THE ALPHANUMERIC PROPERTIES OF HYPERPOLYGRAPHEMIC OLIGOSYLLABLES

T. A. Hall
Bloomington, Indiana

In a recent article in this journal (Word Ways, August 2003), Hugo Brandt Corstius examines the connection between the number of letters in a word and the number of syllables it contains. Basing his observations on Dutch, he coins the term 'oligosyllabic', which he defines as a word in which the number of letters is more than four times greater than the number of its syllables. For example, the Dutch word HANDBOOGSCHUTTERSMAATSCHAPPIJ is 'oligosyllabic' because it contains 29 letters and 7 syllables. By contrast, a Dutch word like OPERA does not qualify as 'oligosyllabic' because it has a mere 5 letters distributed over 3 syllables.

It is clearly the ratio of the number of letters divided by the number of syllables that makes certain words interesting to logophiles. If that number is high – defined arbitrarily by Costius as being greater than 4 – then that word is deemed 'oligosyllabic'.

My first point is a terminological quibble: The word 'oligosyllabic' should mean 'characterized by few syllables', but we are referring here to a class of words that is defined not simply by their having a small number of syllables. The crucial factor involves the relationship between the number of letters in a word and the number of syllables that it contains. Thus, instead of calling a clearly polysyllabic word like HANDBOOGSCHUTTERSMAATSCHAPPIJ as 'oligosyllabic', it makes more sense to refer to it as a 'polygraphemic oligosyllable', i.e. literally 'a word characterized by many letters (graphemes) and few syllables'. True, my terminological emendation fails to capture the actual relationship between the number of syllables and the number of letters, but I believe it is more accurate than the term 'oligosyllabic' alone.

In the remainder of this note I offer some of my own observations concerning the properties of polygraphemic oligosyllables.

Section 1: Some statistics

In the article referred to above, Corstius provides a useful chart in which he lists a number of Dutch words containing between 1 and 12 syllables and the number of letters in those words. Figure 1 below has been constructed on the basis of the data provided by Corstius. In the first column I provide the number of syllables (between 1 and 7) and in the second column the maximum number of letters for a word in each of these seven categories. In the third column I give the ratio of the maximum number of letters to the number of syllables (henceforth the 'LS-ratio'), which I have rounded off to the nearest tenth. In the final column I present the Dutch word with the highest LS-ratio for each syllable type. I have only included words between 1 and 7 syllables because none of the Dutch examples containing 8 or more syllables has an LS-ratio of more than 4.

Figure 1: Polygraphemic examples from Dutch

syllables	*letters*	*LS-ratio*	*example*
1	8	8	SCHREEUW
2	12	6	VOORTSCHREED
3	16	5.3	VASTGESCHROEFDST
4	20	5	SCHOONHEIDSWEDSTRIJD
5	24	4.8	RIJKSVOORLICHTINGSDIENST
6	28	4.7	STAATSLUCHTVAARTMAATSCHAPPIJ
7	29	4.1	HANDBOOGSCHUTTERSMAATSCHAPPIJ

It can be observed from Figure 1 that the LS-ratio is inversely proportional to the number of syllables. Put differently, it becomes less and less likely to obtain an LS-ratio greater than 4 when the number of syllables increases, as observed in this journal by Rex Gooch (Word Ways, November 2003). In fact, these statistics confirm a law of quantitative linguistics called Menzerath's Law, which says that the increase of a unit (e.g. word size in terms of the number of syllables) results in a decrease of its constituents (e.g. the number of letters.) This being said, it is striking that the polysyllabic examples in Figure 1 are compound words. Are there noncompound words containing, say, 5, 6 or 7 syllables which have an LS-ratio greater than 4? I leave this question open for readers of this journal to ponder. Instead, I concentrate for the remainder of this note on a separate issue.

Section 2: Hyperpolygraphemic monosyllables

Given the observation that the LS-ratio increases as the syllable count decreases, it is not surprising that words containing a single syllable (monosyllables) have the highest LS-ratio. In fact, a glance at Figure 1 reveals that the LS-ratio for monosyllables is considerably higher than 4. I propose that words such as these be referred to as 'hyperpolygraphemic', although I refrain here from assigning a specific number to describe this word type. Since it appears that it is only words containing one syllable that possess this property, it is therefore useful to refer to these items as 'hyperpolygraphemic monosyllables'.

