ETAOIN SHRDLU WORDS

ANIL
Perth, Australia

• PURE WORDS
Etaoin shrdlu (e-s), the classic 12 most frequent letters, make up 46% of the alphabet and are the first letters on 55% of Web3 pages, so one would expect lots of words to contain these letters alone. Toward quantifying this I did an indicative study sampling the head word on 4/10 of Web3 pages (every -2, -5 -7 and -0). Of 1063 words just 117 were pure e-s (11% or 20% of e-s initials). 38% (400 = 249 e-s initials + 151 other) had a single miss, so 49% of the words contained no more than one non e-s letter. It’s interesting that there were 3.4 times more single non’s than pure e-s.

Most common single extra letters were C 71, P 63, G 51, M 41. Add C to e-s for a half alphabet and 18% are pure “etaoin shrdluc”. (“A twin-shared luck” beats “a twin-shared loo”!). Conversely, take away successive letters from e-s and the % declines rapidly: e-s = 11.0%, –U leaves 8.6%, –L 5.8%, –D 3.3%, –R 1.5%, –SH 0.4% (only 4 pure etaoin words). From the E end the decline was even faster, as expected: –E 4.7% left, –T 2.8%, –A 1.5%, –O 0.3%, –IN 0.1% (just 1 shrdlu word).

• TRANSPOSALS
The only reported transposal of etaoinshrdlu is outlandisher (OED). I add antishoulder, unlisted but with many Net quotes for the hyphenated spelling, mostly related to countering ‘shoulder surfing’, plus one unhyphenated use, antishoulder policy, meaning no bicycles allowed on road shoulders. Jeff Grant found another transposal cited on the Net, the coinage out-hardlines.

• LONGEST
My longest e-s words were alliterational and Neanderthaloid (14 letters each). Interinstitutionaloid (18), the longest in M-W Pocket Dictionary, oddly isn’t in Web3. DEINSTITUTIONALISATION (22, Chambers; inferred plural 23) is the longest listed e-s word found. Other long words inferred or in Net quotes: disinterestednesses (19, plural inferred + Net); antiauthoritarianists (21, Net + Web3 -ism); interinstitutionalised (22, Net); deinstitutionalisationist (25, Net, adjective); and antideinstitutionalisations (27, Net, plural inferred). Antideinstitutionalisationists, a 30-letter coinage from Jeff, had no hits. Can readers find longer ones, listed, inferred, or googled?
• E-S FREE WORDS
The longest word free of e-s, except for poly-Z and poly-M onomatopoeia, is probably Pygmy, although the unlisted coinage pygmyfy was found on the Net. Next longest is by-by, then eight Web3 headword trigrams, cwm, cyp, gym, gyp, myx-, pyx, zyg-, zym-. Can’t do much without e-s!

• ALPHANUMERIC WEIGHT
Antideinstitutionalisations has an alphanumeric value (AV, where A=1, Z=26, etc) of 342, highest of the above. Highest listed word, deinstitutionalisation, is 279. But their AV densities (both average 12.7 points/letter) are well below my test words trust (19.6) and Usun (18.75). The densest of all e-s words is U (21, upper class; Burmese title), then utu (20.7), ut and tutu (20.5) and tut-tut (20.3).

For comparison, what is the highest AV of all words? Counting nonce OED quotes, the highest dictionary AV is 1118 for the ‘word’ zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz (43z) under the headword Z. The Web3 45-letter pneumonoultramicroscopicsilicovolcanocoanosis is both the longest and the highest AV dictionary headword at 560 (density 12.4), but higher AV chemical names are given under shorter headwords (see Gooch 05-67). The densest words are any length of the snore-bore poly-z’s (26), then zy (25.5, old “see”, OED), XYZ (25, numerous OED et al. citations, eg, unknowns, and the x,y,z Cartesian coordinate system), then XXY (24.7, Chambers, XXY syndrome, a sex chromosome number abnormality).

For this and the following article I thank Ross Eckler for pointers and Jeff Grant for extensive help.

---

WH NDS VWLS?

NL
Prth, Strl

Grk ws th frst lphbt t gv sprt chrctrs fr th vwls. Rlr lphbts gt b wtht thm, . g., ld Smtc lnggs. Y en rd ths t hlps t prv tht w dn’t rll nd thm! N mjr prblm wth lmtng vwls s wth snml ltr wrds lk nd . Vwlss mngrms nd bg rms r ls prblm tc (blw).

• Wlth vwls, mngmng wld b lts sr. Mn wrds wld hv dzns f mngms. N wndrs f mn prs f snmms wld mk vwllss mngms? Hr’s n: n = n (n ndvdl rcl = th nmbr f sd ndvdl s). Cn y thnk f n lnr ns?

• Lng smlr lns, hr’s n pq VVLLSS NGRM WRD-SQR PM: (Trnsln blw.)

    N
    N

• Fnd th CHLLNG: fnd th vwllss mngrm r bg rm wth th mst trnspsls (splngs), xclng bg brrvtns.
Deciphering WHO NEEDS VOWELS?
ANIL, Perth, Australia (or Anile, ‘peareth sterile?)

• Definitive Anagram:
aN = oNe. Several others turned up in the lists below (cf). Another is seen in this article’s title, NDS: NeeDS ≈ eNDS (our needs [or imagined needs] create our aims). The title could then read: “Who Ends Vowels?” That would be your friendly vocaliclastic author.

• Anagram Word-Square Poem:
  ONE
  NEO
  EON
(The poem’s message is: whether it’s a happy new eon depends on one’s actions and viewpoint.)

