1 Question 1

The greedy approach is efficient, but heavily suboptimal since we are picking the most likely symbol each step and we can't go back and change the decision after. According to this [2] ACL tutorial a beam search is better since we maintain $k > 1$ hypotheses at a time. It is asymptotically exact when $k \to +\infty$. Here k must be chosen based on a validation set.

2 Question 2

Like said in the first question, one problem of this method is that it is very suboptimal.
- Onother problem is not being able to capture and recognize the gender like in this example:

$$\text{She is so mean} \to \text{elle est tellement méchant méchant.}$$

So here we have 'méchant' instead of 'méchante', this may be due to the lack of training on examples containing both genders.
- Another problem is over-translation as described in [1] which is translating one word multiple times. This will increase the alignment error like in this example where 'mean' is translated " times.

$$\text{She is so mean} \to \text{elle est tellement méchant méchant.}$$

This can be solved by using coverage vectors that indicates if a word is translated or not so that we can force that a word is translated only once. Another solution is to use local attentions like described in [3].
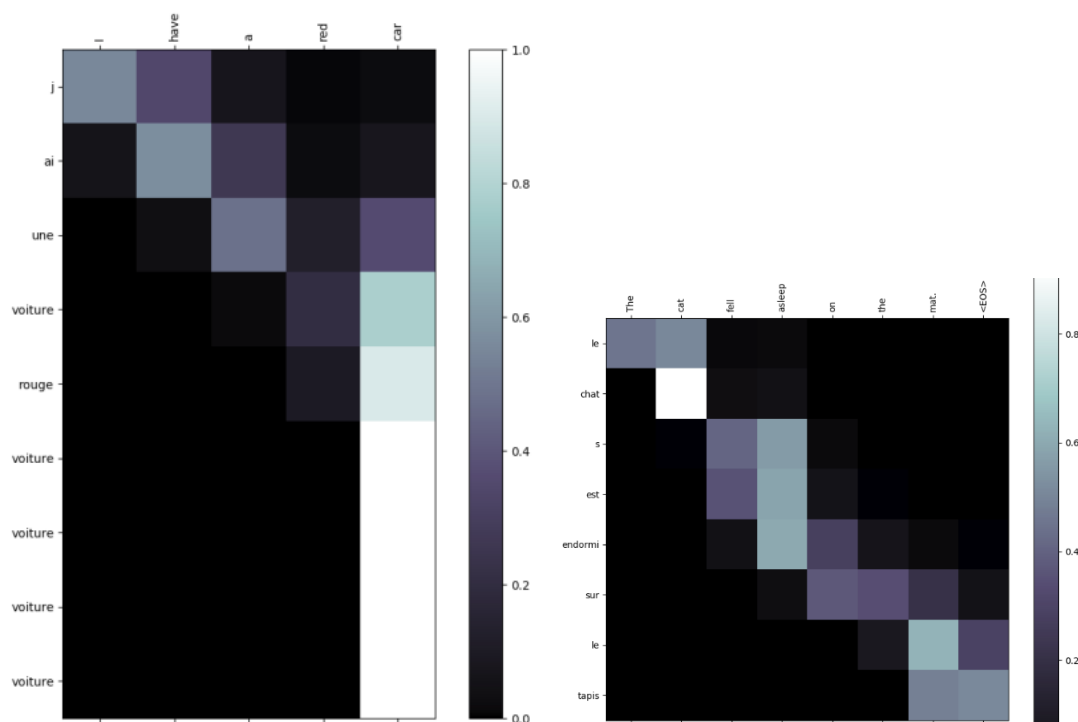
3 Question 3

Figure 1: source/target alignments

The figure 1 shows the alignments of two examples. We can notice the ability of the model to capture the adjective-noun inversion like in the first example on the left (red car $\to$ voiture rouge). The alignment matrix

shows that in general the weights have big values on the diagonal since in general there are no inversions and each translated word correspond to the word at the same position. But is our example and thanks to the attention mechanism it was able to capture the inversion.

In the second example [on the right], we can notice that the alignment is not done with string coefficient with only one word but instead all the words, especially near the end of the sentence, have a role in determining the target word.

4 Question 3

Below are two examples produced by our translation model:

$$\text{She is so mean} \rightarrow \text{'elle est tellement méchant'}$$

$$\text{I did not mean to hurt you} \rightarrow \text{'je n ai pas voulu intention de blesser'}$$

These two examples shows that the model is able to capture the context in which occurs the word. So 'mean' is translated to 'méchant' and 'voulu intention' based on the situation of usage. This is the most important and difficult part in natural language processing and attetion seems to be able to solve it efficiently.

References

[1] Christopher D. Manning Minh-Thang Luong, Hieu Pham. Effective approaches to attention-based neural machine translation. In *arXiv:1508.04025v5 [cs.CL]*.

[2] Christopher Manning Thang Luong, Kyunghyun Cho. Acl tutorial 2016. In *NMT-ACL 2016, slides 87-95*.

[3] Yang Liu Xiaohua Liu Hang Li Zhaopeng Tu, Zhengdong Lu. Modeling coverage for neural machine translation. In *arXiv:1601.04811v6 [cs.CL]*.