# Scoring of Lexic Squares

Erik Nakamura

June 25, 2016

**Abstract**

Lexic squares are presented as a challenge and should therefore have a way to be rated. This document explores how lexic squares are rated and score. No public challenge would be legitimate without some ranking, not only for the satisfaction of those that submit their entries, but also for challengers to have a distinct goal. First, basic scoring guidelines are laid out, then two methods of scoring (word-based and letter-based) are compared and shown to be equally suitable for ranking. Then it is shown that letter-based scoring systems are Superior to word-based scoring systems. This document will not, however, discuss how challengers submit their solutions and how those solutions are scored on a technical level, it only briefly describes the theory of lexic square scoring.

# Introduction

Lexic squares are unique in that there is no perfect solution and that one method of finding solutions can be applied to various sets of words. The premise of the original lexic square is that it must include "every" English word. However, this list of "every" English word is allusive if not non existent; hence the quotations. Questions arise as to whether or not proper nouns should be included, or plurals, etc.. Furthermore not one perfect list of every word exists due to the changing nature of language and countless debatable topics such as spelling and slang. Yet despite this the challenge offers a list of a large number of words for challengers to try to put in the smallest possible lexic square. Keep in mind that lexic squares do not need to contain a large list of words. The challenge of creating an optimum lexic square is just as relevant with lists of words in the thousands, hundreds, and even tens. Given this nature, a way to score lexic squares based on how optimum they are must be flexible.

# The Premis of Scoring

Scoring, in all forms, is done for comparison. The scoring of lexic squares is no different. Given the fact that lexic squares can be made for any list of words, any lexic squares being compared must be using the exact same list of words. That being said, a good scoring system should have a range equal for all lists of words, so that even lexic squares produced from different lists can be compared. If, for example, a list of 100 words is used to form a lexic square with score $x$ and a list of 10000 words forms a lexic square with the same score $x$, it should be clear that both are equally optimum but the solution with 10000 words has a larger square.

# Word-based vs Letter-based Scoring

Let $s$ equal the height of the lexic square
Let $w$ equal the number of words in the list of words in the lexic square
Let $l$ equal the number of letters contained in the list of words in the lexic square
Note that a valid solution for a lexic square must contain all words in its given list.
Let $p$ equal the final score of the lexic square

In word based scoring:

$$p = \frac{w}{s^2}$$

In letter based scoring:

$$p = \frac{l}{s^2}$$

where $s^2$ equals the area of the lexic square

Put plainly, the score is equal to the number of words or letters over the area of the lexic square. The score therefore represents how much of the lexic square is filled with words or letters, and how much of it is empty. Recall that an optimum lexic square has the least amount of empty, i.e., unused space. These scores can also be seen as the density of the lexic square, where a higher density is more optimal. Given the fact that for both word and letter based methods of scoring, $w$ and $l$ do not change between two lexic squares being compared, any increase in square size has an exponential effect on the score.

It is also critical to note that both word and letter based methods of scoring compare lexical squares equally. To prove this:

As before:
Let $w$ equal the number of words in the list of words in the lexic square
Let $l$ equal the number of letters contained in the list of words in the lexic square
Let $p$ equal the final score of the lexic square

Let $s_1$ equal the height of one lexic square and $s_2$ equal the height of another where $s_1 \neq s_2$

Let $n$ represent the ratio of $w$ and $l$ such that $n = \frac{l}{w}$
Note that $n$ subsequently represents the average number of letters per word on the list.

For both types of scoring to be comparativly equal, the following must be true:

$$\frac{\frac{w}{s_1^2}}{\frac{w}{s_2^2}} = \frac{\frac{l}{s_1^2}}{\frac{l}{s_2^2}}$$

In other words, the ratio between two word-based scores must be the same as the ratio between two letter-based scores from the same lexic squares. If this is true, then the ranking of word-based scored lexic squares will be the same as the ranking of letter-based scored lexic squares:

let $p_1 = \dfrac{w}{s_1^2}$

let $p_2 = \dfrac{w}{s_2^2}$

$p_1$ and $p_2$ both represent two distnct scores of lexic squares formed from the same list.

