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 (8497 words) of 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, they are not particularly likely to show up. To account for this, we also have a list of commonly used words (5757. 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

$$C = tG + (1-t)Y$$

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 result instead. Each of yellow letters has $P(Green) = 0.2$, 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!

Letter-Seeking

This is our Yellow-Maxing list, separated by scope of the guesses and scope of the possible words. The first set is our full list, judged on it's ability to guess the words in the full list. The second set of three is the words in the short list, predicting the short list. The last set of three values is for the words in the full list, predicting words in the short list.

Word (FoS)	Y	Word (SoS)	Y	Word (FoS)	Y
arose	1.773	arose	1.909	arose	1.909
orate	1.771	raise	1.884	arise	1.884
arise	1.769	arise	1.884	raise	1.884
raise	1.769	tears	1.871	serai	1.884
serai	1.769	rates	1.871	aster	1.871

Consonants

Word (FoS)	Y	Word (SoS)	Y	Word (FoS)	Y
slart	1.112	turns	1.242	slart	1.280
slirt	1.112	stern	1.242	slirt	1.280
snirt	1.097	rents	1.242	snirt	1.242
snort	1.097	snort	1.242	snort	1.242
snurt	1.097	runts	1.242	snurt	1.242

Position-Seeking

This table follows the same logic, 3 triplets based on their scoring method.

Word (FoS)	G	Word (SoS)	G	Word (FoS)	G
soree	0.651	sores	0.886	soles	0.870
salay	0.607	sales	0.874	cones	0.825
boree	0.592	soles	0.870	sures	0.820
sairy	0.586	sates	0.855	tales	0.813
saily	0.586	sires	0.844	mores	0.812

Consonants

Word (FoS)	G	Word (SoS)	G	Word (FoS)	G
serry	0.433	sorts	0.592	carls	0.528
sorry	0.433	salts	0.577	slops	0.528
surly	0.433	silts	0.577	pants	0.522
sorty	0.429	slits	0.572	turns	0.521
slyly	0.411	slots	0.572	darts	0.520

Combined Weighting

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

Word (FoS)	C	Word (SoS)	C	Word (FoS)	C
arose	0.774	arose	0.749	arose	0.749
orate	0.748	raise	0.805	arise	0.747
arise	0.773	arise	0.747	raise	0.805
raise	0.815	tears	0.951	serai	0.763
serai	0.799	rates	1.050	aster	0.760

Consonants

Word (FoS)	C	Word (SoS)	C	Word (FoS)	C
surly	0.584	carts	0.704	turns	0.701
sorty	0.582	crits	0.702	darts	0.695
stray	0.560	sorts	0.701	dorts	0.695
strey	0.560	turns	0.701	carls	0.692
stroy	0.560	terns	0.701	grits	0.669

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	arose	soree	tarie
... Consonants	slart	serry	surly
Short on Short	arose	sores	tares
... Consonants	turns	sorts	carts
Full on Short	arose	soles	tales
... Consonants	slart	carls	turns