Solving Wordle Heuristically *(DRAFT — February 10, 2022)*

Daniel W. Dichter
Independent Researcher
Cambridge, Massachusetts, U.S.A.
daniel.w.dichter@gmail.com

Wordle is a daily word game in which a player must guess a hidden five-letter word in six or fewer tries. After each guess, they get a color-coded response that indicates how close their guess was. This paper describes a solver that wins all 2,315 possible games, averaging a score of 3.596 with a standard deviation of 0.679. This is comparable to the current-best solvers, yet is achieved heuristically with many fewer calculations.

I. Game Parameters

Wordle has two dictionaries: 2,315 possible solutions, and 12,972 acceptable guesses.[1][2] Solutions are limited to commonly-known, non-S-plural words. The default version of the game is assumed, wherein guesses need not be consistent with known information. Games are assumed to be mutually independent, and thereby the solver cannot recall previous solutions to reduce the set of initial words.

II. Probability and Value

A matrix P measuring 26x5 is defined, reporting the probability of each letter α in each position ρ of the solution. Guessing a tile that is already known to have zero or unity probability[1] provides no new information, or value. Accordingly, maximum value must occur at some intermediate p corresponding to minimum certainty. This is modeled as a transfer function relating probability p to value v:

$$v(p) = \sin(\pi p)^{2/3} \quad \text{for} \quad 0 \leq p \leq 1 \tag{1}$$

Another matrix V, also measuring 26x5, is then similarly defined. P and V are updated after every move as solutions are eliminated.

III. Scoring and Guessing

To score the overall value of a word, interactions between tiles are modeled with a set of heuristics, as follows.

A tile with non-unity probability captures the value of that letter in that position, and one-half the value of that letter in all other positions. This is modeled as a 5x5 weight matrix W containing ones on the diagonal and one-half everywhere else; rows correspond to the tile being scored, and columns correpond to the position being evaluated.

Tiles that letter-match with other tiles in the same word diminish each other's value, as if by *interference*. Similarly, when unity-probability occurs, tiles may return as yellow, indicating the already-known tile(s), thereby providing minimal value, as if by *occlusion*. These phenomena are modeled in a 5x5 quality matrix Q given in Table 2, itself a function of q_m, the quantity of letter-matching tiles, and q_u, the quantity of letter-matching unity-probability tiles.

The scalar score S of a word is then given by:

$$S = \sum_{\rho=1}^{5} \begin{cases} Q(\rho)W(\rho) \cdot V(\alpha) & \text{if} \;\; P(\alpha,\rho) < 1 \\ 0 & \text{if} \;\; P(\alpha,\rho) = 1 \end{cases} \tag{2}$$

[1] Obtained directly by a gray or green tile, or indirectly by deduction.

Game State	Words Left
S O A R E	2,315
D I N E R	117
U L A M A	3
E L D E R	1
S O A R E	2,315
T H I L K	63
S K I L L	1
S O A R E	2,315
T A L O N	44
A L O F T	6
S O A R E	2,315
C L E A N	68
B E P A T	5
P L E A T	1
S O A R E	2,315
T H A C K	11
S H A R D	2

Table 1. Solutions for recent games.

The top-scoring word is picked until there are two or fewer solutions remaining. Ties are broken first by solution probability, then alphabetically.

	$q_m = 1$	$q_m = 2$	$q_m = 3$	$\cdots$
$q_u = 0$	H<u>E</u>RON 1	GR<u>E</u>BE 1/2	G<u>E</u>ESE 1/3	
$q_u = 1$	*n/a* 0	GR<u>E</u>B**E** 1/4	G<u>E</u>**E**SE 1/4	
$q_u = 2$	*n/a* 0	*n/a* 0	G<u>E</u>**E**S**E** 1/3	
$\vdots$				$\ddots$

Table 2. Quality matrix Q modeling interference and occlusion. Underline denotes tile of interest; bold denotes unity probability.

IV. Eliminating Solutions

After each guess, some solutions can be eliminated. This is executed sequentially, tile-by-tile, as follows.

Gray tile. If any and all additional letter-matching tiles are also gray, eliminate words that contain this letter. Otherwise, eliminate words that contain this letter in this position, and eliminate words that do not contain the overall quantity of this letter indicated.

Yellow tile. Eliminate words that contain this letter in this position, and eliminate words that do not contain this letter in non-green non-letter-matching tiles.

Green tile. Eliminate words that do not contain this letter in this position.

