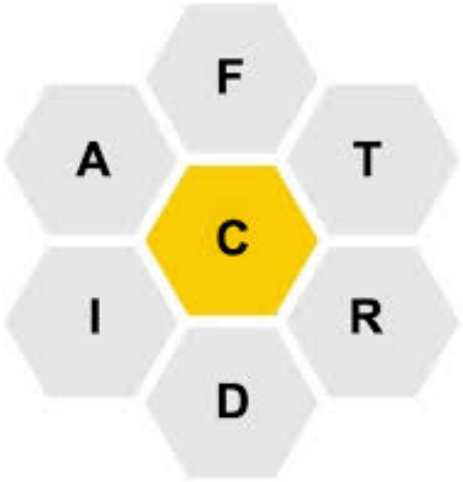
Linguistic Analysis of the New York Times Spelling Bee

Introduction

I started playing the New York Times Spelling Bee a couple years ago and I've always found it to be a relaxing activity, until recently. The rules of the Spelling Bee are relatively simple. The New York Times[1] lists the rules as the following:



> - Construct as many words as you can using at least 4 letters, including the center letter of the puzzle.
> - Words should be at least 4 letters long (no maximum limit).
> - Each Spelling Bee puzzle is curated to focus on relatively common words (with a few tougher ones periodically to keep things challenging). We try to avoid terms that are ultra-specific to any professional field to maintain a level playing field for all of our solvers.

The idea of using common words is important because a vital aspect of language is shared understanding between the listener and the speaker. A 2017 study[2] on the use of colloquial words in classrooms

found that the increasing in colloquial words used correlated with "the increasing of students' fluency, vocabulary and confidence in speaking by teacher's assessment" (Hasanah 33)[2].

Lately, I've been questioning the last rule for the NYT Spelling Bee that all words are "relatively common" with a few exceptions. I've noticed that the last 5 to 10 words in every puzzle for me are words that I've never heard of before. With each puzzle containing roughly 20-40 possible words, only 75% of these are, in my opinion, relatively common. But this is merely a hypothesis and I want to put actual numbers to this idea. In this analysis I will seek to answer the question that's been bugging me in the past few months: How common are the New York Times Spelling Bee words?

The Method

My first course of action in analyzing the "commonness" of the New York Times Spelling Bee words was to see if I could write a function to solve the Spelling Bee using only common words. I decided to use the words from the Brown corpus as my base corpus for all possible common words that could be present in the Spelling Bee. Given that the corpus is comprised of over 1 million words and 56,000 unique tokens, I figured we had a good chance of solving the puzzle using this corpus as a reference.

In [134…
```
import nltk
```

In [135…
```
words = [w.lower() for w in nltk.corpus.brown.words()
         if w.isalpha() and len(w)>3]
freqs = nltk.FreqDist(words)
words = set(words)
```

I then used regular expressions to only return words that followed the rules of the puzzle (at least 4 letters, use only puzzle letters, must use center letter at least once).

In [136…
```
import re
def NYTSpellingBee(letters):
```

```
    letters = letters.lower()
    sub = [w for w in words if letters[0] in w
           and re.fullmatch(r'\b[ ' + letters + r']+\b', w)
           != None]
    return sorted(sub)
```

After making this function, I decided to test it on the puzzle from 11/5/22: **Z**ABELMO

In [137…
```
NYTSpellingBee('ZABELMO')
```

Out[137]:
```
['ablaze',
 'amaze',
 'azalea',
 'blaze',
 'boaz',
 'booze',
 'embezzle',
 'laze',
 'maze',
 'mezzo',
 'ooze',
 'zabel',
 'zeal',
 'zeme']
```

From the Spelling Bee archives, I know that there are 18 words that NYT has deemed possible puzzle answers. My function found 12 of these answers plus two extra words ('boaz' and 'zeme'). However, there were 6 puzzle answers ('bamboozle', 'bazoo', 'bezel', 'bozo', 'mezze', and 'zoom') that the function did not find since they weren't in the Brown corpus. So, assuming that the Brown corpus represents "common" words, the NYT 11/5/22 Spelling Bee had $\frac{12}{18}$ or 67% "common" words in the total puzzle answers. This seems to support my theory that around 75% of the puzzle answers are actually puzzle answers.

The Analysis

So far I've explored just one puzzle. But I want to see if this is reflected across historic data from NYT Spelling Bee answers. However, in order

to test this, I needed historic data. The NYT was not very helpful in providing a dataset like this and I couldn't find one online, so I had to create my own from cross refrencing the archives. Due to limited time and access to the archives, the dataset is limited in that it only contains a few weeks worth of puzzles. But this will still give us a good look into the "commonness" of the Spelling Bee answers.

