

Advanced course on natural language processing
Graduate School of Science and Technology, Keio University

Report Assignment:
DeepL, neural network-based translation tool

Alejandro Vicent Micó

alejandrovicm@keio.jp

2022-2023

Index

1. Introduction.....	3
2. Motivation	3
3. The neural network model.....	4
4. DeepL API translation process.....	5
4.1 Data creation	5
4.2 Using the API for translation.....	6
5. Evaluation of results	8
6. Comparison with Google Translate	8
7. Conclusion	10
8. Bibliography.....	11

1. Introduction

In this paper a natural language tool, more specifically a translation tool, is presented. DeepL is a neural network-based translation tool [1] that is available on the internet for its online use, as well as many others such as Google Translate, Microsoft Translator or Amazon Translate. In addition, DeepL Pro can be purchased and downloaded as an improved version [2], and there is also an API that can be used by developers to translate larger amounts of text in their programs [3].

The DeepL API is the option used in this paper, which consists of the following chapters: a motivation explaining the topic and tool choices, a description of the neural network model, the steps followed to create the data for translating and the translation part. Furthermore, the evaluation of the obtained results, a comparison with the Google Translate tool and the conclusions of the paper.

2. Motivation

Nowadays, people like to travel all around the world and visit new places. It can be because they like the culture of those places, the food, the things they can do there, or for any other reason. Traveling is easier than ever thanks to the internet, where you can find good deals for plane tickets and hotel rooms, and even guides that explain you the things you can do at the target destination.

However, the language barrier is still a problem. Either you are going on a field trip, or you want to study or work abroad, it is going to be hard if you do not speak the native language of that place. In this regard, many companies have invested a lot of resources in translation tools, which are very useful when it comes to traveling. It is still a bit inconvenient to use the translator every time you want to speak or understand the native language, but a lot of progress is being made to improve this kind of issues.

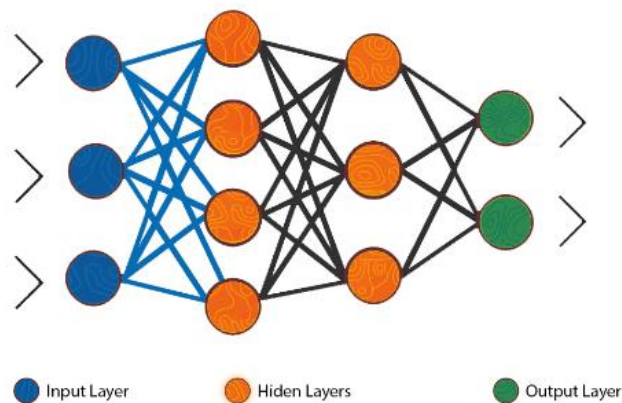
There are many translation tools available on the internet, both free and for purchase. Some of them are easier to use but not so accurate, while others implement advanced techniques, such as machine learning and neural networks, to improve their translation performance.

Knowing the importance of using an efficient and accurate translation tool, we are going to evaluate DeepL, a translation tool based on a neural network model that uses machine learning to improve the quality of their translations [4]. DeepL outperforms other translation systems, being the best translation tool on the internet right now [5]. In addition, it provides an API with different plans, which you can use in a program to translate larger texts. Because of these reasons, DeepL was the chosen tool for this paper.

3. The neural network model

Before going into details about DeepL, we first have to explain the model DeepL is based on: neural networks. A neural network is a simplified model of the way the human brain processes information [6]. A large number of processing units interconnected among them take the role of neurons and are stimulated to make the network work.

These units are arranged in 3 layers: the input layer, representing input fields; the hidden layers and an output layer, representing the target fields. The processing units are distributed among these layers and interconnected with different weights. When the data enters the network through the input layer, it is propagated from unit to unit, until it reaches the output layer.



The connection weights are initially random, but the network learns through training. It examines individual records, generating a prediction for each one, and making adjustments to the weights whenever an incorrect prediction is made. By repeating this process many times, the network improves until the stopping criteria are met. By doing this, the accuracy of the network is improved by replicated known outcomes, so that when an unknown outcome is found, the network can be applied.

