# START OF QUIZ
# Student ID:
# 27841444,Gu,Martin

## Question 1

Topic: Lecture 1
Source: Lecture 1

Describe allomorphy, with an example we did not cover in class. (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

Do you think that we could do Chinese Word Segmentation in a bottom-up way like we do with BPE? Why might this work (or not)? (1)

## Question 4

Topic: Lecture 4
Source: Lecture 4

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

## Question 5

Topic: Lecture 1
Source: Lecture 1

Explain how morphology, phonology, and syntax are tightly bound to each other. (1)

## Question 6

Topic: Lecture 2
Source: Lecture 2

As a thought experiment, how might we build a calculator using an FST? Imagine that the FST reads input on one side of the tape, and generates operations (that are carried out by an algorithm) on the output side. (2)

## Question 7

Topic: Lecture 3
Source: Lecture 3

For a language like Archi, which has extremely productive inflection (a verb can theoretically appear in over 1.5 million different forms), do you think that a larger or smaller BPE vocabulary size would be more beneficial? Explain your assumptions about the morphological structure of the language when making your assessment. (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 3

There is an alternative to BPE that randomly "forgets" to merge together certain subword sequences when it is creating its vocabulary (for example, "forget" will occasionally be represented as "for-get", occasionally as "forget", occasionally as "for-g-et", etc. What impacts do you think this might have on the vocabulary and model performance? Secondly, do you think there is a different impact between forgetting early iteration, mid iteration, and late iteration merges? (3)

# END OF QUIZ