STATISTICS OF WORD NEIGHBOURS

REX GOOCH
Welwyn, Herts, England

INTRODUCTION

The work reported here was undertaken during preparation of an article on Word Ladders: I felt in need of some facts to direct my efforts, and to clarify some remarks others had made. This article presents the frequency of occurrence of some types of words, and compares this to a "theoretical" frequency. The theoretical frequencies are based upon very simple assumptions. Nevertheless in some cases the predictions are quite good: in other cases I seek to find why the predictions are poor.

There is great interest in wordplay in words of the same length which differ by just one letter from each other, not least in the formation of word ladders. We might call these "neighbours": they are adjacent nodes in a graph (network) of words. If you put such links together, you can construct graphs or networks (the former a mathematical term, and the latter used in project planning). Graphs have been discussed many times in Word Ways (e.g. by Leonard Gordon), and also in *Making the Alphabet Dance*. Graphs naturally contain word ladders. I discuss various types of neighbours, introducing a new type.

Throughout this article, the reader should bear in mind that many statements are true only with respect to the vocabulary used. Thus an isolano (especially) may no longer be such if many more words are added to the vocabulary. For example, LLYN (Word Ways May 1970) and BANKRUPTCY (Word Ways Aug 1971) are not isolanos if the vocabulary includes GLYN, LAYN, LEYN, LOYN, LYYN or BANKRUPTLY; neither are ECRU, TPRW, or UMFF. This article uses a larger vocabulary than previous authors have used, but a larger vocabulary still would probably reduce the number of isolanos, increase the number of onalosi and neighbours, etc.

Words quoted appear in the Oxford English Dictionary, the Merriam-Webster Unabridged 2nd Edition, Stedman's Medical Dictionary, Pulliam and Carruth's The Complete Word Game Dictionary, or the Official Scrabble Players Dictionary if no source is cited. Labeled sources include Beyond (Borgmann, *Beyond Language*), Cooper (An Archaic Dictionary), EDD (English Dialect Dictionary), Nobbs (Dictionary of Norfolk Words and Uses), TAW (Times Atlas of the World), Web 3 (Merriam-Webster Unabridged, 3rd Edition) and Thurber (*The Wonderful World of O*). Inferred words are labeled inf'd, and variant OED forms, vf.

WORD OCCUPANCIES

Let us first calculate what we might expect if words were composed of letters purely at random. Using the 26-letter alphabet, there are 26x26 combinations of two letters, or 676 possible words. In general, there is room for 26^n n-letter words. This number grows very rapidly: there are half a million four-letter "words", so a dictionary of these alone would be about as large as the largest dictionaries published.

I now factor my own word list up somewhat to one million words, which I believe is a reasonable guess as to the number of modern words (including technical terms) that exist in English. Next, I calculate what fraction of all possible words of a given length actually exist; this I will call the "occupancy".

Word Length	1	2	3	4	5	6	7	8	9	10	11
Actual Words	26	673	4746	18212	44738	74897	100027	118286	125654	120774	105770
Potential words	26	676	17576	456976	11.88M	309M	8.03G	209G	5.4T	141T	3670T
Occupancy/M	1M	0.996M	270027	39853	3766	242	12.5	0.566	0.023	0.000856	.00000288

Notes:
1 M means million, G (Giga) means thousand million, T (Tera) means million million.
2 The fraction occupancy has been multiplied by one million to avoid many zeroes after the decimal point.
3 The 26 one-letter words have not been scaled up!
4 Jeff Grant has found all 676 two-letter words (The Concise Dictionary of Two-Letter Words) compared to the 673 estimate.
5 In the projected one million word dictionary, there would be sufficient words of each of the lengths seven to eleven to fill the Concise Oxford Dictionary.

I am forced to use a log scale for the occupancy graph (Graph 1) to prevent all the points from length four or five upwards from lying along the axis. From length five upwards, the potential word space is very sparsely occupied. From the table or the graph, if you write four letters at random in succession, there is one chance in about 25 that you will have written a word; but if you do the same with nine letters the chance is only one in 50 million. The word space grows far faster than the numbers of actual words; moreover the actual number of words stops growing and begins to decline after length six or seven. The two factors thus conspire to make random letters extremely unlikely to form long words. Now we know that it is very easy to form word ladders for four-letter words; the section on isolanos below shows how rare it is for four-letter words not to have a neighbour differing in only one letter. It follows that words do not occupy the word space at random, but rather in clusters (even one big cluster). Thus most letters of the alphabet can be placed before -ARE, -EST and -AND to form words. We will examine this "commonality" of letters a little more closely later.

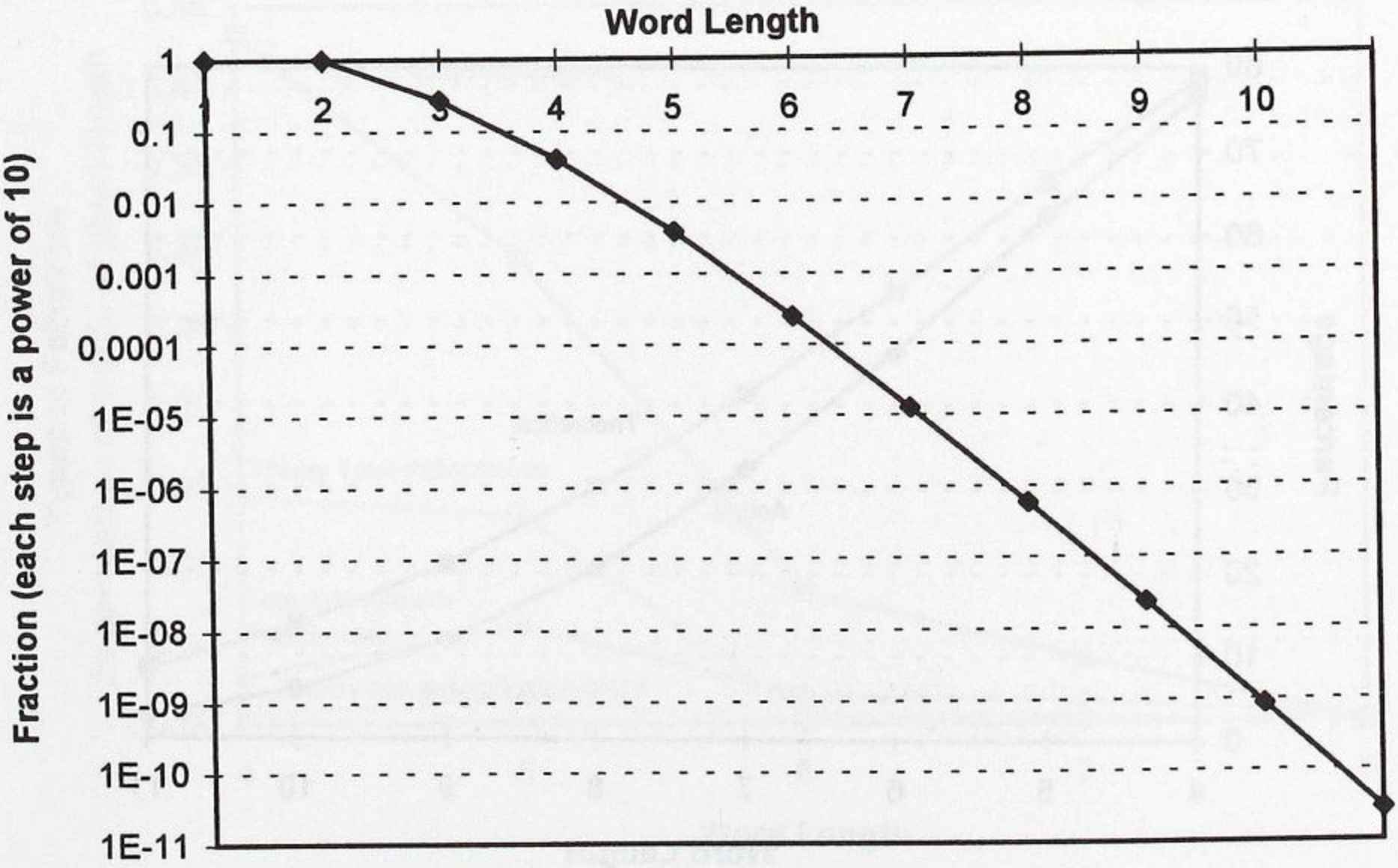
HETEROGRAMS (SOLO ISOGRAMS)