In [138…
```
import pandas as pd
data = pd.read_csv('NYTSpellingBee.csv')
```

I could then test each puzzle answer to see if it was in my "common" words corpus or not.

In [139…
```
def iscommon(word):
    word = word.lower()
    if word in words:
        return 1
    else:
        return 0
Common = [iscommon(str(w)) for w in list(data['Word'])]
data['Common'] = Common
```

From there, I can create a z-statistic confidence interval to generate an estimate for the true proportion of Spelling Bee answers that are "common" words.

In [140…
```
import math
[round(sum(Common)/len(Common) - 1.96 * math.sqrt(
     sum(Common)/len(Common)*(1-sum(Common)/
                              len(Common))/len(Common)),4),
 round(sum(Common)/len(Common) + 1.96 * math.sqrt(
     sum(Common)/len(Common)*(1-sum(Common)/
                              len(Common))/len(Common)),4)]
```

Out[140]:
```
[0.5524, 0.6151]
```

This confidence interval tells me that there's a 95% chance that the true proportion of Spelling Bee answers that are "common" words is between 55.24% to 61.51%. This is statistical evidence to support my hypothesis that less than 75% of the puzzle answers are "common" words.

Day of the Week Analysis

This analysis prompted a second question: Does the "commonness" of puzzle answers depend on the day of the week? NYT is nototious for posting more difficult puzzles on weekends to save its followers from the Sunday scaries. I've noticed from personal experience that the crosswords are certainly more difficult on Sundays and I would imagine that the Spelling Bee would follow this same trend. In order to test this hypothesis that Sunday puzzles are more difficult, I needed to establish a metric for determining difficulty. I decided to use the number of occurances of each puzzle answer in the Brown corpus as a "Commonness Score". A higher number of occurances means the puzzle answer is more common and thus easier to find for the average player. I started by defining a function to return the "Commonness Score" for a given word.

In [141…
```
def CommonScore(word):
    word = word.lower()
    score = freqs[word]
    return score
```

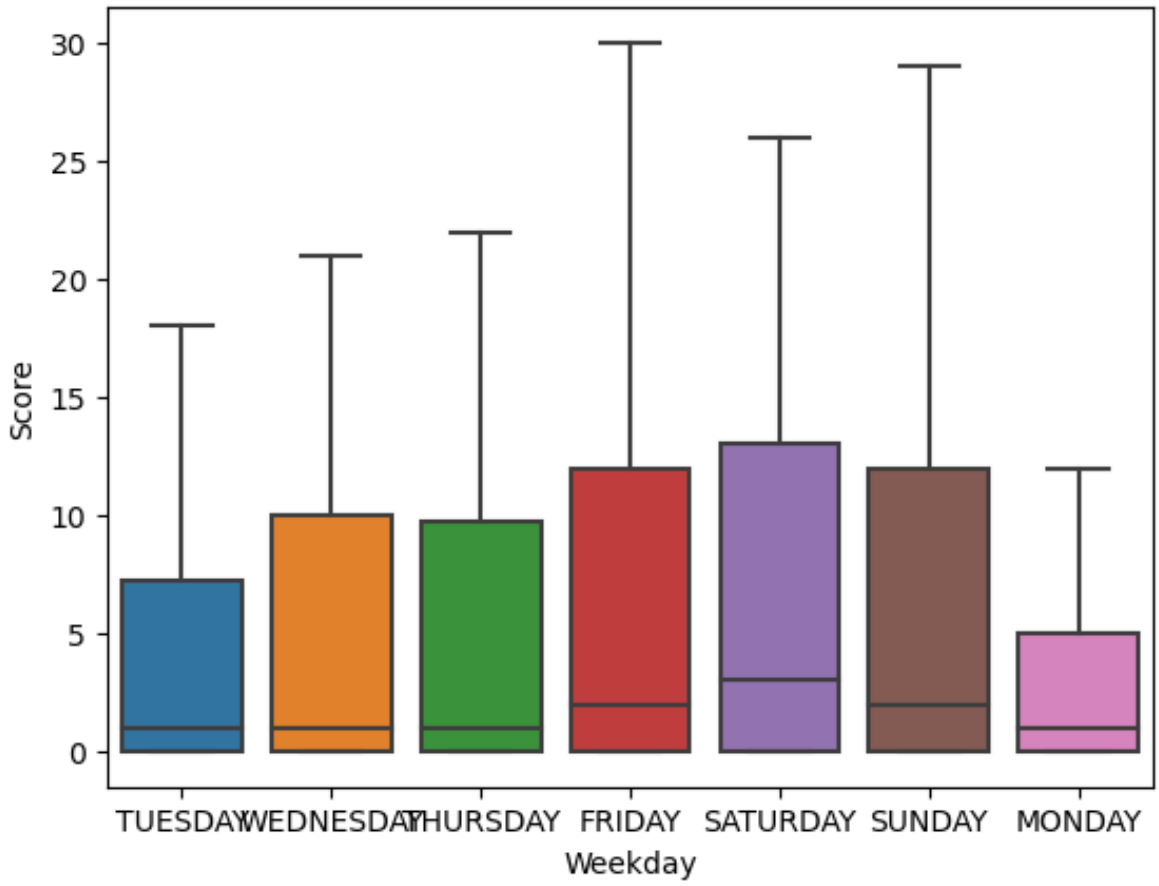
I then added a column to my data frame for the scores of each puzzle answer.