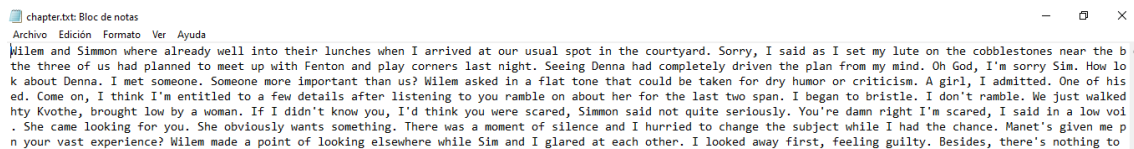
There are many types of neural networks [7], but some of the most important are Artificial Neural Networks (ANN), Convolution Neural Networks (CNN) and Recurrent Neural Networks (RNN). DeepL translates text using artificial neural networks (ANN), which are good for processing text data. This type of networks, also known as Universal Function Approximators, have the capacity to learn weights that map any input to the output.

4. DeepL API translation process

For evaluating the translation tool DeepL, the available DeepL API is going to be used. First, a set of sentences has to be obtained as part of the data creation process. After that, using the DeepL API, the sentences will be translated from English to Spanish. Then, the translations will be manually checked and marked as “Correctly translated” or “Incorrectly translated”, and we will be able to see the accuracy of the tool.

4.1 Data creation

It has been decided to use a chapter of a book as input data for the translation tool, because it reflects a daily situation with different type of sentences, both long and short. The sentences have been taken from a chapter of “The name of the wind”, a fantasy novel I really like. The entire chapter was taken from the book pdf file and pasted into a file named chapter.txt.



After that, a python script was written in order to split the chapter text into single sentences, to make sure that the input data for the API is correct and easily readable. To do that, Python 3.9.2 was used in a Windows 10 machine. The program to split text into sentences is the one shown below.

```
1 import re
2 import sys
3 import os
4 alphabets= "([A-Za-z])"
5 prefixes = "(Mr|St|Mrs|Ms|Dr)[.]"
6 suffixes = "(Inc|Ltd|Jr|Sr|Co)"
7 starters = "(Mr|Mrs|Ms|Dr|He|She|It|They|Us|Our|We|But|However|
8 acronyms = "([A-Z][.][A-Z])?(?=[A-Z][.])?"
9 websites = "([.](com|net|org|io|gov))"
10 digits = "([0-9])"
11
12 #This function is taken from https://stackoverflow.com/questions/4576077/how-can
13 def split_into_sentences(text):
14     text = " " + text + " "
15     text = text.replace("\n", " ")
16     text = re.sub(prefixes, "\\1<prd>",text)
17     text = re.sub(websites,"<prd>\\1",text)
18     text = re.sub(digits + "[.]" + digits,"\\1<prd>\\2",text)
19     if "..." in text: text = text.replace("...", "<prd><prd><prd>")
20     if "Ph.D" in text: text = text.replace("Ph.D.", "Ph<prd>D<prd>")
21     text = re.sub("\s" + alphabets + "[.]" + "\s", "\\1<prd> ",text)
22     text = re.sub(acronyms+" "+starters,"\\1<stop> \\2",text)
23     text = re.sub(alphabets + "[.]" + alphabets + "[.]" + alphabets + "[.]", "\\1: \\2",text)
24     text = re.sub(alphabets + "[.]" + alphabets + "[.]", "\\1<prd>\\2<prd>",text)
25     text = re.sub(" "+suffixes+"[.]" + starters," \\1<stop> \\2",text)
26     text = re.sub(" "+suffixes+"[.]", "\\1<prd>",text)
27     text = re.sub(" " + alphabets + "[.]", "\\1<prd>",text)
28     if "&" in text: text = text.replace("&", "&")
29     if "\"" in text: text = text.replace("\"", "\"")
30
31     if "!" in text: text = text.replace("!", "!")
32     if "?" in text: text = text.replace("?", "?")
33     text = text.replace(".", ".<stop>")
34     text = text.replace("<?>","<stop>")
35     text = text.replace("!", "!<stop>")
36     text = text.replace("<prd>",".")
37     sentences = text.split("<stop>")
38     sentences = sentences[:-1]
39     sentences = [s.strip() for s in sentences]
40     return sentences
41
42 if len(sys.argv) == 1:
43     print("No file provided! Please, provide the path of the file to be translated.")
44 elif not os.path.exists(sys.argv[1]):
45     print("Incorrect file! Please, provide a valid path for the file to be translated.")
46 else:
47     #The file passed as argument is read and split into sentences
48     with open(sys.argv[1]) as f:
49         chapter_text = f.readlines()
50     sentences = split_into_sentences(chapter_text[0])
51     print("Text split into sentences!")
52
53 #The sentences are written into a new file
54 file_name = os.path.splitext(sys.argv[1])[0]
55 with open(file_name + "_sentences.txt", "w") as file:
56     for line in sentences:
57         file.write("'" + line + "'\n")
58     print("Result written in " + file_name + "_sentences.txt")
```