I use Susan Thorpe's term for what have otherwise been called "solo isograms"--words in which all of the letters are different. We are free to choose any of the 26 letters for the first letter of a heterogram, but we are then restricted to 25 choices for the second letter, 24 for the third, and so on. The potential number of heterograms for four-letter words is 26x25x24x23 = 358800, out of 26x26x26x26 = 456976 possible four-letter words, i.e. 78.5 per cent. This compares quite well with the 76.7 per cent for four-letter heterograms in my vocabulary. The same procedure was used to construct Graph 2 (next page) for other lengths. The general shape of the graph is expected, as the longer the word, the more chance that two letters will be the same. However, by length eleven, the prediction is just over three times the actual: in fact the ratio of predicted percentage to actual percentage grows steadily from 1.02 to 3.1. We conclude that letters repeat themselves more commonly than expected in longer words. This may be due the relatively high frequency of some letters (such as E), contrary to our random assumption.

PALINDROMES

Consider a five-letter palindrome. We are free to choose any of the 26 letters for the first three letters, but the last two are then determined, so the possible number of five-letter palindromes is 26x26x26 = 17576. This same figure applies to six-letter palindromes; there are 26x26x26x26 seven- and eight-letter palindromes, and so on. If we wish to insist that

GRAPH 1 - FRACTION OCCUPANCY OF WORD SPACE



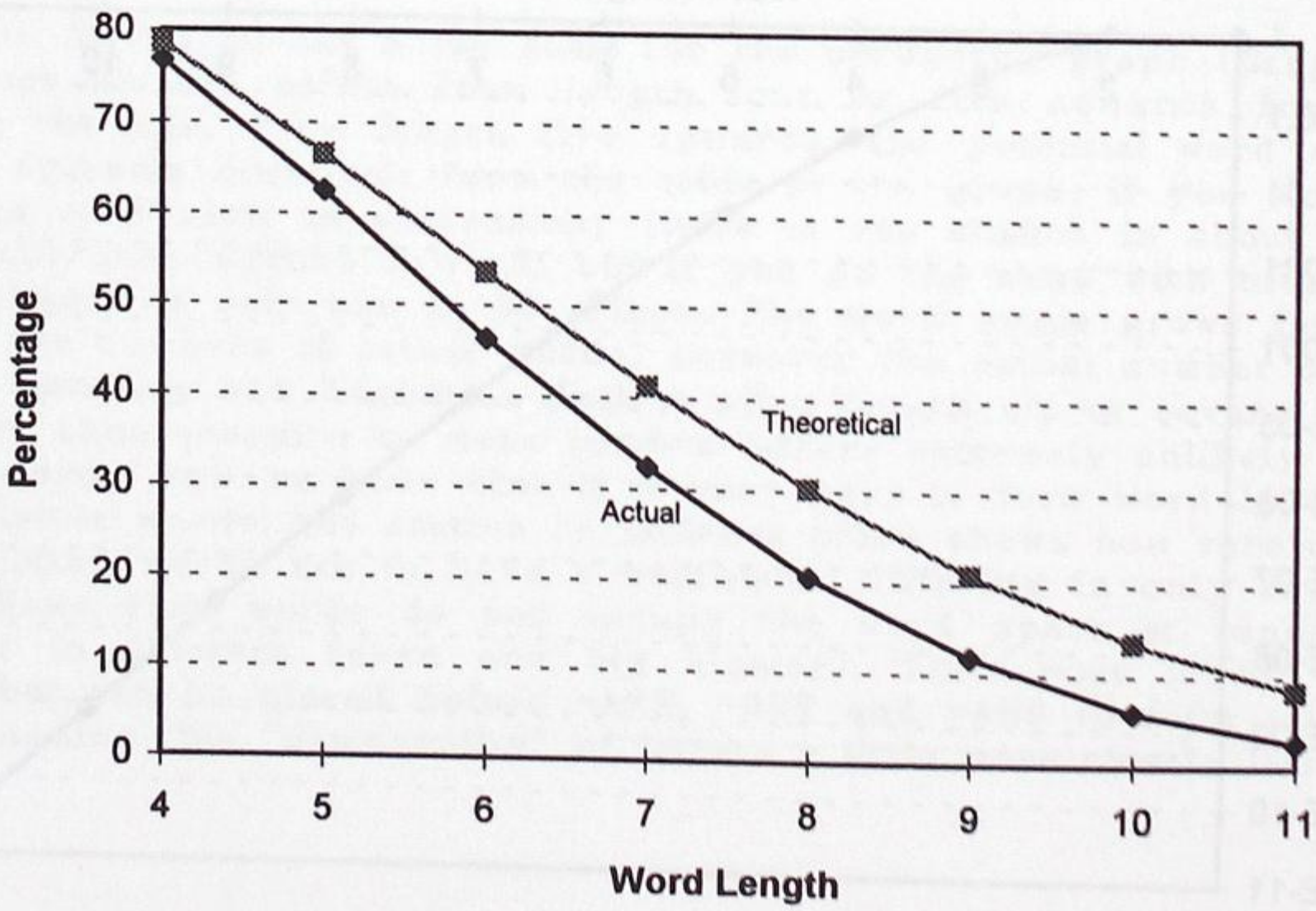
all letters are different (insofar as possible), then we might talk of a heterogrammatic palindrome or heteropalindrome. For five- and six-letter words, there could be 26x25x24 of these; for seven- and eight-letter words, 26x25x24x23.

Graph 3 (next page) shows the results of such calculations, along with actual values. For eight letters or more, the numbers are too small to show, the largest being 6. The heteropalindromes lie in the gap between the two lines: they are therefore easily the dominant type of palindrome. Although the shapes of the theoretical lines mimic reasonably the shapes of the actual lines, the actual number of palindromes (and heteropalindromes) is 2 or 3 times the predicted numbers for words of length four or five, 9 times for words of length seven, and a staggering 18 times for words of length six.

The predictions for the non-heteropalindromes are even worse at between 4 and 28 times higher in actuality than in theory, though we have the excuse that the numbers are very small. The largest set is the following: AASAA (Cooper), ABABA (NZ), AJAJA, ALALA, ANANA (F&W), ARARA, EELEE (F&W), HUH-UH, IGIGI (Web 3), IRIRI, LLULL (Catalan writer), OOLOO (F&W), OOPOO (Nobbs), OOROO (Thurber), SESES (=seizes), SSESS (EDD), SUSUS, ULULU, ZAZAZ (Cooper), ZZZZZ (wake-up service).

