Journal of Quantitative Linguistics

ISSN: 0929-6174 (Print) 1744-5035 (Online) Journal homepage: http://www.tandfonline.com/loi/njql20

How Many Possible Trade Names are There?

Geoffrey Sampson

To cite this article: Geoffrey Sampson (2016) How Many Possible Trade Names are There?, Journal of Quantitative Linguistics, 23:4, 342-360, DOI: 10.1080/09296174.2016.1226429

To link to this article: http://dx.doi.org/10.1080/09296174.2016.1226429

Published online: 21 Sep 2016.

Journal of Quantitative Linguistics, 2016
Vol. 23, No. 4, 342–360, http://dx.doi.org/10.1080/09296174.2016.1226429

How Many Possible Trade Names are There?*

Geoffrey Sampson
Department of Linguistics and Modern Languages, UNISA, South Africa

ABSTRACT

The number of possible distinct names of reasonable length is necessarily finite, and with the heavy demand for new names in some areas of current English-speaking commerce it seems possible that the supply might approach exhaustion. Numbers of possibilities cannot be calculated exactly, because most wordforms theoretically allowed by English phonology are too clumsy to be usable. However, Monte Carlo methods permit estimates of numbers of names which are plausible by present-day standards. These estimates suggest that the prospect of "running out of names" may not be merely fanciful.

1. OUTLINE OF THE PROBLEM

Trade depends on names. Firms need distinctive names for themselves, for their brands, and for their individual products or services. Owners of trade names put considerable effort into protecting their rights in their names. We often read of cases where a firm takes legal action to try to prevent another business, perhaps in an unrelated business sector, using a name which they believe might be confused with their own name.

Fifty years ago, it may have seemed as though there would always be abundant not-yet-used names to go round. Today, in a globalized society with more liberal business régimes, there are indications that it is becoming harder to find names which are available and suitable. Company names were commonly formed from owners' surnames or place names or consisted of ordinary English words. But the days when names of those kinds were enough are behind us. Shortly after the English confectionery manufacturer

*Address correspondence to: Geoffrey Sampson, Department of Linguistics and Modern Languages, Theo van Wijk, 9-88, UNISA, Preller Street, Pretoria 0003, South Africa. E-mail: sampson@cantab.net

Cadbury's was taken over in 2010 by an American company, the section containing it was spun off under the exotic-looking name *Mondelēz*, with a macron to encourage the pronunciation [li:z] for the last syllable. This name was apparently derived from French *monde* and *délice*, but it seems unlikely that many English-speaking consumers will see it as more than an arbitrary orthographic sequence. *The Economist* recently ran an article discussing the many strange types of name which some firms have recently taken to using, "resorting to ever more desperate means in order to stand out from the crowd" (*The Economist*, 2015); the article refers to a story by Arthur C. Clarke (1953) in which an early computer is put to enumerating "all the possible names of God" – when it finishes, listing all nine billion names, the world ends.

One area where the need for invented names has become specially salient is medicine. Drugs used to be given names with meaningful derivations, for instance the name *aspirin* referred to the fact that the active ingredient occurs naturally in a herb then assigned to the genus *Spiraea*. But with the huge outpouring of novel drugs in recent times, each needing a generic name and one or more proprietary names, it has become usual for these names to be meaningless and often rather outlandish-sounding.[1] A few pain-relief drugs have the following generic names, and corresponding capitalized proprietary names: *celecoxib*, *Celebrex*; *diclofenac*, *Cataflam*, *Voltaren*, *Zipsor*; *ibuprofen*, *Advil*, *Haltran*, *Motrin*, *Trendar*, etc.; *sulindac*, *Clinoril*; *meloxicam*, *Mobic*.[2] I have not researched the background of these names, but they appear for the most part to be concatenations of meaningless syllables. The *flam* of *Cataflam* was doubtless chosen because diclofenac is an anti-in*flam*matory, and the *Clin-* of *Clinoril* was perhaps chosen to echo the word "clinical", but etymologizing seemingly cannot go much further than that. Moreover, while all the names are certainly pronounceable, some of their phonological patterns seem unusual, relative to the patterns found in ordinary non-name words. The four-syllable length of the first name, *celecoxib*, for instance, would suggest a derivation from a classical language if the word were part of the ordinary English vocabulary, but so

[1]It is perhaps inappropriate to call a generic drug name a "trade name". But the issue we are concerned with is the supply of new names in general; my title refers to trade names, because it is chiefly commerce which is revealing the limits to that supply.

[2]The names capitalized here are believed to be registered proprietary names, and the respective owners' rights are hereby acknowledged. Later in this article, many examples of hypothetical trade names will be cited; so far as the author knows, none of these are actually in use, but he apologizes in advance if, unknown to him, any of them should coincide with real proprietary names.

far as I know no Greek or Latin word could give rise to an English polysyllable ending in -*ib*.

Another area rich in novel names is information-technology startups, where (unlike in the pharmaceutical industry) new products need distinctive names *before* a decision is taken about launching them on the market. For instance, among the many startup names listed as participating in the 2014 programme of one accelerator organization, MassChallenge, were: *Accel*, *Admetsys*, *Agira*, *Aldatu*, *Anfiro*, *Bubbi*, *ConsortiEX*, *Droplette*, *Eulysis*, *gameblyr*, *Jisto*, *KnipBio*, *Kuona*, *Lengio*, *Lig*, *Medlio*, *noonee*, *Oto*, *Recardo*, *Sano*, *SproutslO*, *Twiage*, *Unima*, *Varada*.

