The Longest Words Using The Fewest Letters

David M. Rheingold / Barnstable, Massachusetts

Austin Byl, Evan Crouch, Elizabeth Krodel, Celia Schiffman, Nico Toutenhoofd / Boulder, Colorado

Claude Shannon may be remembered as the father of information theory for the formula engraved on his tombstone, but it's not the only measure of entropy he helped devise. *Metric entropy* gauges the amount of information proportionate to the number of unique characters in a message.

Within the English language, the maximum metric entropy individual words can attain is 0.5283 binary digits (or bits). This is the value of any three-letter word composed of three different letters (e.g., WHO). Multiplying 0.5283 times the number of distinct letters equals 1.58, and 2 raised to the power of 1.58 yields the same number of letters. These values hold true for any trigram with three different letters in a 26-letter alphabet.

Measurements differ for written text involving other characters including numbers, punctuation, even spaces. Distinguishing between uppercase and lowercase would double the number of letters to 52. Adjusting for word frequency would further alter values as trigrams like WHO occur far more often in written English than, say, KEY. (How much more often? We found 3,598,284 instances of WHO in Project Gutenberg versus 58,863 of KEY, or 61 times as many.)

Our lexicon derives from three primary sources: Merriam-Webster, Webster's, and the Moby II wordlist, with a combined 382,843 words. Of these we find some 2,000 trigrams made up of three different letters, including WHO and KEY, tied for highest metric entropy.

At the opposite extreme we find much longer words with greater redundancy. The single word with the lowest (0.1150) metric entropy, DICHLORODIPHENYLTRICHLOROETHANE, with 31 letters also happens to be the longest word in all three wordsets. It is the chemical name of an insecticide known more commonly by its acronym DDT (occurring 213 times more frequently in Project Gutenberg). While DDT is also a trigram, its metric entropy (0.3061) is lower than that of WHO or KEY as it contains two rather than three unique letters.

Scientific and medical words tend to have the lowest metric entropy, along with palindromes (WOWWOW) and compound words consisting of repeating units (WALLAWALLA). Outside these overlapping categories remain long words consisting of few distinct letters. Here are some of the less technical ones:

Word	Metric entropy	Letters	Unique	Ratio
Humuhumunukunukuapuaa	0.1155 bits	21	7	33%
Antidisestablishmentarianism	0.1192 bits	28	12	43%
Antiinstitutionalists	0.1304 bits	21	8	38%
Possessionlessness	0.1306 bits	18	7	39%
Transubstantiationalist	0.1326 bits	23	10	43%
Senselessness	0.1337 bits	13	4	31%
Antitintinnabularian	0.1352 bits	20	8	40%
Superincomprehensibleness	0.1372 bits	25	13	52%
Antinationalization	0.1384 bits	19	7	37%
Noncondescendingness	0.1385 bits	20	8	40%
Overrepresentativeness	0.1386 bits	22	10	45%
Nonconscientiousness	0.1402 bits	20	8	40%

Successlessness	0.1431 bits	15	6	40%
Sleevelessness	0.1433 bits	14	5	36%
Nonsententiousness	0.1441 bits	18	7	39%

The last word on this list falls into a more rarefied class—autological—and thus makes an appropriate stopping point.

References

Golomb, Solomon W. (1984) "Amalgamate, Chemist!," *Word Ways*: Vol. 17: Iss. 1, Article 6. Available at: https://digitalcommons.butler.edu/wordways/vol17/iss1/6

Merriam-Webster Collegiate Reference Application Programming Interface. Available at: https://www.dictionaryapi.com

Project Gutenberg DVD (2010). Available at: http://www.gutenberg.org/wiki/Gutenberg:The_CD_and_DVD_Project

Project Gutenberg Etext of Moby Word II. Available at https://github.com/dwyl/english-words

Sebastian, Anthony and Merriam, Renee (1988) "On Reflexivity in Words," *Word Ways*: Vol. 21: Iss. 3, Article 2. Available at: https://digitalcommons.butler.edu/wordways/vol21/iss3/2

Shannon, Claude E. and Weaver, Warren (1949) *The Mathematical Theory of Communication*, University of Illinois Press: Urbana and Chicago.

Soni, Jimmy and Goodman, Rob (2017) *A Mind at Play: How Claude Shannon Invented the Information Age*, Simon & Schuster: New York.

Webster's Revised Unabridged Dictionary. Available at: http://www.gutenberg.org/ebooks/673