After executing the program with the command `python split_text.py chapter.txt`, a new file with the split sentences is created with name `chapter_sentences.txt`. The scripts used, as well as the files produced, are available at the Github public repository <https://github.com/TheAlexet/NLP2022-2023.git>.

```
C:\Users\theal\OneDrive\Escritorio\NLP>python split_text.py chapter.txt
Text split into sentences!
Result written in chapter_sentences.txt

C:\Users\theal\OneDrive\Escritorio\NLP>
```

chapter_sentences.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

Wilem and Simmon where already well into their lunches when I arrived at our usual spot in the courtyard. Sorry, I said as I set my lute on the cobblestones near the bench. Got caught up haggling. I had been on the other side of the river buying a dram of quicksilver and a pouch of sea salt. The last had cost me dearly, but for once I wasn't concerned about money. If fortune smiled on me, I would be moving up the ranks in the Fishery soon, and that meant my money troubles would soon be over. While shopping in Imre, I had also, quite by coincidence, wandered past the inn where Denna was staying, but she hadn't been there All the same, I was in a fine mood. I tipped my lute case onto its side and flipped it open so the sun could warm the new strings, helping them stretch. Then I settled onto the stone bench under the pennant pole next to my two friends. So where were you last night? Simmon asked too casually. It was only then I remembered that the three of us had planned to meet up with Fenton and play corners last night. Seeing Denna had completely driven the plan from my mind. Oh God, I'm sorry Sim. How long did you wait for me? He gave me a look.

The chapter had a total of 147 sentences, so the new file also has 147 lines. This will be the input data for the translation API of DeepL.

4.2 Using the API for translation

For using the DeepL API, a DeepL account had to be created. The plan used is DeepL API Free, which allows you to use the API for free during the first month. Once the plan is established, a key is given, which can be used for making requests to the DeepL API. For making such requests, a python script named `translate_sentences.py` was written.

```
1 import requests
2 import sys
3 import os
4 import time
5 import json
6
7 my_key = "4ff64d62-a732-706b-9385-34e341c61eb3:fx"
8 url_translate = "https://api-free.deepl.com/v2/document"
9 url_status = "https://api-free.deepl.com/v2/document/{0}"
10 url_result = "https://api-free.deepl.com/v2/document/{0}/result"
11 source_language = "EN"
12 target_language = "ES"
13
14 #The file is translated using the DeepL API.
15 #The resulting document's id and key are provided as a response.
16 def translate(path):
17     up_file = open(path, "rb")
18     _params = {
19         "source_lang": source_language,
20         "auth_key": my_key,
21         "target_lang": target_language
22     }
23     response = requests.post(url_translate, params=_params, files={"file": up_file})
24     json_response = json.loads(response.text)
25     id = json_response["document_id"]
26     key = json_response["document_key"]
27     print("Waiting for translation...")
28     time.sleep(5)
29     check_status(id, key)
30
31 #Using the document's id and key obtained, the translation status is checked.
32 def check_status(id, key):
33     _params = {
34         "auth_key": my_key,
35         "document_key": key
36     }
37     response = requests.get(url_status.format(id), params=_params)
38     print("Translation done!")
39     print(response.text)
40     get_result(id, key)
41
42 #Using the document's id and key obtained, the translation result is obtained and saved
43 def get_result(id, key):
44     _params = {
45         "auth_key": my_key,
46         "document_key": key
47     }
48     response = requests.get(url_result.format(id), params=_params, allow_redirects=True)
49     file_name = os.path.splitext(sys.argv[1])[0]
50     with open(file_name + " translated.txt", "wb") as file:
51         file.write(response.content)
52     print("Result written in " + file_name + " translated.txt")
53
54 if len(sys.argv) == 1:
55     print("No file provided! Please, provide the path of the file to be translated.")
56 elif not os.path.exists(sys.argv[1]):
57     print("Incorrect file! Please, provide a valid path for the file to be translated.")
58 else:
59     #The file passed as argument is translated and saved into a new file
60     translate(sys.argv[1])
```

This script takes a set of sentences as input and does three types of requests to the DeepL API. First of all, and using the method `translate`, it makes a request to `url_translate` with the set of sentences. This will send the sentences to the API, which will be translated and stored online.

