

Machine Translation — Practical Work

Master MoSIG, AIW, Natural Language and Speech Processing

Deadline: January 21, 2022; 23:59

Foreword

For this practical work you must organize yourselves in groups of students with the same mother tongue. Groups should have a maximum of three students. Individual work is also possible.

On the first page of the document, you will produce you should identify each member of the group with a translator's number.

There are three parts in which results are expected: human translation, machine translation and evaluation; then a conclusion.

In the parts with results, the tasks are identified as individual work or group work. Each member of the group completes the individual tasks (they are compiled in the report). The group meets for the group work which should provide an opportunity for interesting discussion.

The conclusion should be made in group work. You should share your observations and comments.

Goal	2
Data	2
Human Translation	3
Machine Translation.....	4
MT Evaluation	6
Conclusion.....	6
Deliverable.....	Erreur ! Signet non défini.
Module's Website:	7
MT Systems Available online	8

Goal

Experience what human translation, machine translation and evaluation are about.

Data

The group will choose a written text, consisting of 250 words at least, in a language that you know well and that is not your native language. This text should deal with a subject you know well. It may be a newspaper or journal article available in electronic form.

Each translator will translate this text into the native language in two steps: – a translation that you will do yourself (human translation), a translation that you will produce by post-editing machine translation results (machine translation + post-edition).

Human Translation

Each translator will translate by himself the first half of the text (≈ 125 words). The authorized help are:

- Bilingual dictionaries (e.g. WordReference.com*),
- A concordancer or translation memory (e.g. Linear B[†] et/ou Lingee[‡])

Individual work:

Source Sentence	Translation	Time Spent	Remarks
This is the first sentence.	Ceci est la première phrase.	5 seconds	No difficulty
The inner housing is located between the engine valve and camshaft and transmits the zero or partial lift to the valve.	Le boîtier intérieur est situé entre la soupape du moteur et l'arbre à cames et transmet à la soupape la levée nulle ou partielle.	3 minutes	Dictionary lookup for: "housing", "camshaft" and "valve". In particular, for an "engine" (« moteur »), the word "valve" is translated by « soupape »

Table 1 : Presentation of the manual translation sub-task

One row per sentence with its translation, the time you spent and some remarks about your work.

Each translator will begin the translation when the first column of the table is ready. He will calculate the total time spent on the translation of these ≈ 125 words and give the average time per word.

Once the translation of each sentence is complete, he will copy all translations one after the other to make a first raw translation. He will then copy this raw translation version to revise the overall text produced by highlighting what you have changed from the first translation.

Raw Translation

Ceci est la première phrase. Le boîtier intérieur est situé entre la soupape du moteur et l'arbre à cames et transmet à la soupape la levée nulle ou partielle.

Revised Translation

Ceci est la première phrase. Le boîtier intérieur est situé entre la soupape du moteur et l'arbre à cames ; **il** transmet à la soupape **une** levée nulle ou partielle.

Group work:

You will compare your results as far as timing, raw translation and revised translation are concerned.

* <http://www.wordreference.com>

† <http://www.linearb.co.uk/>

‡ <http://www.linguee.fr/>

Machine Translation

The group will choose different machine translation systems for the second half of the text (≈125 words); at least 2 if possible for the concerned mother tongue:

- Google translate* (many language pairs)
- SYSTRANet† (many language pairs)
- DeepL‡ (European languages pairs)

See also section “MT Systems Available online” which lists some of the systems available online.

Individual work:

	Proposals			My Post-Edition			
Source Sentence	Google	Systran	Prompt	Chosen System	Post-Edition	Time Spent	Remarks
This is the first sentence.	Il s'agit de la première phrase.	C'est la première phrase.	C'est la première phrase.	Systran	C'est la première phrase.	0 seconde	No difficulty
The inner housing is located between the engine valve and camshaft and transmits the zero or partial lift to the valve.	Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames et transmet la levée nulle ou partielle <i>de la vanne</i> .	Le logement intérieur est placé entre la valve de réacteurs et l'arbre à cames et transmet l'ascenseur zéro ou partiel à la valve.	Le logement intérieur est localisé entre la valve de locomotive et l'arbre à cames et transmet l'ascenseur zéro ou partiel à la valve.	Google	Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames et transmet la levée nulle ou partielle à la soupape.	30 secondes	Systran and Prompt vocabulary choices are bad: housing, reactor, elevator. « de la vanne » transformed into « à la soupape ». Checking the choice of « soupape » in the dictionary

Table 2: Presentation of the machine translation & post-edition task

* <http://translate.google.com>

† <http://www.systranet.com/>

‡ <https://www.deepl.com/translator>

One row per sentence with translation proposals from different systems (at least 2), chosen system, post-edition (correct translation obtained changing as little as possible the chosen translation), time spent and remarks about your work (reason for your choice of the translation to post-edit, description of modifications made, ...).

