

**START OF QUIZ**

**Student ID:**

**31181845, Yuan, Su**

## Question 1

Topic: Lecture 4

Source: Lecture 4

What kind of tasks are CRFs appropriate for (2 requirements)? Don't list tasks, but rather the general class of tasks. (1)

## Question 2

Topic: Lecture 3

Source: Lecture 3

Describe, with an example (not used in class) the difference between surface and canonical segmentation. (1)

### Question 3

Topic: Lecture 4

Source: Lecture 4

What is the role of the partition function ( $Z(x)$ ) in CRFs? Why is it necessary? (1)

## Question 4

Topic: Lecture 3

Source: Lecture 3

If we were to try to use an HMM for segmentation, describe what the transition and emission probabilities would be. (1)

## Question 5

Topic: Lecture 2

Source: Lecture 2

Do you think that FSTs can work with reduplicative morphology? Explain. (1)

## Question 6

Topic: Lecture 1  
Source: Lecture 1

Suppletion is a process by which morphological patterns (called paradigms) merge to form a mixed paradigm. For example, the past tense of “to go” comes from an older verb, “wendan - to turn”. Describe how syncretic paradigms might impact a machine learning model, and how we can learn to model them accurately. (2)

## Question 7

Topic: Lecture 1

Source: Lecture 1

Vowel harmony is a process by which vowels in affixes must match some of the properties of the vowels in the root. For example, in Turkish, “houses” is “evler”, while “schools” is “okullar”, where the plural suffix must have a front or back vowel, matching the root (“ev” and “okul”). Given the ML models you’ve seen so far, give a specific example of a model that you think can learn this process, and explain why it’s well suited to the task. (2)



## Question 8

Topic: Lecture 2

Source: Lecture 2

In some ways, Statistical Machine Translation (SMT) was similar to an FST modified by a re-ordering model (ie, each input word had a corresponding output translation, and then the words were re-ordered to fit a language model). These models have been supplanted by NMT. What shortcomings of FSTs do you think put a ceiling on SMT performance? (2)

## Question 9

Topic: Long

Source: Lecture 4

In 1799, Napoleon's armies were raiding Egypt when they discovered the Rosetta stone, which contained a parallel corpus of Latin, Greek, and Egyptian (written in hieroglyphics). This was one of the greatest linguistic discoveries, as it allowed linguists to decipher the long unknown Egyptian language. Imagine that we found a similar stone today, but with several languages (you can assume they are well-known languages), including Linear A (a language spoken in ancient Crete). Would you prefer that the languages have high morphological complexity (like, say Finnish or Turkish), simpler morphology (like English or Chinese), or something in the middle, and how would you use this information to inform your automated approach to decrypting Linear A? You can ignore the fact that modern languages didn't exist when Linear A was spoken. Assume that the other languages are completely interpretable.

(3)

**END OF QUIZ**