After waiting for 5 seconds to ensure the translation has been completed, it calls the second method, `check_status`, which will make a request to the API for checking the status of the translation (successful or not). Finally, in the method `get_result`, a third request to

the API will be made, obtaining from the API the translated sentences in Spanish and storing them in a .txt file. For executing the program, we write *python translate_sentences.py chapter_sentences.txt*.

```
C:\Users\theal\OneDrive\Escritorio\NLP>python translate_sentences.py chapter_sentences.txt
Waiting for translation...
Translation done!
{"document_id":"3f5f2454c45998b4e823b344092ed828","status":"done","billed_characters":6589}
Result written in chapter_sentences_translated.txt

C:\Users\theal\OneDrive\Escritorio\NLP>
```

chapter_sentences_translated.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

Wilem y Simmon ya estaban almorzando cuando llegué a nuestro lugar habitual en el patio.
Lo siento -dije mientras dejaba mi laúd sobre los adoquines, cerca del banco-.
Me entretuve regateando.
Había estado al otro lado del río comprando una botella de azogue y una bolsita de sal marina.
Esto último me había costado caro, pero por una vez no me preocupaba el dinero.
Si la fortuna me sonreía, pronto ascendería en el escalafón de la Pesquería, y eso significaba que pronto se a
Mientras compraba en Imre, también había pasado, por casualidad, por la posada donde se alojaba Denna, pero el
En cualquier caso, estaba de buen humor.
Puse el estuche del laúd de lado y lo abrí para que el sol calentara las cuerdas nuevas y las estirara.
Luego me senté en el banco de piedra bajo el mástil junto a mis dos amigos.
¿Dónde estuviste anoche?
preguntó Simmon con demasiada indiferencia.
Fue entonces cuando recordé que los tres habíamos planeado reunirnos con Fenton y jugar a las esquinas anoche.
Ver a Denna me había quitado el plan de la cabeza.
Oh Dios, lo siento Sim.
¿Cuánto tiempo me has esperado?
Me miró.
Lo siento, repetí, esperando parecer tan culpable como me sentía.
Se me había olvidado.
Sim sonrió, encogiéndose de hombros.
No es para tanto.
Cuando nos dimos cuenta de que no ibas a aparecer, fuimos a la Biblioteca a beber y a mirar chicas.
¿Se enfadó Fenton?
Furioso, dijo Wilem con calma, entrando por fin en la conversación.

A file with name *chapter_sentences_translated.txt* will be created with the resulting 147 sentences translated to Spanish.

5. Evaluation of results

The translated sentences to Spanish obtained in the last step were pasted into an excel sheet together with the English sentences. After that, the correct translation was manually checked sentence by sentence, writing “Yes” in a column if the sentence is Spanish was correct grammatically and it meant the same as the English sentence. It was marked as “No” otherwise.

28	Made claims about your parentage and sexual tendency toward animals, Wilem said with	Hizo afirmaciones sobre tu parentesco y tu tendencia sexual hacia los animales, dijo Wilem con cara ser	Yes
29	...in the Tehlin's cassock!	...¡con la sotana de Tehlin!	No
30	Simmon sang with his mouth full.	cantó Simmon con la boca llena.	Yes
31	Then he laughed and started to choke.	Luego se rió y empezó a ahogarse.	Yes
32	I pounded him on the back.	Le golpeé en la espalda.	Yes
33	Where were you?	¿Dónde estabas?	Yes
34	Wilem asked while Sim tried to get his breathback.	preguntó Wilem mientras Sim intentaba recuperar el aliento.	Yes
35	Anker said you left early.	Anker dijo que te habías ido pronto.	Yes
36	For some reason, I found myself reluctant to talk about Denna.	Por alguna razón, me encontré reacio a hablar de Denna.	Yes
37	I met someone.	Conoci a alguien.	Yes
38	Someone more important than us?	¿Alguien más importante que nosotros?	Yes
39	Wilem asked in a flat tone that could be taken for dry humor or criticism.	preguntó Wilem en un tono plano que podía tomarse por humor seco o crítica.	Yes
40	A girl, I admitted.	Una chica, admití.	Yes
41	One of his eyebrows went up.	Una de sus cejas se alzó.	Yes
42	The one you've been chasing around?	¿La que has estado persiguiendo?	Yes
43	I haven't been chasing anyone, I protested.	No he estado persiguiendo a nadie -protesté-.	Yes
44	She found me, at Anker's.	Ella me encontró, en casa de Anker.	Yes
45	Good sign, Wilem said.	Buena señal, dijo Wilem.	Yes
46	Simmon nodded wisely then looked up with a playful glint in his eye.	Simmon asintió sabiamente y luego levantó la vista con un brillo juguetón en los ojos.	Yes
47	So did you make any music?	¿Has hecho algo de música?	No
48	He nudged me with an elbow and wagged his eyebrows up and down.	Me dio un codazo y movió las cejas de arriba abajo.	Yes
49	A little duet?	¿Un pequeño dúo?	Yes
50	He looked too ridiculous for me to be offended.	¡Parecía demasiado ridículo para que me ofendiera!	Yes