The frequency of words of the form AxAxA alone is enough to make nonsense of the theory. I have noted before the relatively high frequency of words with repetitive sounds. The same explanation also works for words of length six: ANA-ANA, ESSSSE, KAKKAK, MAM-MAM,

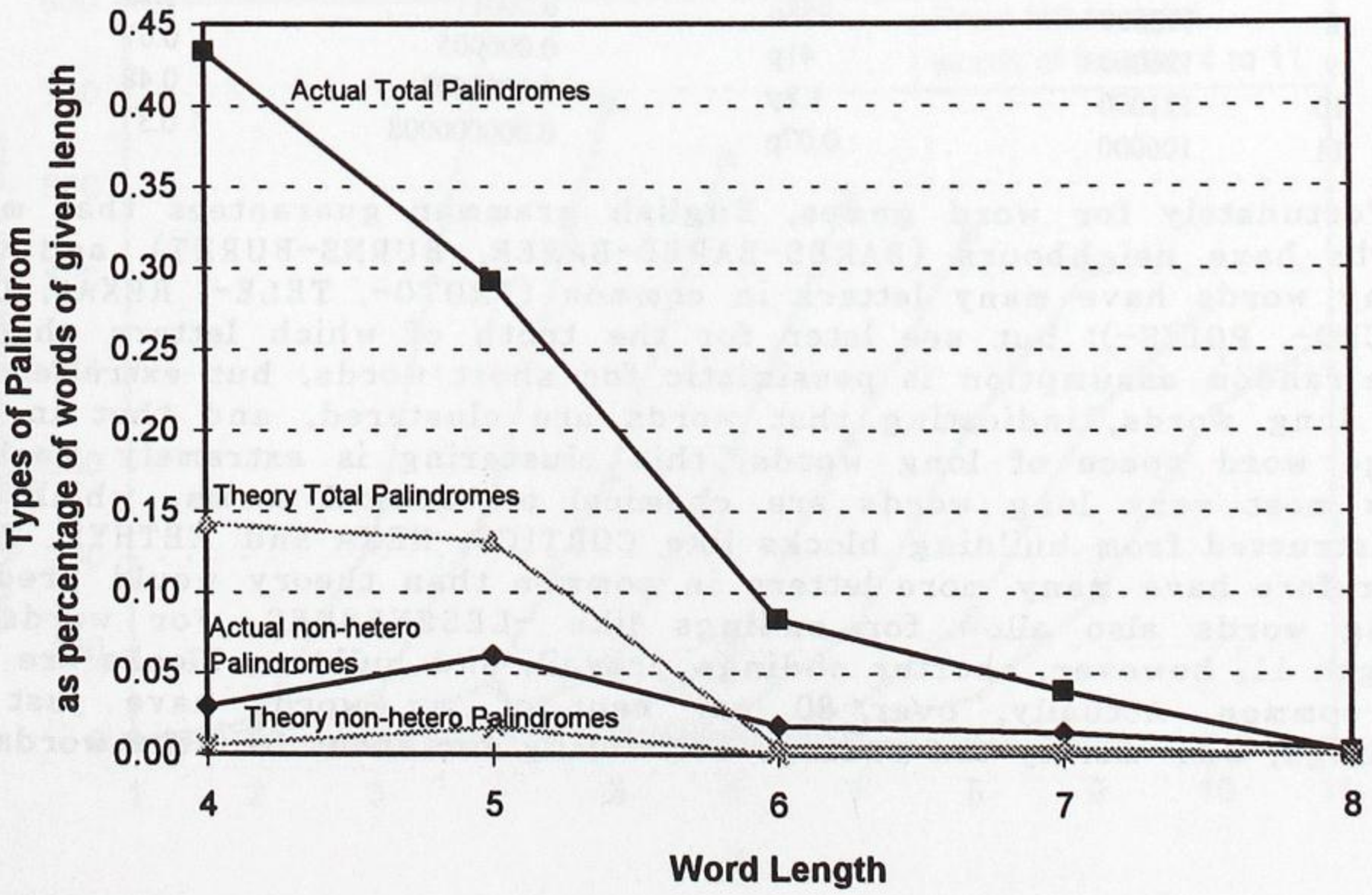
GRAPH 2 - HETEROGRAM FREQUENCIES



NAN-NAN, PIP-PIP, TAT-TAT, TUT-TUT (vf), WAW-WAW, WOW-WOW, and also works for words of length seven. The popularity of tautonyms explains the problem with the prediction for length six non-heteropalindromes.

Returning to the largest category of palindromes--those which have all letters different (as much as possible)--we should note firstly that the absolute numbers are still quite small, the most popular group containing only 96 words. Words of length four seem to have their own explanation: over one-third of them have just E or O as the second letter (PEEP, TOOT, etc.), with over half having one of the five vowels in that position. Somewhat fewer than half have a good variety of consonants as the second letter, typically with a vowel preceding (OPPO). So certainly the letters for four-letter words are chosen in a very selective fashion. Turning now to five- and six-letter words, it seems that common endings readily make words when preceded by their reversals: two-thirds of five-letter words and over half of six-letter heteropalindromes are accounted for by respectively 19 and 12 hetero-bigrams out of the 650 possible. Definitely nonrandom! The commonest of these endings for both lengths are -ES, -ER and -ET, though the commonest for five letters is -EN , and the commonest for six letters is -IT. Some of these are due to grammar (plurals, comparatives, past and present tenses). I now give one sample word for each of the commonest endings in order of their popularity: NEWEN, REFER, TENET, SEXES, KAYAK, DEWED, MARAM, RADAR, SOLOS, STOTS, WOROW; TIBBIT, REDDER, SELLES, TARRAT (EDD), TEBBET,

GRAPH 3 - PALINDROME FREQUENCIES



WORROW. Thus, although the predictions for palindromes are far too low, there are good reasons to be found in the nature of the English language.

To lend some kind of perspective, the actual number of palindromes in the graph is just under 300. Compare this to Jeff Grant's *Palindromicon*, which has about 3000. Unfortunately I could not do this study on those words because (apart from foreign words and proper names) I have no idea of the total number of different words in the 130 or so sources that he references.

COMMONALITY OF LETTERS IN WORDS

Start with a four-letter word. If words are made from letters at random, the chance of the first letter in another word being different is 25/26, and the chance of each of the other letters being the same is 1/26, so the chance of two words differing only in their first letters is 25 divided by 26x26x26x26. The chance of the two words differing in any one of the four letters is four times this, or 0.00022. If there are 18,000 four-letter words, each has on average four other words differing by just one letter, i.e. "neighbours". Here are the calculations for other word lengths, based on our hypothetical one million word vocabulary, compared with the findings from my word list.

Word Length	Estimated no words	Prob of word with 1 letter different	Calculated neighbours per word	Actual neighbours per word
4	18200	219µ	4	20
5	44800	11µ	0.47	10
6	74900	486000p	0.036	4.6
7	100000	2179p	0.0022	2.2
8	118000	958p	0.00011	1.15
9	126000	41p	0.000005	0.67
10	121000	1.8p	0.0000002	0.48
11	106000	0.07p	0.000000008	0.37