The lexic squares can also be given a letter-based score

let $p_{l1} = \dfrac{l}{s_1^2}$

let $p_{l2} = \dfrac{l}{s_2^2}$

given $s_1 > s_2$ and $\frac{\frac{w}{s_1^2}}{\frac{w}{s_2^2}} = \frac{\frac{l}{s_1^2}}{\frac{l}{s_2^2}}$

$p_1 > p_2$

and

$p_{l1} > p_{l2}$

thus the ranking remains the same for both scoring methods if $\frac{\frac{w}{s_1^2}}{\frac{w}{s_2^2}} = \frac{\frac{l}{s_1^2}}{\frac{l}{s_2^2}}$

which can be proven to be true for any given $w$, $l$, and corrosponding $n$ where:

$$n = \frac{l}{w}$$
$$w = \frac{l}{n}$$
$$l = nw$$

$$\frac{\frac{w}{s_1^2}}{\frac{w}{s_2^2}}$$
$$= \frac{\frac{\frac{l}{n}}{s_1^2}}{\frac{\frac{l}{n}}{s_2^2}}$$
$$= \frac{\frac{1}{n} \times \frac{l}{s_1^2}}{\frac{1}{n} \times \frac{l}{s_2^2}}$$
$$= \frac{\frac{l}{s_1^2}}{\frac{l}{s_2^2}}$$
$$\frac{\frac{w}{s_1^2}}{\frac{w}{s_2^2}} = \frac{\frac{l}{s_1^2}}{\frac{l}{s_2^2}}$$

This shows that either scoring method will achieve the same rankings given lexic squares produced with the same list. Hoever it does not dismiss the fact that one scoring method may be better than the other.

## Details on $w$, $l$, and $n$

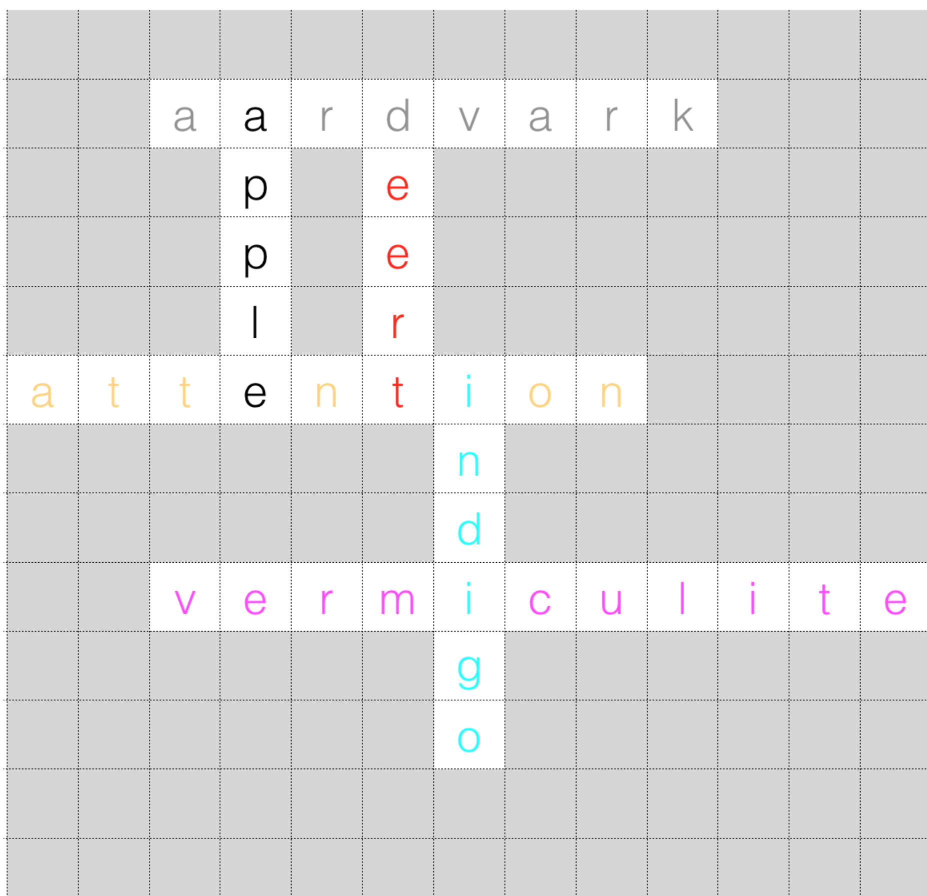
It is important to note that the values $w$, $l$, and $n$ do not change for different lexic squares using the same list of words even though a single letter can be used to form more than one word. For example, given the list of words "a", "as", "ass", and "assign", where $w = 4$ and $l = 12$, a lexic square solution could be the square containing the word "assign". This square could contain all 4 words within the space of only 1 word, and all 12 letters within the space of only 6 letters. Therefore the lexic square may apear to have a $w$ of 1 and an $l$ of 6, but the scoring method dictates that $w$ and $l$ still remain 4 and 12 respectivly. The variebles $w$, $l$, and $n$ should be viewed as belonging to the list of words the lexic square is using, and not to the lexic square itself. The only variable that belongs to the lexic square and that changes in the score calculation is $s$.

