OLANTO

myMT v2.0

Installation Manual for the Translation (Back-End) Server

CONTENT

1. Context

myMT is a machine translation system based on the open source Statistical Machine Translation (SMT) engine called Moses[1]. In order to turn Moses into a production system, it was necessary to add several tools:

- A preparation system called **myMT Preparator** which turns bi-text documents into translation memories; and
- A management system called **myMT Manager** to cut a document to be translated into sentences, distribute the sentences over the various translation nodes available, re-build the final translated text, and manages the translation nodes in the various language pairs and directions.

myMT requires two application servers to run: a Back-End server running Ubuntu (for Moses), and a Front-End server running Windows (for myMT Preparator and myMT Manager).

This document is the installation manual to set up the Back-End (Ubuntu) server.

2. Installation Procedure

This procedure aims at simplifying as much as possible the installation of the Moses SMT engine in a virtual machine, and to make that virtual machine re-usable by allowing its replication. Thus it will be possible to have a large number of translation engines running in parallel if the appropriate hardware is available.

It is of course possible to install Ubuntu natively on a server (instead of encapsulating it in a virtual environment) but this will not facilitate its replication when more translating power is needed.

First you should download from Olanto's web site the following file:

- Back_End.zip

2.1. Creating the Ubuntu 14.04 64-bit Virtual Machine

(This section refers to Ubuntu 10.04, but the procedure is exactly the same for Ubuntu 12.04 and Ubuntu 14.04.)

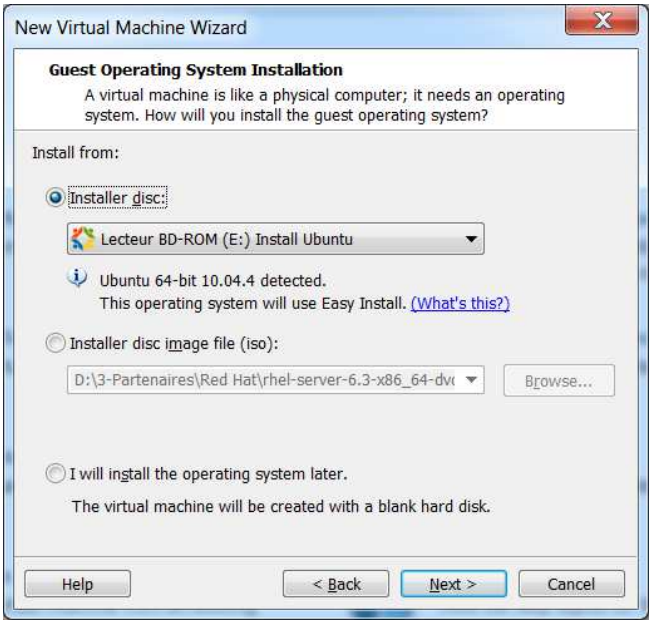
The first step in installing Ubuntu is to create a virtual machine (VM) with the required minimum hardware resources. The VM building environment will install Ubuntu while creating the VM.

The VM environment used in this manual is VMware Workstation v9.0. However any other VM environment may be used too (we tested it also under VirtualBox).

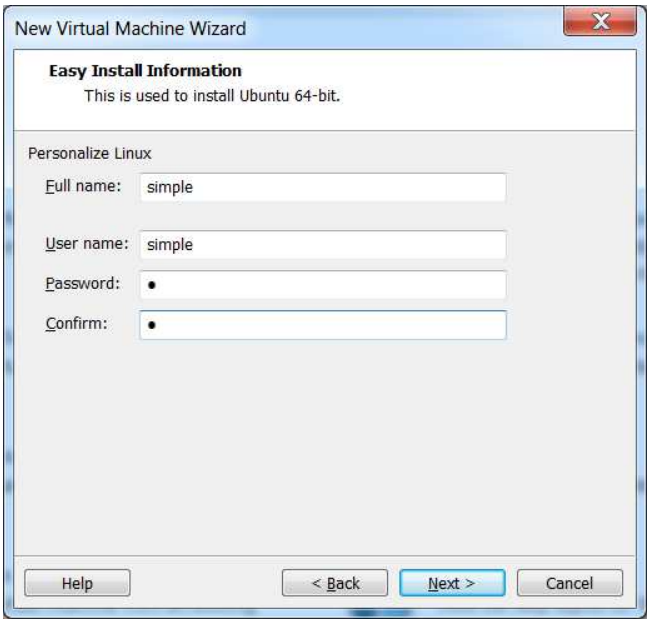
- Start VMware Workstation 9 and choose Menu, New Virtual Machine (CTRL+N). This launches the VM creation wizard. Choose the "Typical" mode and click on "Next":

[1] http://www.statmt.org/moses/

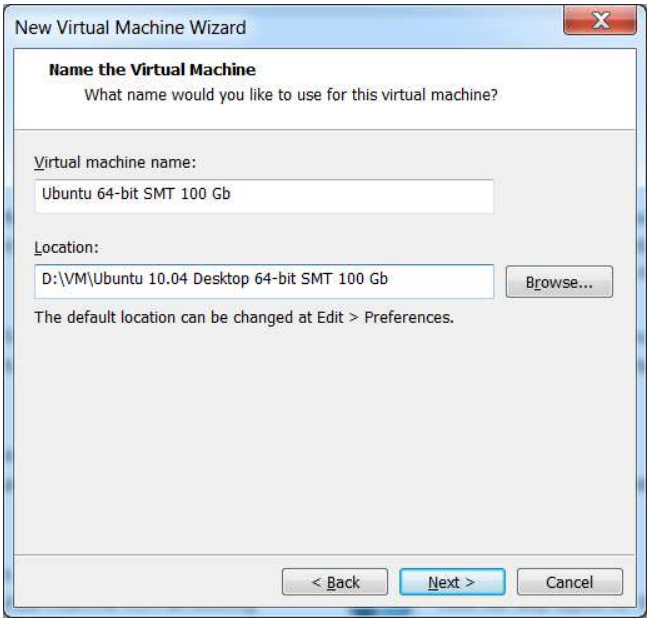
- You must define which operating system you want to run in the new VM. In the example below we provided an ISO image of the Ubuntu Desktop 10.04 installer on a CD-ROM; alternatively you can choose to provide the ISO file on the hard disk of your host computer. When the OS is defined click on “Next”:



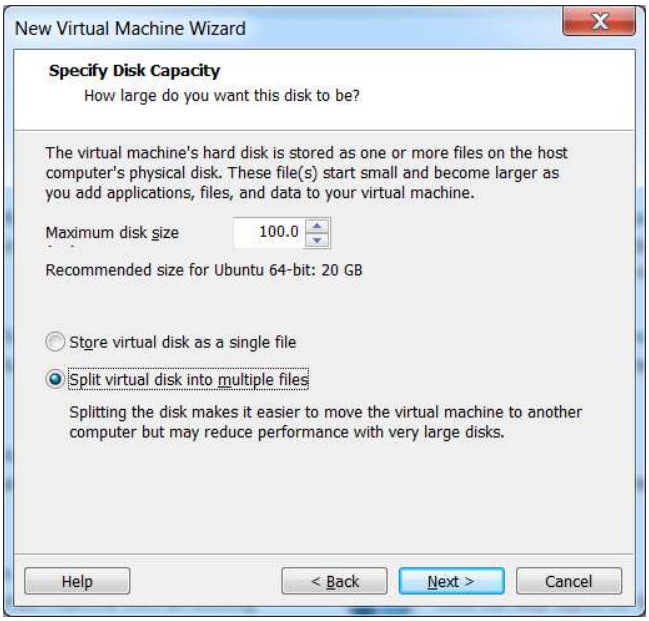
- In the next box enter “simple” (no capitals) as both the Full name and the User name, and choose a password, then click on “Next”:



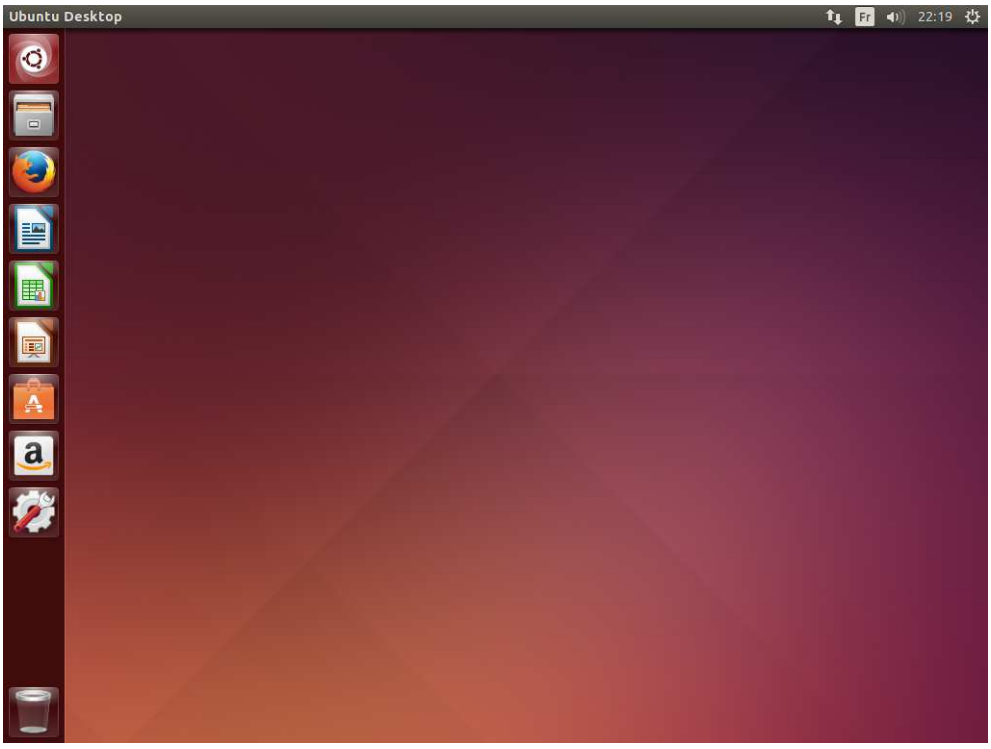
- Give a name to the VM and define the location where the files will be created. Remember to create a dedicated folder beforehand (in this example it is called “Ubuntu 10.04 Desktop 64-bit SMT 100 Gb” and it is located on the D:/VM path). Then click on “Next”:



