# START OF QUIZ
# Student ID:
# 18643544,Zhang,Cindy

## Question 1

Topic: Lecture 3
Source: Lecture 3

In class, we saw that the entropy of a fair coin toss is 1 (bit), but that was because we were using a binary logarithm. Similarly, the entropy of an 8-sided fair die is 1, if we use an octal logarithm. What is the entropy of an 8-sided die using a binary logarithm? Either show your work or explain the relation. (1)

## Question 2

Topic: Lecture 1
Source: Lecture 1

Knowing what you know about parsing, describe how compounding could be considered syntax, instead of morphology. In other words, how might we parse compounds? (1)

## Question 3

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 4

Topic: Lecture 2
Source: Lecture 2

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

## Question 5

Topic: Lecture 1
Source: Lecture 1

English is often described as an "analytic language with some fusional properties". Describe what that means, with an example. (1)

## Question 6

Topic: Lecture 4
Source: Lecture 4

Imagine that we had a language like Chinese that doesn't use spaces for word segmentation, but is considerably more morphologically complex than Chinese. Do you think that it would be easier or harder to segment? Give an explanation (and include any assumptions). (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 4
Source: Lecture 4

Garden path sentences are sentences that start with one parse, but need to be reparsed in the middle of the sentence ("The old man the boats." - 'old' changes from an adjective to a noun, and 'man' from a noun to a verb). A bad Chinese word segmentation could result in the same need to re-parse our segmentation after encountering a new word. Of the methods we looked at, which do you think is the most likely to be able to "correct" a segmentation? Explain. (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