## The Superiority of Letter-based Scoring

Recall that letter based scoring is calculated as such:

$$p = \frac{l}{s^2}$$

Give the nature of lexic squares where one unit of area is occupied by one letter, a lexic square where all units of area map to exactly one letter within a word, will have the score 1. That is, every unit of area is being used to host a letter. Note the figure below:



The figure illustrates how difficult it is to utilize every unit of area in a lexic square. In the figure, which demonstrates a lexic square for a list of words where $w = 6$ and $l = 38$, most units of area are unoccupied. For this reason the score can be expected to be well under 1.

$$s = 13$$
$$l = 38$$

$$\begin{aligned} p &= \frac{l}{s^2} \\ &= \frac{38}{13^2} \\ &= \frac{38}{169} \\ &= 0.224852071 \end{aligned}$$

Using a letter-based scoring system for any list of words will create scores that approach 1 as the lexic squares become more optimum. It is even possible for a score to exceed 1. Yet approaching and exceeding a score of 1 is extremely difficult. Therefore a letter based system presents a range of roughly $(0, 1]$ where a very optimum lexic square can be quickly identified based on how close its score is to 1. Since the range applies to lists of all sizes, one can judge a score without any knowledge of what list it originated from. For example, one may be presented with the score 0.9, and without any knowledge of how large the list the lexic square originated from, it could be said that the lexic square is very optimum. Similarly, one can observe a score of 0.03 and immediately know that the lexic square is not optimum, without any knowledge of what the list looks like.

The roughly $(0, 1]$ range also allows for scores to be easily visualized. Since lexic squares are geometric they lend themselves for visualization, so their scores should take advantage of this. For example, a score of 0.5 means that roughly half of the square is occupied by letters. This allows a visual representation to be quickly matched with how optimum a lexic square is.

A letter-based scoring system not only provides a consistent ranking system, but also an objective representation of the lexic square. A score should not be abstract, it should be a condensed representation of what it is scoring. Letter-based scoring achieves this.

## Portability of Letter-based Scoring into Three Dimentions

The concept of lexic squares can be developed into lexic cubes where letters can be formed not only upwards, downwards, left, right, and diagonally on one plane, but on a stack of any number of planes.

For these lexic cubes the same scoring system can be used where the volume of the lexic cube is used to divide the number of letters within it.