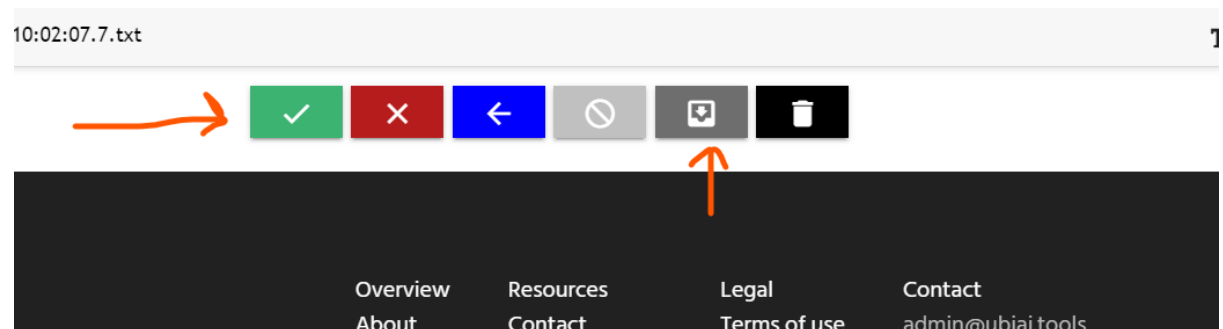
5. Go next on relations and submit on step 5
6. Now upload the data.txt file (Select no-pre annotation)
7. Now annotate word wise. First select the entity and then select the word to annotate it.



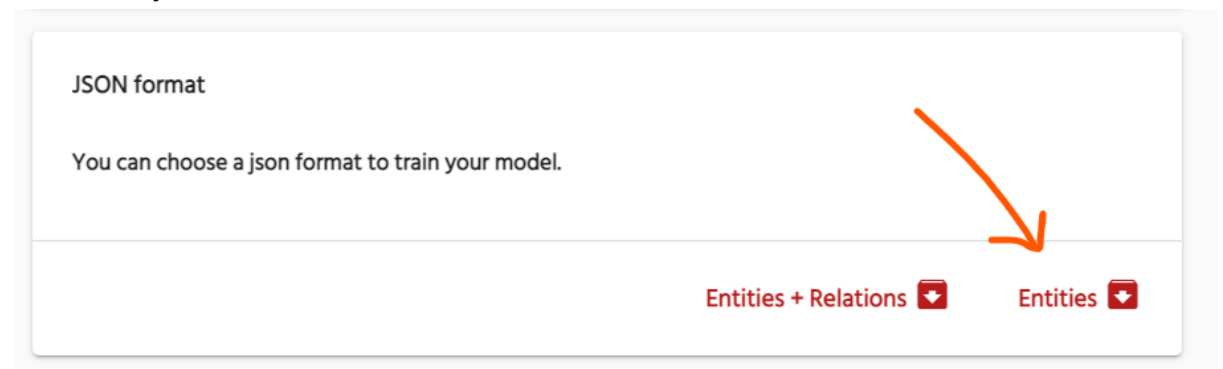
Notice the tick on "Unidentified" entity

- After annotation all the sentences, scroll down to the bottom of the page and select the green box and then click on the install button

shio ki umid iss jit se Jura hay,2050 tak desh will be world Top in position .



And select json entities



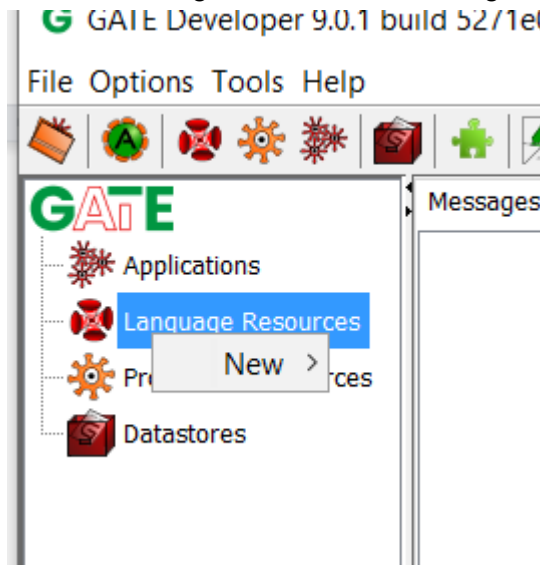
- Unzip the folder and upload the json file on drive.

