## **Data exploration / linguistic indicators**

**Results**

**(2) z-scores of ngrams for avoidance / non-avoidance and different strategies (using FightingWords)**

**2.1 Avoidance / non-avoidance**

* Non-avoidance:
  + 10 most frequent bigrams in non-avoidance answers: ['do know', 'you know', 'little bit', 'it was', 'do think', 'at all', 'what do', 'think about', 'what are', 'out there']
* Avoidance
  + 10 most frequent bigrams in avoidance answers: ['hon friend', 'hon lady', 'does not', 'secretary state', 'will minister', 'we are', 'those who', 'will he', 'does hon', 'am sure']

→ all with significant z-scores

**2.2 ‘Fight’ vs. ‘Flight’**

* Fight:
  + 10 most frequent bigrams in ‘Fight’ answers: ['the hon', 'hon lady', 'will right', 'do not', 'per cent', 'right hon', 'hon gentleman', 'does hon', 'secretary state', 'we are']

→ only last three with significant z-score

* Flight
  + 10 most frequent bigrams in ‘Flight’ answers: ['hon friend', 'will he', 'my hon', 'they are', 'does he', 'such as', 'he will', 'will minister', 'has been', 'last year']

→ none with significant z-score

### **(3) Levenshtein distance between question and answer**

Threshold refers to the length difference between question and answer. When one of the two is much longer than the other, the Levenshtein distance will also increase.

**3.1 Avoidance vs. non-avoidance**

* Average Levenshtein distance for non-avoidance Q-A pairs: 314.91
  + - 10% percentile: 138.4
    - 90% percentile: 491.8
* Average Levenshtein distance for avoidance Q-A pairs: 372.79
  + - 10% percentile: 127.7
    - 90% percentile: 466.8

→ non-avoidance answers are more similar to their questions w.r.t. sequential form

**3.2 ‘Fight’ vs. ‘Flight’**

* Average Levenshtein distance for 'Flight' Q-A pairs: 303.16
  + - 10% percentile: 127.7
    - 90% percentile: 466.8
* Average Levenshtein distance for 'Fight' Q-A pairs: 375.96
  + - 10% percentile: 152.4
    - 90% percentile: 502.4

→ ‘Fight’ pairs are less similar w.r.t. their surface form