INF5820, Assignment 3: Machine Translation

Some clarifications

2 Alignment

Part c

The questions:

> Repeat the experiment from the lecture and see that you get the same results after 1,2, 5, 25, 100 iterations. (You may do this by running several experiments with different parameters.)
>
> Compare the results to the output of model3, i.e. with the default settings (5 iterations of model1 followed by 5 iterations of model3).

should be read:

> What are the translation probabilities after 1, 2, 5, 25 and 100 iterations of model 1? Check that you get the same results as at the lecture 26 October (last slide).
>
> What are the translation probabilities after 5 iterations of model 1 followed by 5 iterations of model 3? How does it compare to the results from model 1?

5.2 The training

The text:

> We are going to run two series of experiments, one on a smaller text and one on the whole newstest2011. You should extract a subtext of your training corpus consisting of 15000 sentences. Call it "small.en" and "small.es" (or "small.fr", etc.). You will get an individual number n and you should extract from sentence number n to sentence number $n + 15000$. This may be done on the model of

```
~/corpus $ head -n10 newstest2011.true.en | tail -n5 > small.en
```

should have been:

We are going to run two series of experiments, one on a smaller text and one on the whole news-commentary-v7. You should extract a subtext of your training corpus consisting of 15000 sentences. Call it "small.en" and "small.es" (or "small.fr", etc.). You will get an individual number n and you should extract from sentence number n to sentence number $n + 15000$. This may be done on the model of

```
~/corpus $ head -n10 news-commentary-v7.es-en.en | tail -n5 > small.en
```

Or you may do the extraction after tokenization and truecasing.