## GATE

<https://gate.ac.uk/download/#latest> (Use windows-specific installer)

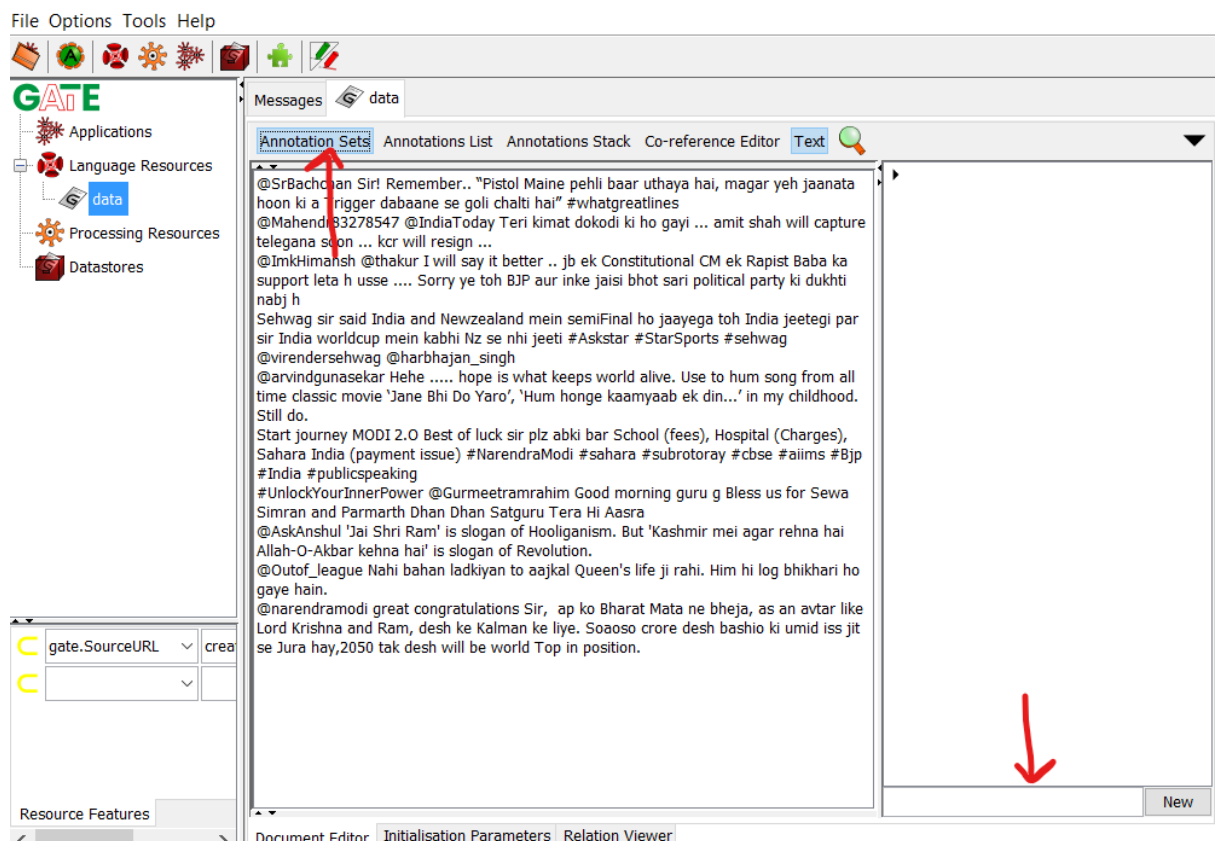
- Follow all the installation procedures as the program driver asks. (**Note:** At the start of the this step, start the timer and end the timer when you finish the installation step note down the timing under header, "Installation Time")

2. Add the data to the GATE application (**Note:** At the start of the this step, start the timer and end the timer when you finish the configuration step to step 4 and note down the timing under header, “Configuration Time”)



Click on the new and then gate document, name it and click on the newly created document.

