UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in **INF5820 Language technological applications**
Day of exam: **8 December 2014**
Exam hours: **1430-1830**
This examination paper consists of 5 pages.
Appendices: **0**
Permitted materials: **None**

Make sure that your copy of this examination paper is complete before answering.

- **You may answer in English, Norwegian, Danish or Swedish.**
- **You should answer all questions. The weight of the various questions are indicated.**
- **You should read through the whole set to see whether anything is unclear so that you can ask your questions to the teachers when they arrive.**
- **If you think some assumptions are missing, make your own and explain them!**

1 Word-based Statistical Machine Translation (20%)

The simplest models for Statistical Machine Translation we have considered are word-based. Central to this model is the concept of a word alignment. For example, in the so-called IBM1 model one calculates $P(\mathbf{f} \mid \mathbf{e})$ by summing over all alignments $\mathbf{a}$:

(1) $$P(\mathbf{f} \mid \mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

Furthermore, each addend is calculated by

(2) $$P(\mathbf{f}, \mathbf{a}|\mathbf{e}) = \frac{\epsilon}{(k+1)^m} \prod_{j=1}^{m} t(f_j|e_{a_j})$$

a) (10%) How is a word alignment defined in this model? Illustrate with a pair of sentences from a language pair of your choice. Each sentence should be at least five words long. Consider two different alignments between the two sentences. What correspond to $\mathbf{f}, \mathbf{a}, \mathbf{e}, k, m, f_3, e_{a_3}$ in the examples?

b) (10%) Such word alignments have been critized for not always giving the correct relationship between the two sentences. Why is that so? Explain by giving examples of relationships you think cannot be represented properly with this concept of a word alignment.

2 Phrase-based Translation (30%)

The main model for phrase-based statistical machine translation may be expressed by the following formula.

(3) $$e_{\text{best}} = \text{argmax}_e \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\ d(start_i - end_{i-1} - 1)\ \prod_{i=1}^{|\mathbf{e}|} p_{LM}(e_i|e_1 \dots e_{i-1})$$

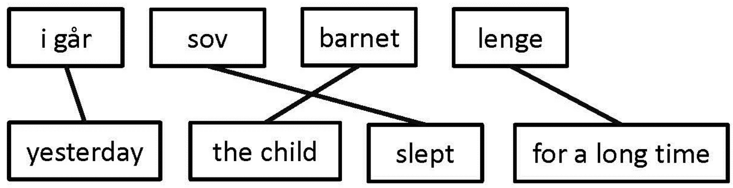

Figure 1:

To understand this formula, we may consider an example where we are to translate the following sentence from Norwegian to English.

(4) i går sov barnet lenge

One possible phrase alignment for this sentence is illustrated by figure 1.

a) (10%) Given this alignment, explain the following parts of the formula.

   i) What is I in the example?
   ii) What is the sequence: $(\bar{f}_1, \bar{e}_1), (\bar{f}_2, \bar{e}_2) \ldots (\bar{f}_I, \bar{e}_I)$ in the example?
   iii) For $i = 1, \ldots I$, what is $(start_i - end_{i-1} - 1)$ in the example?
   iv) Given a pair $\bar{f}_i, \bar{e}_i$, what does $\phi(\bar{f}_i|\bar{e}_i)$ express? (You are not supposed to know the actual values here.)

b) (20%) We assume we have trained a model for phrase-based SMT and we will translate a Norwegian sentence, like (4). We have to use an algorithm searching for the best English translation based on formula (3). This is called decoding. An exhaustive search for the best translation would be intractable. We therefore use various techniques for reducing the search space. One of them is called *pruning*. One way to implement decoding with pruning is to use a stack decoder. Explain the main steps in the stack decoding algorithm.

   (You do not have to include other optimization steps like *Recombination* or the *Estimation of future cost*.)

3 Speech recognition (15%)

a) What is the fundamental frequency F0 of a speech sound? (5%)

b) Calculate the Word Error Rate (WER) of the following utterance (10%):

Gold standard: "please turn a little bit to the left"
Actual utterance: "turn hi little bee to the left hey"

4 Noisy-channel models (20%)

You wish to develop a module to automatically filter out disfluent repetitions such as "the the" from a speech corpus. To this end, you adopt a noisy-channel approach to search for the most likely cleaned-up utterance X^* for the observed (possibly disfluent) utterance Y:

(5) $X^* = \arg\max P(Y|X)P(X)$

The noisy-channel model above is a combination of two models:

- The probabilistic model for $P(X)$ is a standard bigram language model. Here is a small excerpt of this model (where `<s>` and `</s>` denote the start and end symbol):

Word w_{i-1}	Word w_i	Log-probability
`<s>`	skoene	-5.2
boksen	`</s>`	-4.7
den	lille	-2.6
er	i	-1.7
i	den	-1.8
lille	boksen	-4.9
lille	lille	-6.9
skoene	er	-4.4

- The channel model $P(Y|X)$ assumes that every word has a probability 0.99 of generating the exact same word in the observed utterance, and a probability 0.01 of generating a repetition (that is, generating the repetition "$w_i\, w_i$" in the observed utterance Y from the initial word "w_i" in X).

Based on this noisy-channel model, you wish to process the following utterance (representing the observation Y):

"skoene er i den lille lille boksen"

Questions:

a) Analyse the above utterance according to Shriberg's disfluency model (5%).

b) Apply the noisy-channel model to the two following hypothesis for the output X:
Hypothesis 1: $X =$"skoene er i den lille lille boksen"
Hypothesis 2: $X =$"skoene er i den lille boksen"

For each of the two hypotheses, calculate its corresponding value $\log(P(Y|X)P(X))$. Which hypotheses has the highest value and will therefore be selected according to Eq. (5)? Explain your steps (15%).

NB: $\log(0.99) \approx -0.004$ and $\log(0.01) = -2$.

5 Evaluation (15%)

For the last obligatory assignment, you had to design a simple spoken dialogue system for a domain of your choice.

a) How would you evaluate your dialogue system based on the PARADISE framework? Describe the general methodology followed by PARADISE (7%).

b) List a few metrics (at least four) that could be used to evaluate your dialogue system (8%).

END