- Define the disk capacity (typically 20 Gb for a test application, and at least 100 Gb for a production system – it all depends on the volume of the training corpus) and choose “Split virtual disk into multiple files” before clicking on “Next”:



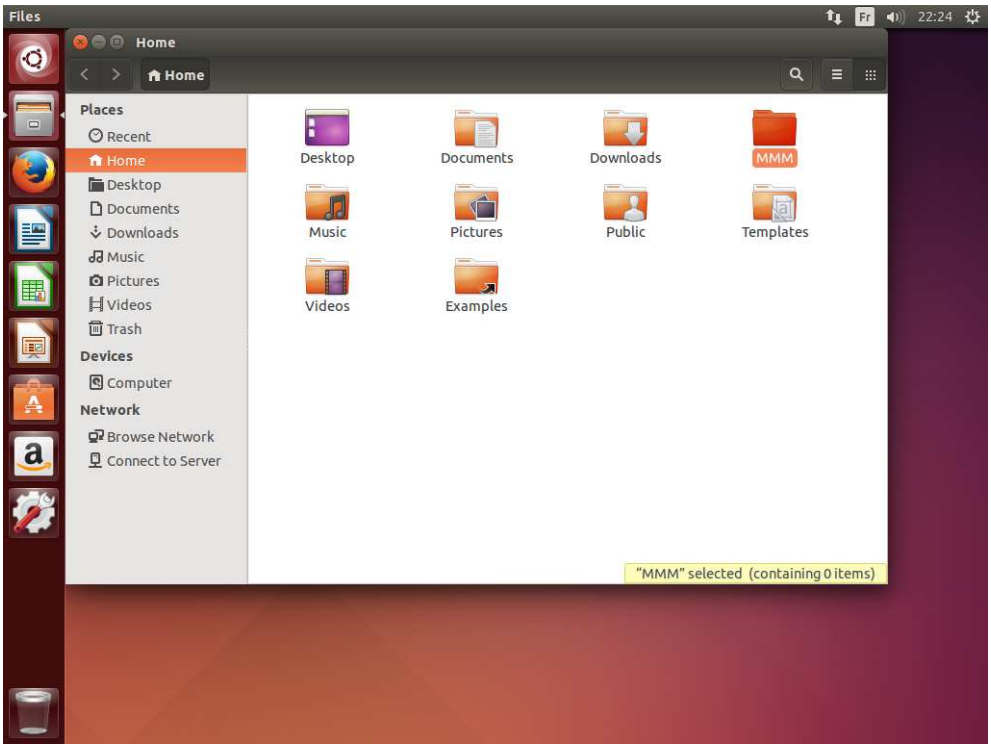
- At this stage the installation process is launched (this does take some time). When it is completed you will see the default Ubuntu Desktop 14.04 environment:



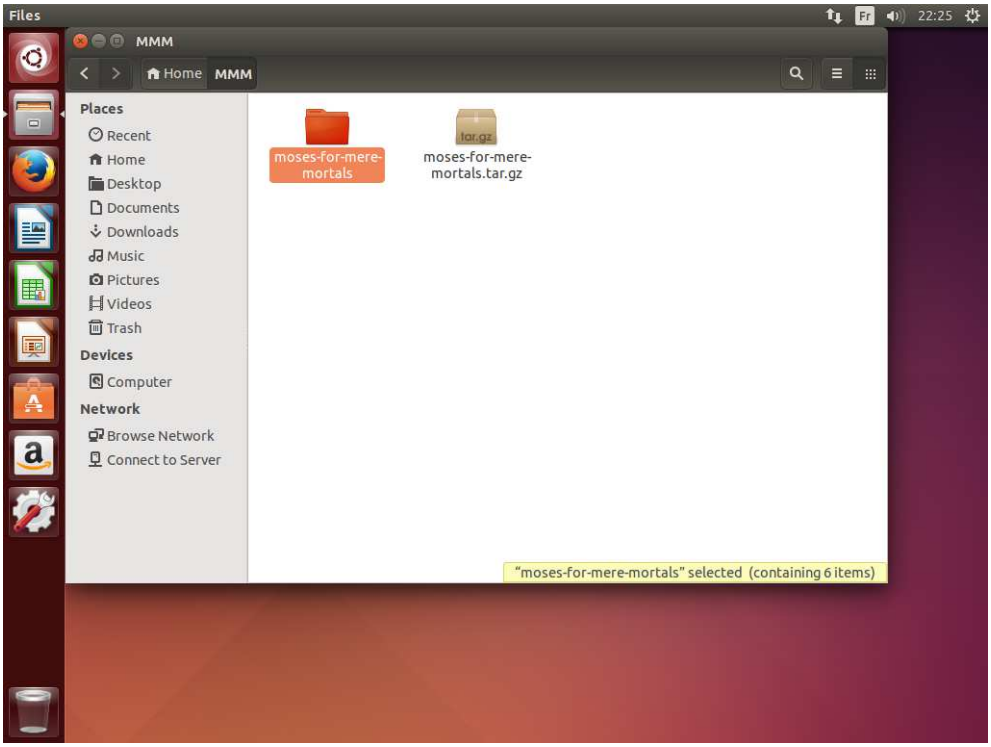
- Change the keyboard setting if necessary (the default one is US-English)
- Give a fix IP address to your VM and check out the web connection
- Shutdown Ubuntu. Then in the VMware menu, go to the Edit, Preferences (CTRL+P) menu and define the size of the RAM and the CPUs available for the VM. For a test version it should be at least 4 Gb of RAM and 4 CPU cores. For a production system there should be at least 8 Gb of RAM and 8 or 16 CPU cores.
- Restart Ubuntu.

2.2. Installing Moses

- Under the “simple” folder, create an “MMM” folder (MMM stands for “Moses for Mere Mortals”[2] which is a set of simplified installation scripts for Moses):

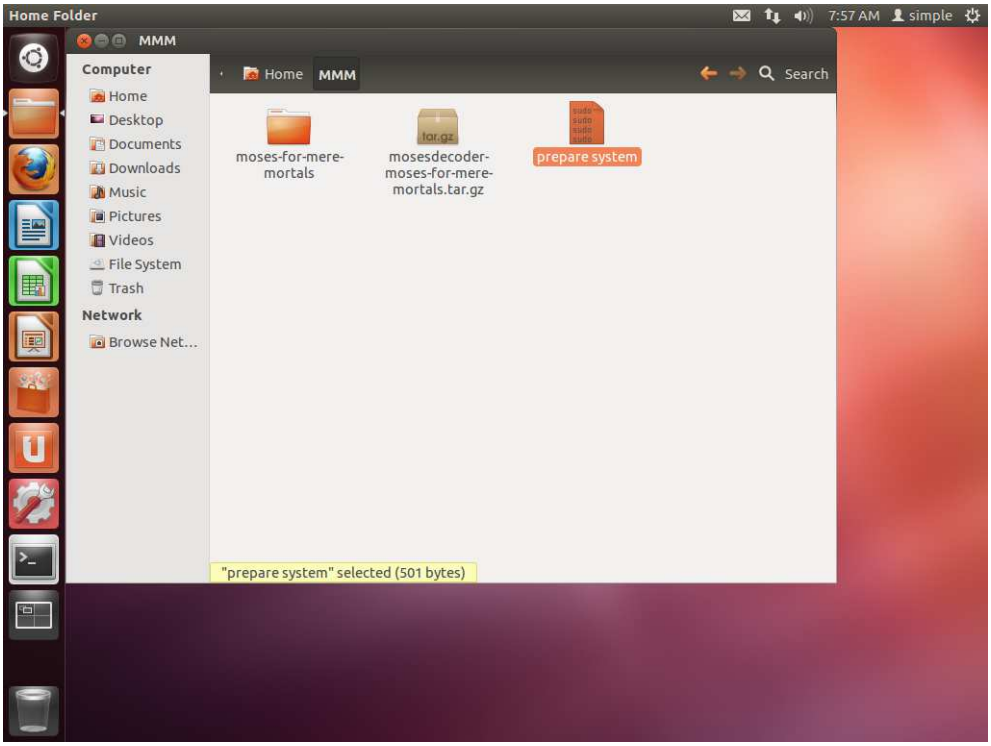


- From the “1-install moses” installation package, copy the “mosesdecoder-moses-for-mere-mortals.tar.gz” file to the MMM folder and extract it at the same location:

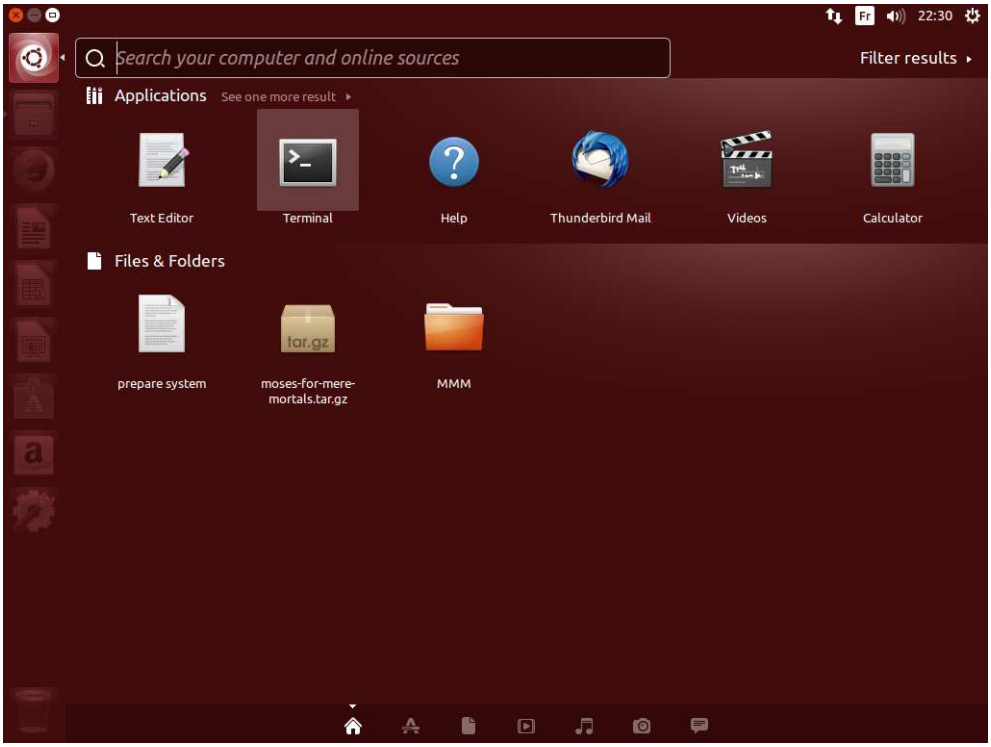


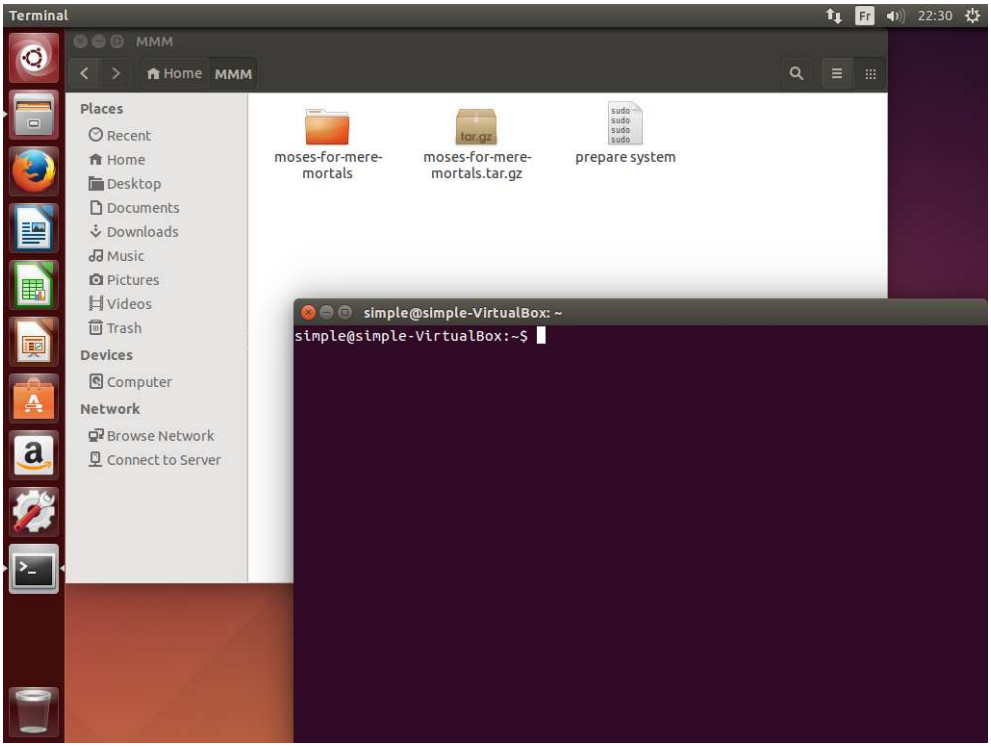
[2] http://code.google.com/p/moses-for-mere-mortals/

- From the same “1-install moses” installation package, copy the “prepare_system” file into the MMM folder:

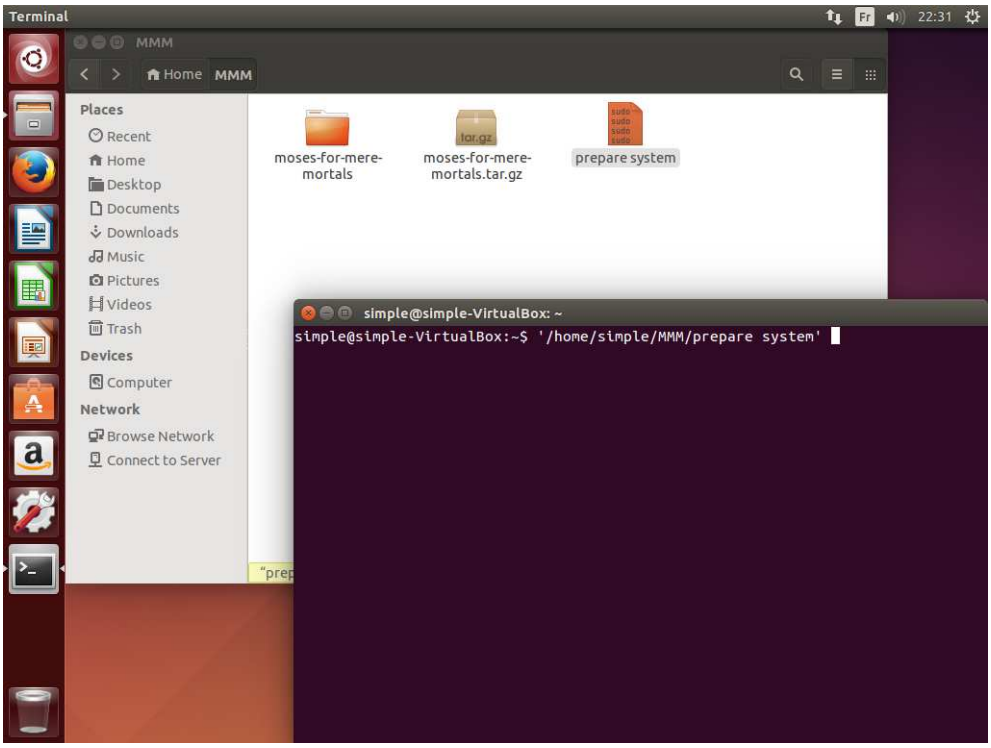


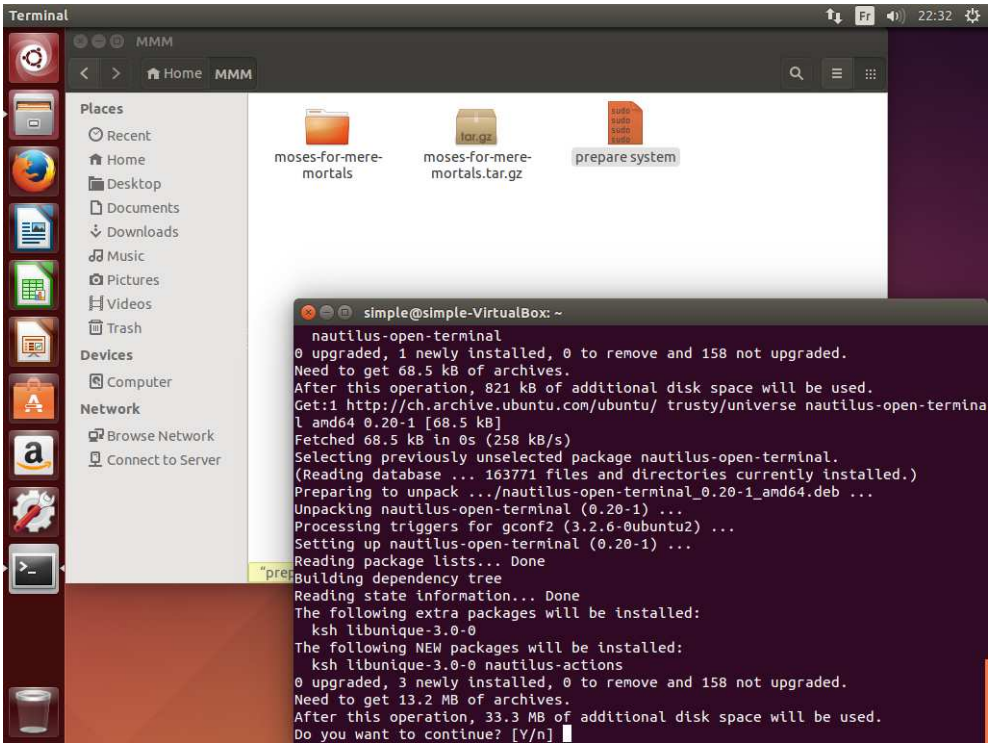
- Click on the Ubuntu icon at the top of the side bar and select “Terminal”:





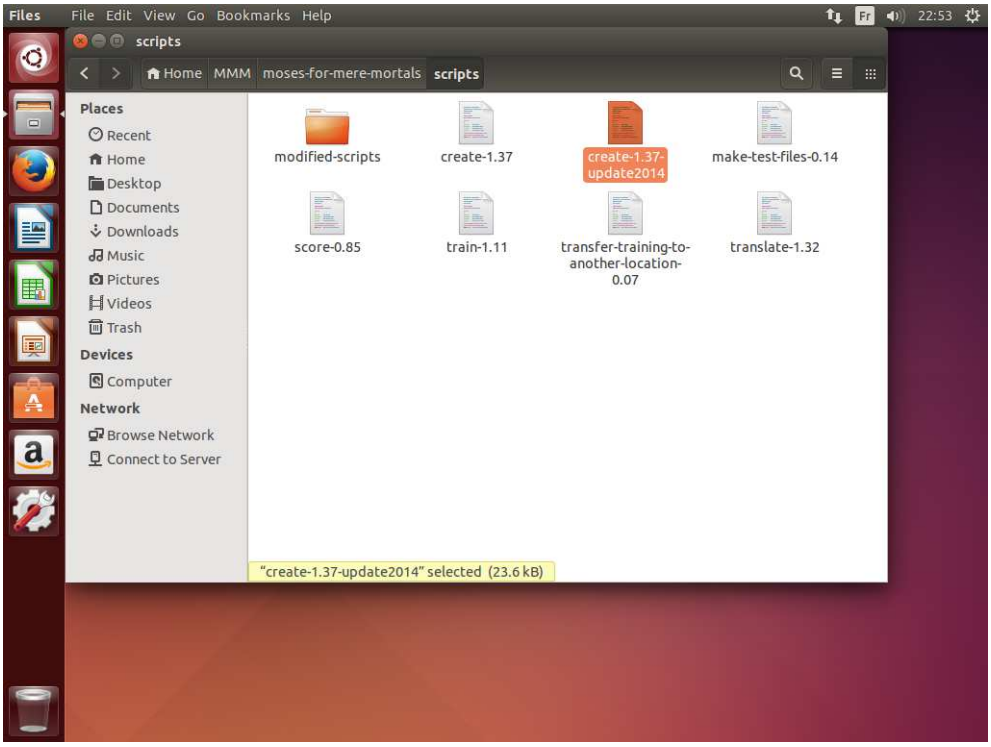
- Execute the “prepare_system” file by dragging and dropping it into the terminal window. Answer yes to all questions:



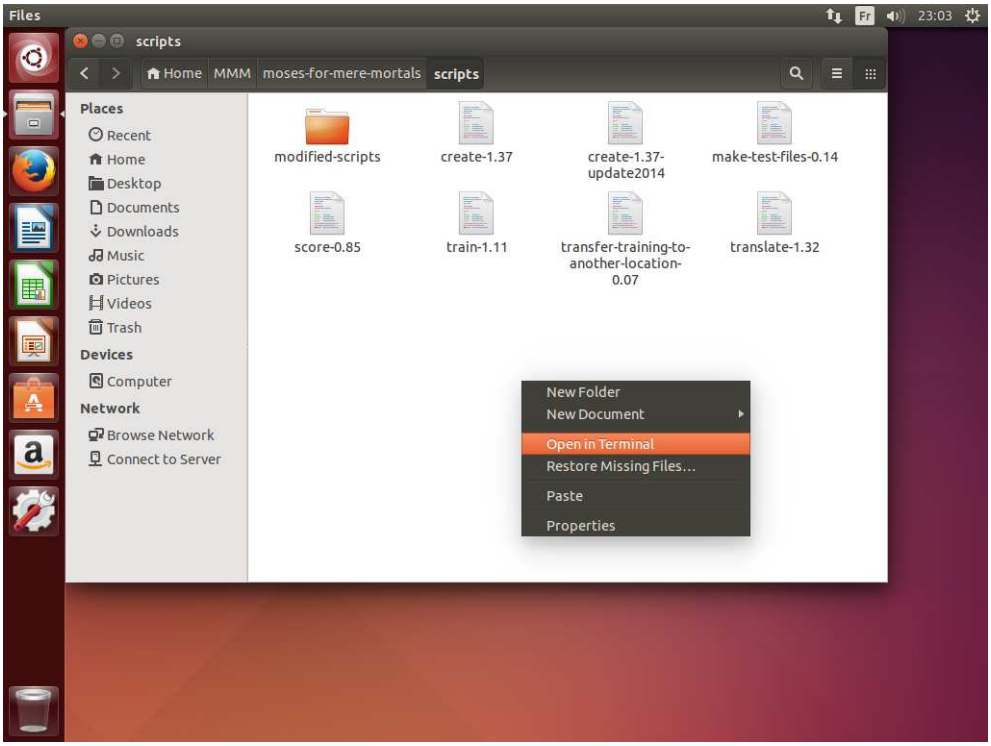


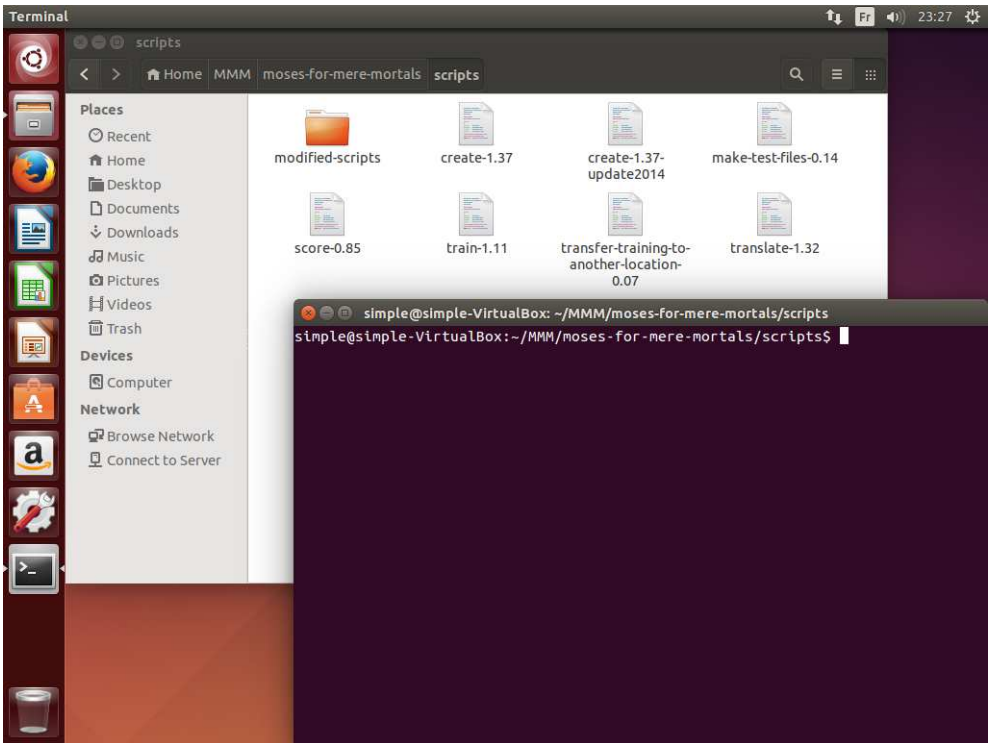
Wait until all the scripts are completed. This operation takes quite some time.

- From the “1-install moses” installation package, copy the “create1.37-update2012” file into the MMM/moses-for-mere-mortals/scripts/ folder:

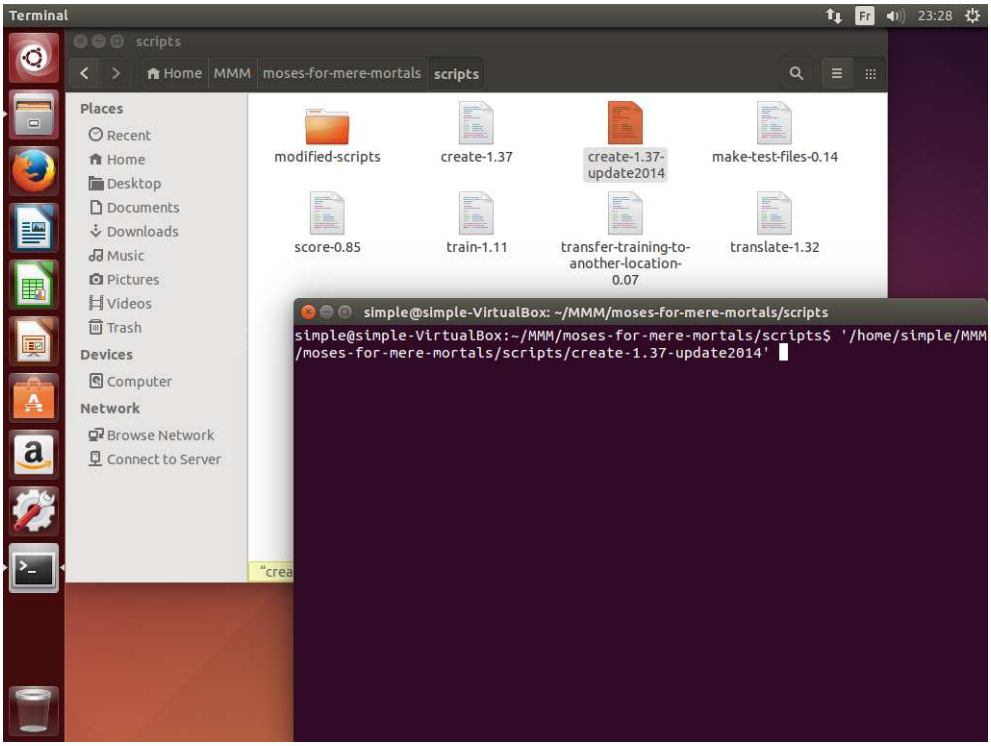


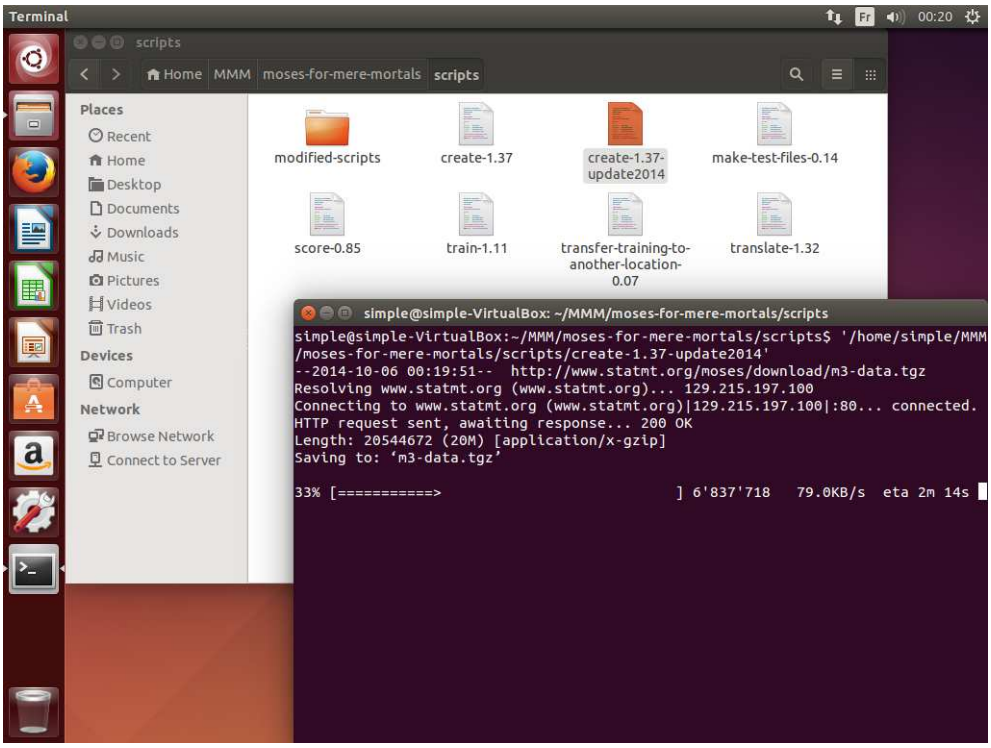
- Right-click anywhere in the folder window and choose “Open in Terminal”:





- Execute the “create1.37-update2012” file by dragging and dropping it into the terminal window:





Wait until all the recompilation tasks are completed. This is also a rather lengthy process.