In this situation, it is of interest to ask how large the universe of potential names is. Could a time come when we exhaust the supply of possible names? From a legal point of view, debates about whether a new trade name is too close to an existing proprietary name ought to be informed by data about how much separation between names is in principle possible. This paper can do no more than broach the issue and offer tentative answers, but even these may be informative when the questions are new.

Of course, we can have as many different names as we want, if we do not care how long they are. But owners of trade names do care. A trade name needs to be memorable and convenient to use, so a six-syllable name might scarcely be worth considering. The question to be addressed below is how many potential names of a given length exist, and the lengths considered in detail will be reasonably short.

2. SPOKEN AND WRITTEN WORDFORMS

We shall consider this question from the viewpoint of the English language. We aim to count potential words that would be pronounceable in English, and could be given a written form which would identify that pronunciation in terms of the norms of English orthography.

When computing numbers of possible distinct English words, we have a choice between basing the computation on written wordforms (letter sequences) or on spoken forms (phoneme sequences). In a legal context the main focus would normally be on written forms (a registered trade name would standardly be defined as a letter sequence), however I assume that in order to be satisfactorily distinctive, a trade name should be unique in pronunciation as well as spelling. The owner of a hypothetical name *Meelok* might not be happy to find others using the names *Mealok* or *Meelock*, since the natural way to pronounce all three would be /mi:lQk/. How much this mattered, commercially or legally, would no doubt depend in practice on how far the firms or products in question were in competition with one another, but that consideration lies outside the purview of this paper. For our purposes, I assume that names are required to be distinct in both sound and spelling, irrespective of what they refer to.

(Spoken wordforms will be transcribed in this paper using the computer-oriented SAMPA system, see Gibbon, Moore, & Winski, 1997, pp. 699–702, which replaces the special characters of the International Phonetic Association alphabet with characters drawn from the ASCII character-set. Of the alternative SAMPA symbols for the DRESS vowel, we shall use /E/, since /e/ will be used for another purpose below; and, as an exception for the sake of ease of reading, the TRAP vowel will be transcribed as "&" rather than as SAMPA "{".)

In order to count possibilities, it is easiest to treat spoken forms as basic. The boundary between pronounceable and unpronounceable phoneme sequences is more determinate than the boundary between possible and impossible written letter sequences. Furthermore, while the unsystematic nature of English spelling means that correspondences between letter sequences and phoneme sequences are one-to-many in both directions, for instance, a name *Ledonac* might be read as /li:dQn&k/, /lEdQn&k/, or /lEd@Un&k/, while the first of these possibilities, /li:dQn&k/, might be spelled *Ledonac*, *Leedonac*, *Leedonack*, or *Leadonac*, there usually seem to be more alternative spellings for a given pronunciation than alternative pronunciations for a given written form. Both of these considerations suggest that a suitable strategy will be to aim initially to calculate numbers of pronounceable phoneme sequences, and to modify the results in the light of properties of English spelling, rather than to begin from spellings.

Alternative spellings for the same phoneme sequence will often yield names which differ greatly in "flavour". For instance, the sequence /r&lih&k/ could be spelled *Rallyhack*, suggesting a compound of native Germanic roots and hence a "relaxed" name appropriate for, say, the leisure sector; as a drug name the same phoneme sequence would more likely be given a Latinate spelling such as *Ralihac*, to suggest a "scientific" background. Nevertheless, for our investigation the single phoneme sequence will be counted once only.

3. THE BOUNDS OF THE PRONOUNCEABLE

Even the boundary between pronounceable and unpronounceable phoneme sequences is not perfectly determinate. Some sequences (for instance, those which correspond to actual words of the language) are clearly pronounceable, and some, e.g. /rbgtp/, /Ni:bE/ are clearly unpronounceable, as English words. (What sequences of sounds are pronounceable varies greatly from language to language – there are many languages in which /Ni:bE/, beginning with the velar nasal and ending with the DRESS vowel, would be a very normal-sounding word.) But there is also a penumbra of marginal possibilities. For instance, probably all English-speakers know the word *psst*, which might be glossed as something like "Over here, but keep quiet!" It is quite conceivable that *Psst!* might be adopted as a trade name, perhaps for some product associated with secrecy. Yet at the same time it is clear that the phoneme sequence /pst/ falls well outside the patterns used for words of the ordinary vocabulary. The approach adopted here will be to omit the "penumbra" and count only phoneme sequences which fit the pattern of ordinary English vocabulary.

Inevitably there are debatable cases. Thus, I have counted word-initial /sf-/ as a permissible cluster because of *sphere*, *sphincter*, but it is very rare. I have assumed that diphthongs resemble one another with respect to the range of consonant(-cluster)s which can follow them, but although /-aUT/ is certainly possible (*mouth*), and /-OIT/ *-oyth* seems utterable with ease by phonetically-naive English speakers, I am not aware of any English words that contain /-OIT/. We also find words from foreign languages used as trade names in the English-speaking world, some of which contain un-English phonology; for instance, Dior markets a perfume under the name *J'adore*, beginning with a sound /Z/ that does not occur initially in English (and which my calculations exclude accordingly). There can scarcely be any hard-and-fast line drawn between foreignisms like this which English-speakers find acceptable and other sounds or sound-combinations which seem too alien to be used in trade names.[3]

[3] This issue is only partly a matter of the intrinsic "alienness" of sounds. *J'adore* is evidently acceptable although initial /Z-/ does not occur in English. On the other hand the popular Korean car marque Hyundai (Sino-Korean 現代 /hjɤndæ/, "modern") seems invariably to be pronounced in English as three syllables, /haIVndaI/, even though /hj/ is a normal English initial sequence (*huge*, *humour*). Relevant factors in this case are probably that English /hj/ is not spelled using the letter *y*, and also that to English-speakers French is a much more familiar language than Korean.

