

A Little Wordle Play

Robert Petit

March 2022

Introduction

The first guess for a WordleWord is an interesting microcosm of arbitrary, hyper-optimizable strategy decisions. One of the interesting catches of Wordle as a whole is trying to work through the results that you obtain throughout your guesses. The first word is unique: much like a chess opening, the game hasn't shaped up. Here, we will attempt to provide a small selection of "best first words" for a number of different strategies:

- Letter Hunting
- Position Hunting
- Letter & Position Combo
- Vowel Minimization

This provides you with the choice: do you want to get as many yellow letters or green letters as possible. We can weight and combine these scores to get a ranking that compromises between yellow and green results as well. Since vowels are fairly trivial to include on later guesses, we may be more interested in maximizing our correctly-guessed consonants and ignoring the contribution of vowels to our score.

On Judgement

In order to do this, we are going to use two different sets of words to treat as our "potential guesses" for the game. We have a large list (15577 words) of almost all 5 letter words in the English language. This is *technically* the full breadth of possible words in Wordle, but it includes words like "aalii" (a type of bush found in Hawaii) and "ganch" (to execute via impaling) in the list. While these are possible to enter, they are not particularly likely to show up. To account for this, we also have a list of Wordle-allowed words (10640). We will use both the full short word lists to judge our results.

This will yield a set of results for each strategy: results when our scope is *all* words and results when our scope is *likely* words. This is four combinations of possible results:

- All words judged on all words
- All words judged on likely words
- Likely words judged on likely words
- Likely words judged on all words

We are only concerned with the first three. The last of these is useless, after all, there is no world in which you only want to *guess* a common word, but a possible result is *any* word.

Letter Hunting

We make the simple observation that

$$P(1\{x \in W\}) = E[1\{x \in W\}] = \frac{1}{n} \sum_{i=1}^n 1\{x \in w_i\}$$

Where x is a given letter, W is a random word and w_i is a specific word within our target sample. We then calculate $P(1\{x \in W\})$ for each unique letter in a word a form the expected yellow results, Y as

$$Y = \sum_{j=1}^u P(1\{x_j \in W\})$$

where the subscript j denotes a specific letter and u is the number of unique letters in that word. This gives us the expected number of yellow results any one word should have; naturally, we simply want this value to be the highest possible.

Position Hunting

This is largely similiar to letter hunting, but instead of being concerned with $P(1\{x \in W\})$ we now need to calculate the letter being in the correct spot, which is given by

$$P(1\{x_j = W_j\}) = E[1\{x_j = W_j\}] = \frac{1}{n} \sum_{i=1}^n 1\{x_j \in w_{ij}\}$$

We then calculate the expected number of green results, G , as

$$G = \sum_{j=1}^5 P(1\{x_j = w_{ij}\})$$

Combined Scoring & Vowel Omission

Once we have calculated a Y and G score for a word, we can combine these two scores based on a given weight t to give us the combined score C

$$C = t \times G + (1 - t) \times Y$$

Of note with the C score is to remember that we do discredit green results when searching for yellow results. A yellow letter and a green letter will both count for our purposes as a “yellow outcome”, but only a green letter will count as a “green outcome”. When we look at our full score, it’s important to remember that the trade-off is between having results hit in general and having a specifically green result instead. Each of yellow letters has $P(\text{Green}) = 0.2$, so there is a *very* strong relationship between them. This becomes consequential when we are choosing our weighting factor t , since at low values of t we would expect remarkable ranking similarity between different choices.

Finally, we can reconsider the Y , G , and C values but ignore the vowels when calculating each score in order to get a separate vowel-omitted score.

Results

The fun stuff! All of these tables share a common format: three sets of two columns. The first set of three is the long list, judged on the long list. The second set is the short list, judged on the short list. The last set is the long list, judged on the short list. Likely, the columns of interest are the last three, since they give all possible guesses predicting any reasonable result. The tables show the top 5 performers, where ties are broken by the other scores: ties in yellow-seeking scores are broken by their green-seeking performance and vice-versa.

Letter-Seeking

Word (FoF)	Y	Word (SoS)	Y	Word (FoS)	Y
aeros	1.883	aeros	1.943	aeros	1.943
soare	1.883	soare	1.943	soare	1.943
arose	1.883	reais	1.915	arose	1.943
reais	1.865	serai	1.915	reais	1.915
raise	1.865	aesir	1.915	serai	1.915

Consonants

Word (FoF)	Y	Word (SoS)	Y	Word (FoS)	Y
rynds	1.274	rynds	1.329	rynds	1.329
tryps	1.260	tryps	1.305	tryps	1.305
byrls	1.243	byrls	1.287	byrls	1.287
tirls	1.202	tirls	1.241	tirls	1.241
rotls	1.202	rotls	1.241	rotls	1.241

Position-Seeking

Word (FoF)	G	Word (SoS)	G	Word (FoS)	G
sanes	0.799	sores	0.936	sores	0.936
sores	0.797	sanes	0.930	sanes	0.930
sales	0.789	sales	0.921	sales	0.921
sones	0.776	sones	0.916	sones	0.916
sates	0.772	soles	0.908	soles	0.908

Consonants

Word (FoF)	G	Word (SoS)	G	Word (FoS)	G
sorts	0.542	sorts	0.648	sorts	0.648
sorns	0.534	grrls	0.634	grrls	0.634
grrls	0.530	sorns	0.634	sorns	0.634
serrs	0.530	serrs	0.630	serrs	0.630
sants	0.521	sants	0.629	sants	0.629

Combined Weighting

Here is the combined values for $t = 0.75$.

Word (FoF)	C	Word (SoS)	C	Word (FoS)	C
aeros	0.919	aeros	1.014	aeros	1.014
soare	0.880	soare	0.880	soare	0.880
arose	0.748	aesir	0.733	arose	0.729
aesir	0.733	reais	1.005	aesir	0.733
reais	0.917	serai	0.816	reais	1.005

Consonants

Word (FoF)	C	Word (SoS)	C	Word (FoS)	C
byrls	0.688	byrls	0.782	byrls	0.782
grrls	0.668	grrls	0.760	grrls	0.760
tirls	0.659	tirls	0.749	tirls	0.749
tarns	0.655	rynds	0.749	rynds	0.749
terns	0.655	tarns	0.743	tarns	0.743

Conclusion

So, depending on your and desired strategy, one of the words from the following table is certainly the best option. Based purely on estimation, it is likely that the consonant-filtered long-on-short combined score is the most efficient word to pick, but doing this hinges on a balanced strategy and considering words that are not commonly used which minimizes your chances of having a *correct* first guess. By picking the combined-score short-on-short word, you alternatively maximize your chance of having a correct first guess.

	Y	G	C
Full on Full	aeros	sanes	tares
... Consonants	rynds	sorts	byrls
Short on Short	aeros	sores	tares
... Consonants	rynds	sorts	byrls
Full on Short	aeros	sores	tares
... Consonants	rynds	sorts	byrls