3. Copy the contents from the data.txt file to the new gate document.
4. Click on annotation sets



And add the following annotation entities

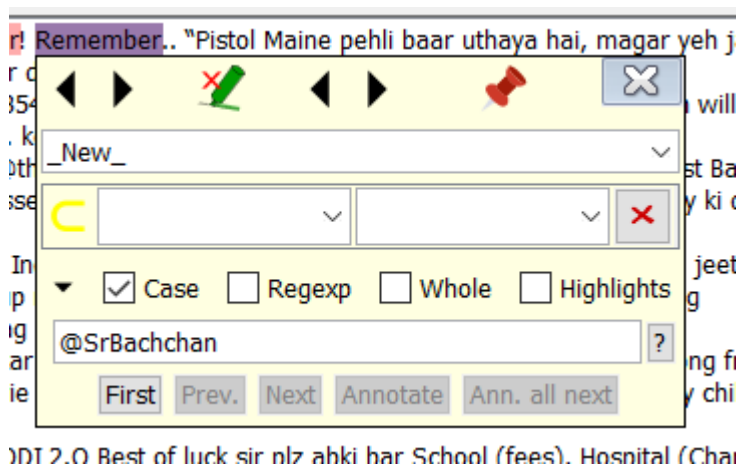
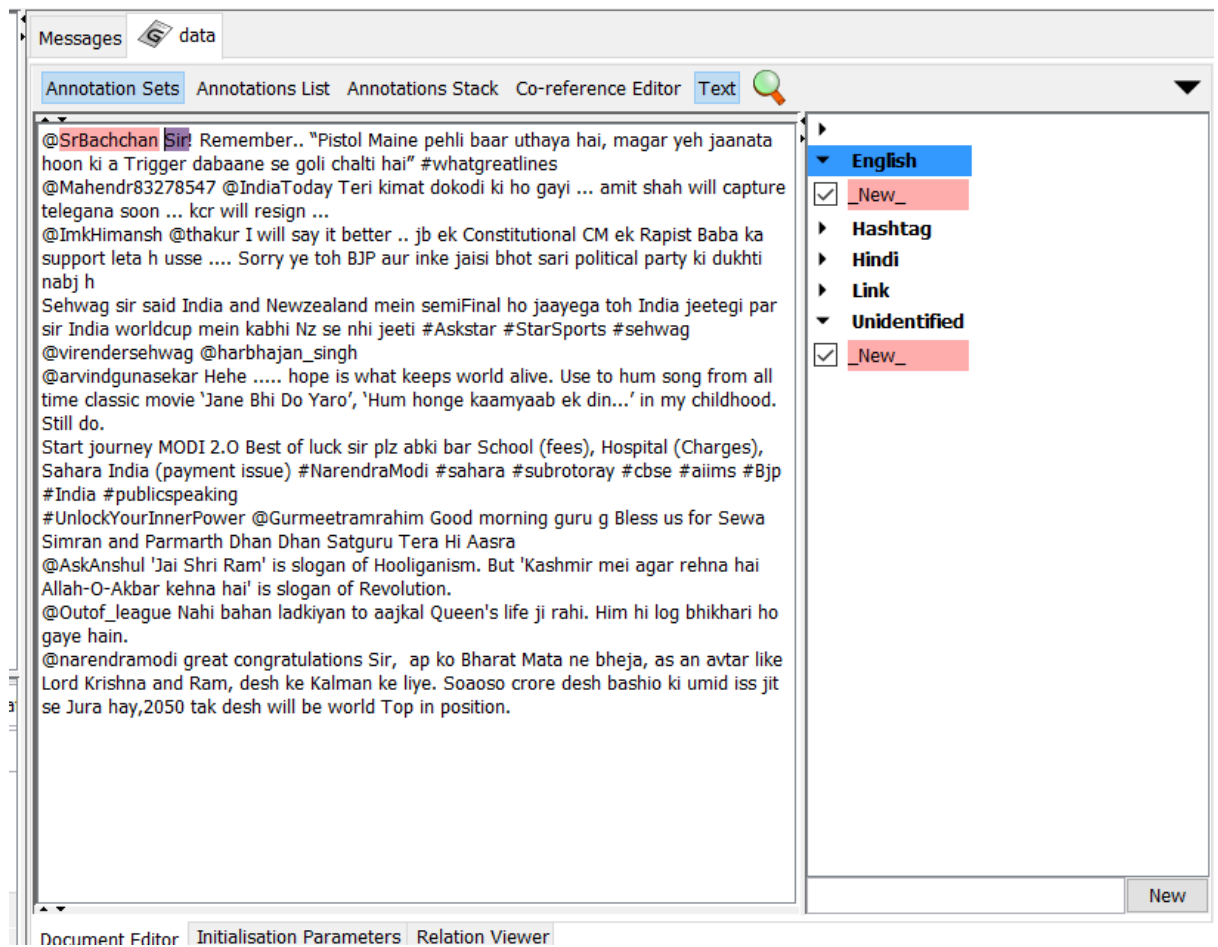
- a. Hindi
- b. English