A special problem is that some phoneme sequences are common in inflected forms but never occur in uninflected words. Thus /-md/ occurs in past tenses, e.g. *seemed* /si:md/, and /-bz/ occurs in plurals, e.g. *ribs* /rIbz/, but one will not find uninflected words ending in these clusters. It is not entirely clear what is the "right" way to handle these cases for present purposes (there certainly are businesses whose names are plural nouns, though I have not encountered one whose name is a past tense). But on the whole trade names tend to be uninflected forms, so rightly or wrongly the decision was to exclude these special phoneme sequences from the calculations. On the other hand, final /-ps, -pt/ are treated as valid possibilities because of a handful of words such as *apse*, *crypt*, though the great majority of words ending in these clusters are inflected forms.

For debatable cases like these, my calculations are based on rough-and-ready common-sense decisions. Rather than taking space here to detail all these decisions, I have placed online at <www.grsampson.net/SWordGen.html> the software which I wrote for purposes of the Monte Carlo experiments to be discussed below. This contains explicit lists of the sounds and sound-combinations I deemed permissible; readers who disagree with my decisions can easily check how the results would be affected by making those decisions differently. My surmise is that, at the level of precision relevant to this enquiry, other reasonable decisions would not give vastly different results.

4. PHONEMES AND GRAPHONES

Looking at English phonology in more detail, another source of indeterminacy relates to reduced vowels in unstressed syllables. The successive syllables of polysyllabic words receive different degrees of stress, and syllables having least stress also have fewer contrasts between vowels than other syllables: rather than a six-way contrast between the "checked" vowels /I E & Q V U/ one finds just a two-way contrast between the "obscure" vowels /i @/, neither of which occurs in stressed syllables. For our purposes there is no perfect way to incorporate this issue into calculations: English spelling does not explicitly indicate stress patterns, so one cannot expect that individuals who encounter an invented polysyllabic trade name in written rather than spoken form will agree on how to stress it. The solution I have chosen is to assume that all syllables in a trade name are given sufficient stress to contain stressed-syllable vowels rather than obscure vowels. As compared with a solution which attempts to recognize the possibility of obscure vowels, this decision will in one way reduce the calculated number of possible wordforms (two vowels which English does in reality use will be omitted from the calculated possibilities), but in another way will increase that number (in some longer words, one or more syllables would in practice have to be stressless, but our calculations will counterfactually pretend that the full range of alternative stressed-syllable vowels are available in those syllables). The consequence is that the numbers emerging from the calculations below can only be regarded as order-of-magnitude estimates rather than exact; but we shall see that order-of-magnitude figures are the best one could hope for in any case, and arguably for commercial and legal purposes it is ballpark estimates rather than precise numbers which are of most interest.

A further source of indeterminacy has to do with the English spelling conventions for vowels and diphthongs. English phonology has a contrast between checked vowels, which must be followed by a consonant, and "long" or "free" vowels, which can occur finally in a syllable or a word. (Many of the "free vowels" are diphthongs: for the purposes of this investigation, English diphthongs – and affricate consonants /tS dZ/ – are regarded as single phoneme units, and references to "vowels" below will include diphthongs.) The vowel letters are each regularly ambiguous between a checked and a free vowel; for instance, the letter *o* can stand for the checked vowel of *cot*, /kQt/, or for the diphthong of *cold*, /k@Uld/. Furthermore, the letter *u* can stand for either of the distinct checked vowels /V U/. In monosyllables these distinctions can often be made explicit in the spelling, for instance by using "silent *e*": the invented form *fot* could only be /fQt/, while *fote* is unambiguously /f@Ut/. But even for monosyllables English orthography is not systematic enough to eliminate all these ambiguities (e.g. the spelling -*ut* can represent /Ut/ as in *put*, or /Vt/ as in *gut*). And in polysyllables there will very often be no recognized way of showing which of the alternative regular phonetic values of a vowel letter is intended by a particular spelling.[4]

[4]One might object that this is not entirely true, if we consider devices like the Mondelēz macron. But, apart from the fact that this orthography is a large departure from the norms of English naming, I surmise that Mondelēz would not be happy if a competitor named itself Mondelĕz, and pointed to the breve to establish that its name was different because pronounced with /-lEz/ rather than /li:z/. In practice, identity of letter-sequences is probably a sufficient condition for perceived identity of names.