Out of the 147 sentences, 5 were detected to be incorrectly translated and then marked with a “No”. The other 142 were correctly translated. As a result 142/147 sentences in Spanish were correct, which means that, in this case, DeepL’s API had an accuracy of 96,59%, which is very high, taking into account that the translation of a sentences does not have any information about the context of those sentences.

6. Comparison with Google Translate

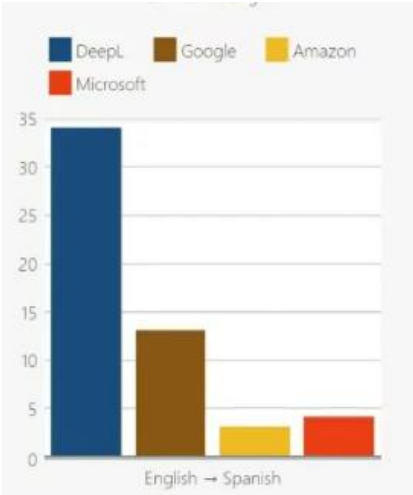
In addition, a translation comparison with Google Translate, the most used translation tool on the internet, has been done. The same sentences in English were translated into Spanish using Google Translate, and then pasted in an Excel for comparison.

1	Wilem and Simmon where already well into their lunches when I arrived at our usual spot	Wilem y Simmon ya estaban en sus almuerzos cuando llegué a nuestro lugar habitual en el patio.	Yes
2	Sorry, I said as I set my lute on the cobblestones near the bench.	Lo siento, dije mientras dejaba mi laúd en los adoquines cerca del banco.	Yes
3	Got caught up haggling.	Quedó atrapado regateando.	No
4	I had been on the other side of the river buying a dram of quicksilver and a pouch of sea	Había estado al otro lado del río comprando un trago de mercurio y una bolsa de sal marina.	No
5	The last had cost me dearly, but for once I wasn't concerned about money.	El último me había costado caro, pero por una vez no me preocupaba el dinero.	Yes
6	If fortune smiled on me, I would be moving up the ranks in the Fishery soon, and that me	Si la fortuna me sonreía, pronto ascendería de rango en la Pesca, y eso significaba que mis problemas de	Yes
7	While shopping in Imre, I had also, quite by coincidence, wandered past the inn where D	Mientras hacía compras en Imre, también, por pura coincidencia, pasé por delante de la posada donde s	Yes
8	All the same, I was in a fine mood.	De todos modos, estaba de buen humor.	Yes
9	I tipped my lute case onto its side and flipped it open so the sun could warm the new str	Incliné el estuche de mi laúd de lado y lo abrí para que el sol pudiera calentar las cuerdas nuevas, ayudá	Yes
10	Then I settled onto the stone bench under the pennant pole next to my two friends.	Luego me acomodé en el banco de piedra debajo del poste del banderín junto a mis dos amigos.	Yes
11	So where were you last night?	Entonces, ¿dónde estuviste anoche?	Yes
12	Simmon asked too casually.	Simmon preguntó demasiado casualmente.	Yes
13	It was only then I remembered that the three of us had planned to meet up with Fenton	Fue entonces cuando recordé que los tres habíamos planeado reunirnos con Fenton y jugar en las esqui	Yes
14	Seeing Denna had completely driven the plan from my mind.	Ver a Denna había borrado por completo el plan de mi mente.	Yes
15	Oh God, I'm sorry Sim.	Oh Dios, lo siento Sim.	Yes
16	How long did you wait for me?	¿Cuánto tiempo me esperaste?	Yes
17	He gave me a look.	Me dio una mirada.	No
18	I'm sorry, I repeated, hoping I looked as guilty as I felt.	Lo siento, repetí, esperando parecer tan culpable como me sentía.	Yes
19	I forgot.	Me olvidé.	Yes
20	Sim grinned, shrugging it off.	Sim sonrió, encogiéndose de hombros.	Yes
21	It's not a big deal.	No es gran cosa.	Yes
22	When we figured out you weren't going to show, we went to the Library to drink and loo	Cuando nos dimos cuenta de que no ibas a aparecer, fuimos a la biblioteca a beber y mirar chicas.	Yes
23	Was Fenton mad?	¿Fenton estaba loco?	No

