

**Natural Language Processing
Final Exam**

February 4th, 2026

1. **[2 points]** In the context of the linguistic study of word meaning, introduce the two notions of internal semantic structure and external semantic structure of a word. Provide some simple examples.
2. **[1 points]** Outline the historical periods of the field of natural language processing from years 60's onward, as presented in lecture 1. Discuss the dominant research paradigms associated to each of these periods.
3. **[6 points]** Some English document d has been tokenized based on white spaces. The resulting dictionary and word frequencies are reported in the following table

word	playing	staying	played	stayed	strayed
freq	15	10	12	8	5

Apply the byte pair encoding algorithm to derive subword tokens for d , using the character ‘_’ to mark the end of each word. Report and comment each of the first twelve iterations (merge operations) in a run of the algorithm, and include the frequency updates **for each step**.

4. **[6 points]** In the context of large language models (LLMs), answer the following questions.
 - (a) Introduce the decoder-only architecture for LLM, and explain the role of causal masking in the self-attention layer.
 - (b) Explain how the model transforms the final hidden state of a token into a probability distribution over the vocabulary.
 - (c) Explain the method of weight tying and provide one reason why this is beneficial for the model.
 - (d) Define the objective function for next token prediction. During training, do we feed the model its own previous predictions or the actual ground-truth tokens?
 - (e) Explain how a decoder-only model performs text completion. Provide examples of how the two tasks of sentiment analysis and translation can be reformulated as a simple text completion prompt.

(see next page)

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

head	w_3	w_3	$\langle \text{ROOT} \rangle$	w_3	w_6	w_7	w_3	w_7	w_7
dependent	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9

Answer the following questions.

- Draw a graphical representation of the dependency tree above, with arcs directed from the head to the dependent.
 - Apply to the above tree the oracle presented in syntactic parsing lecture, to construct a sequence of training instances for the arc-standard parser.
6. **[6 points]** In the context of the task of machine reading, answer the following questions.
- Introduce the notions of query, passage and span and the associated notation.
 - Define the main equations of the neural model known as Stanford attentive reader, and explain how the model implements the attention mechanism.
7. **[3 points]** Introduce the architecture of retrieval-augmented generation (RAG) systems, and discuss the important advantages of RAG with respect to standard chatbots.
8. **[3 points]** Introduce the so-called large reasoning models (LRMs) and explain how LRMs achieve their reasoning capabilities.