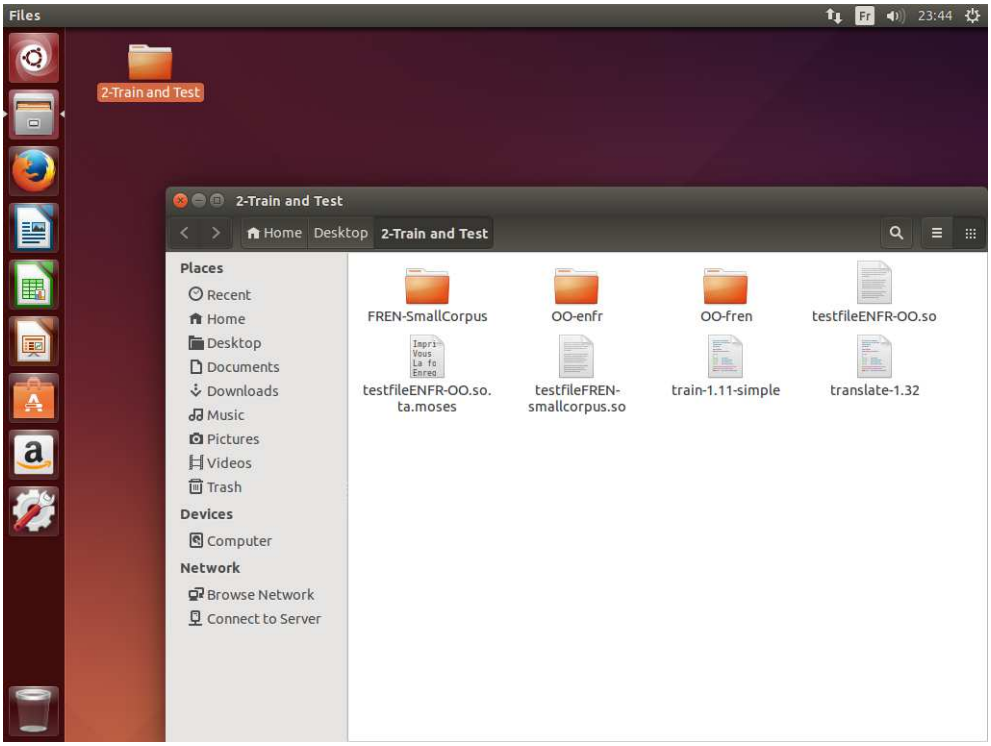
- Check out that a new folder called “moses-irstlm-randlm” was created (with all its contents) into the root of the “simple” account.

2.3. Training and Testing Moses

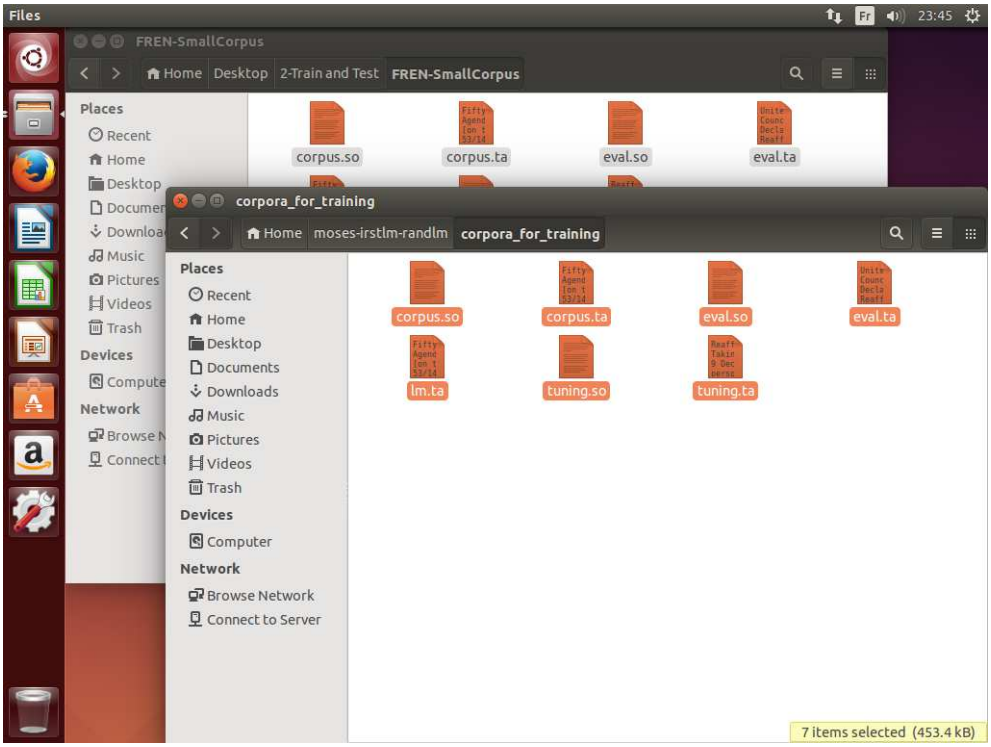
This procedure is based on the assumption that your VM has enough resources to build a complete translation model. If such is not the case you should raise the RAM and CPU capacities.

The training and testing documents used in this example were copied from the Open Office documentation in English and French. This demo version is meant to show how an application like myMT may be used to help translate or localize software functions or elements of a Graphic User Interface. Of course the training and testing material may be replaced with your own documents in source and target languages.

- Copy the entire installation package called “2-Train and Test” to the desktop and open it:

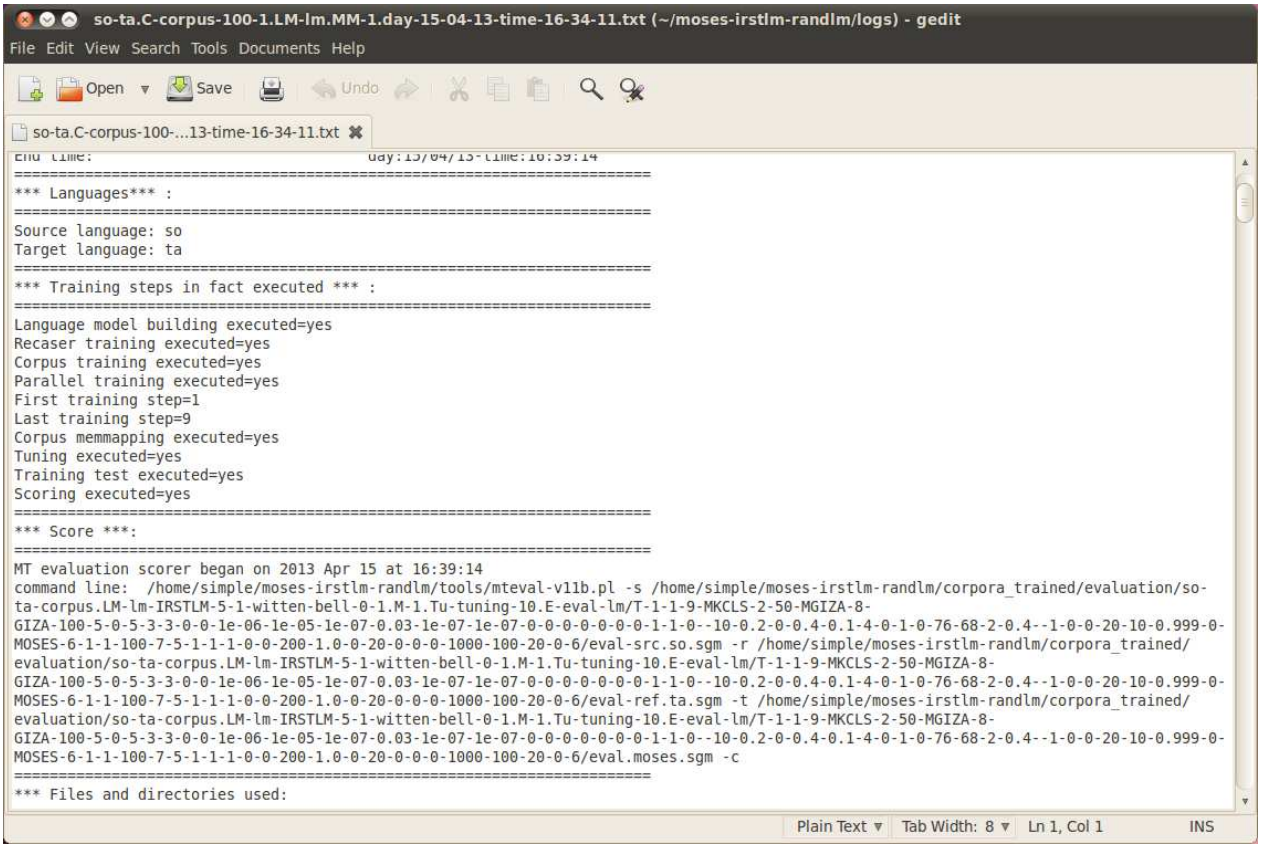


- Prepare the training corpus:
    - Empty the ~/ moses-irstlm-randlm /corpora_for_training/ folder
    - Copy the content of the “FREN-SmallCorpus” folder into ~/ moses-irstlm-randlm /corpora_for_training/ :