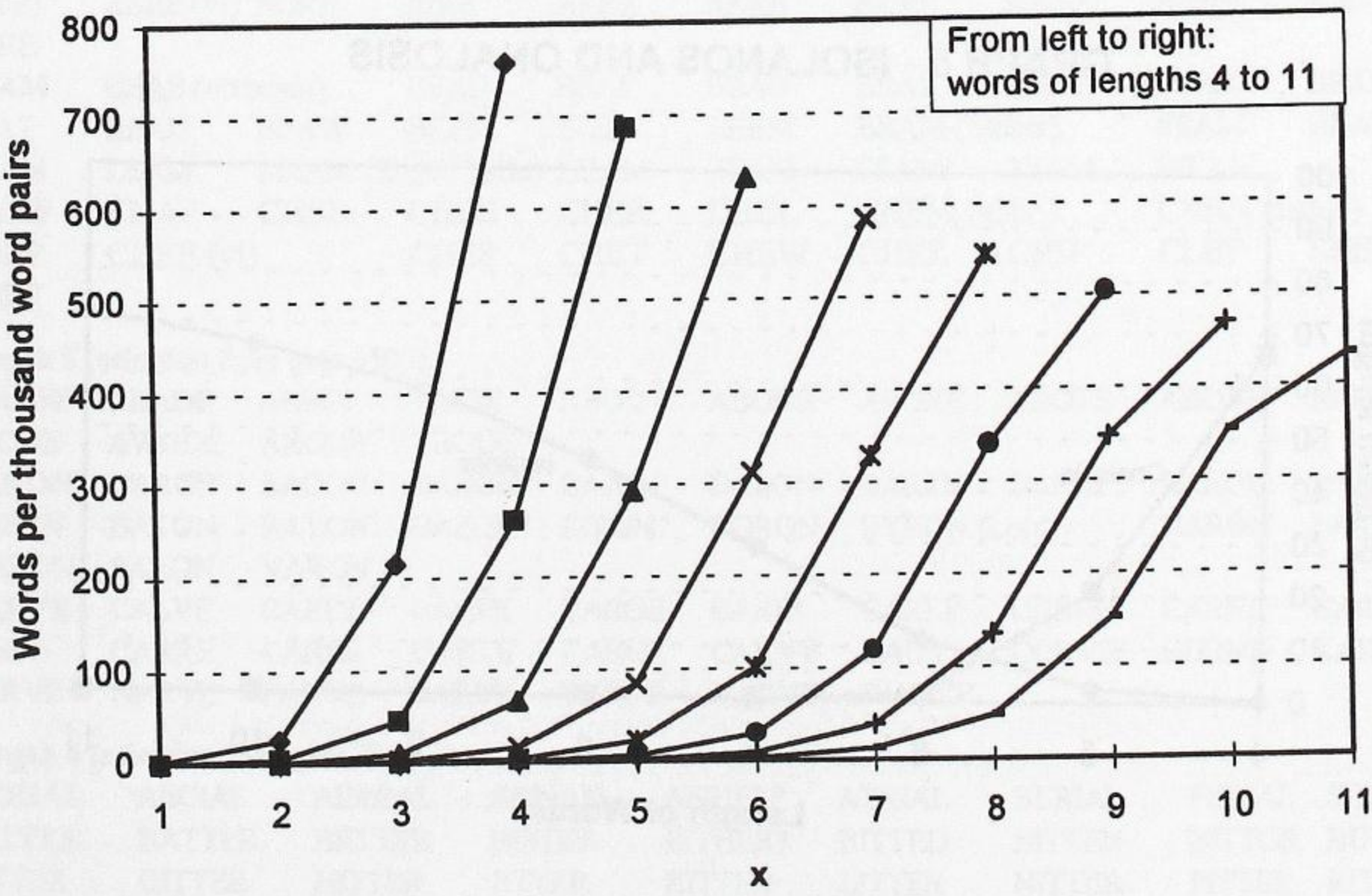
Fortunately for word games, English grammar guarantees that many verbs have neighbours (BARES-BARED-BARER, BURNS-BURNT), and that many words have many letters in common (PROTO-, TELE-, HEXA-, UN-, NECRO-, POLIS-); but see later for the truth of which letters change. The random assumption is pessimistic for short words, but extremely so for long words, indicating that words are clustered, and that in the huge word space of long words, this clustering is extremely marked. Now most very long words are chemical or medical terms, which are constructed from building blocks like CORTICO, HEXA and METHYL. They therefore have many more letters in common than theory would predict. Long words also allow for endings like -LESSNESSES. For words of length 11, however, shorter endings prevail, and building blocks are not so common. Actually, over 80 per cent of my words have just 54 endings, with merely ten endings accounting for about half the words.

-ED 10.1%, -ES (excluding -ITES,-ATES,-IZES) 7.9%, -IC 6.0%, -ING 5.9%, -ICS 4.0%, -ER 3.7%, -IA 3.4%, -ION 2.7%, -US 2.7%, -LY (excluding -ALLY) 2.5%

Note that my actual figures are based on a smaller vocabulary than the theoretical one; were the vocabulary larger, the gap between calculation and experience would be a little wider.

Here is a more general approach to propinquity of words in the word space. Imagine taking a set of words of the same length, say four letters. Compare each word to every other word. For each comparison, note how many letters of the word pair do not match. Average the results. If you had a vocabulary of, say, 15,000 four-letter words, you would have made 15000x14999 comparisons, so divide by this. For convenience, I show the results multiplied by 1000 in Graph 4 (below). For example, each four-letter word has on average 1.5 words with one non-matching position per thousand word pairs--"neighbours".Also, each four-letter word is shown as having 25 mates per thousand with two non-matching positions, 215 with three non-matching positions, and 758 with four non-matching positions (this last figure includes the effect of all isolanos). Note that each pair is counted twice. Now if each word is paired with 15000 others (say), then 15000x1.5 per thousand = 22.5 will be neighbours, suitable for ladder construction. For nine-letter words, only 0.007 per thousand have neighbours, and assuming 100,000 words, each has 0.7 neighbours on average. Among other things, these figures

