Master Degree in Computer Engineering

Natural Language Processing
Final Exam

June 20th, 2025

1. **[2 points]** According to lecture 2 'Essentials of linguistics', the field of generative linguistics posits five levels of linguistic description. Introduce and explain each of these five levels.

2. **[6 points]** In the context of POS tagging for Sweden, consider the hidden Markov model consisting of the transition and emission matrices reported below. These matrices are specified in terms of **costs** (negative log probabilities). As an example, the transition from PP to PL has cost 12, and the transition from PL to PP has cost 2.

	PL	PN	PP	VB	⟨EOS⟩
⟨BOS⟩	11	2	3	4	19
PL	17	3	2	5	7
PN	5	4	3	1	8
PP	12	4	6	7	9
VB	3	2	3	3	7

	hen	vilar	ut
PL	17	17	4
PN	3	19	19
PP	19	19	3
VB	19	8	19

Consider the Swedish sentence fragment 'hen vilar ut' (literal translation: she rests out). In the following table each entry represents the **lowest cost** of reaching the associated word/POS pair, starting from ⟨BOS⟩. Use the Viterbi algorithm to fill in the table and to calculate the lowest cost POS tag sequence. Report the intermediate computation for each entry in the table.

	⟨BOS⟩	hen	vilar	ut	⟨EOS⟩
⟨BOS⟩	0	–	–	–	–
PL	–				–
PN	–				–
PP	–				–
VB	–				–
⟨EOS⟩	–	–	–	–	

3. **[6 points]** Considering contextualized word embeddings, answer the following questions.
   (a) Introduce the general architecture of the model known as ELMo (embeddings from language model).
   (b) With reference to the char-CNN layer, specify and discuss the so-called highway block.
   (c) Specify how the output embeddings at each of the three layers are combined to produce contextualized word embeddings.

(see next page)

4. **[3 points]** In the context of decoder-only large language models, answer the following questions.

   (a) Introduce and explain the most common strategies used for the task of generating the next word (sampling).

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

5. **[5 points]** In the context of transition-based parsing, answer the following questions.

   (a) Define the notion of spurious ambiguity, as we have introduced it in the course lectures.

   (b) State two different sequences of transitions that make an arc-standard parser produce the projective dependency tree consisting of the following unlabeled dependency relations

head	w_3	w_3	$\langle \text{ROOT} \rangle$	w_3
dependent	w_1	w_2	w_3	w_4

6. **[2 points]** In the context of large language models, introduce the process of instruction tuning, also called supervised fine tuning, and explain its main rationale. Mention also some of the most popular datasets for instruction tuning.

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

   (a) Introduce the encoder-decoder architecture using recurrent neural network (RNN) and static context vector $\mathbf{c}$, and draw the model recurrent equations.

   (b) Augment the model in (a) with the use of dynamic context vector $\mathbf{c}_t$ computed under bilinear model attention.

8. **[2 points]** In the context of large reasoning models, discuss the paradigm shift from scaling train-time compute to scaling test-time compute.