First Midterm Exam

601.467/667 Introduction to Human Language Technology

Fall 2026
Johns Hopkins University
Co-ordinator: Philipp Koehn

8 October 2026

Complete all questions.

Use additional paper if needed.

Time: 75 minutes.

Name of student: _____

Q1. Syntax *20 points*

Formal Language Theory

1. **Reading a regular grammar (4 points)**

   Consider the grammar below, with start symbol S and alphabet $\Sigma = \{a, b\}$:

   $$S \to aS \mid bA \qquad A \to bA \mid \epsilon$$

   a. Describe the language generated by this grammar in words, and give a regular expression for it. *(2 points)*

   b. Can the grammar generate `b`? Can it generate `aba`? Why or why not? *(1 point)*

   c. Give a sequence of rule applications that generates `aabb`. *(1 point)*

2. **Applying the Chomsky hierarchy (4 points)**

   For each production below, identify the **most restrictive Chomsky-hierarchy rule format** (regular, context-free, context-sensitive, or unrestricted) that permits it, and explain your choice. Standard conventions apply: capital letters denote nonterminals, and lowercase letters denote terminals.

   a. $A \to aB$ *(1 point)*

   b. $A \to BC$ *(1 point)*

   c. $bAc \to bDc$ *(1 point)*

   d. $bA \to c$ *(1 point)*

Constituency and Parsing

3. **From a treebank tree to a grammar (5 points)**

   A treebank contains the following constituent tree:

```
(S
  (NP (DT the) (NN student))
  (VP (VBD saw)
      (NP (DT a) (NN bird))))
```

   a. Extract all distinct CFG productions represented by this tree, including productions that generate words. List a repeated production only once. *(3 points)*

   b. Use the substitution test to provide evidence that `a bird` is a constituent. *(1 point)*

   c. Explain why a CFG extracted from a large treebank can generate sentences that native speakers judge ungrammatical. Hint: use the rules you extracted above (feel free to add some). *(1 point)*

4. **Binarization and a small CKY parse (7 points)**

Consider this grammar, with start symbol S:

$$\begin{aligned} \mathrm{S} &\rightarrow \mathrm{NP}\ \mathrm{V}\ \mathrm{NP} \\ \mathrm{NP} &\rightarrow \mathrm{D}\ \mathrm{N} \\ \mathrm{D} &\rightarrow \text{a} \mid \text{the} \\ \mathrm{N} &\rightarrow \text{robot} \mid \text{bird} \\ \mathrm{V} &\rightarrow \text{sees} \end{aligned}$$

a. Binarize the first production by grouping $\mathrm{V}\,\mathrm{NP}$ under a new nonterminal X. Write the two resulting productions. *(2 points)*

b. Using the binarized grammar, apply CKY to:

$$\text{a}_1 \quad \text{robot}_2 \quad \text{sees}_3 \quad \text{the}_4 \quad \text{bird}_5$$

List **every nonempty chart entry spanning at least two words**. Use inclusive word positions: for example, $\mathrm{NP}[1, 2]$ means an NP covering words 1 through 2. *(4 points)*

c. Does the grammar recognize the sentence? Identify the chart entry that establishes your answer. *(1 point)*

Q2. Semantics *15 points*

Consider the following document-term count matrix:

Doc #	cat	tail	grep	dog	python	leash
1	12	8	0	0	3	7
2	0	0	0	10	0	11
3	17	0	15	0	1	1
4	5	13	7	0	9	0

1. Treating the matrix as our initial knowledge of similarity between words and between documents, which pairs of words and pairs of documents have zero similarity? Explain why. *(5 points)*

2. Under a semantic model trained using the distributional information in the table above, which of the initially-zero-similarity words would ”dog” be closest to? Explain why. *(5 points)*

3. Alongside the ”animal”-semantics, the table shows a ”programming”-semantics, where some words have particular meanings in each. Such words are *polysemous*, like ”tail” and ”cat”. What difference between *how* these two words are polysemous does WordNet *fail* to capture? *(5 points)*

Q3. Deep Learning *10 points*

1. Why do we use a softmax layer? *(2 points)*

2. What is supervised learning? *(2 points)*

3. Please give 2 examples of Supervised Learning Tasks in HLT. *(2 points)*

4. Why have deep learning models gotten better in the last 15 years? Name 2 reasons for this. *(4 points)*

Q4. Information Retrieval *10 points*

1. What is the difference between queries and topics? *(3 points)*

2. What is the benefit of learned sparse retrieval models compared to dense retrieval models? *(3 points)*

3. What is the benefit of retrieval-augmented generation compared to only using an LLM to respond to users? *(4 points)*

Q5. Machine Translation *10 points*

1. Define the three main application scenarios for machine translation: *(3 points)*

   - Dissemination:
   - Assimilation:
   - Communication:

2. Explain how each application scenario imposes different requirements on the translation accuracy and speed/memory for the MT engine. *(2 points)*

3. Please demonstrate your understanding of the attention mechanism in a Transformer encoder block. Suppose I have an input of 3 tokens with embeddings e_1, e_2, e_3, where $e_j \in \mathcal{R}^d$. I would like to compute the representations for the next layer l_1, l_2, l_3, where $l_i \in \mathcal{R}^d$. Using the idea of scaled dot-product (self)-attention in Transformers, write down the equation for one of these vectors l_1 as a function of the other terms. Also, indicate which term is query, key, and value. You do not need to use multi-head attention; just assume a single head for simplicity. *(5 points)*

Q6. Large Language Models *10 points*

1. What training strategy is used to make models less harmful and why? *(3 points)*

2. The design of today's transformer large language models is called *decoder-only*. Explain the origin of this term. *(3 points)*

3. Why is Low-Rank Adaptation more efficient? Compute the savings in number of weights that are trained for a 1000-dimensional matrix when using a LoRA adapter of rank 20. *(4 points)*

Q7. Neural and Symbolic Language Processing *10 points*

1. **Prompt sensitivity.** An LLM gives the wrong answer to a question. You rephrase the question without adding any new information, and it now gives the correct answer. What does this tell us about how LLMs process knowledge? *(3 points)*

2. **Combining neural and symbolic methods.** Consider a task in which an LLM must produce an answer that obeys a strict set of logical constraints. What advantage could a symbolic solver provide if it is combined with the LLM? *(3 points)*

3. **Agentic system design.** A task can be implemented either as a LangFlow-style fixed workflow or as a ReAct-style agent that decides which action to take next based on intermediate results. Give one type of task for which each approach would be preferable, and briefly explain why. *(4 points)*

Extra Space