When the first columns of the table are ready, each translator will choose and post-edit. he will compute the total time spent for the translation of these ≈ 125 words and give the average time per word.

He will compare this latter figure to the one you found for the first part of the experiment. What can he conclude?

Once the translation of each sentence is complete, you will copy all translations one after the other to make a first raw translation. You will then copy this raw translation version to revise the overall text produced by highlighting what you have changed from the first translation.

First Translation (from post-edition)

C'est la première phrase. Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames et transmet la levée nulle ou partielle à la soupape.

Revised Translation

C'est la première phrase. Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames ; **il** transmet **une** levée nulle ou partielle à la soupape.

Group work:

You will compare your results as far as timing, raw translation and revised translation are concerned.

MT Evaluation

Individual work:

Each translator will compute the Translation Error Rate (TER) for each of the post-editions he produced. In order to do this, he must use the software provided in the "Practical" section of the module website.

He will first download the file "TER-Masters.zip". After decompression, he will get a folder that contains the java executable program and a file "README.txt" which contains all instructions for use.

He will produce 2 aligned files (Eval_Ter_Hypos.txt & Eval_Ter_Refs.txt) with the translation proposal he post-edited and the post-edition he proposed on each line. He can then calculate the TER score for each sentence.

He will report these results in a table as follows:

Source Sentence	Chosen MT hypothesis*	Post-edited MT hypothesis†	TER Score
Sentence 1	Machine Translation for Sentence 1	Post-edited MT for Sentence 1	TER1 %
Sentence 2	Machine Translation for Sentence 2	Post-edited MT for Sentence 2	TER2 %

Table 3: TER Score Presentation

Group work:

What do these numbers tell you about the translations you have post-edited?

Conclusion

Group work:

In a brief conclusion, you will tell what you have learned through this experience.

* Translation you chose in Table 2.

† Post-edition you proposed in Table 2.

Deliverable

Before uploading your compressed folder, please complete the table below. Whether you are working in a group or alone (one line per group). Locate your group number and name your compressed file according to the standard below. Use the next available group number.

Groups List file location:

<https://docs.google.com/spreadsheets/d/1e1FHfFPmvH40EhURcEgWsGaNdz3Sb0yv3t4JBE6m-Z4/edit?usp=sharing>

You will upload a zipped folder containing your report (pdf) plus the input files (hypos & post-editions) you used for TER computation and the result file. You will name the zip file as follows:

WTrad_Groupe_Number (e.g. for me: **WTrad_Groupe_1.zip**).

Do not forget to put your name(s) in your report and give it a title.

Delivery deadline: January 22th, 2021; 23:59.

Upload folder location:

<https://drive.google.com/drive/folders/1tay7pG1Vz78-XR2-ueWkhbgkOw6hz2Tj?usp=sharing>

Module's Website:

<http://lig-membres.imag.fr/blanchon/SitesEns/NLSP/>

MT Systems Available online*

Provider	URL	Source Languages	Target Languages
Google	https://translate.google.com/	105 languages	105 languages
DeepL	https://www.deepl.com/translator	French, English, German, Spanish, Italian, Dutch, Polish	
Systran	http://www.systranet.com/translate	Arabic, Chinese, Dutch, English, French, German, Greek, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Swedish	<i>Source language dependent</i>
PROMT [†]	http://www.online-translator.com	Arabic, Catalan, Chinese, English, French, Finnish, German, Greek, Italian, Japanese, Kazakh, Korean, Polish, Portuguese, Russian, Spanish, Turkish	<i>Source language dependent</i>
Lingue.tech*	https://www.lingue.tech/en/personal-translator-demo/	Arabic, Chinese, English, French, German, Italian, Portuguese, Spanish	<i>Source language dependent</i>
Morphologic*	http://www.webforditas.hu/?show=textTab&lang=english	Bulgarian, Danish, English, French, German, Hungarian, Latvian, Norwegian, Polish, Portuguese, Russian, Spanish, Ukrainian	
SkyCode	http://webtrance.skycode.com/online.asp?setlang=en	Bulgarian, English, French, German, Italian, Spanish	Bulgarian, Chinese, English, French, German, Italian, Spanish, Turkish
Amebis	http://presis.amebis.si/prevajanje/?jezik=en&	German ⇒ Slovenian, English ⇒ Slovenian, Slovenian ⇒ English	
PWN	http://translatca.pl	English, German, Polish, Russian	
Lingenio	http://www.translate.eu	English ⇔ German, English ⇔ French, English ⇔ Chinese	

* All the free text translation web sites are not listed here, you can also look for yourself

† The user may select a domain for the source text