CAN ALL WORDS BE EXPRESSED AS SUMAGRAMS?

DANIEL AUSTIN
Cambridge, England
(daniel@fun-with-words.com)

At the place where mathematics and the world of anagrams overlap, many interesting phenomena arise, as regular readers of Word Ways are frequently reminded. Undoubtedly the most famous example is the anagram ELEVEN PLUS TWO = TWELVE PLUS ONE discovered by Melvin Wellman over half a century ago, with equality both mathematically and as an anagram. Two more of these appeared in Word Ways in 1992. They are Spanish examples by Lee Sallows: CATORCE + UNO = ONCE + CUATRO and DOCE + TRES = TRECE + DOS.

Sumagrams are a special breed of anagram which allows letters not only to be added, but subtracted too. They are a hybrid of anagrams and mathematical sums. This is best illustrated with an example.

SIXTEEN + TEN - NINE = SEXTET

That is to say the letters of SIXTEEN and TEN, with the letters of NINE removed, can be rearranged to form the word SEXTET. Here are some more sumagrams. I have written the numbers numerically for brevity.

FORTUNE = 10 + 4
ENVOYS = 61 + 7 – 16
FEIGN = 48 + 9 – 31

There is just one rule; you may add and subtract as many numbers as you wish, provided you use only natural numbers (non-negative, whole numbers). Here are some more words which can be formed by this method. Several of them were created by Mark Tothill for a puzzle he posted to the *rec.puzzles* newsgroup in 2001. Try to find a sumagram equation for each of them (answers at the back). They are all possible using six or fewer numbers in each case.

GIFT
INTENT
NOTIFYING
THEORY
EXISTENT
WRITHE

So just how rare are these words? Are they as rare as the true anagrams equations in the opening paragraph? Or can we find a sumagram for any word we choose? The aim of this article is to answer these questions.

Intuitively you might feel that the set of words that can be formed this way would be very limited. On the other hand, there are infinitely many natural numbers, which we can add and subtract in an equation of any length, to form any of several million words in an English dictionary. Does this massive flexibility of sumagrams allow one to form any word one chooses, albeit probably by adding and subtracting an inelegantly long string of numbers?

One way to approach this problem is to see which individual letters can be created with a sumagram equation. Simple inspection yields a few.

Y = 80 - 8
T = 68 - 86
E = 16 - 10 - 6

These are not the only possible solutions for E, T, and Y, but we are not concerned with *how many* ways they can be made, only *whether* they can. Since we know that these three letters are possible solutions to sumagram equations, we know that any words made using only a subset of these letters are too.

YET = Y + E + T
= (80 - 8) + (16 - 10 - 6) + (68 - 86)
= 80 + 68 + 16 - 86 - 10 - 8 - 6

Because of the method we have used to arrive at this solution, it is unlikely that it is the simplest possible sumagram for YET. But again, we are only interested in whether a solution exists, no matter how complex. The challenge is to see how many single letters we can find a sumagram equation for. If there exists a solution for every letter of the alphabet, we're home and dry; that would mean any word you care to choose can be expressed as a sumagram.

Immediately we run into a problem. The letters B, C, J, K, P, Q, and Z do not occur in the English name for any number before ONE BILLION. I will look at where we might be able to find these letters later. For the time being, I am going to concentrate on searching sumagram solutions for the other 19 letters of the alphabet.

You will recall that we have already found solutions for E and T above. We can use this to help us find a solution for N because TEN = T + E + N. Or to put it another way:

N = 10 - T - E
= 10 - (68 - 86) - (16 - 10 - 6)
= 86 + 10 + 10 + 6 - 68 - 16

Now we can look for numbers containing just E, N, T, and Y, and one other letter. If we find such a number, we are guaranteed to have a sumagram solution for the extra letter. TWENTY is the perfect candidate.

W = 20 - T - E - N - T - Y

Now we're on a roll. Every time we find a solution for a new letter, we gain more power to find solutions for others. We have found sums for each of T and W, so next we can use TWO to find a solution for O. We can find a solution for I in a similar way using NINE.

O = 2 - T - W
I = 9 - N - N - E

We can also exploit a quirk of English spelling, namely the fact that the FOUR in FORTY is spelled without a U. It is words with slight differences like this that are most valuable in the hunt for sumagrams of single letters.

U = 4 + T + Y - 40

But here the fun stops. There are no numbers spelled using E, I, N, O, T, U, W, Y and just one other letter. It turns out that we are now completely stuck. Unless, that is, we go beyond a billion and enlist the help of the gigantic NONILLION. Before I explain how this number rescues the situation, I want to look at whether using such large numbers is really legitimate.

I was initially reluctant to use numbers beyond a billion, mainly because there is no universal agreement on how numbers beyond 999,999,999 are named. Under the American system, the next natural number is called ONE BILLION. A British or Australian speaker of English, on the other hand, would use this term to refer to a number a factor of a thousand larger. (Although, in recent years, the American number-naming system has been taking hold in these countries too.)

However, the actual numerical value of a billion is irrelevant. Whatever system you choose to use, it is a legitimate number with a consistent spelling, which is included in any English dictionary. The same argument goes for these numbers too: TRILLION, QUADRILLION, QUINTILLION, SEXTILLION, SEPTILLION, OCTILLION, NONILLION, DECILLION, UNDECILLION, DUODECILLION, TREDECILLION, QUATTUORDECILLION, QUINDECILLION, SEXDECILLION, SEPTENDECILLION, OCTODECILLION, NOVEMDECILLION, VIGINTILLION... CENTILLION. In every case the actual value is not universally agreed; but fortunately for us the spelling is.

