

## Installing GATE:

If you are at the University of Pittsburgh, go here for GATE installation instructions.

All others, please visit the GATE site to download the latest version of GATE.

Most users here currently run GATE 4.0 and we recommend that you do so too. In our experience it has been more stable and has better rendering than previous versions that we we have used on Windows and Linux platforms.

## Pointing GATE on Windows to our XML Schemas:

### GATE 3.1 or 4.0

When you set up GATE for the first time to annotate, then once you have started GATE, use the menu to go to File->Manage CREOLE plugins. If the mpqa-annotation schema (the exact name may vary) is already present in the list of "Known CREOLE directories", select it and check the radio button for *Load now*. If you want GATE to be set up automatically for future annotation sessions, too, then select *Load always* as well. Click the OK button and the schema is loaded. At the moment, the correct creole

repository to use is:

<http://www.cs.pitt.edu//mpqa/opinion-annotations/gate-annotation-new/>

If you didn't find an mpqa CREOLE directory in the list, then use the button at the bottom of the popped up window to "Add a new CREOLE repository" and type in the above address.

If this still doesn't work for you, there might be a problem with the web-server. If you're a local, you can ask somebody in the lab for help.

### GATE 2.0

1. Find gate.bat and open it in your favorite text editor. The likely location for the gate.bat file is C:\Program Files\gate\bin. The path could also be like this: C:\Program Files\GATE-2.0\bin .

2. Go to the last line of the file. It should look like this:  

```
start "GATE" "%JAVA%" -Xmx200m -Djava.ext.dirs="%EXTDIR%"
-classpath %CLASSPATH% gate.Main %FLAGS% %1 %2 %3 %4 %5
%6 %7 %8 %9
```
3. Add:  

```
-d http://www.cs.pitt.edu/mpqa/opinion-annotations/gate-annotation-new
after gate.Main and before %FLAGS%.
```
4. The resulting line should look like this:  

```
start "GATE" "%JAVA%" -Xmx200m -Djava.ext.dirs="%EXTDIR%"
-classpath %CLASSPATH% gate.Main -d http://www.cs.pitt.edu/mpqa/
opinion-annotations/gate-annotation-new %FLAGS% %1 %2 %3 %4 %5 %6
%7 %8 %9
```
5. Save gate.bat. If you happened to use something like Word to edit the file, make sure that you save gate.bat as TEXT ONLY!
6. If you've run GATE before and experimented with loading xml-schemas, you may also want to delete your gate.session file.

## Starting GATE:

You must be connected to the internet when you start GATE, or GATE will be unable to load the xml-schemas that specify the MPQA annotation types. However, you do not need to remain connected to the internet as you continue to work in GATE.

**Windows:** Double click on the gate shortcut on your desktop or find it via Start->All Programs->GATE-3.1 (or similar).

**UNIX:** The hard-core way is to cd inside the GATE installation directory and then type "bin/ant run". You may be able to do this more easily, say with a desktop icon, depending on the linux version and your knowledge of linux.