• Vowelless Monogram/Bigram Transposals “Chilling Challenge”:
I only did the monogram N and bigram pair R/S (guessing them to be the richest), as a benchmark to be bettered. Most words are from Web 3 and Collins Scrabble Dict. and exclude affixes, abbrvs. and capitalised and non-English words. N had 70 “transposals”, 12 in M-W Pocket Dict. (bold):
N = aeon/eon ain aine aine anoi an an’a ane ani anoa anoa anu/anu any aume ayn ‘ayn een e’en ëeene/ëeen/ëeene/ëeene/ëeny eina eine ene eooan ëeoooea/ëeoooea eyen/eyne in inia ion ‘n/’n’ na-nae naio naio naio nay ne ne/nee/née ne nie noa noo noy nu ny nye on one onie ono ony on un unai/unia uni.
† OED, OED2, or Shorter Oxford / § Wiktionary/Wikipedia

R/S had 179 transposals (125 RS+54 SR), or 157 (104+53) excluding /** noted; 43 in MWPD.
* nonce plural of an adjective, etc, as a word (noun) in itself; eg, “There are not two aerys in this sentence,” aery being an adjective treated as a noun;
** nonce noun of action made by adding -er to a verb, sanctioned by Fowler for any English verb.

RS = aerais aeries †aeris †aerious a eros aers *aerys aeries *aires airs/ayres air-sea *airys aras araise arayse aeras arius ariose ariso aris arise/arose arouse ars arse arsey/arsy, aruis arusa aryas *arys auras *aureae/aureas auroeus aures/aurus aurous **ayers ears *e’ers *eeries/*eerys eras erase eres *eries eros erose ers euros eyers eyras eyres eyries ioros iras iris oars *oarys *o’ers *oors *oories oras orse *orys ours *ouries oyers rais/reis raise ras rase rays reas reis res resai resay reseau reseey re-sue reus/*reais/*reaes reuse *reys reyes rias rise roes rois roosa roose rose/roso rosé rosy roués rouse r’s rues rusa ruse ryas ryes *urs uraeus uraos ureas urgeuse ures †uroos ursa/ursae urus.

SR = asar asor easer easier eyesore oasar osier sair/saie sar sari/sarie saury sayee sayer see-air, see-ear sear seare seeer seer ser sera serai seraya sire sir/sire/siree siri soar soiree soor sora soree soorie soori souri souari sour sra sri suer sur sura sure **syer usar user usura/usurae/usure/usury.

Since the bigram far outweighed the monogram, there are probly trigrams (eg, N/R/S or R/S/T) with even more transposals. In any case it seems we do need vowels for very short consonant groupings. How large a group is necessary to get by without vowels? Generally that is, allowing exceptions of course. Re-reading the vowelless text above I’d guess four consonants will usually suffice, even three in most cases. But most ones and twos need vowels. Yet in the context of a sentence even they can very often be read without their vowels, barring exotic words.

262
A somewhat similar idea to this was explored in Ross Eckler’s 82-119 “Compression of English Text”. Instead of vowels he eliminated different percentages of the most superfluous letters to create the least ambiguity.

- All-Vowel Words:
My search ran into some of these oddities. Most are in Borgmann’s Language on Vacation (p. 156) but not aye-aye and Iouea. Iouea is the shortest word with all five vowels and the shortest four-syllable word, but it was coined to be such by the wordsmith zoologist who named this sponge genus (Wiktionary). Borgmann also cites the one-consonant word Aaeaean, which I omitted from my N list as it’s capitalised.

- Cognate to Near Definitive Vowelled Anagrams:
Notes: † = also anagrams with their vowels; †† = reversals.

I. etymologically related:
aine, aine (elders: M, F) ayn/ain, ‘ayn/ain (Hebrew, Arabic alphabet letters)
airs, arias † raise, arise, rise
an, one (noted above) sear, sere (scorch, dry)
areas, ares sir, sire
aures/auris, ears reyes, rois (kings)
aureus, aureous, aurous

Plus several alternative spellings of the same word, like ain/ane/one/un, na/nae/nay/no, sere/sear (dry), usura/usurae/usure/usury.

II. unrelated (some flawed by being different parts of speech):
airs, raise (voice/s opinions) raise, rouse
airs, soar (fly/s), †† reus, suer (antigrams, defendant/plaintiff; picture
ares, ures (land area units) the two facing off in court mirroring each other)
auras, rays (haloes) roes, rusa (deer) *
† tears, ‘ears (hears—dialect) sayer, sure (commonly the same)
† erase, easier † sear, rase (burn, burn down)
† eros, rose (poetic, slang; + pun!) † sir, sri
-eyres, oyers (courts, trials) sore, sour (adj. of people)
in, on (occasionally interchangeable; otherwise usually antigrams)

* There are also eight other mammals: aruis (sheep), euros/roos/uuros (kangaroos), eyras (wildcats), ursa/ursae (bear/s), urus (aurochs);
two + six plants: anu/anuy, naio / aras, arusa, osier, roosa, seraya, souari;
two + three birds: ani, nye / ioras, soor, sora;
five moneys: aureus, euros, toras, reis, †rosa;
eight related to vision(s): iris, see/seir/ser, eyers, eyesore, rays, resee;
and five related to anger or bitterness: †ires/rise, rouse, sore, sour.

Years (= eras) isn’t listed as that y is a consonant, which I believe it isn’t in any of the list words.