And what is the LS-ratio for English hyperpolygraphemic monosyllables? This is a question that has been posed in many different ways through the years (see below). In Figure 2 I have listed English monosyllables containing between 7 and 10 letters. Since the number of syllables is one, the LS-ratio for all of these examples is identical to the number of letters they contain. I have given examples for each of the four categories, although I make no claims that this list is exhaustive.

Figure 2: English monosyllables

letters	*examples*
7	SQUEEZE, SCREECH, STRETCH, SCRATCH, STRANGE, SQUELCH, SCRUNCH, THOUGHT, BROUGHT
8	SCROUNGE, SCRAUNCH, BROUGHAM, STRAIGHT, STRENGTH, SQUEEZED, THOUGHTS, STRESSED, STRIPPED
9	SQUELCHED, STRETCHED, SCROUNGED, STRENGTHS, SCREECHED, SCRATCHED, SCRUNCHED, STRAIGHTS
10	SCRAUNCHED

Considerable quantities of ink have been spilled through the years in an attempt to prove that English really does have more examples in the 10 letter category. Consider Ralph G. Beaman. Over forty years ago (Word Ways, November 1970) he took the liberty of promoting the 9 letter past tense verbs listed above into nouns and then adding the –S plural, thereby forming words like SCRATCHEDS and STRETCHEDS. But no serious dictionary recognizes these obviously contrived monstrosities. A similar point can be made for the example BROUGHAMMED listed in the Wikipedia article 'List of the longest English words with one syllable'. According to the authoritative sources for English with which I am familiar, the base BROUGHAM is a noun ('type of horse-drawn carriage') and not a verb. The verb 'to brougham' is a creation attributed to the poet William Harman in a competition to find the longest monosyllable. I would be a hypocrite if I were to reject neologisms out of principle, but I would argue that this particular coinage has not yet established itself into the English tongue to take it into serious consideration. The same point holds for the verb SCHTROUMPF (and its inflected form SCHTROUMPFED), cited in the Wikipedia article cited above. The word SQUIRRELLED is sometimes claimed to be the longest monosyllable (with 11 letters), but the lexica with which I am familiar list this item as disyllabic, with only an optional pronunciation as a monosyllable. One might be inclined to add SCROOTCHED to the 10 letter category, but this item is usually considered to be a regionalism which is restricted in its usage to the Midland and Southern United States.

These questionable items aside, there is a generalization that was not appreciated by the authors cited above: Many of the English examples listed in Figure 2 consist of a root and a suffix, while others are simply roots by themselves (monomorphemes). In fact, all of the English examples containing 9 letters listed above belong to the former category, e.g. STRETCHED = STRETCH+ED). The one example in the 10 letter category listed in Figure 2 is also a suffixed word.

So what is the maximal length of letters in an English monosyllable? There is more than one answer depending on the type of monosyllable one takes into consideration. The answer is 10 if one wants to include English words with a suffix, although it needs to be stressed that SCRAUNCHED is one of the few truly legitimate monosyllables of this length. However, if one only wants to consider monomorphemes, then the more honest answer would be 8.

A language that beats English in terms of the LS-ratio for monosyllables is German. Some examples of hyperpolygraphemic monosyllables in that language are listed in Figure 3. These words consist of common, everyday nouns, adjectives and verbs (in the imperative singular).

Figure 3: German monosyllables

letters	*examples*
7	STRÜMPF, STREICH, STRAUCH, SCHLAMM, SCHLANK, SCHWACH, SCHWARZ, GRABSCH, TRATSCH, DEUTSCH, PFLICHT
8	SCHLAUCH, SCHLACHT, SCHLECHT, SCHLEICH, SCHLICHT, SCHLUCHT, SCHRUMPF, SCHNARCH, SCHWULST
9	SCHLAUCHS,SCHRUMPFT, SCHNARCHT, SCHLEICHT, HERRSCHST, KNIRSCHST, SCHIMPFST, STREICHST, QUETSCHST
10	SCHRUMPFST, SCHNARCHST, SCHLEICHST

The examples in Figure 3 reveal that German beats English in two respects: First, there are many highly frequent monomorphemes consisting of 8 letters. Second, German monosyllables consisting of 9 and 10 letters are also plentiful, although it should be added that these examples have in common that they all contain an inflectional suffix, e.g. –S, –T, –ST.