## Loading a Document into GATE

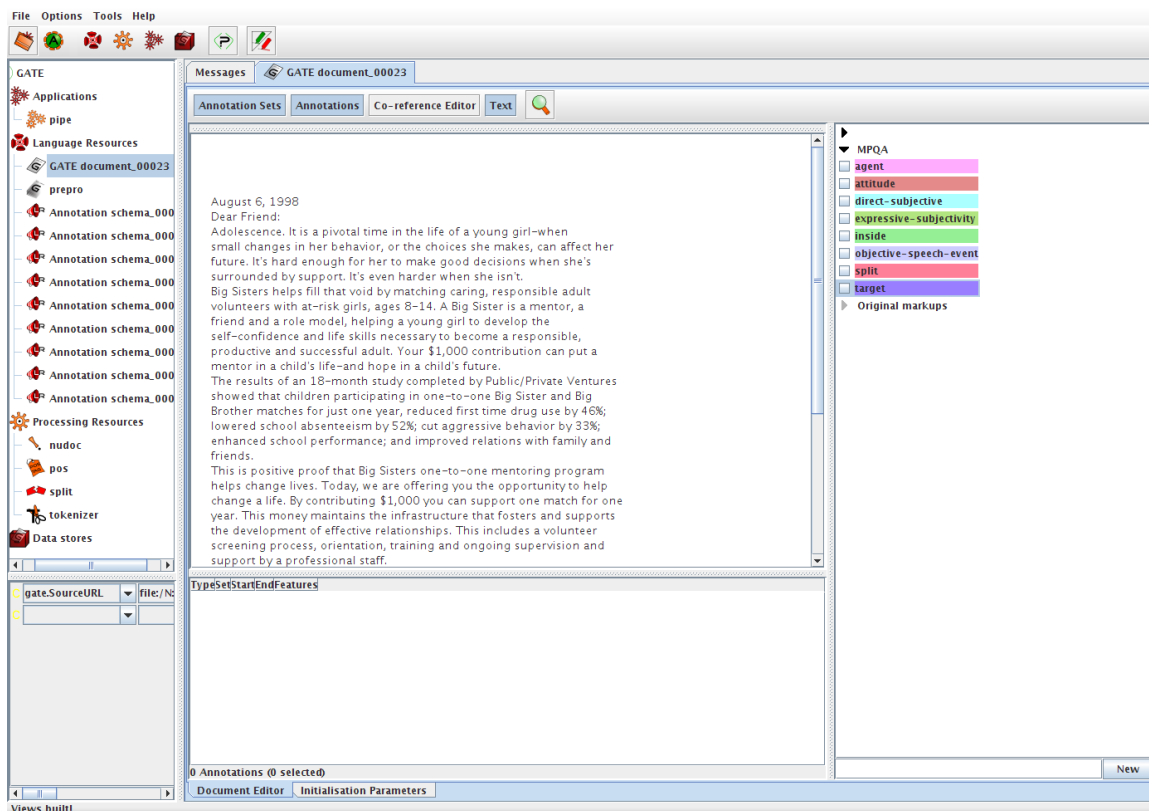
There are two ways to do it:

1. In the left navigation frame, right click on Language Resources -> New -> GATE document.
2. Alternatively, use the menu and go to: File ->New Language Resource-> New GATE document

The "Parameters for the new GATE document" window will open. In the window:

- Leave or set `preserveOriginalContent=true`.
  - Click the open-folder button (end of row beginning with `sourceUrl`).
  - Find the directory containing the file that you will annotate.
  - Select the file that you want to open.
  - Give the file a name ending with your initials. Example: `hr37-taw`.
  - Click OK.
- A gate document with the name that you gave it should show up in the frame on the left side of the GATE window, under Language Resources.
  - Double-click on the xml document that you added under Language Resources. The file will show up in the center frame of the GATE window.
  - Click on the Annotations and Annotation Sets buttons. This will open up the Annotation frame (middle bottom of the GATE window) and the Annotation Sets frame (right side of the GATE window).
  - In the Annotation Sets frame two or three sets of annotations should be listed:
    1. Default annotations
    2. MPQA annotations
    3. Original markup annotations (may or may not be listed)

- If the MPQA annotation set is not listed, type "MPQA" into the text field at the bottom of the Annotation Sets frame and click New.
- GATE should now look much like the image below.



You are now ready to begin annotating the document. If no annotation labels are available when you select text and hover the mouse over the selection, then check that you have set up the correct CREOLE repository at <http://www.cs.pitt.edu/mpqa/opinion-annotations/gate-instructions/startgate.html>

in which the MPQA annotation scheme is defined.

## Some Default Annotations

During a document preparation stage, a number of annotations were added to the document. You can verify that the preprocessing went ok as follows:

- Click on the check box to the left of 'agent' under MPQA annotations. Two zero-length agent annotations, for agents with id=implicit and id=w (writer), will show up in the Annotations frame at the bottom of your GATE window. They will be difficult to see in the upper text box.
- Click on the check box for 'direct-subjective'. You will see one zero-length annotation, starting and ending at 0.
- Hide these annotations from view by unmarking the checkboxes.
- Click on the check boxes for the 'objective-speech-event' and 'inside' annotation types. Default 'objective-speech-event' and 'inside' annotations were added for the writer only. These annotations should now be listed in the Annotations frame at the bottom of the screen.
- Click on the "Start" column heading in the Annotations frame. This will sort the listed annotations in ascending order by starting byte. The initial default annotations of 'objective-speech event' for the writer are zero-span annotations at the beginning of each sentence. Each 'inside' annotation for the writer spans an entire sentence (or at least a sentence as delineated by GATE.)
- Select an 'inside' annotation from the list. The span for that annotation will flash in the document window.
- Hide these annotations from view by unchecking the 'on' and 'in' checkboxes.
- There are also 'split' annotations added by GATE's sentence splitter. Show these now by clicking on the 'split' checkbox. On what to do about bad sentence splits, Pitt users may consult the annotation FAQ on the Pitt Wiki.

