Master Degree in Computer Engineering

Natural Language Processing
Final Exam

September 11th, 2026

1. **[2 points]** List the most common tasks involved in text normalization. Can you provide examples of tasks that should be applied in some fixed order?

2. **[6 points]** Consider the task of POS tagging and the following hidden Markov model, consisting of transition and emission matrices specified in terms of costs (negative log probabilities). As an example, the transition from PN to VB has score 7, and the transition from VB to PN has score 9.

	DT	PN	PP	VB	⟨EOS⟩
⟨BOS⟩	7	7	13	17	22
DT	22	11	19	19	19
PN	11	7	9	7	9
PP	7	9	17	14	19
VB	7	9	9	8	9

	ils	sont	allés
DT	15	17	21
PN	5	19	18
PP	19	17	18
VB	17	6	6

   (a) Draw a finite state representation of the model, including the special states ⟨BOS⟩ and ⟨EOS⟩.

   (b) Consider the French sentence fragment 'ils sont allés'. In the following table each entry represents the lowest cost of reaching the associate word/POS pair, starting from ⟨BOS⟩. Use the Viterbi algorithm to fill in the table and to calculate the lowest cost POS tag sequence.

	⟨BOS⟩	ils	sont	allés	⟨EOS⟩
⟨BOS⟩	0	–	–	–	–
DT	–				–
PN	–				–
PP	–				–
VB	–				–
⟨EOS⟩	–	–	–	–	

3. **[5 points]** In the context of the BERT neural model, answer the following questions.

   (a) Describe the bidirectional encoding implemented by BERT.

   (b) Describe and motivate the masked language modeling task.

   (c) Describe and motivate the next sentence prediction task.

4. **[2 points]** Define the general problem of structured prediction and express it as an optimization problem in a mathematical way. Introduce and discuss an example of a natural language processing task that can be formulated as a structured prediction problem.

(see next page)

5. **[7 points]** In the context of large language models (LLMs) and pretraining, answer the following questions.

   (a) Explain how the decoder component of a transformer can be turned into a language model and display the full stack view of the language modeling head.

   (b) Introduce the decoding problem for LLM and discuss the sampling techniques for its solution that have been presented in class.

   (c) Discuss the trade-off between quality and diversity in text generation.

6. **[4 points]** Introduce the mathematical relations known as scaling laws. Discuss how these laws can be practically used to improve pre-training strategies.

7. **[5 points]** In the context of machine translation, answer the following questions.

   (a) Introduce the neural translation model based on Transformer that has been presented in class, providing its basic equations and explaining how the model works.

   (b) Define the notion of masked self-attention and specify in which part of the model this is used.

   (c) Define the notion of cross-attention and specify in which part of the model this is used.

8. **[2 points]** What is the general idea underlying transfer learning? Why is transfer learning relevant to the LLM pretraining/fine-tuning paradigm?