# START OF QUIZ
# Student ID: 12582557,li Chen,Mariana

## Question 1

Topic: Lecture 3
Source: Lecture 3

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

## Question 2

Topic: Lecture 4
Source: Lecture 4

Transition-based segmentation is very similar to the SR parser we saw last block, except it uses 2 FIFO structures, and doesn't require a stack. What is different about segmentation so that it doesn't require a stack? (1)

## Question 3

Topic: Lecture 4
Source: Lecture 4

Literate Chinese speakers have no difficulty parsing words in Chinese text (outside normal ambiguities). What advantages do you think they have that our algorithms are lacking? (1)

## Question 4

Topic: Lecture 2
Source: Lecture 2

In class, I was very careful to only have deletion and substitution in our rewrite rules. What implications might insertion have on rewrites? (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 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 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 1

In English (and other stress-timed languages, such as German, Russian, Arabic, Greek, Hindi, Thai, etc.), stress tends to be strongest on the root of the word, and is softer along affixes (and in English, on periphrastic necessities like auxiliary verbs - try it!). What implications might this have on an ASR system, do you think they are a significant issue, and can you envision any way of moderating them with morphological knowledge? (3)

# END OF QUIZ