## General Instructions for How to Annotate a Span

- Highlight the span of text that you want to annotate. Make sure that you do NOT accidentally include any spaces at the beginning or end of the span of text you are annotating.

**EXAMPLE:** "China" in the sentence,

"**China** said on Tuesday a U.S. State Department report that accused Beijing of suppressing religious freedom was full of lies and urged Washington not to hold double standard in the war on terrorism."

- In the frame that pops up, go to the scroll list at the very top and select an annotation type. In our example, we want to select 'agent'. As a result, the frame should look like this:



- To make certain that the annotation type is properly selected, you should actually click on the highlighted word in the text rather than just hitting the return key on your keyboard.
- If you select an annotation type that's different from the first one listed, double check that the desired label and only the desired label appears over the text span. Sometimes the software responds so quickly that it will apply a label for the first-listed annotation type before you have a

chance to select the one you really want. If that happens, remove the undesired extra label later.

- Start filling in the attributes for the annotation frame you're working on. This may involve selecting from a list (in the agent frame, the agent-uncertain field is a drop-down list) or typing in information (in the agent frame, you need to type in source or target ids)

## General Instructions for How to Add Features to an Annotation

**For GATE 3.1 and 4.0, follow these instructions:**

1. You can either scan the text in the GATE editing window or find a particular annotation in the Annotations frame at the bottom of the GATE window. In the latter case, you can scroll and sort by Type and Starting byte number to help you find the correct annotation. You can also use the arrow keys to maneuver.
2. When you have found the right annotation to edit, hover the mouse over it and the annotation frame for it will pop up. If there are two or more annotations covering the text over which your mouse is hovering, then select the type of annotation that you want to edit. Its annotation frame will then open.
3. If the feature you want to specify is not listed in the annotation frame, you can type its name into an empty field at the bottom of the frame that has a yellow letter "C" to its left. Once you hit return after typing in the name, the feature will appear in its proper alphabetized place within the feature list. Look for it and then specify the value. If you expect a drop down list with legal feature values, but it doesn't show up, type in the feature value manually.
4. Repeat the above for each feature that you want to add.
5. Click the Dismiss button in the top right corner of the annotation frame when you are finished.

6. Check that the features for the new annotation now appear under Features for the annotation in the Annotations frame of the GATE window. If they don't show up right away, select a different annotation and the feature list will refresh.

**In GATE 2.0, use the instructions below:**

1. Find the annotation in the Annotations frame at the bottom of the GATE window. You can scroll and sort by Type and Starting byte number to help you find the correct annotation. You can also use the arrow keys to maneuver.
2. Right click on the annotation, select Edit. The Edit Annotation window will open. (You can also double click on the annotation to edit it.)
3. Select a feature from the Possible features list on the left side of the window and click the << button to move the feature to the Current features list.
4. Set the value for the feature. Depending on the feature, you can either select the value from a pull-down list, or you type in the value and hit [enter].
5. Repeat steps 3 and 4 for each feature that you want to add.
6. Click the OK button when you are finished.
7. The features for the annotation will show under Features for the annotation in the Annotations frame, at the bottom of the GATE window. If they don't show up right away, select a different annotation and the feature list will refresh.

## **Saving a Document**

Save your document reasonably often as you annotate. GATE has no auto-save feature!

1. Right click on the document name under Language Resources

OR

right click on the appropriate tab in the list of open documents at the top of the middle frame.

2. Select: Save As XML.
3. Type in the file name that you want to give it. Example: hr7-taw.xml. Make sure that it was saved with an .xml extension.
4. When you are completely done with your annotations and have saved the document for the last time, you may want to try closing the document in GATE (right-click -> Close), and reopening it to check that all of you annotations were saved properly.
5. Please rename your completed, final annotated document. Unless instructed otherwise, use the original document name, extended by your login (or initials) and the word "final":  
e.g. CWY098.josefr.final.xml

Finally, click on the Messages tab to see if there was an error saving the file.

## Other recommendations

### Accidental edits of the XML document

In the context of the Pitt group, one thing to be careful about when performing annotations in GATE is that you might click into the text area and introduce characters or white space.

This causes problems when other people are annotating the same documents in parallel and one wants to perform an automatic comparison of the two annotations. It also could cause a problem if the additional material gets introduced after the gate\_default file that stores the tokenization for the xml document was created and is not updated.

The upshot is: be really, really careful not to modify the original text!