GRAPH 4 - AVERAGE NUMBER OF WORDS WITH x NON-MATCHING POSITIONS

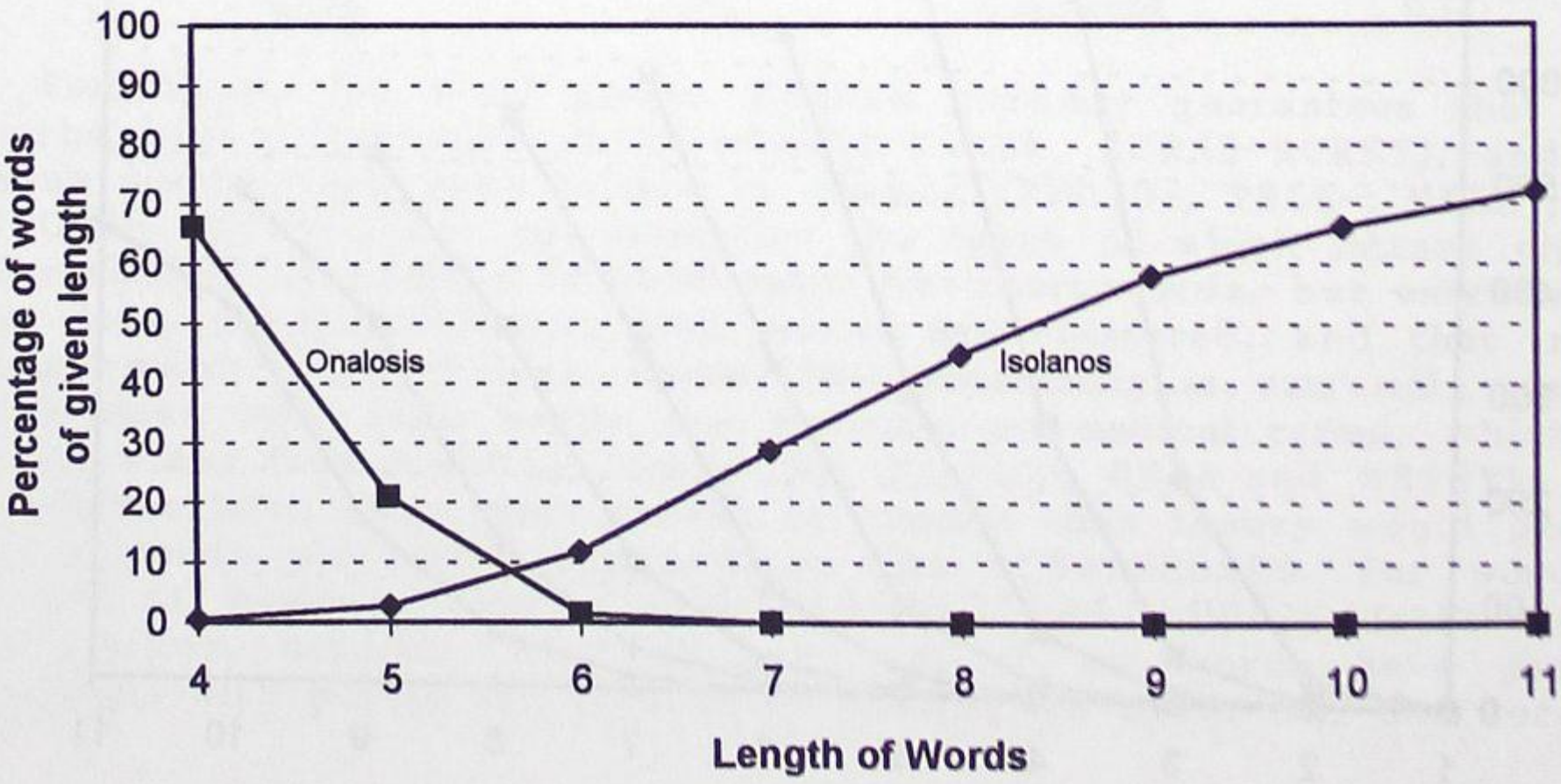


confirm the results reported in the previous paragraph. You may think it significant that there is fewer than one neighbour per word for words of length nine or more. Certainly, to make an ideal word ladder we have to find eight neighbours in succession for nine-letter words, compared with only three for four-letter words, and this makes such an enterprise most difficult.

ISOLANOS

Some words have no neighbours, and by 1971 these were commonly called isolanos, though they were first termed singularities by Rudolph Castown (Word Ways February 1969), who actually used the term isolano to describe a word not in the mainstream (from isolated Queen's pawn). The words that Castown called isolanos might be called terminal words, as they connect to only one other word. The only problem remaining is my tendency to use isolani as the plural! Graph 5 (below) shows how the number of isolanos increases inexorably with word length. There is scarcely a modern isolano of length four (though Dave Silverman conjectured there were 100, in the May 1970 Word Ways); however, there are likely many acronym-isolanos such as USMC, RSVP, GMBH. Isolanos of length five include many foreign words (TOKYO, SOYUZ) which would perhaps not be isolated if we included other words from the language concerned. (Another example is QAZAQ, an old-fashioned transliteration, which is not an isolano in Persian or some Inuit dialect.) Above length eight, isolanos are the norm. Beyond length nine, the combined effect of fewer words with a higher percentage of isolanos makes word ladders of a reasonable length very rare. In the August 1973 Word Ways, Eckler reported that the MWPD words of length sixteen were the shortest set containing 100 per cent isolanos.

GRAPH 5 - ISOLANOS AND ONALOSIS



Length 4: (selection from 66): AJJA (Jordan, TAW) AKOV ANKH AWFU AXXA (personal name, Paris) EJOO EPHA ESOX ICBM JUJU OGPU OMSK OPPO RIQQ (Beyond) SYBO UHUH UPPE YHWH YMPN YSYB ZIIM ZNAK ZYGA

Length 5 (selection from 856):

ALEPH	ATAXY	AWFUL	AXIAL	BIJOU	COYPU	DIAZO	DILDO	EBOLA
EDIFY	EPOXY	FLYBY	FOEHN	FOVEA	GAVOT	HAVOC	HOKKU	ICILY
JOEYS	JUGUM	KHMER	KIOSK	KLUTZ	LDOPA	LISZT	MIASM	MNEME
MYOMA	NGWEE	OMEGA	OOMPH	OPERA	ORCZY	ORTHO	OXLIP	PIZZA
PSALM	PSYCH	QAZAQ	RALPH	RIMPI	RIOJA	SYBIL	SYLPH	TISN'T
UH-HUH	UHLAN	ULTRA	USUAL	UZZLE	VEXIL	VODKA	VROUW	WALDO
WIZZO	XENYL	YEGGS	ZLOTY	ZOOID				

ONALOSIS (FRIENDLY WORDS)

The reverse, letter-wise, of the isolano is the onalosi. This term is used of words which can produce other words no matter which (single one) of their letters is changed. In the August 1973 Word Ways, Eckler expressed this by saying that onalosis have a garble in every position. Most four-letter words are onalosis, and Graph 5 shows how their occurrence decreases sharply as word length increases. I give a small selection of onalosis of lengths four to six, and all those of length seven I found. The onalosi is in boldface, and the supporting words (in excess) follow. I found no onalosis beyond length seven. In comparison, the longest Merriam-Webster Pocket Dictionary onalosi is SHORE or CANTER, and the longest one in Webster's Second is reported to be PASTERS. Tom Pulliam used the term friendly words for onalosis.

Length 4 (selection from over 8500):

ACRE	AARE	ACHE	ACKE	ACLE	ACME	ACNE	ACRY	ACSE	ACTE
(Latin)	AERE (vf)	AGRE	AIRE	AKRE	ARRE	AURE	AWRE	AYRE	ICRE
OCRE									
BEAM	BEAB (acronym)		BEAD	BEAF	BEAG	BEAK	BEAL	BEAN	BEAR
BEAT	BEAU	BEAW	BEEM	BEIM	BERM	BRAM (Stoker)		FEAM	HEAM
KEAM	LEAM	MEAM (Roget, Latin)		NEAM	REAM	SEAM	TEAM	WEAM	
CHEF	CHAF	CHEB	CHEE	CHEK	CHEL	CHEM (abbr.)		CHEN (name)	
CHEP	CHER (vf)		CHES	CHET	CHEW	CHEZ	CHIF	CLEF	SHEF
THEF									

Length 5 (selection from over 6500):

ABODE	ABADE	ABEDE	ABIDE	ABODY	ABONE	ABORE	ABOTE	ABOVE	ANODE
APODE	AWODE	AXODE	YBODE						
BARON	AARON	BACON	BADON?	BAJON	BALON	BAREN	BARIN	BAROI	BAROS?
BASON	BATON	BAYON	BAZON?	BIRON?	BORON	BYRON (Lord)		HARON	LARON
MARON	SARON	VARON							
CARVE	CALVE	CARDE	CARFE	CARGE	CARIE	CARLE	CARME	CARNE	CARPE
(Latin)	CARRE	CARSE	CARTE	CARUE	CARVY	CAUVE	CORVE	CURVE	KARVE
LARVE	NARVE	PARVE	SARVE	TARVE	VARVE	WARVE			

Length 6 (selection from over 750):

AERIAL	AECIAL	AEREAL	AERIAN	AERIEL?	ATRIAL	BERIAL	FERIAL	SERIAL
BITTER	BATTER	BETTER	BISTER	BITHER?	BITTED	BITTEN	BITTOR	BUTTER
FITTER	GITTER	HITTER	JITTER	KITTER	LITTER	NITTER	PITTER	RITTER
SITTER	TITTER	WITTER						

CHEESE	CHEASE	CHEEFE	CHEERE	CHEESA	CHEESY	CHEEVE	CHEISE	CHERSE
CHESSE	CHIESE	CLEESE (vf)						
CREESE	PHEESE	WHEESE						

Length 7 (all 9 found):

BRANDER	BLANDER	BRAIDER	BRANDED	BRANDEL?	BRANDUR	BRANNER	BRONDER
GRANDER							
DELATER	DEBATER	DELATED	DELATES	DELATOR	DELAYER	DELETER	DERATER
DEWATER	DILATER	RELATER					
GRANTER	GRAFTER	GRANDER	GRANFER	GRANGER	GRANTED	GRANTEE	GRANTOR
GRINTER	GRUNTER	GYANTER (W79-077)		TRANTER			
HARRIER	BARRIER	CARRIER	FARRIER	HAIRIER	HARDIER	HARPIER	HARRIAR
HARRIED	HARRIES	HARRIET	HARROER	HURRIER	MARRIER	PARRIER	TARRIER
MARRINE	CARRINE	MARLINE	MARRING	MARRISE	MARRITE	MARRONE	MARZINE
MAURINE?	MURRINE						
MILLINE	KILLINE	MIDLINE	MILLANE	MILLILE	MILLIME	MILLING	MILLINK
MILLITE	MILRINE	MILVINE	MISLINE	MOLLINE			
PASTERS	BASTERS	CASTERS	EASTERS	FASTERS	GASTERS	LASTERS	MASTERS
PAITERS (EDD)		PALTERS	PARTERS	PASSERS	PASTELS	PASTERN	PASTERY
PASTORS	PATTERS	PESTERS	POSTERS	RASTERS	TASTERS	WASTERS	
TAINTER	FAINTER	PAINTER	TAINDER	TAINTED	TAINTOR	TAISTER	TAUNTER
TAYNTER	TEINTER	TWINTER					
TERRANE	AERRANE?	FERRANE	TERPANE	TERRACE	TERRAGE	TERRANS	TERRENE
TERRINE	TERTANE	TETRANE	TIRRANE				

