1 Question 1

The presentation [1] mentions 4 strategies :

- exhaustive search : choose the best translation among all possible translation, which is practically infeasible
- ancestral sampling : we sample each word from the softmax distribution from previous step
- greedy search : we don't sample but just take the argmax of the previous step softmax distribution
- beam search : at each step we update the k most likely hypotheses of sequences being the right translation

Our greedy decoding strategy is memory and computationally efficient because at each step we choose the best solution. The drawback is that we are forced to take a short-term view which leads to poor quality translation. In the other hand the presentation show that even though beam search gives better results, it's in fact computationally expensive because the k hypothesis that need to be compute and update at each step are hard to parallelize.

2 Question 2

We can see two major problems: sentences that do not finish and some words that repeat themselves. This problem can be explain by the attention mechanism we use. The mechanism allow to put attention on every word of the source no matter where we are is the translation process. Thus the most important words stay important during the whole translation and the NMT tends to repeat them to insists on their importance. That being said we can propose a local attention mechanism which would slide the attention focus along the source while translating. This way would permit to go across an important word in the source but not focus the attention on it too many times. Therefore the word would not be repeated. The drawback of local attention is that in some languages sentences are not constructed the same way and words can come in a total opposite order which would lead to miss the attention on the right words.
Other problems like the translator not following grammar rules can probably be fixed by stacking GRU. Stacking GRU would complexify the network and help it to compute more challenging tasks.

3 Question 3

Here are two figures showing on which source word the attention mechanism focus in the choice of each target word. In the figure 1, english and french words are in the same order which lead to a diagonal graph.

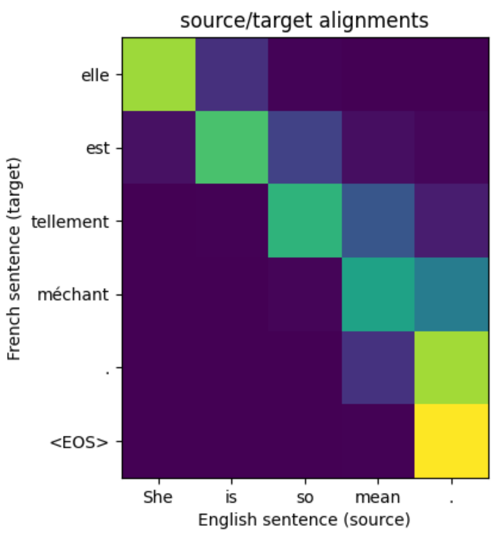

Figure 1: Example of attention in source/target alignment without adjective-noun inversion.

When we try with a sentence containing adjective-noun inversion we might expect to see the inversion in the attention graph. For example red car becomes voiture rouge in french. In the figure 2 we observe that attention mechanism focus on the word 'car' when translating 'voiture' eventhough the word is not in the same position, that is the diagonal stop being a true diagonal when arriving at red car.

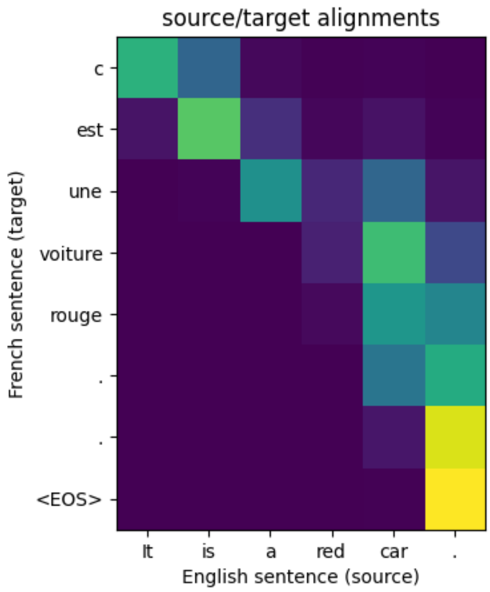

Figure 2: Example of attention in source/target alignment with adjective-noun inversion.

4 Question 4

We get the following results:

- 'I did not mean to hurt you' $\longmapsto$ 'je n ai pas voulu intention de blesser blesser blesser blesser blesser blesser . blesser . blesser . . . .'
- 'She is so mean' $\longmapsto$ 'elle est tellement méchant méchant . <EOS>'

We observe that "mean" translation depends on the context. In the first sentence it means "intention" whereas in the second it means "méchant". It shows that the model can handle polysemy.

References

[1] Christopher Manning Thang Luong, Kyunghyun Cho. Neural machine translation. pages 87–95, 2016.