BRAC UNIVERSITY
Department of Computer Science and Engineering

Examination: Final   Semester: Fall 2024
Duration: 80 minutes   Full Marks: 30

CSE 440: Natural Language Processing II
Figures in the right margin indicate marks.

Answer all 3.

1. A. You have three sentence pairs for English-Spanish translation: [6]
        a. the blue house → la casa azul
        b. house → casa
        c. blue → azul
        Perform 3 rounds of EM algorithm simulation on this.

    B. Why are CPUs not ideal for large neural model training/inference? Give   [2]
       two reasons.

    C. Write a short note on Gated Recurrent Units.   [2]

2. A. What are the major challenges of a dense yet static word representation   [2]
       like word2vec?

    B. Provide an example of a real-world incident where gender bias in word   [2]
       embeddings can cause harm or controversy.

    C. Derive the equation for an HMM based PoS tagger using appropriate   [6]
       approximations. Show (with examples and equations) how you can
       calculate tag transition probabilities and emission probabilities.

3. A. Write short notes on: [2]
        a. Optimizer
        b. Regularizer

    A. Compare MEMM and HMM. You have to write two points where MEMM is   [4]
       better than HMM, and two more points where HMM is more useful than
       MEMM.

    B. Explain in detail the attention mechanism of an encoder decoder   [4]
       translation model.