Assignment
Noisy Channel Model and Machine Translation

Problem 1

Questions:

- What is statistical natural language processing?

- What is the definition of independence?

- What is the definition of conditional probability?

- What is Bayes formula?

- What is the empty cept and why do we need it in the IBM model?

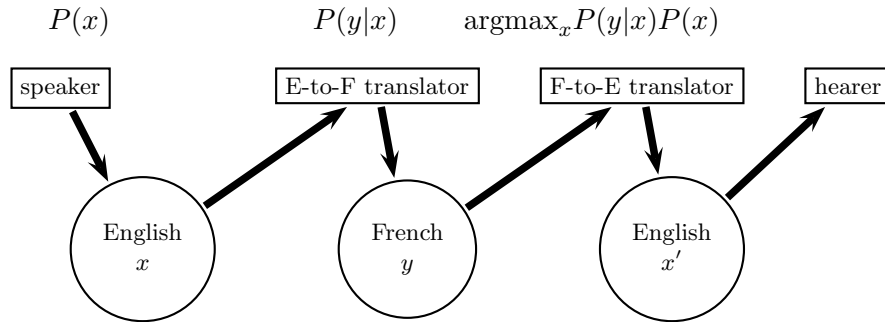- Why is morphology a challenge for the IBM model?

Problem 2

Give examples for the following.

- It is incorrect to assume that the translation of one word in a sentence is independent of the translations of the other words in the sentence.

- It is incorrect to assume that each target word is generated by a single source word.

- It is incorrect to assume that the following two are independent: (i) the words occurring in a sentence pair taken from a parallel corpus and (ii) the alignment between the words.

Problem 3

Below you see a picture of a noisy channel model for machine translation.

$$P(x) \qquad\qquad P(y|x) \qquad \mathrm{argmax}_x P(y|x)P(x)$$

```
[ speaker ]          [ E-to-F translator ]   [ F-to-E translator ]   [ hearer ]
```


This is the corresponding formula:

$$\mathrm{argmax}_{\text{word-sequence}} P(\text{word-sequence}|\text{evidence})$$

$$= \mathrm{argmax}_{\text{word-sequence}} \frac{P(\text{evidence}|\text{word-sequence})P(\text{word-sequence})}{P(\text{evidence})}$$

$$= \mathrm{argmax}_{\text{word-sequence}} \quad P(\text{evidence}|\text{word-sequence}) \quad P(\text{word-sequence})$$

Assume that the translations "Star" (eine berühmte Künstlerpersönlich-keit) and "Stern" of "star" are equally likely and that the translations "ausleihen" und "engagieren" of "hire" are equally likely. In this case, how will a translation system based on the noisy channel model be able to select the correct translation of "will hire a star" from the four possibilities "wird einen Star engagieren" "wird einen Star ausleihen" "wird einen Stern engagieren" "wird einen Stern ausleihen"?

In your answer refer either to the formula or to the graphical depiction of the model.

Problem 4

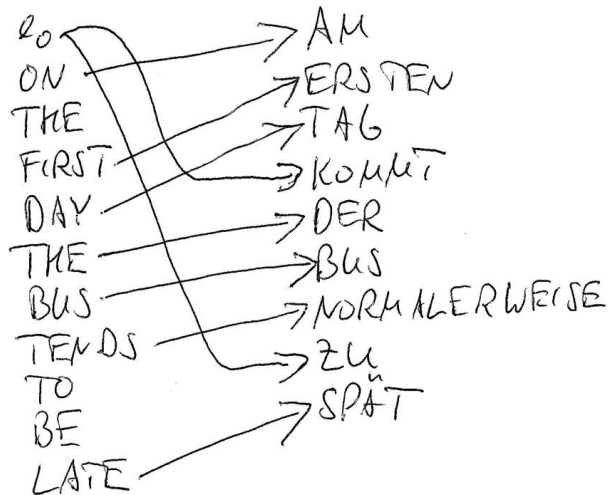Given these probabilities:

$$P(A) \quad = \quad 0.01$$

$$P(B) = 0.01$$
$$P(AB) = 0.001$$

are A and B dependent or independent? Explain why.

Problem 5

Estimate $P(\text{der}|\text{the})$ and $P(\text{kommt}|e_0)$ based on the shown alignment. Give both the ML and Laplace estimates.

Problem 6

Consider this parallel corpus consisting of one sentence pair.

Mary recorded a folk song

Maria nahm ein Volkslied auf

    (i) Draw the alignment between the two sentences by hand. Each line should be an arrow starting at an English word and pointing to a German word. (If there are several equally good or equally bad possibilities, we recommend that you only give one to simplify your answer.)

    (ii) Give the IBM alignment model notation for your alignment, e.g., $< 4, 2, 1, 3 >$.