Let me remind you where we had got to; NONILLION was just about to save the day. But not without the help of another very special number: TRILLION. (I should mention that it does not matter that we are omitting the ONE from ONE BILLION, ONE TRILLION, etc. because we have already found solutions for O, N, and E.)

NONILLION is made entirely from letters for which we have found a solution and, in addition, L. In fact, it contains two Ls, and this poses a bigger problem than might at first appear to be the case. Sumagram equations are not *ordinary* mathematics; you cannot simply *divide by two*; nor is subtracting L an option, after all it is a solution for L that we are seeking.

L + L = NONILLION - N - O - N - I - I - O - N

Here is where TRILLION gets to play its important role.

R + L + L = TRILLION - T - I - I - O - N

Subtract the latter equation from the former, and hey presto, we have a solution for R!

R = (TRILLION - T - I - I - O - N) - (NONILLION - N - O - N - I - I - O - N)

Having a solution for R triggers a cascade of other solutions. To remind you, we have now found solutions for E, I, N, O, R, T, U, W, and Y.

F = 4 - O - U - R
H = 3 - T - R - E - E
G = 8 - E - I - H - T
V = 5 - F - I - E
S = 7 - E - V - E - N
L = 11 - E - E - V - E - N
X = 6 - S - I
M = MILLION - I - L - L - I - O - N

Only A and D remain unsolved from our list of 19 possible letters. Additionally, the letters B, C, P, and Q occur in the post-billion numbers that we have decided to allow, giving us the potential to find solutions for those too. And solutions for these six all come easily.

B = BILLION - I - L - L - I - O - N
Q = QUINTILLION - U - I - N - T - I - L - L - I - O - N
P = SEPTILLION - S - E - T - I - L - L - I - O - N
C = OCTILLION - O - T - I - L - L - I - O - N
D = DECILLION - E - C - I - L - L - I - O - N
A = THOUSAND - T - H - O - U - S - N - D

Only the troublesome J, K, and Z remain. Rick Rothstein suggested allowing ZERO. True, it is mathematically fruitless to add or subtract zero, but it is certainly not illegal, and it is far from fruitless for the purposes of sumagrams.

Z = 0 - E - R - O

And these solutions for all letters bar J and K are the best that can be done. Some dictionaries include JILLION, but only vaguely defined as 'a very large number'. Ben Zimmer brought LAKH (containing a K) to my attention, an Anglo-Indian term synonymous with a hundred thousand, which is found in many English dictionaries. The sad truth, though, is that under no generally accepted English number-naming system is there a word for any integer that contains J or K.

But we have conquered the problem and demonstrated that words that can be expressed as sumagrams are far from rare. Any word not containing J or K has a solution. This, of course, is the vast majority of English words – about 99%.

I wonder if there is a perfect sumagrammatical language in which *all* words – without exception – have a corresponding sumagram. With several thousand languages in the world, I'd be willing to bet there is. I will conclude by considering what properties we might look for when seeking such a language.

The most obvious requirement is that every letter in the language's alphabet must appear in at least one of its number terms. So perhaps the most likely candidates are languages with very small alphabets. The smallest alphabet of any known language belongs to Pirahã, with just seven consonants and three vowels, spoken by no more than about 300 people in an isolated area of Amazonia.

Like many of the more primitive languages of the world, Pirahã is unwritten (save for the transcriptions of the few linguists who have studied it). However, in principle it can be written by assigning a symbol to each of the ten spoken phonemes, so this need not worry us unduly. A far more problematic fact about Pirahã becomes apparent when we ask what its numbers words are. Pirahã has no number terms! Indeed, it is claimed that the language has no concept of counting at all.

Pirahã may be unique in its complete lack of number terms, but many of the world's languages are almost as numerically primitive (e.g. having only words corresponding to "one", "two", and "many"). Of course, such languages are of no use to us in our search for the sumagram holy grail.

Hawaiian has a very small alphabet too. And it is a written language including words for the natural numbers. But unfortunately our prayers do not seem to be answered here either, for I know of no Hawaiian number term containing the letter P.

Furthermore, we cannot appeal to languages like Chinese, in which the standard writing system is incompatible with the notion of anagrams. Of course, any language may be represented in Latin orthography, but then the question arises of which system of transliteration system should be employed. It is also an issue whether we should disregard diacritics; should Á be regarded as a separate letter from À, for example?

Within the subset of languages that have a writing system with a standard alphabet, and have words for the natural numbers, many are historically related. For example English and German come from the same roots. It is therefore no surprise that, as we found in English, German number terms do not use the letters J or K either.

We found the plentiful supply of novel English words for gigantic numbers like QUINTILLION to be very useful. But, the vast majority of the world's languages do not have many such words. (It is arguable that they are not valid words in English either, for many dictionaries do not include them, and they are encountered incredibly rarely.)

There is one final issue of importance: the regularity (of spelling) of the number terms. As we discovered earlier, small spelling discrepancies (like the facts that the FOUR in FORTY loses its U, and that the morpheme TEN in numbers like SIXTEEN has become corrupted to TEEN) are handy for our purposes. A language which is either too regular or too irregular in this respect may be impossible to crack sumagrammatically.

All this leaves me uncertain of how likely it is that there exists a language in which all words can be expressed as sumagrams. I am inclined to be optimistic, but I leave it to the reader to take up the challenge of finding it.