Therefore, to ensure that we do not separately count wordforms which are phonologically distinct but could not reliably be distinguished in writing, our calculations will treat sets of vowels which are standardly written with the same vowel letter as if they were a single vowel – the calculations will recognize no distinction between, e.g. the /Q/ of *rot* and the /@U/ of *rote*. (For monosyllables this approach would severely distort the facts, but monosyllables are such a tiny fraction of all possible words of reasonable length that for our purposes the distortion will be negligible.) My software implements this decision by using the symbols /a e i o u/ as cover-symbols for the sets of vowels regularly written with the respective letters, as follows:

a = /&/ or /eI/
e = /E/ or /i:/
i = /I/ or /aI/
o = /Q/ or /@U/
u = /V/, /U/, /u:/, or /ju:/

(In the same way – though this point will make far less difference to the eventual results – since the voiceless and voiced fricatives /T D/ of *thigh*, *thy* are both spelled *th*, they will be treated as if they were the same sound.) I shall identify distinct wordforms as distinct sequences of what I shall call "graphones": units which contrast phonologically and can also be reasonably reliably distinguished orthographically.[5] Thus /a e i o u/ will be graphones with the alternative phoneme values just listed, and /T/ will be a graphone with the values /T/ and /D/. The symbols /C J W Y/ will be graphones representing /tS dZ aU OI/ respectively. (The diphthongs /aU OI/ differ from the other English diphthongs in that they cannot be spelled with single letters, and their spellings are not also regularly used for simple vowels.) In all other cases graphones and phonemes will be interchangeable.

Finally, the range of pronunciation contrasts in English differs to some extent between regional accents. For practical commercial and legal purposes it will make sense to treat spoken forms as distinct only if they would contrast no matter whether they were spoken in standard British "Received Pronunciation" or in "General American English". In particular, my calculations will not separately count spoken forms differing only in presence versus absence of postvocalic /r/, since this distinction does not occur in RP: a pair of words such as *taught* and *tort* are distinct for most American speakers but, in RP, they are homophones. On the other hand,

[5]The irregularity of English spelling means that perfect reliability is not achievable.

no account is taken of more localized accents. For many speakers in the US South, words like *pin* and *pen* are homophones, but my calculations will treat /I/ and /E/ as distinct vowels in all environments.

5. AN INITIAL CALCULATION

For a first approximation, the numbers of possible names of a given length in syllables can be calculated algebraically. Let us represent the sizes of various classes of phonemes and phoneme-sequences as follows:

a initials
z finals
v vowels regularly spellable by single letters
d diphthongs not so spellable
Z post-free finals

The terms "initial" and "final" refer to the consonants and consonant clusters which can occur respectively at the beginning and at the end of a syllable (including zero, in cases where a syllable begins or ends with a vowel). Every English consonant phoneme other than the velar nasal /N/ can occur as an initial, and every consonant other than /h j w/ can occur as a final, but there are also sequences of two or three consonants (e.g. /sp, pl, skr/) which can occur initially, and sequences (e.g. /mp, rd/) which can occur finally.[6] "Post-free finals" refers to the fact that the range of finals which can occur after a free vowel is very limited by comparison with those that occur after checked vowels. We find no English words like /li:mp/ *leemp* or /leINk/ *laink*, for instance. "Post-free finals" are the small set of finals (various dental consonants and consonant clusters) which can occur in these positions (and hence are the only finals that can follow /W Y/).

Then it seems to follow that the number of possible names having *s* syllables should be $(a\ .\ (vz + dZ))^s$. The values *v* and *d* will be 5 and 2, for the graphone-sets /a e i o u/ and /W Y/ respectively. According to my analysis (again, see the associated software), *a*, *z*, and *Z* have the values 55, 69, and 14 respectively. The formula just given thus yields the results:

[6]In RP English, a postvocalic /r/ is realized as modification of the quality of the preceding vowel, e.g. the sequence treated for present purposes as /kard/ *card* is pronounced [kA:d]. But our calculations will be simplified by treating postvocalic /r/ as a separate phoneme, as it is in General American English.

Table 1. Numbers of possible wordforms.

syllables	wordforms
1	20,515
2	421 million
3	8.6e12 (8.6 trillion)[a]
4	1.8e17 (180 quadrillion)

[a]The notation *n* e *y* means $n \times 10^y$.

However, as it stands the formula overestimates numbers of possibilities. First, it involves some double counting. For instance, allowable syllable-finals include zero, /s/, and /sk/, while allowable initials include /skr/, /kr/, and /r/, so, medially between successive syllable nuclei, the sequence /-skr-/ could arise in three different ways. In all such cases, the single resulting sequence should be counted once only. Furthermore, not all the different graphone sequences implied by the formula will be genuinely distinct. English does not usually distinguish geminate from single consonants in speech: the surname *Hatrick* would commonly be a homophone of the word *hat-trick*. (Some speakers may distinguish them, but the distinction seems too evanescent to be a satisfactory basis for a distinctive name.) Hence graphone sequences including geminates should be discounted. Also, before /r/ followed by a consonant, or word-final /r/, the phonemes /e i u/ are neutralized (as /3:/). And, since one of the values of the graphone /u/ is /ju:/, we should not count /-u-/ and /-ju-/ as separate possibilities.

It would be difficult to allow for these issues by modifying the algebraic formula, but it is easy to take account of them experimentally. The software is set to generate wordforms of a given length randomly, to filter out all but one alternative in cases like these (for instance wordforms containing /-skr-/ between two vowels are accepted only when generated with zero final followed by /skr-/ initial), and to keep a running tally of the proportion of forms generated which pass the filter. After a few hundred iterations the tally converges on a value constant to a couple of significant figures, sufficient precision for our purposes, and this value can be multiplied into the relevant figure in Table 1 to give a corrected result.

The proportions of forms of one, two, three, and four syllables accepted by this filtering process are respectively 0.89, 0.67, 0.53, and 0.43. Thus Table 1 can be replaced by the corrected Table 2.

Table 2. Numbers of possible wordforms (corrected).

syllables	wordforms
1	18,300
2	280 million
3	4.6 trillion
4	76 quadrillion

The numbers in Table 2 are lower than those they replace, but from a practical point of view they are hardly significantly lower. Product names are often three syllables long, and we have seen that drugs can have four-syllable names. So, in view of figures like these, it may seem that there is not the slightest danger of running out of names.