When checking the correctness of the translations, it was easy to notice that the ones done using Google Translate were not as accurate as the DeepL translated sentences. Out of

147 sentences, 25 were found to be incorrectly translated. Five more times the incorrect sentences of DeepL. This left 122/147 correctly translated sentences, and an accuracy of 82,99%.

As a result, we can see that DeepL is far more accurate than Google Translate, which is the most used translation tool right now. This can be seen in the following image, where the accuracy of DeepL is put together with the accuracies of Google Translate, Amazon Translate and Microsoft Translator.



The more accurate and natural-sounding translations of DeepL are the tool’s strongest point. However, it is not perfect, and it has a couple of weak points. First of all, it supports only 27 languages [8], which is far less than its competitors. Nevertheless, these list covers the majority of European languages, plus Japanese and Chinese. In addition, the amount of text you can translate using the free plan of DeepL is limited, and the premium plans are not as cheap as other options. This would be the second drawback of DeepL.

	DeepL Translator free	Starter \$10.49 <small>per user per month</small> Try for free <small>Free for 30 days. Cancel anytime.</small>	RECOMMENDED Advanced \$34.49 <small>per user per month</small> Try for free <small>Free for 30 days. Cancel anytime.</small>	Ultimate \$68.99 <small>per user per month</small> Buy now
Maximum data security <small>i</small>		✓	✓	✓
Translate text <small>i</small>	Limited characters & volume	Unlimited *	Unlimited *	Unlimited *
Translate whole files (per user) <small>i</small>	Limited characters & editing ✓ Up to 5 MB file size	5 files/month ✓ Up to 10 MB file size	20 files/month ✓ Up to 10 MB file size	100 files/month ✓ Up to 10 MB file size
Formal/informal tone <small>i</small>		✓	✓	✓
Glossary (limited languages) <small>i</small>	1 glossary limited to 10 entries	1 glossary with 5,000 entries	2,000 glossaries with 5,000 entries	2,000 glossaries with 5,000 entries
Team administration <small>i</small>		For 2 or more users	For 2 or more users	For 2 or more users
Single sign-on (SSO) integration <small>i</small>			For 35 or more users	For 35 or more users
CAT tool integration <small>i</small>			✓	✓

7. Conclusion

To sum up, in this paper a natural language tool, in this case a translation tool, has been studied and analyzed. The tool, named DeepL, is the most accurate translation tool on the Internet, even if it is not the most used. To perform this task, the neural network model in which DeepL is based has been studied. In addition, there has been a process of data creation. A program has been created to parse all the sentences of a book's chapter.

Secondly, the DeepL API has been integrated in a Python program for translating the data created. After that, the results obtained from the translation have been evaluated, to check the performance of DeepL. Finally, the tool has been compared to another translation tool, Google Translate, showing that DeepL is the most accurate one.

This work allowed me to realize how important translation tools are nowadays and helped me to get a better understanding of how they work and perform in real situations. Among all the things done, learning how the tool (API in this case) worked and generating the results was the hardest part in my opinion, because I wanted to use it for my specific case. Nevertheless, the processes of data creation and analysis of results were also time-consuming. All in all, I think that this work was really enriching and I enjoyed learning about translation tools, its usefulness and processes behind the translations.

8. Bibliography

- [1] DeepL official website. <<https://www.deepl.com/translator>>
- [2] DeepL official website: DeepL Pro download.
<<https://www.deepl.com/pro?cta=header-pro>>
- [3] DeepL official website: API. <<https://www.deepl.com/pro-api?cta=header-pro-api/>>
- [4] DeepL official blog: How does DeepL work? <<https://www.deepl.com/en/blog/how-does-deepl-work>>
- [5] DeepL official website: Quality. <<https://www.deepl.com/quality.html>>
- [6] The neural networks model. <<https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=networks-neural-model>>
- [7] CNN vs RNN vs ANN – Analyzing 3 Types of Neural Networks in Deep Learning.
<<https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/>>
- [8] DeepL official website: Supported languages. <<https://support.deepl.com/hc/en-us/articles/360019925219-Languages-included-in-DeepL-Pro>>