In [142…
```
scores = [CommonScore(str(w)) for w in list(data['Word'])]
data['Score'] = scores
```

This allowed me to make a histogram of the distribution of commonness for puzzle answers by weekday.

In [143…
```
import seaborn as sns
sns.boxplot(x = data['Weekday'],
            y = data['Score'],
            showfliers = False)
```

Out[143]:
```
<AxesSubplot: xlabel='Weekday', ylabel='Score'>
```



A wide histogram reveals a spread out distribution. Conversely, a thin band indicates a centered distribution. Recall that higher scores correlate to an easier puzzle. So, a thin band close to the lower end of the scores indicates a day of the week with consistently uncommon puzzle answers. For the sake of this analysis, this will be the indicator of a difficult puzzle. This histogram indicates that Monday Spelling Bee puzzles have a centered distribution around uncommon words. This is surprising to me as Monday crosswords and other puzzles tend to be easier. Puzzles during the workweek in general are typically easier, but we can see from the histogram that the mean common word scores for Monday-Thursday puzzles are the lowest of all the days of the week. There are certainly other metrics aside from commonness that comes into play for a puzzles overall difficulty, and if we were to explore that we might find that weekend puzzles are harder. But from the nature of my analysis, I do not have statistical evidence to support the idea that Sunday puzzles are the most difficult by the nature of having the most uncommon words. The data actually suggest the opposite; weekday puzzles have the most uncommon words.

Conclusion

Commonness is an important concept to these Spelling Bee puzzles because it makes the goal of finding all possible words attainable. But is it really a bad thing to challenge our vocabulary? In her journal article *The Power of Words*[3], Kathleen M. Roe, PhD writes that "when we don't have a word for something, we can't talk about it." This is why it's important for us to challenge and expand our vocabulary often. If we only speak the words we've always spoken, then we'll only think the thoughts we've always thought. The most important aspect of human existence is growth. We start as nothing and, without growth, we remain nothing.

Part of me wanted to be upset with the New York Times for making their Spelling Bee puzzles so impossibly hard, but as I've pondered with how to interpret my findings I keep circling back to one main takeaway: opportunity. At the end of the day, the puzzles are meant to be a challenge. They're meant to be an opportunity to learn something new, expand your vocabulary, and - most importantly - *grow*.

I'll be the first to admit that the New York Times has epically failed to follow the third rule of commonness in their puzzles with a mere 60% of answers being "common" words. But I'll also be the first to admit that maybe this was a rule that was meant to be broken. The word "bamboozle" doesn't show up in the Brown corpus, but that doesn't mean that I should never use that word in my own vocabulary. Maybe the Spelling Bee succeeded in a bigger way in that it grants NYT followers the opportunity to expand their vocabulary by roughly 40% every puzzle.

As I contemplate the successes and failures of the NYT Spelling Bee, I can't help but think about my 70 year old grandma who consistently solves the puzzle everyday. For her, these words are all common. Her life experience and her commitment to grow her vocabulary everyday aids her in solving even the most difficult of Spelling Bee puzzles. Truthfully, the best takeaway from this analysis, for me, is to hope that

the Spelling Bee remains ever challenging so that one day I might be as verbose, open minded, and self-actualized as my grandma.

References

1. New York Times . Word Games and Logic Puzzles. 2022; https://help.nytimes.com/hc/en-us/articles/360029050872-Word-Games-and-Logic-Puzzles
2. Hasanah, Dhia. (2020). The Use of Colloquial Words in Improving Students' Speaking Through Teacher's Daily Assessment. 10.2991/assehr.k.200406.007.
3. Roe KM, Mata HJ. The Power of Words. Health Promotion Practice. 2019;20(2):153-156. doi:10.1177/1524839919827900