6. FILTERING OUT ABSURD COMBINATIONS

However, the numbers here are upper bounds. Readers may feel less optimistic about trade-name abundance when they appreciate the nature of the bulk of wordforms making up these numbers.

As a sample, Table 3 shows the first 20 forms output when the software was set to generate disyllables and trisyllables randomly. The left-hand

Table 3. Randomly-generated disyllables and trisyllables.

Trelsprels	*Threlceprelce*	sulCSiskhorp	*Sulchshiskhaup*
berswarks	*Birswarx*	TrarCtrondsprol	*Thrarchtrondsproll*
TreNkJond	*Threnkjond*	blabsweJgwanT	*Blabswedjgwanth*
dwoJspruls	*Dwodgesprulce*	swordzulkstam	*Swordzoolxtame*
swikwild	*Swickwild*	ganzTrulbskons	*Ganzthrulbsconce*
fargwuN	*Fargwung*	vanTsklartwunJ	*Vanthsclartwunge*
dwolksblel	*Dwolxblel*	wuskspektskrord	*Wooskspectscraud*
skerksprompt	*Skirxprompt*	sendglontsnaf	*Sendglontsnaff*
twYnsbulJ	*Twoinsbulge*	TwilSklerksnarg	*Thwilshclerxnarg*
skeldTeN	*Skeldtheng*	hilpbubsplal	*Hilpbubsplale*
smondtols	*Smondtolce*	snavmunJklarJ	*Snavmungeclarge*
flolTploNks	*Flolthplonx*	dwofzafskiT	*Dwofzafskith*
Colflif	*Cholflife*	golpmolnkworf	*Golpmolnquorf*
TwenzskoC	*Thwenzscotch*	flYztrikCab	*Floyztrickchab*
jampfokt	*Yampfoct*	CinstrolCbraNks	*Chinstrolchbranx*
glistpriNk	*Glistprink*	splilsmoJgipt	*Splilsmodjgipt*
TrizSroJ	*Thrizshrodge*	spolkspriNksdwerT	*Spolksprinxdworth*
sfiCwuT	*Sphitchwooth*	skilpgrolfsklid	*Skilpgrolfsclyde*
krosbon	*Crossbone*	tuntwelglarS	*Toontwellglarsh*
praSmarT	*Prashmarth*	sersSinCspem	*Sirceshinchspeem*

columns show the forms as sequences of graphones; the right-hand columns give the same forms in one of the ways in which they might be spelled in practice. (We have seen that a given phoneme sequence will typically be spellable in many different ways. Here and below, I have deliberately varied the spellings used to render given graphones orthographically, in order to draw attention to the fact that the "obvious" spellings for given phoneme sequences are not the only, or sometimes the most likely, possible spellings if that sequence is used as a name.)

Clearly, in the main these are fairly absurd "words". *Crossbone*, and perhaps one or two of the other disyllables, might be plausible names, but most of the wordforms feel Martian, or like the deliberately repulsive names of devils in C.S. Lewis's *Screwtape Letters*. They are all pronounceable: English phonology allows quite complex sequences of consonants, for instance /mpsTw/ in the name of the village Hampsthwaite near Harrogate. But in practice complex sequences are infrequent, and one certainly does not expect to find them at more than one place in a word. However, because there are far more possible consonant combinations than there are single consonants, a system like the present one which chooses among alternatives treated as equally probable is bound to make unnaturally heavy use of complex combinations. For a more realistic count of forms which would make plausible names, we need to rein in the use of consonant sequences. But any restriction of that kind will greatly reduce the numbers of possible wordforms.

Monosyllables generated randomly in the same fashion perhaps seem on the whole more plausible; see Table 4. A higher proportion of possible monosyllables coincide with actual words (e.g. *Tub*, *Hath*), and even in monosyllables that are not real words the possibilities for complex consonant

Table 4. Randomly-generated monosyllables.

TulT	*Thulth*	tub	*Tub*
haT	*Hath*	prost	*Prost*
swiJ	*Swidge*	Twenz	*Thwenze*
Traks	*Thrax*	lort	*Lort*
werz	*Wurze*	brun	*Broon*
skreg	*Screeg*	sklaz	*Sclaze*
swunC	*Swinch*	skulC	*Skulch*
splart	*Splart*	strad	*Strade*
jipt	*Yipt*	walks	*Walks*
brorv	*Brauve*	niz	*Nize*

sequences where one syllable meets another do not arise. However, since (as already mentioned) the number of possible monosyllables is tiny relative to longer words, we shall not consider monosyllables further.

