**Deliverable 4**

Search for perplexity measures (Section 3.7 in J&M 3rd2018) in Python and compare perplexity to lexical diversity, as specified on page 9 of BKL, on the corpus downloaded from the Python NLTK in the previous exercise. What does each measure? **Is there a potential relationship and, if so, what is it? Is there a relationship of these notions to information content?** This is not a pure programming exercise.

**Perplexity**

* **Perplexity** is a measurement of **how well a probability model predicts a test data**. In the context of Natural Language Processing, perplexity is one way to **evaluate language models**.
* Perplexity is just an *exponentiation of the entropy*!
* Low perplexity is good and high perplexity is bad since the perplexity is the exponentiation of the entropy

**Lexical Diversity**

* **First,** We need to understand Lexical Richness of a Text, the basic idea behind that measure is that if the text is more complex or more varied, the writer of the text uses a more varied vocabulary so there’s a larger number of types or we can say a large number of unique words.
* In order to Calculate Lexical Richness, we use a metric called **Lexical Diversity.**Lexical Diversity, According To [Wikipedia](https://en.wikipedia.org/wiki/Lexical_diversity) is the ratio of different unique word stems (types) to the total number of words (tokens).
* It is an indicator about a text as in how complex and difficult to read it is.