Beyond 7 letters, the maximum number of letters giving rise to neighbours in a given word is:

Length 8:	7 letters	CRAMPERS
Length 9	6 letters	MESSELINE, UNCREATED
Length 10	5 letters	CONFIDENCE, DELECTABLE, FORMALISES (-zes in OED), HOMOLOGIES, MENOSTASIS, NECROMANCY, PONOPHOBIA, UNBREACHED
Length 11	5 letters	SYNTHETISES (-zes in OED)

To give more feel for the kind of words that are onalosis, some others of length six are AERIAL, BADGER, BALDER, BANGER, BARGER, BEATER, BITTER, BRAZES, BULLET, BURDEN, BURGER, BUTTER, CALLED, CANKER...

MULTIPLE ONALOSIS

If an onalosi is a word in which every letter may be substituted by another, then if each letter can be replaced by two other letters, I call it a double onalosi, or two-fold onalosi, and similarly for triple or three-fold. Since there are no seven-letter onalosis, there are naturally no seven-letter double onalosis. If we accept all Jeff Grant's 676 two-letter words, then every two-letter word is a 25-fold onalosi! I found

-fold	2	3	4	5	6	7	8	9	10
4 letters	2418	1503	914	512	112	21	9	2	4
5 letters	1266	260	63	4					
6 letters	26	2							

10-fold 4-letter onalosis: SAIE,SAIT,SEIE,SEIT; 9-fold 4-letter onalosis: SEAT,SEET; 5-fold 4-letter onalosis: MOLES,PALLE,SAULE,SOULE; 3-fold 6-letter onalosis: COLLER,SANDER.

Examples of two-fold six-letter onalosis: BALDER, BURGER, CANTER, CASTER, COSTER, MANGER, MARINE, MUSTER, PORTER, SHARES, SHORES, WARDER, WASTER.

SAIT BAIT DAIT EAIT (vf) FAIT GAIT HAIT LAIT MAIT NAIT PAIT RAIT SAAT SACT
SAET SAFT SAHT SAIC SAID SAIE SAIF SAIH SAIK SAIL SAIM SAIN SAIP SAIR SAIS
SAIV SALT SANT SAPT (vf) SART SATT SAUT SAWT SAXT SEIT SHIT SKIT SLIT SMIT
SNIT SOIT SPIT SUIT SWIT TAIT VAIT WAIT YAIT

SEAT BEAT FEAT GEAT HEAT JEAT LEAT MEAT NEAT PEAT REAT SAAT SCAT SEAC SEAD
SEAH SEAK SEAL SEAM SEAN SEAR SEAS SEAU SEAW SEAX SEAY (vf) SECT SEET SEIT
SEKT SELT SENT SEPT SERT SEST SETT SEXT SEYT SHAT SKAT SLAT SNAT SPAT STAT
SWAT TEAT WEAT YEAT

SAULE BAULE CAULE FAULE GAULE HAULE MAULE NAULE PAULE SAALE
SABLE SALLE SAPLE SATLE SAUCE SAUFE SAULD SAULF SAULL SAULM
SAULS SAULT SAUME SAUNE SAUSE SAUTE SAUVE SCULE SHULE SOULE
SPULE STULE WAULE

COLLER (see below)

MARINE CARINE FARINE GARINE? KARINE (forename) LARINE MADINE MAGINE MALINE
MARANE MARICE MARINA MARINO MARITE MARONE MAXINE (forename) MORINE MURINE
NARINE PARINE WARINE

FRIENDLIER AND MULTIPLY FRIENDLIER WORDS

Tom Pulliam introduced a subset of onalosis called friendlier words, illustrated by CRIMP in the August 1992 Kickshaws. Replacing the first letter by P we get PRIMP, the second by H, CHIMP, then A, S and E, the essential feature being that the replacement letters also spell a word, PHASE. He thought that six-letter examples existed, and in fact Mary Lois Dennison found one: CANTER to PERNOD (November 1992 Kickshaws). There is also CANTER-PERIOD, BRACKS-CLINTY and PASTER-CONSOL. Because of the shortage of seven-letter onalosis, I could find no seven-letter friendlier words. A superior example would be one in which the two words are related.

The reader might well ask whether there exist words that are friendlier in more than one way. As the term friendliest (see below) is already used, perhaps we should call such words multiply friendlier (or promiscuous?). The answer is as one might suspect. The four-letter word BARE is friendlier in no fewer than 1922 ways (and this is not likely to be maximal); examples of words formed from the replacement letters include AULD, AUNT, AUTO, CELT, CENT, COCK, CODA, COIN, COLT... Some (BARE-YONI) are even apposite (not to mention lubricious!).

One would expect five-letter examples to be less friendly. I found 844 (only!) for PALES. Examples include BEGAN, BOSUN, BUILD, BURMA, CIGAR, COULD, COYLY, CUPID, DECAY, DIVAN, DOGMA, DUCAT...

Somewhat at random, I found 66 derivatives for the six-letter COLLER, including BABION, BEYDOM, GUIDON, HABION, HAAKON, HABEAS, HAYDAY, HEYDAY, KEYWAY, PAYDAY, PEYTON, REBEAT, SEAWAY, SUBWAY and TEUTON.

Friendliest words are a kind of reflexive or reversible friendlier word introduced by Peter Newby in the November 1992 Kickshaws. They are friendlier words (e.g., CAT-HOD) in which the target (HOD) can be converted back to the source (CAT) by the same mechanism: CAT-hat-cot-cad-HOD, HOD-cod-had-hot-CAT. In the May 1993 Kickshaws, Leonard Gordon suggests an even more complex sort of friendliest word.

WORDS WITH MOST NEIGHBOURS

The most neighbours that a word can have is 25 per letter, making 100 for four-letter words or 250 for ten-letter ones. (In the August 1973 Word Ways, Eckler said that the word in a network having the most neighbours had maximum ambiguity.) The attentive reader should by now expect far fewer than the maximum possible number of neighbours for the longer words. This is so: the most neighbourly four-letter word has over 14 neighbours per letter, whereas no eleven-letter word has as many as one neighbour per letter.