Different methods could be used to force the system to generate more plausible polysyllables. The approach taken here is based on the fact that, probably in all languages, the most natural pattern for polysyllabic words is alternation of single consonants and vowels: (C)VCV…(C). For each wordform generated, the system counts the number of violations of that pattern (points where a consonant is immediately followed by another consonant, or a vowel by another vowel), and filters out forms in which the number of violations exceeds a threshold. (Likewise, the computer in Arthur C. Clarke's story was made to "eliminate ridiculous combinations".) Since the set of graphones divides straightforwardly into vocalic and consonantal subsets, defining CVCV violations in a phonetic sense is straightforward. But one can also think of cases where a graphone of either subset is realized as a pair of letters as a "visual CVCV violation", so the count of violations is incremented by one for each occurrence of the graphones /S C J N T W Y/ in a wordform.[7]

The total number of possible wordforms having no more CVCV violations than a given threshold is estimated as before by applying an empirically-determined "filter factor" to the relevant figure from Table 1.[8]

Tables 5, 6, and 7 give samples of 2-, 3-, and 4-syllable wordforms randomly generated using the thresholds 3, 2, and 1 respectively. Table 8 shows the filter factors observed for various combinations of wordform-length and CVCV violation threshold, and Table 9 gives the figures obtained by applying those factors to the relevant entries in Table 1. Most wordforms in Table 5 feel scarcely more plausible than those of Table 3,

[7]The phrase "visual CVCV violation" may be misleading. The reason why the graphones listed above tend to reduce the plausibility of wordforms containing them may not be that alternation of consonant and vowel letters is somehow "natural", but that long English words usually have classical derivations, and the reason why some sounds are spelled with digraphs in English is that those sounds did not occur in the classical languages and hence were not provided with single letters in the alphabet which we inherited from the Romans and, ultimately, from the Greeks. Be that as it may, experimentation shows that including this factor improves the alignment between numerical counts of "CVCV violations" and perceived implausibility of wordforms.

[8]One might think that the threshold should be defined relative to wordform-length, with more violations acceptable in long wordforms. However, a little experimentation has suggested to me that absolute number of violations is the more important factor in deciding how realistic a graphone-sequence feels as a potential name.

Table 5. Randomly-generated polysyllables, max 3 CVCV violations.

iJdrok	*Idjdroke*	huvmiNhiz	*Huvminghease*	ribelguntrig	*Ribbleguntrig*
lintmapt	*Lintmapt*	udkilmlid	*Eudkilmlide*	farsfamanwo	*Farcefaymanwoe*
martklel	*Martclele*	handgutdip	*Handgutdip*	gakudpizkrog	*Gacudpizcrog*
gespret	*Guessprete*	skakwabrov	*Scaquabrove*	ferpkopaljep	*Firpcopalyep*
otjek	*Ottyek*	brilkremun	*Brylcreamun*	sominswekert	*Sominswekert*
hilsrak	*Hilsrake*	povgolkdes	*Povgolkdess*	flavenhegulk	*Flavenhegulk*
melmhupt	*Melmhupt*	zazhotbeln	*Zazzhotbeln*	buvadwarblil	*Buvadwarblil*
flupleN	*Flupleng*	septkutvar	*Septcutvar*	tokripragzeg	*Tocripragzeeg*
huvgrop	*Huvgrope*	sladrugtos	*Sladrugtoss*	hataldfaghag	*Hataldfaghag*
Japkark	*Japcark*	boskrismov	*Boscrissmov*	metvodmakrek	*Metvodmacreek*
forvberd	*Fauvbeard*	gislusfult	*Gislussfult*	jezmukugfaN	*Yezmucugfang*
spunsmen	*Spuncemen*	sfedmegmer	*Sphedmegmere*	jodalsjegnot	*Yodalsyegnote*
mulglolf	*Mulglolph*	rugzalkdut	*Rugzaldut*	kodkakglefum	*Codcackglefum*
spalbzep	*Spalbzeep*	kimwokdeld	*Kimwokdeld*	tasukilsjof	*Tassukilsyoff*
zisdworv	*Zisdwarve*	zolkbupdun	*Zolkbupdune*	funedzumprus	*Fewnedzumprus*
lizgroln	*Lizgroln*	gultgumded	*Gultgumdeed*	spomevjipbop	*Spomevyipbop*
sukgreS	*Suckgresh*	vizumtiT	*Vizzumtith*	sputbubkapop	*Sputbubkapop*
zolprol	*Zollproll*	kotajelf	*Cotayelph*	bemrodlotpak	*Beamrodlotpack*
buntguks	*Buntgux*	Cugersbi	*Chugerceby*	rubdergsodav	*Rubdurgsodave*
swudgror	*Swudgroar*	dwabkobsol	*Dwabcobsole*	tiglidoTkim	*Tiglidothkim*

Table 6. Randomly-generated polysyllables, max 2 CVCV violations.

zentjaf	*Zentyaff*	fiknusef	*Ficknewseff*	uhokefsfil	*Euhokephsphile*
kivswom	*Kivswom*	zizuskler	*Zizuskler*	nelwobjakos	*Nelwobyakose*
tusweJ	*Tusswedge*	zednotrap	*Zednotrap*	pudsogduguk	*Pudsogduguck*
twisjam	*Twissyam*	omdifki	*Omdiffky*	jamreljorug	*Yamrelyorug*
smongal	*Smongal*	suzweslu	*Suzewesslu*	hudupblenot	*Hudupblenote*
geptrep	*Geptrep*	pesarbrer	*Pessarbrer*	rosinilspos	*Rossinilsposs*
datgem	*Datgeam*	muspostan	*Muspostan*	bimnozlevok	*Bimnozlevoc*
baswon	*Bayswan*	hofilbzat	*Hoffilbzat*	nadefargwuf	*Naydefargwoof*
spakgas	*Spackgas*	snejaknam	*Sneyakname*	rabuforsnis	*Rabeuphorsniss*
futswol	*Footswoll*	tivmuprel	*Tivmuprel*	pahodlufbil	*Pahodluffbil*
hulvolp	*Hullvolp*	guvdwugag	*Guvdwugag*	ibviwetbi	*Ibvy-wetby*
gotrot	*Gotrot*	bonozpind	*Bonozpind*	tuJuzunfos	*Tujuzunfoss*
kugseJ	*Cugsedge*	jemstabeg	*Yemstabeg*	feklijafpul	*Feckliyaffpull*
lentsag	*Lentsague*	klofradup	*Clofradoop*	magpelgokot	*Magpelgocot*
lertpoz	*Lertpose*	nubipdwiv	*Nubipdwive*	egumvuhiln	*Eggumvuhiln*
febteN	*Febteng*	vukrabmif	*Viewcrabmiff*	pobetwufrog	*Pobetwoofrog*
fradhut	*Fradhoot*	ilswohov	*Ilcewohove*	diljawafdek	*Dilyawafdeck*
mubopt	*Mewbopt*	uvgevorg	*Euvgevaug*	dukrefizles	*Duckrefizzless*
trodwom	*Trodwom*	gamtumo	*Gamtumo*	fonmiruzjim	*Phonmiruzyim*
sanblal	*Sanblall*	totwozolf	*Totwozolph*	bafnerhozun	*Baffnerhosen*

