Master Degree in Computer Engineering

Natural Language Processing
Final Exam

July 10th, 2025

1. **[2 points]** Introduce the general problem of ambiguity for natural language processing. Provide examples of ambiguity at several layers of linguistic descriptions, such as morphology, syntax, lexical semantics, etc.

2. **[6 points]** An English corpus T has been tokenized based on white spaces. The resulting dictionary and word frequencies are reported in the following table

word	hot	hottest	host	shot	shortest
freq	17	9	7	11	5

   Apply the byte-pair encoding (BPE) algorithm to derive subword tokens for T, using the character '_' to mark the end of each word. Report and comment each of the first eleven iterations (merge operations) in a run of the algorithm, including the frequency updates.

3. **[6 points]** In the context of N-gram techniques for statistical language modeling, answer the following questions.

   (a) Introduce the general idea underlying backoff techniques. Define and explain the so-called stupid backoff.

   (b) Define the technique of simple linear interpolation, and discuss possible strategies for the choice of the interpolation values.

4. **[2 points]** Introduce and discuss the basic ideas underlying concatenative morphology and template morphology, and provide some examples.

(see next page)

5. [**3 points**] In the context of large language models, introduce the relationships known as scaling laws. Discuss how these laws can be practically used to improve pre-training strategies.

6. [**5 points**] Consider the projective dependency tree consisting of the following unlabeled dependency relations

head	w_3	w_1	$\langle$ROOT$\rangle$	w_3	w_7	w_5	w_3	w_7
dependent	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8

   Answer the following questions.

   (a) Draw a graphical representation of the dependency tree above, with arcs directed from the head to the dependent.

   (b) Apply to the above tree the oracle presented in syntactic parsing lecture, to construct a sequence of training instances for the arc-standard parser.

7. [**7 points**] In the course lectures we have presented two deep learning models that derive an attention distribution by computing a vector of so-called bilinear products. Describe the main idea of bilinear attention and the two applications in which we have exploited it, and report the equations of the two models.

8. [**2 points**] Discuss some of the good practices in prompt engineering we have presented in the LLM's lectures, and provide some examples.