4 WARE (57), BARE (55), HARE,HERE,HOLE,SERE (54)
5 CARES,PALES (46), MANES,PARES (45)
6 COLLER (38), COSTER (32)
7 BETTERS,SEALING (28), BARLING,SEARING (27)
8 SLATTERS (20), STARLING (16), BATTERED,NOTATION (15)
9 SLATTERED (14), BATTERING,REVELLING (13), TETTERING (11)
10 SLATTERING (15), SLUTTERING,TONOGRAPHY,XENOPHOBIA (9)
11 MUSTINESS (9), CRYSTALLINE (8)

In 1979, Jeff Grant did some excellent work on this subject, though it was never published. The results for BETTERS, SLATTERS and SLATTERED below are his; the results for SLATTERING are nearly all his and WARE is the result of a combined operation. As there are only 25 possibilities for each letter, improving upon these results might not be too onerous. The actual neighbours for the most neighbourly words are:

4: WARE		AARE	BARE	CARE	DARE	EARE (vf)		FARE	GARE	HARE	IARE
KARE	LARE	MARE	NARE	OARE	PARE	RARE	SARE	TARE	VARE	WABE	WACE
WADE	WAFE	WAGE	WAIE	WAKE	WALE	WAME	WANE	WAPE (vf)		WARB	WARD
WARF	WARG	WARI	WARK	WARL	WARM	WARN	WARP	WARR	WARS	WART	WARY
WASE	WATE	WAUE?	WAVE	WAWE	WAXE (vf)		WAYE (vf)		WAZE (vf)		WERE
WIRE	WORE	WYRE	YARE								

4: BARE		AARE	BABE	BACE	BADE	BAGE	BAIE	BAKE	BALE	BAME	BANE
BARA	BARB	BARC (vf)		BARD	BARF	BARI	BARK	BARM	BARN	BARO	BARP?
BARR	BARS	BART	BARU	BARY	BASE	BATE	BAVE	BAWE	BAYE	BAZE	BERE
BIRE	BORE	BURE	BYRE	CARE	DARE	FARE	GARE	HARE	IARE	KARE	LARE
MARE	NARE	OARE	PARE	RARE	SARE	TARE	VARE	WARE	YARE		

5: CARES	BARES	CADES	CAFES	CAGES	CAKES	CALES	CAMES
CANES	CAPES	CARBS (abbr.)		CARDS	CARED	CAREK	CAREL
CAREN	CARER	CARET	CAREW	CAREX	CAREY?	CARKS	CARLS
CARNS	CARPS	CARRS (vf)	CARTS	CARUS	CARYS (vf)	CASES	CATES
CAVES	CAWES	CERES	CIRES	CORES	CURES	DARES	FARES
HARES	LARES	MARES	NARES	PARES	TARES	VARES	WARES

5: PALES	BALES	CALES	DALES	EALES (vf)	GALES	HALES	KALES
MALES	PACES	PAGES	PAIES (vf)	PALAS	PALEA	PALED	PALEN
PALEO (prefix)		PALER	PALET	PALEW	PALEY	PALIS	PALLS
PALMS	PALPS	PALUS	PALYS	PANES	PAPES	PARES	PASES
PATES	PAUES	PAVES	PAWES	PAXES	PAYES	PELES	PILES
POLES	PULES	RALES	SALES	TALES	VALES	WALES	YALES

6: COLLER	BOLLER	CALLER	CELLER	COALER	COBLER	COILER
COLDER	COLEER	COLIER	COLKER	COLLAR	COLLEM	COLLEN
COLLEP	COLLES	COLLET	COLLEY	COLLOR	COLLYR?	COLTER
COLVER	COLWER	COOLER	COULER	COWLER	COYLER	CULLER
FOLLER	GOLLER	HOLLER	KOLLER	LOLLER	MOLLER	POLLER
ROLLER	SOLLER	TOLLER	VOLLER			

7: SEALING	BEALING	DEALING	FEALING (vf)	GEALING (vf)	HEALING
MEALING	NEALING	PEALING	SCALING	SEAKING	SEALINE
SEALINK (Coy name)		SEAMING	SEANING (vf)	SEARING	SEASING
SEATING	SEAWING	SEELING	SEILING (vf)	SELLING	SETLING
SHALING	SPALING	STALING	SWALING	VEALING	YEALING

7: BETTERS	BATTERS	BEATERS	BEETERS	BELTERS	BENTERS
BESTERS	BETTEES	BETTELS (vf)	BETTERE (vf)	BETTORS	BEWTERS
BITTERS	BUTTERS	BYTTERS (vf)	DETTERS (vf)	FETTERS	GETTERS
IETTERS (vf)	JETTERS	LETTERS	METTERS	NETTERS	PETTERS
RETTERS	SETTERS	TETTERS	WETTERS	YETTERS	

8: SLATTERS	BLATTERS	CLATTERS	FLATTERS	PLATTERS	SCATTERS
SHATTERS	SKATTERS (vf)	SLAHTERS (vf)	SLAITERS (vf)	SLAUTERS (vf)	SLAWTERS (vf)
SLATHERS	SLATTERN	SLATTERY	SLITTERS	SLOTTERS	SMATTERS
SNATTERS	SPATTERS	SWATTERS			

8: STARLING	SCARLING	SHARLING	SNARLING	SPARLING	STABLING
STALLING	STAPLING	STARKING (vf)	STARLINK?	STARNING (inf'd)	STARRING
STARTING	STARVING	STAWLING	STERLING	STIRLING	

9: SLATTERED	BLATTERED	CLATTERED	FLATTERED	SCATTERED	SHATTERED
SKATTERED (vf)	SLATHERED	SLITTERED	SLOTTERED	SLUTTERED	SMATTERED
SNATTERED	SPATTERED	SWATTERED			

9: BATTERING	BANTERING	BARTERING	BATHERING	BATTELING	BATTENING
BETTERING	BITTERING	BUTTERING	MATTERING	NATTERING	PATTERING
TATTERING	YATTERING				

9: REVELLING	BEVELLING	DEVELLING (inf'd)		LEVELLING	NEVELLING
RAVELLING	REBELLING	REFELLING	REPELLING	RESELLING	RETELLING
REVEALING	RIVELLING	ROVELLING			

10: SLATTERING		BLATTERING	CLATTERING	FLATTERING	GLATTERING
SCATTERING	SHATTERING	SKATTERING (vf)		SLATHERING	SLITTERING
SLOTTERING	SLUTTERING	SMATTERING	SNATTERING	SPATTERING	SWATTERING

11: MUSTINESSES	DUSTINESSES	FUSTINESSES	GUSTINESSES	LUSTINESSES
MISTINESSES	MUSHINESSES	MUSKINESSES	MUSSINESSES	RUSTINESSES