Table 7. Randomly-generated polysyllables, max 1 CVCV violation.

vinluk	*Vinluck*	gitretin	*Gitreatin*	mupogozbok	*Mupogozboke*
nufkos	*Nufcoss*	wuzmokor	*Wuzmocore*	fednefonon	*Phednephonon*
Coniv	*Chonive*	itpukas	*Itpucase*	gasivofdep	*Gasivoffdeep*
fasfib	*Fasfib*	lenekvot	*Lenecvote*	nikubuzvem	*Nicubuzveme*
nohilm	*Nohilm*	tuboskol	*Tuboscol*	pibewolger	*Pibewolgur*
rovkid	*Rovekid*	ruJefam	*Rudgeffam*	kubwihugak	*Cubeweeheugac*
wadgud	*Wadegood*	samudhom	*Sammudhome*	nafitumhap	*Nafitumhap*
asvos	*Assvoss*	hodzedil	*Hodzedil*	nelemhovep	*Nelemhovep*
mugluv	*Mewglove*	hefuzjad	*Hefuzzyad*	bulutrizoz	*Bulutrizose*
fozget	*Fozget*	wusitsob	*Wusitsobe*	kubulatag	*Cubulatag*
vegrot	*Vegroat*	dagitwik	*Daggitwick*	koskadopob	*Coscadopob*
bitril	*Bitril*	monefwes	*Monefwess*	gotigeslut	*Gotigeslute*
zuwoT	*Zoowoth*	runogfif	*Runnogfife*	mulgepizul	*Mulgepizzle*
laflis	*Laughlis*	havitrig	*Havitrig*	ihizakwez	*I-hi-zackwheeze*
okzib	*Oakzib*	sosguma	*Sosguma*	riwekuwam	*Ryeweekuwam*
fasner	*Fasner*	bagwulam	*Bagwoollam*	hihodniles	*Hihodnyless*
hubhus	*Hubhuss*	ekbakif	*Ekbakif*	dazfakivun	*Dazefacivune*
wikpad	*Wickpad*	hadozbis	*Hadozbiss*	usefulzid	*Eusephulzide*
hafhaz	*Halfhaze*	polakgev	*Pollackgeve*	tuzarebant	*Tuzarebant*
gimla	*Gimla*	betlukev	*Betlookeve*	kemavnusig	*Chemavneusig*

Table 8. Filter factors.

	1 syll	2 sylls	3 sylls	4 sylls
threshold 3	0.85	0.13	0.0041	5.4e–5
threshold 2	0.69	0.038	0.00049	5.1e–6
threshold 1	0.32	0.0049	3.8e–5	2.1e–7

Table 9. Filter factors applied to unfiltered figures.

	1 syll	2 sylls	3 sylls	4 sylls
threshold 3	17,400	55 million	35 billion	9.6 trillion
threshold 2	14,200	16 million	4.2 billion	900 billion
threshold 1	6600	2.1 million	330 million	37 billion

where no constraint on CVCV violations was applied. But readers will agree, I believe, that on balance the wordforms become more plausible as names, as the CVCV violation threshold is reduced from three to one.

To my mind the crossover from a point where the bulk of examples are too phonologically clumsy to be usable in practice, to a point where most examples are imaginable in use, falls between thresholds 2 and 1. (Perhaps, for disyllables, many of the examples at threshold 2 are plausible; but if names can be *n* syllables long, combinations of fewer than *n* syllables will never be numerous enough to affect the calculations significantly.) This is certainly not to say that all longer wordforms with more than one violation must be implausible. One can easily imagine a German manufacturer of specialist trousers marketing them in the English-speaking world as *Baffner-hosen*, for instance (see the bottom right-hand entry in Table 6). For that matter, we have seen that *Hampsthwaite* exists as a place name, and hence presumably could serve as a trade name of a quite traditional type, yet it scores five violations. But among all three- and four-syllable forms, I would argue that those with more than one violation but which are imaginable as names are quite rare.

Conversely I do not suggest that all forms with one or no violation would make good names. Traders want names to be not just memorable and pronounceable but euphonious; if they echo actual words these should be words with positive (or at least not negative) associations. However, euphony is a matter of personal taste, and it is not clear that tastes in this area are widely enough shared or strong enough to be decisive against a wordform. To my ear, voiced obstruents and labial obstruents both tend to make for lack of euphony. The monosyllable /f@Up/, containing two labials, strikes me as notably blunt and ugly, yet *Fope* is the name of a manufacturer and retailer of elegant jewellery, which presumably wishes to be associated with the reverse of these qualities. The verb *meddle* has negative associations, but that evidently did not dissuade a startup group from naming their virtual health insurance card *Medlio* (see Section 1).

