1 Question 1

The greedy decoding approach we used is efficient in term of memory usage and computation, but not quite optimal, and while it could generate the sequence of words, the quality of output is low when compared to the other decoding approaches. For example, when it comes to Negative Log Likelihood (NLL) which is the criterion we want to minimize, the NLL obtainted is the highest out of all the decoding approaches used, which shows how suboptimal this approach is[1].

2 Question 2

The major problem observed with the translations done by the model is that some words or symbols get replicated, which is due to over translation. We could correct this issue by introducing a coverage model which keeps track of the coverage of the words by a translation rule, this can avoid under-translation by focusing attention on untranslated word, and over-translation by discarding attention from over visited words[2].

3 Question 3

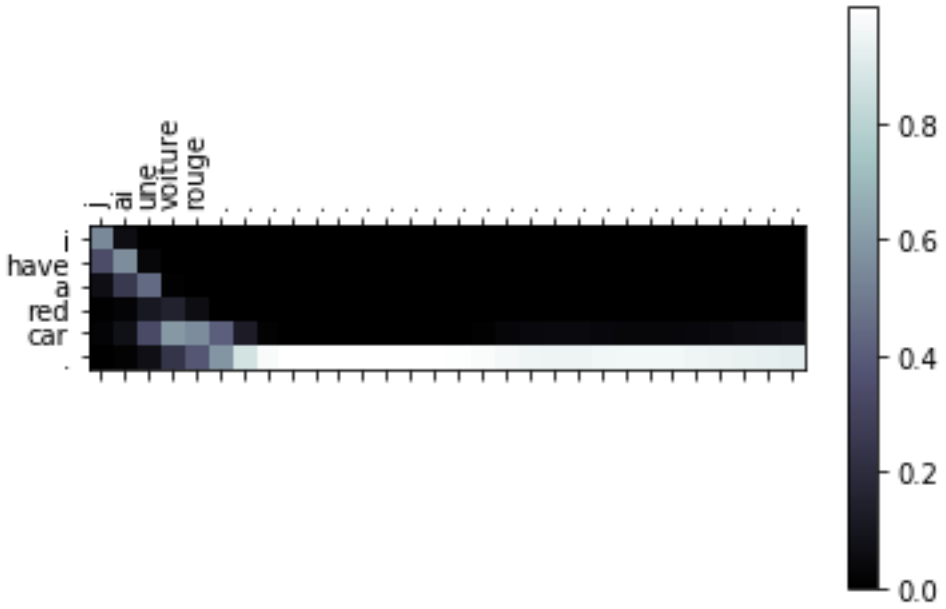

Figure 1: Heatmap of the attention coefficients

We notice that we detected the noun-adjective inversion from ”red car” to ”voiture rouge”.

4 Question 4

The translation of ”she is so mean.” is ”elle est tellement méchant méchant . ¡EOS¿” and the translation of ”I did not mean to hurt you” is ”je n ai pas voulu intention de blesser blesser blesser blesser blesser blesser . blesser . blesser . . . . . .”. Aside from overtranlation of the last words, we can note in the first sentence that it does not take into context the gender, and for the second, it uses two words having nearly the same meaning (intention , voulu) nexts to each other. However, it succeeds to detect semantics as it succeeded in translating the word ”mean” correctly in both sentences.

References

[1] T. L. K. C. C. Manning, “Neural machine translation.”

[2] Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, “Modeling coverage for neural machine translation,” 2016.