What is striking is that German is able to fill the four categories of monosyllables above with honesty and integrity and without resorting to trickery. The words in Figure 3 contain no contrived examples, no questionable neologisms, no regionalisms and no disyllables trying to pass themselves off as monosyllables. German simply wins fair and square.

Section 3: Prelude to perfection

So why does German trump English? As I demonstrate below, the answer to this question can only be revealed once we investigate the alphanumeric properties of the German examples listed in Figure 3. It is striking that so much has been written on the length of English and German monosyllables, but none of this previous literature has taken isopsephy into consideration.

An obvious property shared by virtually all of the German examples is that they contain the trigraph SCH. This is true for all of the words with 8, 9 and 10 letters, and it is almost always true for the examples with 7 letters. By definition, a trigraph is a sequence of three letters which represent a single sound. This is clearly the case for German, in which SCH represents the postalveolar fricative (corresponding to English digraph SH), as in an English word like SHIP. By contrast, English SCH in words like SCHOOL is not a trigraph because it represents two sounds. The same point holds for Dutch

SCH (in words like SCHREEUW from Figure 1), which represents the same fricative as in German plus the CH sound as in German BACH. Note that German SCH is typically situated at the left edge of the word, although there are other examples in which SCH is at the right edge or before the inflectional suffix.

The German language has a plethora of digraphs (e.g. CH, PH, IE), but only one highly-frequent trigraph. It is also striking that the one tetragraph (TSCH in words like TRATSCH) contains the S, C, H sequence. Words like SCHNAPPS, SCHMALTZ and SCHMALTZED are sometimes listed as as long 'English' monosyllables, but these examples all have in common that they are words English has simply snatched away from the German language. Note that this egregious act of thievery included the confiscation of the trigraph SCH *in situ* without the expected transformation into the regular digraph SH. 'English' words like SCHLOCK and SCHMEAR were likewise stolen from Yiddish. But is it not peculiar that English decided to spell these newly acquired items with SCH, even though the fricative sound in Yiddish is traditionally romanized in that language as SH and not as SCH?

Section 4: Gematria àla Germania

So what is it that is so special about the trigraph SCH? Assigning numerical values to S, C and H, where A = 1, B = 2….Z = 26 etc., we obtain 19, 3 and 8. (The German alphabet consists of the same 26 cardinal letters as the English alphabet; the three umlauted vowels Ä, Ö, Ü, as well as ß, are all placed in that order after Z by convention). The three integers 19, 3 and 8 are all examples of what mathematicians refer to as 'deficient' in the sense that they are greater than the sum of their proper divisors. (The proper divisors of an integer *n* are those integers dividing evenly into *n* with the exception of *n* itself). For example, the only proper divisor of the prime number 19 is 1, which is less than 19. The same point holds for 3, which is greater than 1, its only proper divisor. The proper divisors of 8 are 1, 2, and 4, which add up to 7, and since 8 is greater than 7, the former integer is deficient.

The S, C and H in the trigraph SCH might therefore be thought of as being deficient when they occur separately, but when in unison, something wonderful happens: Adding together 19, 3 and 8, we obtain 30, which is an abundant integer and not a deficient integer because it is less than the sum of its proper divisors, i.e. 30 is less than 1 + 2 + 3 + 5 + 6 + 10 + 15 = 42.

It is therefore the 'abundant trigraph' SCH which makes the German examples special, but there are deeper reasons for considering 30 to be unique. As noted above, the proper divisors of 30 are 1, 2, 3, 5, 6, 10, and 15, which add up to 42. Repeating this procedure with 42 as the input and adding together its proper divisors yields 54. In Figure 4 I have presented data for 30 and the next 6 steps. The list of integers in the left hand column is a recursive sequence in which each term is the sum of the proper divisors of the previous term (an 'aliquot sequence' in mathematics). The first seven members of the sequence with 30 as the input are {30, 42, 54, 66, 78, 90, 144}.

Figure 4: The first seven members of the aliquot sequence of 30:

integer	*proper divisors*	*sum of proper divisors*
30	1, 2, 3, 5, 6, 10, 15	42
42	1, 2, 3, 6, 7, 14, 21	54
54	1, 2, 3, 6, 9, 18, 27	66
66	1, 2, 3, 6, 11, 22, 33	78
78	1, 2, 3, 6, 13, 26, 39	90
90	1, 2, 3, 5, 6, 9, 10, 15, 18, 30, 45	144
144	1, 2, 3, 4, 6, 8, 9, 12, 16, 18, 24, 36, 48, 72	259

What is striking about the sequence {30, 42, 54, 66, 78, 90, 144} is that the Greatest Common Divisor (CGD) for its 7 members is 6, which is a 'perfect number' because it equals the sum of its proper divisors (1+2+3=6). Not surprisingly, each of the 7 members of the sequence also shares the three proper divisors of 6: 1, 2 and 3.

