1 Question 1

The greedy decoding approach is efficient but gives suboptimal outputs. This is because it picks the most likely word at each time and assumes that the next word that is the most likely given the previous word was picked gives the highest global probability, which is not always true. As a comparison, with beam search in [3], it takes multiple most likely words at each time, calculates the probability of the next words given each of the previously chosen words and compares them. This gives a better final result but also makes it less efficient compare to the greedy approach.

2 Question 2

The major problem with the translations is that there are words that are unnecessarily translated multiple times, which is known as "over-translation". To overcome this problem, [4] proposes using coverage vectors, which keep track whether a word is already translated and limit the number of times a word is translated. Other solutions to this problem described in [2] are the usage of local attentions and input-feeding approach.

3 Question 3

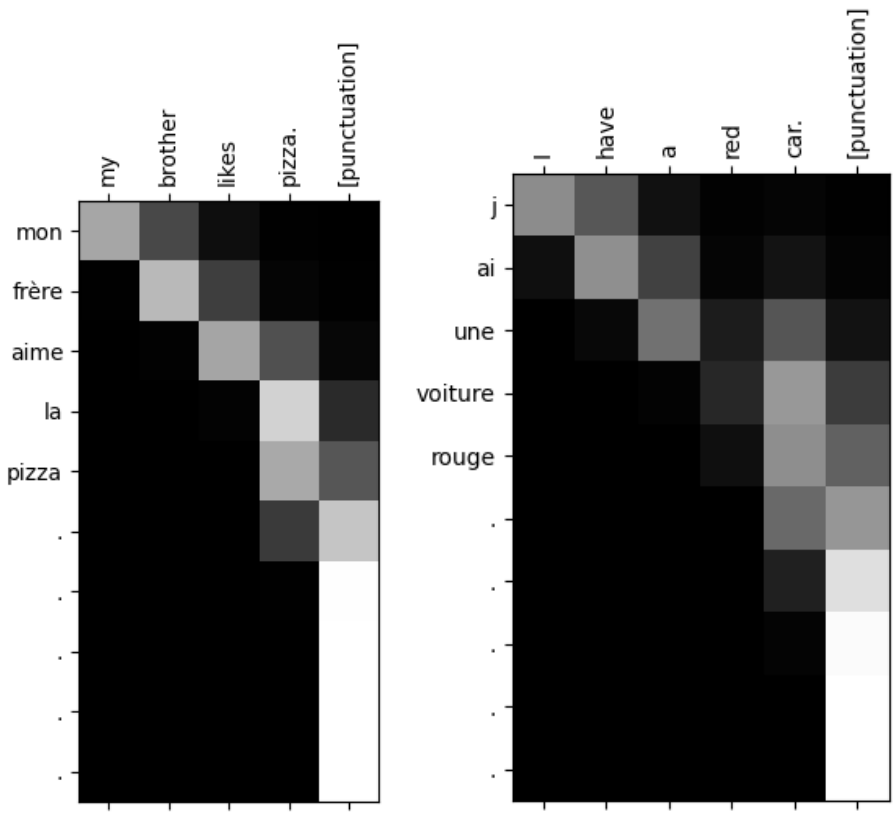

Figure 1: Source/target alignment visualization

In figure 1, on the left, the translation is done by order of the words and thus the weights are big on the diagonal. We can also notice that the word "la" has its biggest weight on "pizza", showing that the model knows that it needs to add an article for the noun for the translation despite not having any article in the source sentence. On the right, we can see the ability of the model to invert the adjective and noun between the two languages.

4 Question 4

The word ‘mean’ in the two given sentences has different meaning and the model is able to differentiate between the two and correctly translates both sentences. In [1], the presented model is designed to consider the left and right context of words, which is based on the attention mechanism that is used by the model in this lab.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[3] Thang Luong, Kyunghyun Cho, and Christopher Manning. Neural machine translation. In *Neural Machine Translation ACL Tutorial*, pages 92–93, 2016.

[4] Zhaopeng Tu, Zhengdong Lu, Yang Liu, and Hang Li. Modeling coverage for neural machine translation. *arXiv preprint arXiv:1601.04811*, 2016.