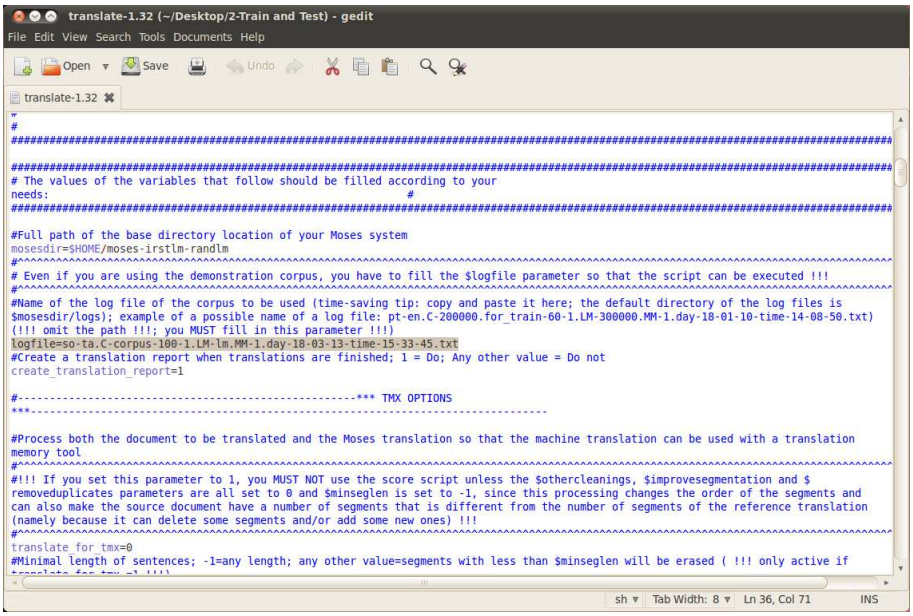
- Train the translation model:
    - Execute the “train-1.11-simple” file located in the simple/Desktop/2-Train and Test folder and choose “Run in Terminal”.
    - Wait until the training operations are completed. This may take about 10 minutes.
    - All the models are located in ~/ moses-irstlm-randlm /corpora_trained/
    - All the logs are located in ~/ moses-irstlm-randlm /logs/

- To check out the BLEU score, edit the file “so-ta.C-corpus-100-1.LM-lm.MM-1.day-[…]time-[…].txt” which is located in ~/moses-irstlm-randlm /logs/. You can view its content by editing the file with the “gedit” application available in Ubuntu:



- To translate a test file:
    - Copy the “testfileFREN-smallcorpus.so” file, which is located in the simple/Desktop/2-Train and Test/ folder, into the ~/moses-irstlm-randlm /translation_input/ folder *(this test file includes a number of English messages from the Open Office application. You can view its content by editing the file with the “gedit” application)*.
    - Edit (with the gedit application) the “translate1.32” file which is in the simple/Desktop/2-Train and Test/ folder.
    - Find this line, which contains the path to the log:

      `logfile=so-ta.C-corpus-100-1.LM-lm.MM-1.day-18-03-13-time-15-33-45.txt`



    - Replace “logfile=” with so-ta.C-corpus-100-1.LM-lm.MM-1…. by the name of your log to specify the model.
    - Execute the “translate1.32” file (Run in Terminal).
    - The result is located in the ~/ moses-irstlm-randlm /translation_output/ folder. Note that models not usable in prod. Used only to check system works.

2.4. Building Up a Real-World Example

This example is also based on the Open Office documentation (see http://opus.lingfil.uu.se/) which is a small corpus.

- Delete the folder ~/ moses-irstlm-randlm /logs/
- Delete the folder ~/ moses-irstlm-randlm /corpora_trained/
- Empty the folder ~/ moses-irstlm-randlm/corpora_for_training/
- Copy the content of folder simple/Desktop/2-Train and Test/OO-fren into ~/ moses-irstlm-randlm /corpora_for_training/
- Train the translation model by executing the simple/Desktop/2-Train and Test/train-1.11-simple file (“Run in Terminal”).

The newly-built language and translation models should be saved so they can be used later:

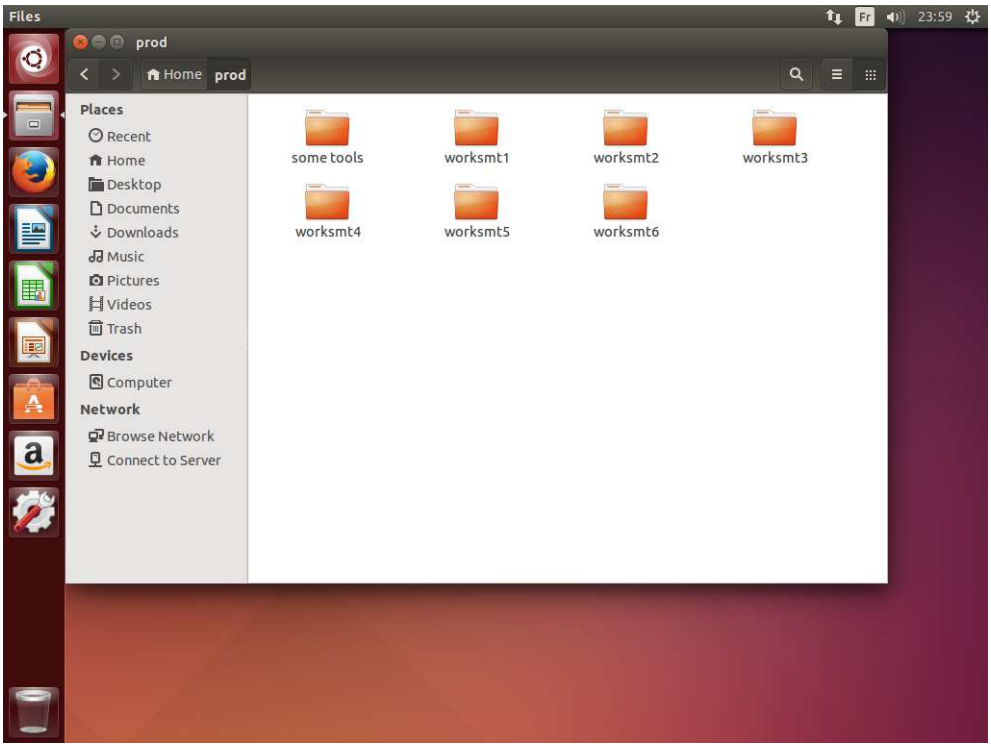
- On Ubuntu’s Desktop, create a new folder called “OO-fren-Models”.

- Copy the content of ~/ moses-irstlm-randlm/logs/ into that new folder.
- Copy the content of ~/ moses-irstlm-randlm/corpora_trained/ into that new folder.
- Compress the new folder (right-click, choose “Compress…”, then “Create”). You can then delete the “OO-fren-Models” folder.

2.5. Installing Production Model

The structure for the 6 possible models should first be installed:

- A translation model = phrase model + recaser + language model
- Basic installation allows 6 different models. By default, models 1 and 2 are filled with OO models ENFR and FREN From the “3-install prod” installation package, copy the file “prod.tar.gz” into Ubuntu’s “simple” folder
- Extract the file into that same location, it becomes a “prod” folder as shown below:



- You can now delete the copy of the “prod.tar.gz” file in order to save disk space.

The system’s translation and recasing capacity is ready and can be tested for SMT1 and SMT2 (models 1 and 2 are filled with OO models ENFR and FREN).

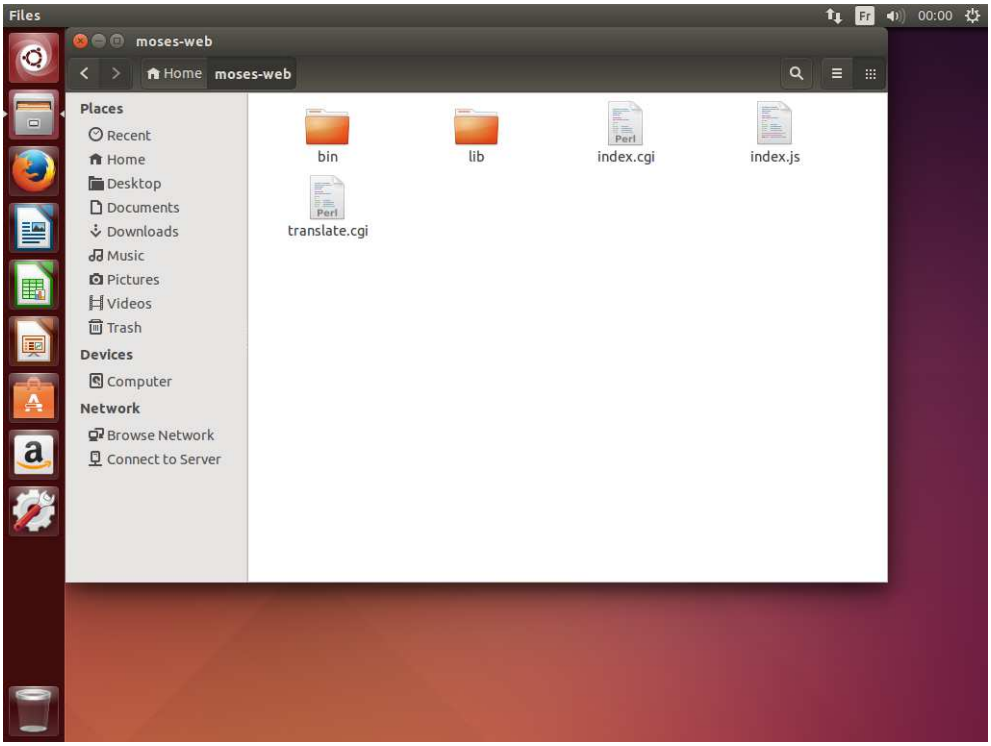
Edit “worksmt1/Moses Testing”, paste the commands in Terminal and check.

