

INF5820, fall 2016

Additional exercise to assignment 2

Non-mandatory

1 Phrase alignment

This is an offer to they who like to implement or feel that they learn better by making implementations. Also, some of you found it boring to count the BLEU scores manually in assignment 1 and implemented the counting instead. I assume you might try the same in exercise 3 on assignment 2.

Here is the task: Implement exercise 5.6.3.6. at page 153 of Koehn: Statistical Machine Translation. The solution should take as input two alignment files (A-files) produced by GIZA. You can assume that the two files result from successfully running GIZA on two sentence-aligned files in both directions.

Then run your program on the two files produced by exercise 3A on the ordinary “Assignment 2: Alignment for Stat. MT”, and for each sentence pair produce all the phrase pairs that may be extracted from them. Write the results to a file

This can replace (3C) on the obligatory assignment, but you should still do point (3B) and make sure that your implemented solution is consistent with your figures.

Submission: The program. The output file with the phrase pairs.

End