That there is something special about 30 becomes even more evident when additional facts are taken into consideration. First, when the sequence illustrated in Figure 4 is repeated for input integers other than 30, it is difficult to find the perfect number 6 playing such a central role. To be more precise: Although I can offer no formal proof, I submit that 30 is the lowest integer in the sequence defined above in which the GCD 6 occurs in 7 or more of its members.

One might assume that the multiples of 6 would yield similar results, but this is not the case, at least not for the multiples of 6 between 12 and 96. Consider first 12: The sum of the proper divisors is 16, which does not have 6 as a divisor. For 24 we obtain 36 (i.e. 1 + 2 + 3 + 4 + 6 + 8 + 12 = 36), and from there we get 55, but 55 does not have 6 as a divisor. For 18 we get 21, which does not have 6 as a divisor. Ditto for 36, 48, 72, and 84. One might suspect that 60 behaves like 30, but that is not the case: The first three members of the aliquot sequence of 60 are {60, 108, 172}, but 172 does not have 6 as a divisor. A similar situation obtains for 96: We have {96, 156, 236}, but 6 does not divide evenly into 236.

It is not until we consider 102 as the input that we encounter a sequence consisting of 7 members, as in Figure 4 for 30: {102, 114, 126, 186, 198, 270, 450}. But it is interesting to compare {30, 42, 54, 66, 78, 90, 144} with {102, 114, 126, 186, 198, 270, 450}. In the former, the difference between each member of the sequence with the preceding one is either 12 (i.e. 42-30, 54-42, 66-54, 78-66, 90-78) or 54 (i.e. 144-90). The GCD of 12 and 54 is 6. By contrast, the difference between each member of the sequence beginning with 102 with the preceding one is either 12 (i.e. 114-102, 126-114, 198-186), 50 (i.e. 186-126), 72 (i.e. 270-198) or 180 (i.e. 450-270). The GCD of 12, 50, 72 and 180 is 2 and not 6. The sequence in Figure 4 (beginning with 30) therefore bears a closer connection to the perfect number 6 than the one beginning with 102.

Inquiring minds might want to know what happens when we repeat the process of adding the proper divisors to the last integer in the right hand column of Figure 4. These data have been presented in Figure 5:

Figure 5: The final eight members of the aliquot sequence of 30:

integer	*proper divisors*	*sum of proper divisors*
259	1, 7, 37	45
45	1, 3, 5, 9, 15	33
33	1, 3, 11	15
15	1, 3, 5	9
9	1, 3	4
4	1, 2	3
3	1	1
1	0	0

Clearly the continuation of the sequence begun in Figure 4 with 30 as the original input (i.e. {259, 45, 33, 15, 9, 4, 3, 1}) is very different from the first part of the sequence (i.e. {30, 42, 54, 66, 78, 90, 144}) because the members of the former do not have 6 as one of the proper divisors. The reason 6 is the GCF in all of the members of the latter and not in the former is a function of the abundant vs. deficient distinction described above: The seven integers in the sequence in Figure 4 are all abundant, in contrast to the integers in the continuation of the same sequence in Figure 5, which are all deficient.

So what do we conclude? German hyperpolygraphemic monosyllables are special because of the prominent role played by the trigraph SCH, which itself can be transposed into 30 – a highly unique integer in the sense that it is so intimately related to the perfect number 6.

A POEM

MARTIN GARDNER

This is an excerpt from Gardner's 1969 book *Never Make Fun Of A Turtle, My Son* (Simon and Schuster, illustrated by John Alcorn).

Twiddle-bug

A television twiddler
　　Is a most annoying pest.
The programs other people like,
　　She's certain to detest.

When Father is excited
　　By a fight on Channel Three,
The twiddler turns to Channel One
　　To see what she can see.

And just when Mother settles down
　　To watch her favorite play,
The twiddle-bug will twist a knob
　　And twiddle it away.

The only thing for Dad to do
　　With such a twiddly-head,
Is grab her firmly by the ear
　　And twiddle her to bed!