2.6. Installing the Daemon

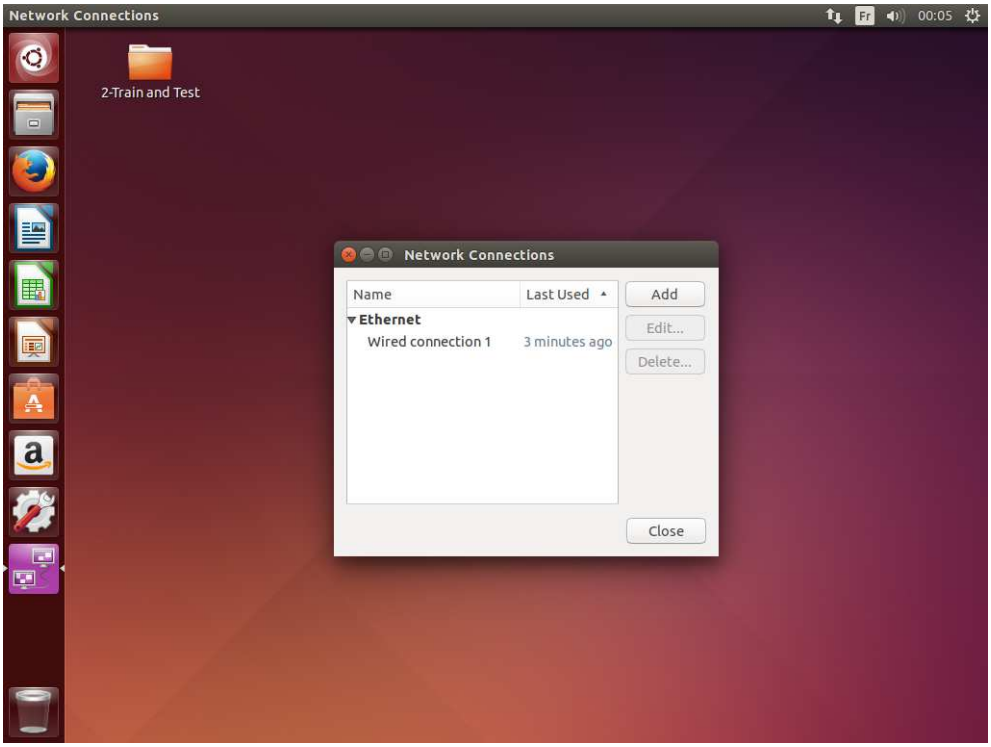
The daemon is a program which constantly checks out whether a translation node (*i.e.* a translation engine) is alive and available.

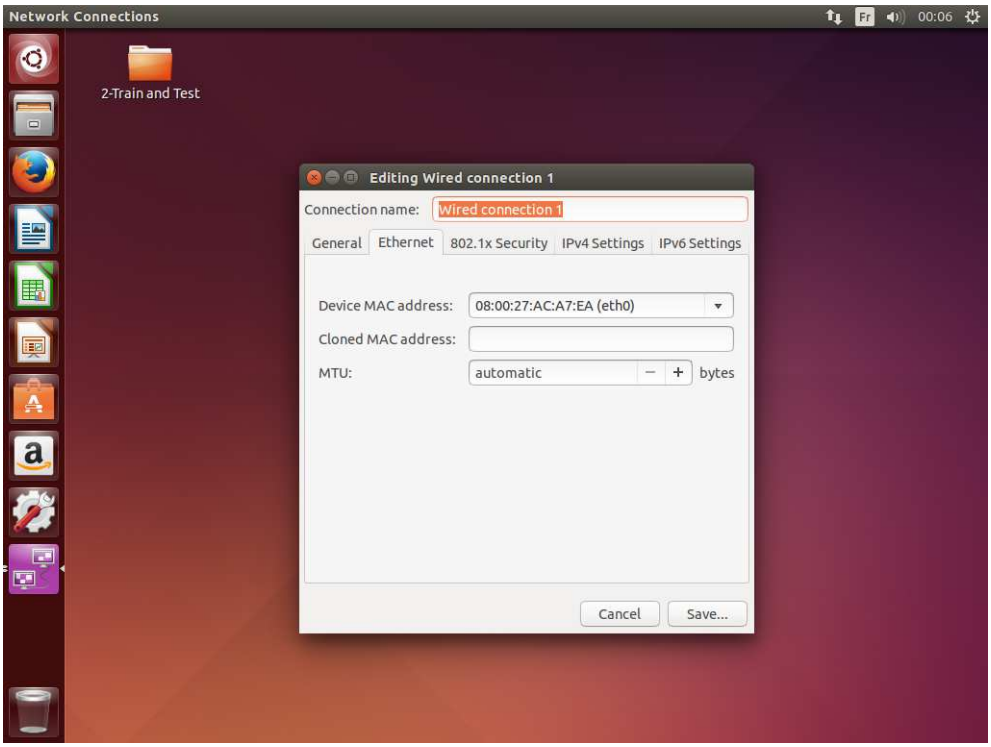
- From the “4-install daemon” installation package, copy the file “moses-web.tar.gz” into the “simple” folder.

- Extract the file into that same location, it becomes a “moses-web” folder as shown below:

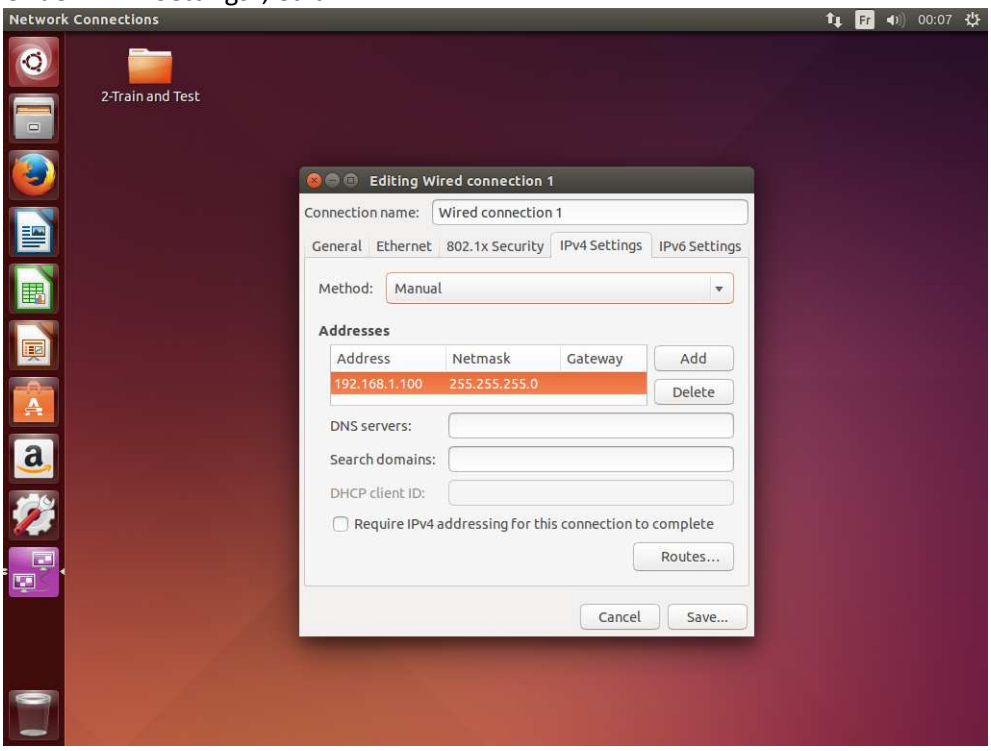


- You can now delete the copy of the “moses-web.tar.gz” file in order to save disk space.
- Set a fix IP address (bridged mode in VMware; in our example that IP is 192.168.1.100):





Under “IPv4 Settings”, edit:



Click on “Save…”
This is only an example!

- Test the new connection: ping the IP from another computer.

- Test the daemon (in prod/some tools/test daemon) edit the file, paste the first two commands into a terminal, then the third command into another terminal. The translation should appear in the first terminal.

2.7. Installing the Local Master

The Local Master is the program which checks out the daemon status and distributes the translation tasks.

2.7.1. Installing Java 1.6.0.27

The default Java JRE must be installed (normally it should be OpenJDK) by opening a Terminal session and typing the following command:

```
# sudo apt-get install default-jre
```

Enter the password you created for the “simple” account and type “Y” when prompted for confirmation.

After the Java installation is completed, it can be tested with the following command:

```
# java -version
```

The reply should be a message like this:

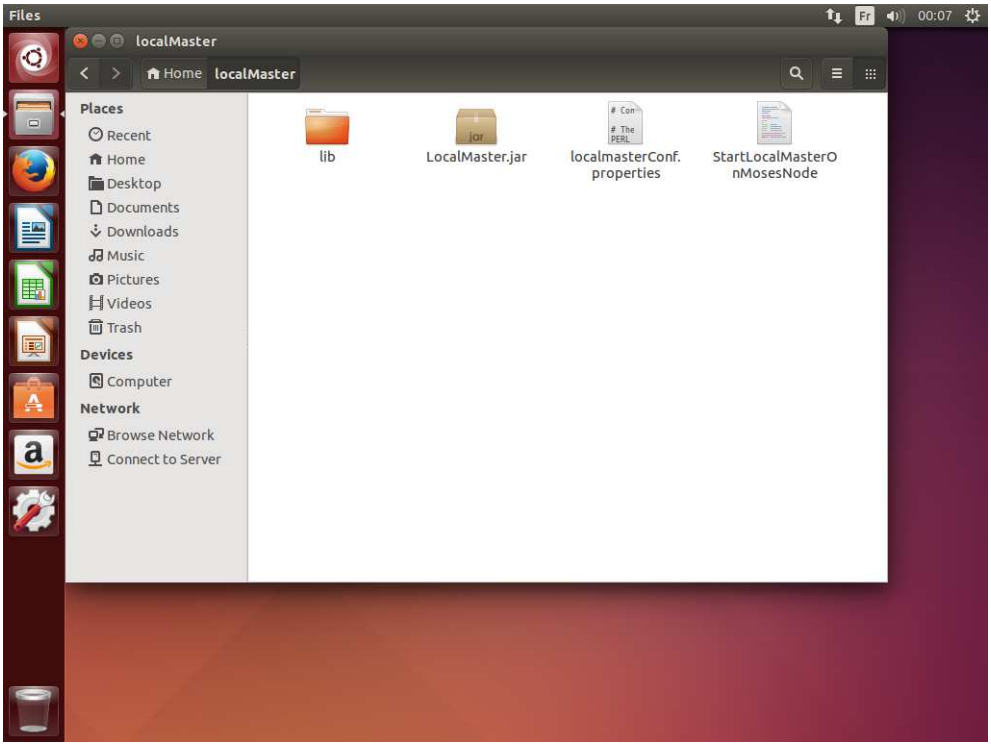
```
java version "1.7.0_65"
OpenJDK Runtime Environment (IcedTea 2.5.2) (7u65-2.5.2-3~14.04)
OpenJDK 64-Bit Server VM (build 24.65-b04, mixed mode)
```

2.7.2. Deploying the Local Master

The location and the name of the Local Master Jar file are not important as long as they are defined in the script variables.