V. Results

Top-scoring first words are shown in Table 3. Solutions for recent games are shown in Table 1. Results for all games are shown in Table 4. A comparison with other solvers including unassisted play by proxy of Twitter[2] is shown in Table 5 and Figure 1.[3] Source code and complete game logs are also provided.[4]

Rank	Word	Score
1	SOARE	1.00000
2	ROATE	0.98966
3	ORATE	0.97787
4	TALER	0.97782
5	STARE	0.97748

Table 3. Top-scoring first words.

Score	Games	PDF	CDF
1	0	0.000	0.000
2	52	0.022	0.022
3	1026	0.443	0.466
4	1050	0.454	0.919
5	179	0.077	0.997
6	8	0.003	1.000
x	0	0.000	1.000

Table 4. Results for all games.

Solver	Win %	Mean	Stdev.	Worst	Opener
Author[4]	100.0	3.596	0.679	6	SOARE
Sanderson[5]	100.0	3.438	0.645	6	CRANE
Glaiel[2]	100.0	3.494	-	5	ROATE
Chao[6]	99.7	3.420	-	x	OPERA
Filion[7]	94.2	-	-	x	AROSE
Twitter[3]	97.9	4.061	1.208	x	–

Table 5. Comparison of solver benchmarks.

[2] 29 unique games, 7,087,784 tweets total.

VI. Discussion

All games are opened with the top-scoring word SOARE. In the middlegame, emergent strategies can be seen, such as abandoning green tiles until the endgame, shifting yellow tiles to alternate positions, and reusing gray tiles for high-value words. The cumulative density function (CDF) shows that 91.9% of games are won within four moves, 99.7% within five, and 100% within six. The probability density function (PDF) shows a mode of four moves, with an associated probability of 45.4%. The average score is within 5% of the current-best solvers, which require significantly more computation.

VII. Acknowledgments

The author gratefully thanks Jess Hall, Julius Dichter, Orion Taylor, Adam Filion, and Tyler Smith for their support.

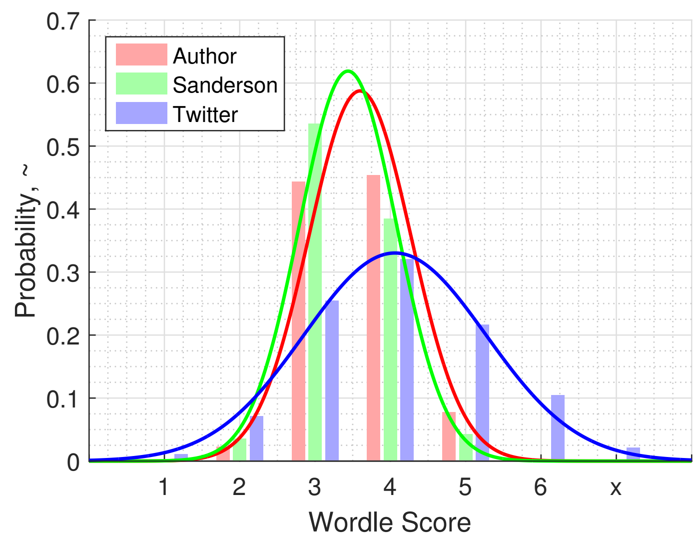

Figure 1. Comparison of solver distributions. Typical and current-best play are shown as limiting cases.

References

[1] Wardle, Josh. *Wordle - a daily word game.* Accessed February 6, 2022. https://www.powerlanguage.co.uk/wordle/.

[2] Glaiel, Tyler. *The mathematically optimal first guess in Wordle.* Accessed February 6, 2022. https://medium.com/@tglaiel/the-mathematically-optimal-first-guess-in-wordle-cbcb03c19b0a.

[3] O'Connor, Kevin. *Wordle Stats (@WordleStats) / Twitter.* Accessed February 6, 2022. https://twitter.com/wordlestats.

[4] Dichter, Daniel. *wordle_solver.* Accessed February 6, 2022. https://github.com/kindofdoon/wordle_solver.

[5] Sanderson, Grant. *The mathematically optimal Wordle strategy.* 3Blue1Brown. Accessed February 6, 2022. https://www.youtube.com/watch?v=v68zYyaEmEA.

[6] Chao, Jason. *Wordle-Solver.* Accessed February 6, 2022. https://github.com/jason-chao/wordle-solver.

[7] Filion, Adam. *Building a Wordle Solver.* Accessed February 6, 2022. https://blogs.mathworks.com/loren/2022/01/18/building-a-wordle-solver/.