Ling 473 Assignment 1
Due 4:30pm on Thursday July 27, 2017

1. (25 points) Write a paragraph describing how you became interested in Computational Linguistics, any projects or specific areas you're interested in, and/or career goals. How would you characterize your experience in linguistics, math, or computer programming (or other relevant engineering)? Recalling the lecture slides, which of the subfields or subtasks of Computational Linguistics are you particularly interested in?

2. (25 points) Consider the following sentence:

   I saw that gas can explode.

   a. How many phrase structure trees can you find for this sentence? Do not include pragmatically odd interpretations. Draw each tree and provide a discriminating explanation of the situation modeled by the interpretation.

   b. Write the phrase structure trees from the previous question using Penn Treebank notation. That is, write it with brackets and parentheses: (S (NP (NNP Kim)) (VP (VBZ sleeps)))

3. (10 points) How many six-letter "words" can be formed from the alphabet { a – z }? A "word" for this question must have at least one vowel { a e i o u }, and may not contain all vowels. Show your work and explain your answer.

4. (10 points) How many ways can the characters in the following tuple be arranged?

   (萄 萄 萄 萄 橙 橙 苹 梨 蕉)

5. (30 points) Consider a document processing system which performs pairwise comparisons and a corpus containing 19 documents as follows:

Topic	Count
Conference Proceedings	7
Journal Articles	9
Workshop Abstracts	3

   a. How many pairwise comparisons are possible between documents on the same topic?

   b. How many pairwise comparisons are possible between documents on different topics?

** (10 points, **extra credit**) In the lecture, we showed that you can form

$$\frac{n!}{(n-k)! \; k!}$$

different unordered sets of k distinct items from a set of n distinct items. Write an expression that gives the number of unordered sets of k items that can be formed from a set of n distinct items while allowing repetition in the output set.