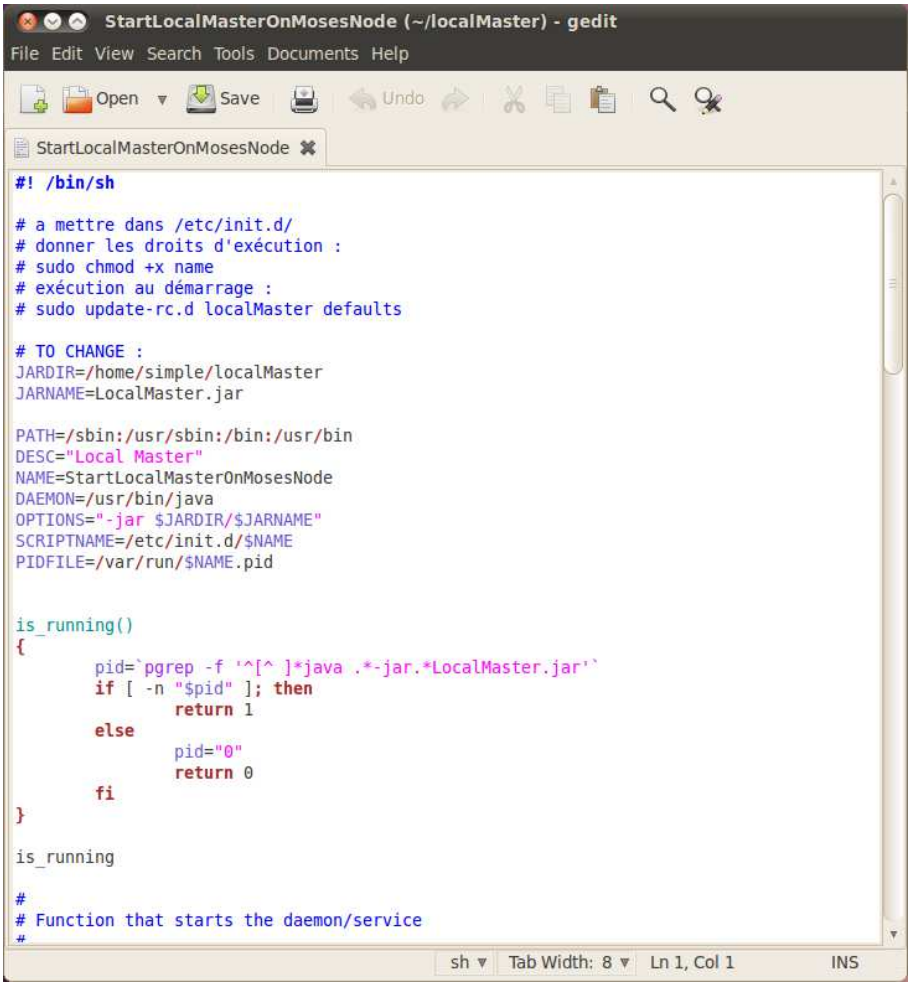
The “lib” sub-folder and the “jar” file must be located in the same folder.

- From the “5-install localmaster” installation package, copy the file “localMaster.tar.gz” to the “simple” folder
- Extract the file into that same location, it becomes a “localMaster” folder as shown below:



- You can now delete the copy of the “localMaster.tar.gz” file in order to save disk space.

- In the “localMaster” folder there is a script file called “StartLocalMasterOnMosesNode”. You can edit it with the “gedit” utility:



If you want to, you can modify the following variables:

NAME : Name of the StartLocalMasterOnMosesNode script (by default: “localMaster”)

DAEMON : Location of the Java executable program (by default: "/usr/bin/java")

JARDIR : Location of the LocalMaster.jar file

JARNAME : Name of the jar file (by default "LocalMaster.jar")

- Move that script file to the “/etc/init.d/” folder by typing the following command in a Terminal window:

```
# sudo mv /home/simple/localMaster/StartLocalMasterOnMosesNode /etc/init.d/
```

- Check out that the script is indeed executable (if not, check properties and click the Execute case:

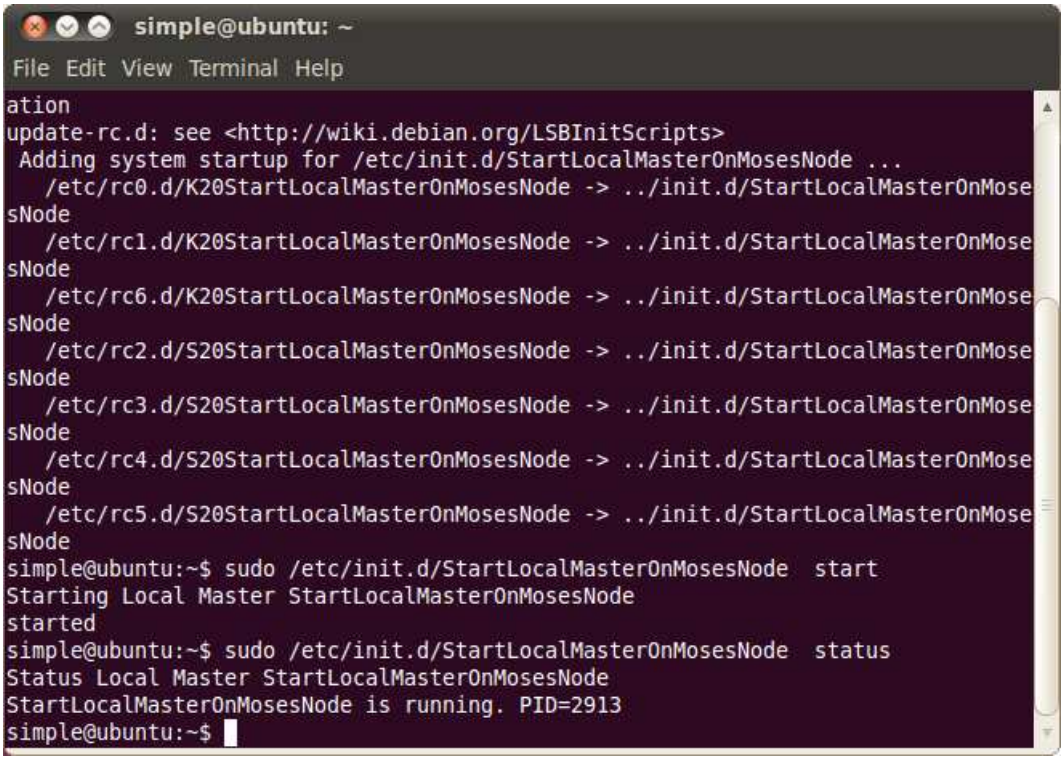
It is compusory to execute the following script so it starts on the OS startup:

```
# sudo update-rc.d StartLocalMasterOnMosesNode defaults
```

The script can also be started and stopped manually with the following commands :

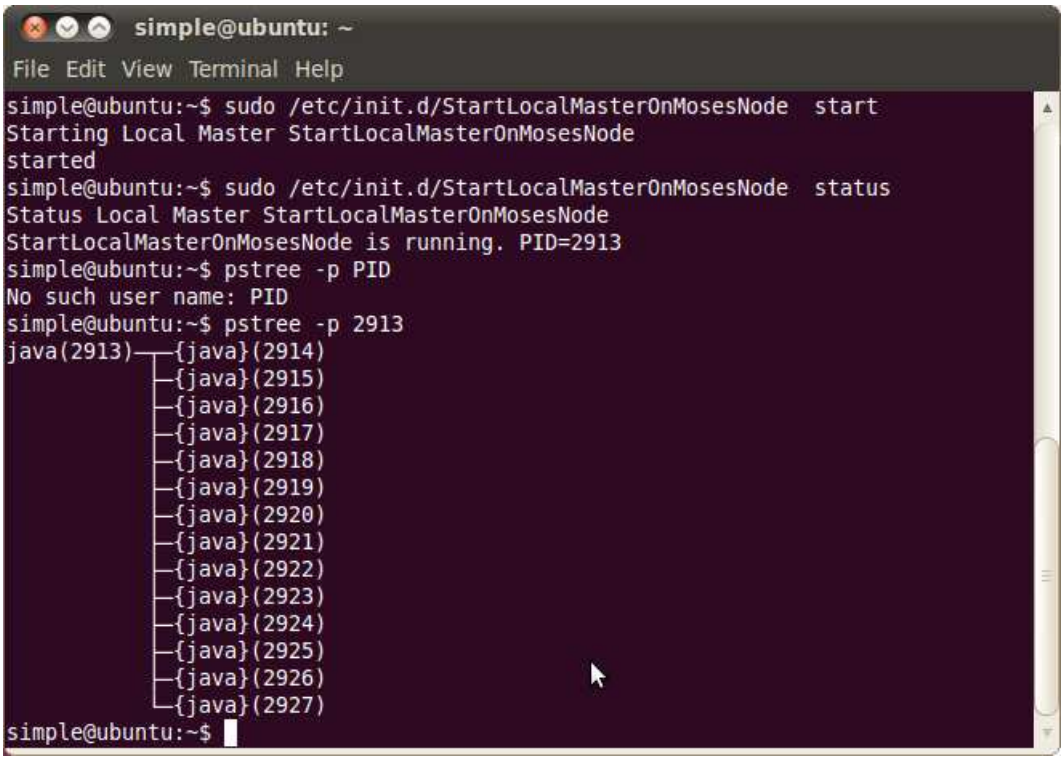
```
# sudo /etc/init.d/StartLocalMasterOnMosesNode  start
# sudo /etc/init.d/StartLocalMasterOnMosesNode stop
# sudo /etc/init.d/StartLocalMasterOnMosesNode status
```

The “status” command returns the PID (process identifier) of the LocalMaster if it is running.



It is then possible to see all the sub-processes (the ones which are launched by the Front-End server) with the following command:

```
# pstree -p [PID]  (replace PID with value, here: 2913)
```



By now we have installed a translation node which is completely autonomous and which can be called from the Front-End server.

3. Next Steps

- Install the Windows Front-End server

For this step please refer to the Front-End Manual.

- Deploy new models (which files need to be copied from training/test into production)

For this step please refer to the New Model manual

- Replicate this node translation (to add more translation node to your users)
    - In your VM management environment, copy the VM and paste it with another name
    - Start the new VM and specified that you copied it (not moved it)
    - Change the IP address of the new VM
    - Add the new IP address in the Front-end Manager (configState.xlm)