This reasoning suggests that the answer to the question of my title might be in the region of the figure in the lowest and rightmost cell of Table 9, namely 37 billion.

However, the decision to treat one CVCV violation as a cutoff was admittedly subjective, so I have approached the question in a second way as a cross-check. I randomly selected a sample of 239 proprietary drug names from the lists on the eMedExpert website (www.emedexpert.com), and noted how many of them contain various numbers of CVCV violations. The results are shown in Table 10.

The large difference between the figures for two and for three violations might suggest that this is where the crossover between plausible and implausible names should be placed, but that overlooks the fact that there are far more ways to construct graphone sequences containing more

Table 10. CVCV violations in real drug names.

0 violations	56
1 violation	98
2 violations	71
3 violations	11
4 violations	0.3

violations than fewer violations. What Table 10 is telling us is that each CVCV violation in a graphone string (including the first) reduces its plausibility as a name. The figure 98 is greater than 56, but 98 is a much smaller proportion of all pronounceable wordforms containing one violation than 56 is of all forms containing no violation.

If we accept the figures of Table 10 as an estimate of the relative numbers of acceptable names with different numbers of CVCV violations, then we might estimate the total number of acceptable names by assuming that *all* wordforms with zero violations are acceptable. The number of pronounceable wordforms of a given length in syllables and containing no CVCV violations can be calculated algebraically.[9] With one exception which should probably be discounted,[10] all zero-CVCV-violation names in the sample contain four syllables or fewer. Possible wordforms one to four syllables long and containing no CVCV violations total 840 million; adding numbers in the proportions of Table 10 to allow for wordforms containing one or more violations gives us a total estimate of 3.6 billion plausible names.

Because the sample represents only drug names, this figure may well underestimate the total number of plausible names. We saw in Section 2 that drug names tend to be chosen to suggest a “scientific” flavour, and this could militate against inclusion of some kinds of graphone sequence which would be perfectly acceptable for names to be used in other contexts. But

[9]The total number of pronounceable wordforms having s syllables is $(c+1) \,.\, v \,.\, (cv)^{s-1} \,.\, (c-2)$, where c is the number of single consonants excluding /S C J N T/ (i.e. 17) and v is 5 as before. The terms $c+1$ and $c-2$ refer to the fact that a word can begin or end with a vowel, but cannot end with any of the consonants /h j w/.

[10]The exception is *Phenylalanine mustard*; *phenylalanine* has five syllables. But although the two-word phrase is evidently a proprietary name, *phenylalanine* is a formal chemical name, and such words belong to a system of their own in which commercial considerations such as memorability play little part. Many names of chemicals are much longer than five syllables. I do not see this single case as strong evidence for counting wordforms longer than four syllables as plausible trade names.

on the other hand this calculation does suggest that the previous estimate of 37 billion, ten times larger, is unlikely to be too low. (Parenthetically it is interesting to note that these two estimates bracket Arthur C. Clarke's figure of "nine billion". If this was purely a guess by Clarke, it seems to have been a lucky guess.)

7. CONCLUSION

If the answer to the title question is "several tens of billions", or even if it is only "several billions", one might feel that humanity is in no imminent danger of running out of distinctive trade names.

However, available names run into billions only provided four-syllable names are acceptable. Drug names are often four syllables long, but laymen do not find such long names easy to remember. In many trade sectors I believe the preference might be for names no longer than two syllables, in which case Table 9 would suggest that the possibilities are a few million only. In the present context that is by no means a large quantity. It is far smaller than one well-informed estimate I have read of the total number of businesses in the world, namely "more than 235 million".[11] If this last figure is about right, then even allowing three-syllable names Table 9 would imply that there are only just enough possibilities for businesses to own one distinctive name each – in practice a business often needs to use many names. And if the second calculation, based on the sample of drug names, were appropriate, then totals would be really tiny: 490,000 disyllables, or 42 million trisyllables.

In a legal context it is also relevant to consider that trade-name disputes are often not about separate businesses using identical names. Frequently the complaint is that a newcomer has chosen a name which differs slightly from that of an incumbent, but which is similar enough for the two to be confusable. My calculations have assumed that two names are distinct provided they differ by as little as one graphone, but we could also ask how many distinct names there would be if each pair had to differ in two, three, or more graphones, or in some given proportion of their graphones. I have not attempted to investigate the arithmetical consequences of requirements

[11] Worapong Smithirittha quoting Dun & Bradstreet data on Quora,<www.quora.com/How-many-companies-there-are-in-the-world>, dated 8th November 2014, accessed 28th November 2015.

like those, but it is obvious that they would lead to figures much smaller than those of Table 9.

I conclude that, although possible names are certainly very numerous, they are not so numerous as to make the idea of "running out of distinctive names" merely fanciful.

ACKNOWLEDGEMENT

I have no financial or similar interest in possible applications of this research.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

REFERENCES

Clarke, A. C. (1953). The nine billion names of God. In F. Pohl (Ed), *Star Science Fiction Stories* (pp. 195–202). New York, NY: Ballantine Books.

Economist (2015). Nine billion company names. *The Economist* (24 Oct), 70.

Gibbon, D., Moore, R., & Winski, R. (Eds) (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.