- c. Unidentified
- d. Links
- e. Hashtags



5. Now perform **word-level** annotations, first select the entity on the right, then select the word and double click to annotate it.

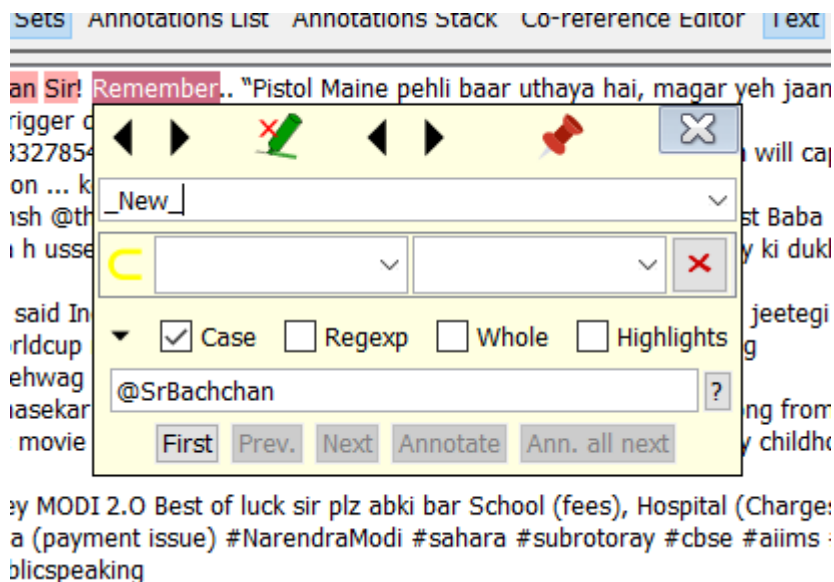




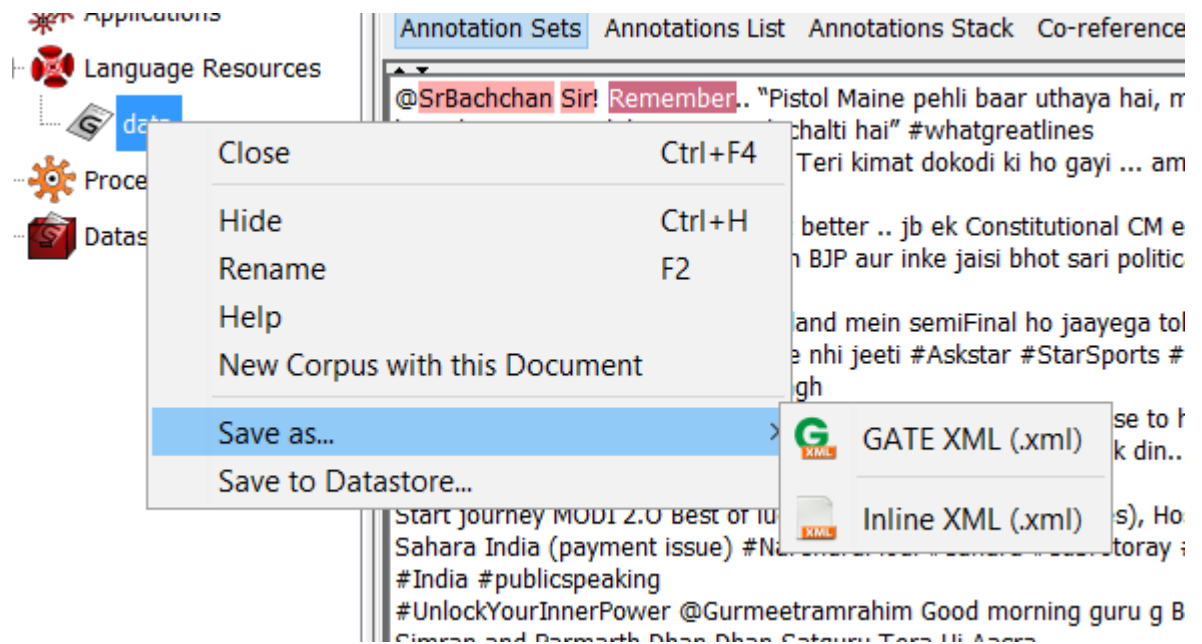
6. If you perform wrong annotation, double click on that word



Click on the wrong annotation and then click on green penicillin with cancel icon to delete that annotation.



7. Post annotation export the annotation as the following gate xml document



8. Upload the xml file on the drive folder for GATE.

## COMA

1. Download the zip folder from the shared google drive, navigate to COMA. Download the zip file. Extract the folder. Note the installation time from step 1 & step 2.
2. Navigate to the frontend folder. Inside the frontend folder, open the terminal and hit the command "npm i" or "npm install". Make sure you have node installed on your machine. (<https://nodejs.org/en/download/>)
3. Now type command "npm start". A new tab should open in your default browser, if not go to url, <http://localhost:3000/login>

### Login

Don't have an account? [Sign Up](#)

4. Click on the signup button and create an account for yourself. Note the configuration time at this step, i.e. step 4.

5. Login and click on continue, you shall see the sentences preloaded for you.
6. Select overall grammar of the sentence, for example “Hi mera naam Aarav hai” can be marked overall Hindi here

COMA

test

⏻

#1

Steps to Follow!

1. Select Root Grammar for the sentence.
2. The selected grammar turns purple.
3. Individual word tags get a default color.
4. 'English' - green, 'Hindi' - yellow, 'Unidentified' - blue .
5. Update individual word tags.

@SrBachchan Sir! Remember.. "Pistol Maine pehli baar uthaya hai, magar yeh jaanata hoon ki a Trigger dabaane se goli chalti hai" #whatgreatlines

English

Hindi

←

Individual Word Tags

@SrBachchan

Sir

!

Remember

..

"Pistol

Maine

pehli

baar

uthaya

hai

,

magar

yeh

jaanata

hoon

ki

a

Trigger

dabaane

se

goli

chalti

hai"

#whatgreatlines

SUBMIT

7. Next update individual word tags wherever necessary, the colour strategy has been defined on the screen in right along with the rules. Note the annotation time per sentence.
8. Hit submit and inform post annotation of the 10 sentences.