**Stimuli Generation**

Working Memory Training as Treatment for Post-Stroke Aphasia

**What’s Required:**

* 180 words
  + Split into 10 lists of 18 each
    - Balanced for length and lexical frequency
  + Between 1-2 syllables long
  + Zipf value between 4 and 5
* 180 non-words
  + Split into 10 lists of 18 each
    - Balanced for length
  + Between 1-2 syllables long

*\*\*See lines 282-292 within the attached manuscript (NRH-ES 42.22.R1\_Proof\_hi) for a description of how this sorting was completed for a previous experiment.*

**Steps:**

1. Narrow down SUBTLEX excel document to only words that are within the required frequency (Zipf between 4 and 5).

2. Use NLTK package within Python (use Anaconda interface):

NLTK: <https://www.nltk.org/install.html>

Anaconda: <https://www.anaconda.com/products/distribution>

a. Determine number of syllables for each word

- Use CMU Pronunciation Dictionary from NLTK

b. Convert everything to bare lemmas

- Use WordNet Lemmatizer from NLTK

3. Within the excel document, using the new column with syllable information, eliminate any rows with greater than two syllables.

4. Run remaining words through SOS (in MatLab) to sort into 18 lists that are balanced for length and lexical frequency.

- SOS scripts are in stim\_creation folder

- You can use these scripts as a starting point, and edit them to produce the correct number of lists with correct number of items and balanced length and frequency distribution

5. Input the completed lists into Wuggy to develop matching non-words lists.

Wuggy: <http://crr.ugent.be/programs-data/wuggy>