1 Question 1

Our greedy decoding method is sub-optimal since it only chooses the most likely word at each step which will often lead to a incorrect translation because the most probable word is generated without considering the whole set of previous outputs. A better way would be to keep track of K different candidate sentences, and at each step, expand these candidates with new probable words keeping only the K most probable generated sentences for translation so far. This is called beam search and yields better translations but is computationally more expensive.

2 Question 2

In our translations, some words are often repeated many times and others are absent. It's a sign that we do not keep track of which source words have already been translated. One way to solve this is to keep track of previous alignments via a dependency from the attention hidden states as an input for the next decoding step. This can be done for example by concatenating the attention output with the next input [2] or by concatenating the input with a vector keeping track of translation coverage in a deterministic way [3] or by learning a vector [1].

3 Question 3

As we see in the figure, the model is sometimes able to detect word inversions as in the first example. The alignment however does not match the appropriate words very well, as the highest score for red is car not $rouge$. After extensive search, I was not able to find a sentence where the word inversion was properly reflected in the alignment. Most of the time the model failed to detect word inversion, as in the second example. Maybe the model would work better with more training data. Another option would be to use local attention, or more sophisticated models like the ones in question 2.

4 Question 4

Let's observe the translations:
'I did not mean to hurt you' $->$ 'je n ai pas voulu intention de blesser blesser blesser blesser blesser blesser . blesser . blesser .'
'She is so mean' $->$ 'elle est tellement méchant méchant . $< EOS >$'

Apart from the problems mentionned in question 2, we see that our translation takes into account the polysemy of the word 'mean' when used in different contexts, which means our word representations are contextualized unlike word2vec embeddings.

References

[1] Ryan Kiros Kyunghyun Cho Aaron C. Courville Ruslan Salakhutdinov Richard S. Zemel Kelvin Xu, Jimmy Ba and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *ICML*, 2015.

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

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

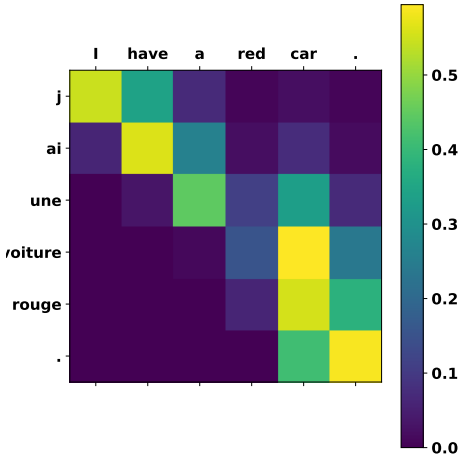

Figure 1: Alignments for *I have a red car* visualized

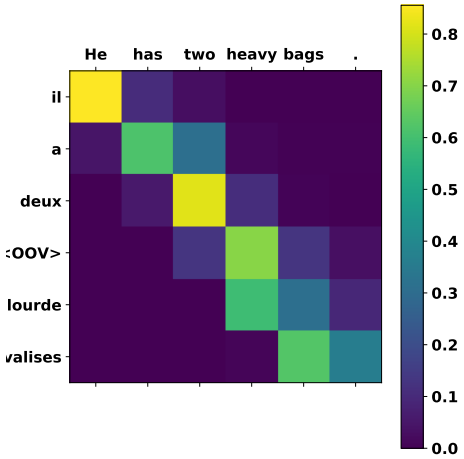

Figure 2: Alignments for *He has two heavy bags* visualized