Machine Learning is an important tool for language translating models as it is almost impossible to translate all possible combinations by ourselves. Through Machine Learning models can now translate words with a very high accuracy making it easier to translate huge texts. Some advantages of such ML models are that :

* They are fast and accurate.
* They can be used to translate multiple languages simultaneously.
* They are cheaper than human translators.

However before we can train our model to translate sentences, it needs to go through certain preprocessing to make it easier for our model to understand huge sentences. For this we import a sample dataset that contains English sentences and their translation in the language we are targeting. Now first we need to remove all the punctuation, special characters and any other thing not useful in translating. We can also convert all the words into either uppercase or lowercase. We can also use an a library to convert longer phrases into shorter one which only convey the meaning. Since the dataset contains English sentences and their translation we need to split it into 2 columns, one with English phrases and other with its translation. Next we can split the phrases into words separated by spaces to help in easier translation.

Since computer won’t understand words we need to convert the words into binary to make it readable by the model. We can do this by comparing the words against a vector and putting ‘1’ where the word is present to get a matrix like [1,0,0] using One Hot Encoding. Next we can apply a train\_test\_split of 0.2 or 0.3 so that for example if we have a dataset of 1000 words, we can use 800 for training and 200 for testing the model for accuracy. We can use sklearn library for the training part. After training the model we need to test it using the test cases we prepared and take the results. We take the output and compare it against the training outputs to check how accurate our model is. Some metrics being used are accuracy, precision etc. We can also use visual means like graphs with one line showing training results and other test results or also use a Scatter graph.