**STEP 1 – CREATE A WORKING DIRECTORY:**

Create a working directory (mine is text\_mining\_project/). Ensure you have the following sub-folders (I have sent you an example that is pre-prepared 😊)

* mate\_tools\_working/
  + anna/
  + transition/
  + parser\_tagger/
* tmv\_tool/
* data/
  + corpus/
  + sent\_corpus/
  + sent\_ref\_corpus/
  + sent\_dates\_corpus/
  + corpus\_prepared\_for\_mate/
  + mate\_parsed\_corpus/
    - split/
    - parsed/
  + merged\_data/
* output/
  + ecb\_speeches/
    - tmv/
      * raw/
      * xml/

**STEP 2 – Save Down Your Corpus:**

Save this in data/corpus/

**STEP 3 – Run the Python Processing codes – be sure to change where the working directories are:**

1. **PYTHON RUN:** python run\_SQA\_1a\_preprocessing\_minimal.py (updating manually the parent filepath at the start. This code populates sent\_corpus, and renames the files to a consistent format)
2. **PYTHON RUN:** Run: **python run\_SQA\_1b\_reference\_drop\_loop1.py –** this will populate another folder sent\_ref\_corpus
3. **PYTHON RUN: python run\_SQA\_1c\_preprocessing\_for\_mate\_loop.py -** This populates "corpus\_prepared\_for\_mate" which you should have made earlier.
4. **PYTHON RUN: run\_SQA\_1ca\_dates\_preprocessing\_for\_sutime.py** – This creates the input files for SUTime, and populates “sent\_dates\_corpus”.

**STEP 4 - Set Up TMV and MATE**

**FIRST SET UP MATE:**

* Download the file **anna-3.61.jar** from [here](https://code.google.com/archive/p/mate-tools/downloads) and put it in a folder called **mate\_tools\_working/anna/**.
* Download transition-1.30.jar from [here](https://code.google.com/archive/p/mate-tools/wikis/ParserAndModels.wiki) and put it in a folder called **mate\_tools\_working/transition/**.
* Download the parser+tagger from [here](https://code.google.com/archive/p/mate-tools/wikis/ParserAndModels.wiki), under “English Models”, which is a .mdl file called **per-eng-S2b-40.mdl** and put it in a folder called **mate\_tools\_working/parser\_tagger/**.
* Download the parser .csh script from [here](https://code.google.com/archive/p/mate-tools/wikis/ParserAndModels.wiki), which is a tiny example script called **parse-eng**, and put it in **mate\_tools\_working/parser\_tagger/**.

**THEN SET UP TMV:**

* Create a different working directory called **tmv\_tool/**. Download the TMV tools (from the GitHub repository [here](https://github.com/aniramm/tmv-annotator)) into this folder and unzip. There should be folders called europarl, example-outputs, and tmv-annotator-tool. There should also be .gitignore, LICENSE, README.md.
* To use the command line, you need to make sure the folder containing python.exe on your computer is added to your Path.
* Note that you need the de-bugged version of the English variant of the tool. You should place the file TMV-EN\_ecb\_test\_david.py (found in **./supplements/** in the directory **./data/tmv\_tool/tmv-annotator-tool/**).
* You also need the correct version of the bash file. Place the file **annoTMV\_2\_loop.sh** in the same directory.

**Run the MATE Parser**

Now **run the parser!** - run **run\_SQA\_mate\_loop\_full.sh**, and ***edit the file to make sure the correct directories are pointed at***. The code is only a couple of lines long and you should be able to figure out whats going on. Two folders are populated - **parsed and tagged.**

**STEP 5 – Run TMV**

**Run “annoTMV\_2\_SQA\_loop.sh”**

*Due to a permissions error in bash on my pc, I added the line "alias python='winpty python.exe'" to my code. this is only for windows however. and due to some issues on my end.*

This populates the folders /output/tmv/raw and /xml with our TMV data. A side note from robs code is that the *code does not automatically throw out the tmv stuff.*

**STEP 6 – Run SUTime – We will discuss this when you have run TMV**

My advice is to familiarise yourself with the Stanford corenlp.

The should be ok – the youtube video I sent is quite nice!

In addition, Rob has compiled a guide to all this, that I have also attached.