1.

(a)

English, alignment and French so far to predict the alignment, English, alignment and French so far to predict the French.

We solve the problem of do not know both alignment and transform probability (parameters and alignment are both unknown)

Length difference

(b)

RNN could capture the dependency of the history context.

History represented as a vector

(c)

I\*J

(d)

In some situation, French will have some words like ‘the’ in English, which doesn’t have any meaning but sentence is not complete without them. However, it cannot find any alignment with words in English, and vice versa.

Coherent

(e)

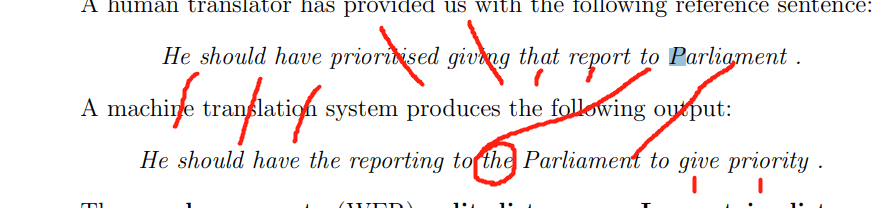
No relationship.

(g)

3.

(a)

WER= 7/9



(b)

Pros: WER can accurately capture the difference between system output sentence and reference.

Cons:

The part of speech changes usually associate with the word change, but the Article like that, the, they are mostly same in meaning or meaningless. But WER count those as errors, where will even only article different the model still reports bad performance of the translation result, which are fine from the point of human’s perspective.

English has flexible order of grammar structure, and WER cannot capture the equality between those expression difference, even the same meaning is expressed.

(c)