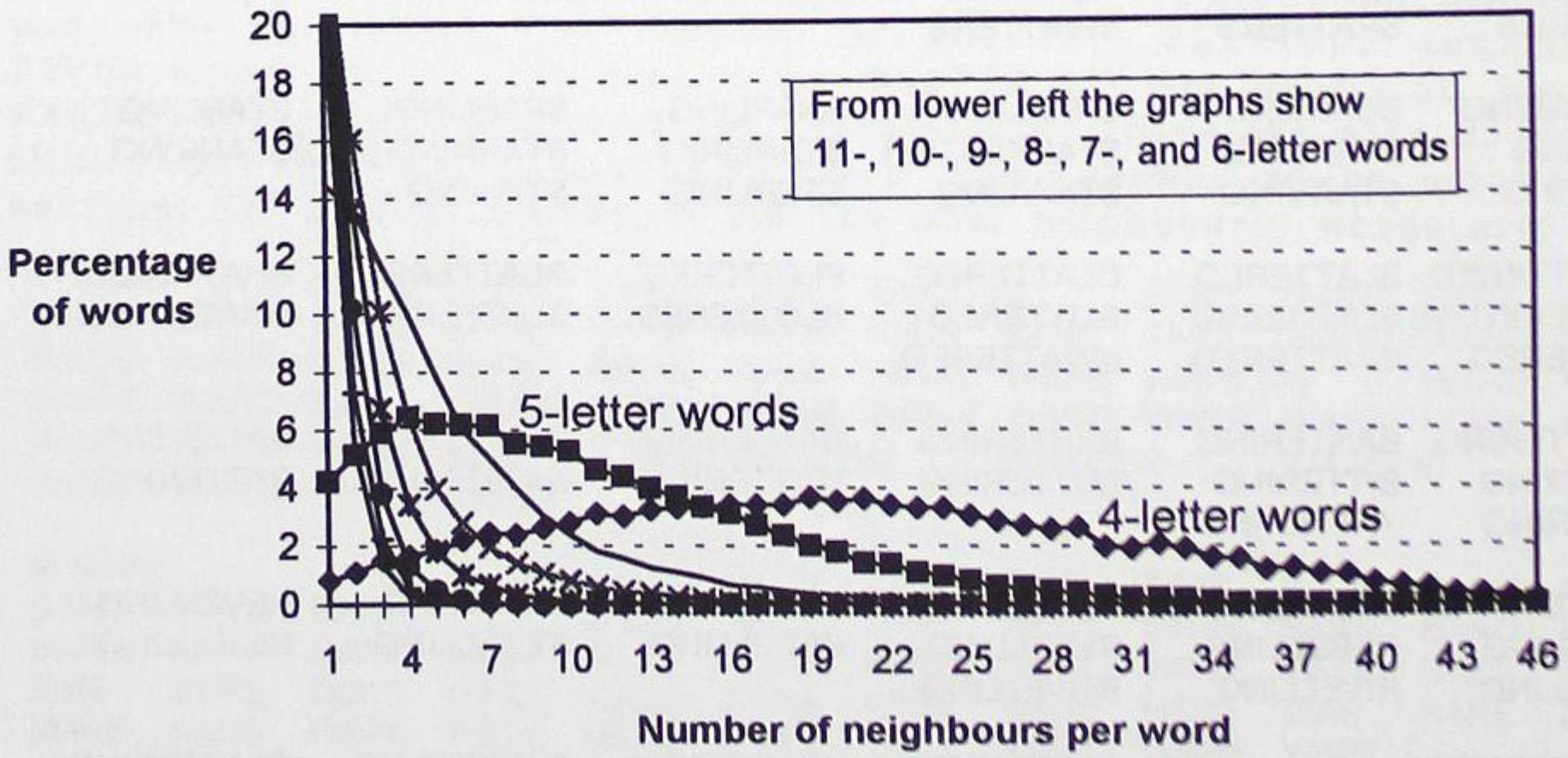
Graph 6 shows the overall picture: it is clear that four- and five-letter words have a neighborliness that differs in kind as well as sheer volume. They have relatively few cases of one, two and three neighbors simply because most words have many more!

English grammar leads to the terminal letter of many words easily giving rise to neighbours, e.g. BRAISES, BRAISER, BRAISED. It is also well-known that the first letter can change, as with BARES, CARES, DARES, FARES... The first letter change proves to be the most common, but the last letter change is not the second most common in the case of shorter words. The average number of neighbours for each position in a word is given in the table below.

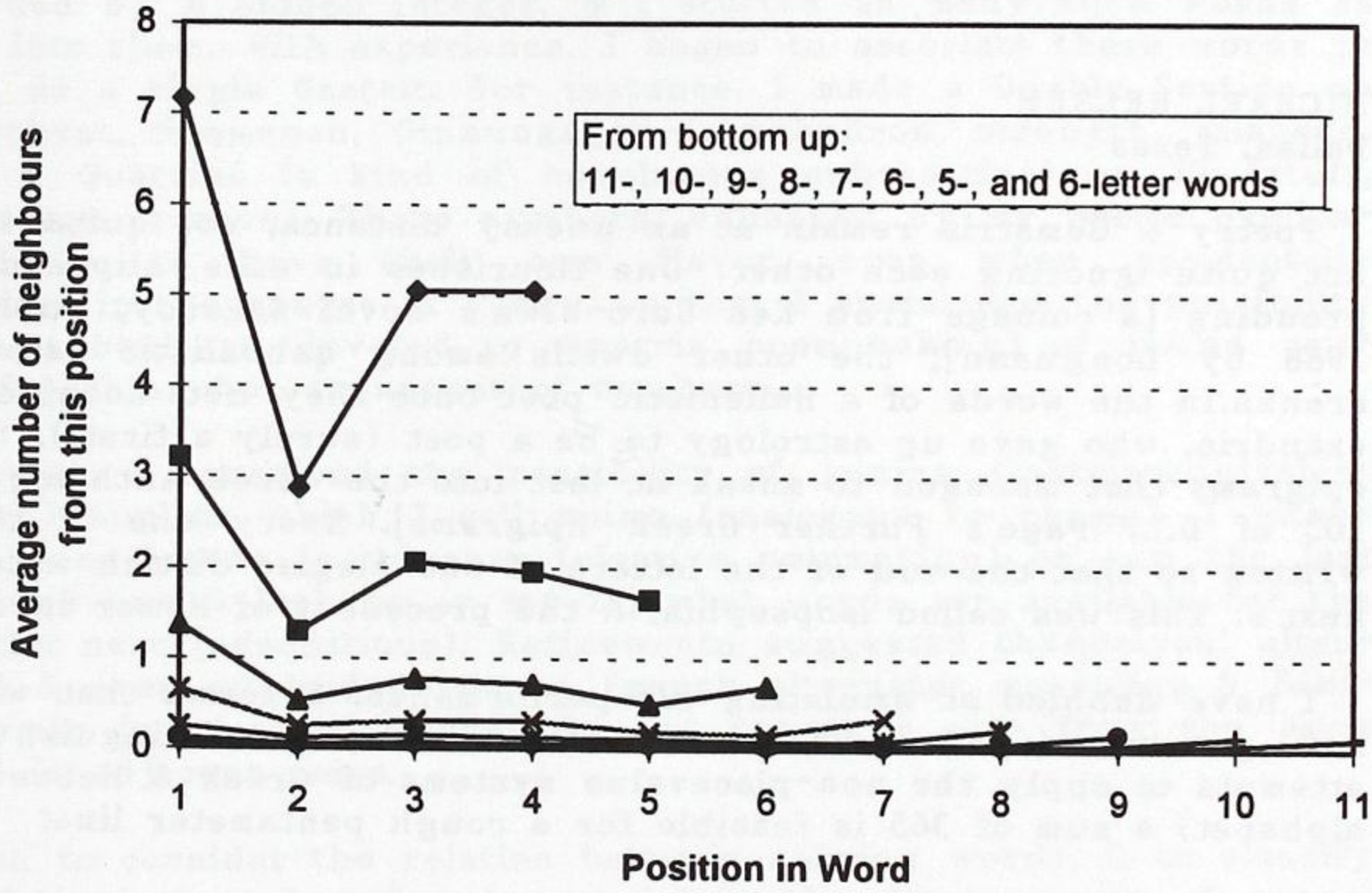
		Position in word										
		1	2	3	4	5	6	7	8	9	10	11	TOT
	4	7.152	2.868	5.044	5.043								20.106
	5	3.201	1.289	2.050	1.945	1.649							10.134
	6	1.363	0.531	0.768	0.713	0.499	0.700						4.574
Word	7	0.651	0.253	0.313	0.307	0.196	0.167	0.349					2.238
Length	8	0.249	0.111	0.152	0.126	0.116	0.085	0.090	0.224				1.153
	9	0.113	0.049	0.066	0.058	0.053	0.051	0.044	0.069	0.169			0.672
	10	0.064	0.029	0.040	0.033	0.030	0.026	0.029	0.034	0.055	0.144		0.482
	11	0.042	0.017	0.024	0.018	0.019	0.019	0.012	0.022	0.024	0.052	0.120	0.370

The ease of making four-letter word ladders with an average of 20 letters per word (more if we exclude isolanos) is evident, but beyond eight letters there is fewer than one neighbour per word, so it is not surprising that only one ideal ladder of nine letters or more has been reported.

GRAPH 6 - PERCENTAGE OF WORDS WITH 1, 2, 3... NEIGHBOURS



GRAPH 7a - NUMBER OF NEIGHBOURS GENERATED FROM POSITIONS 1, 2, 3... OF A WORD (PER WORD)



GRAPH 7b - NUMBER OF NEIGHBOURS GENERATED FROM POSITIONS 1, 2, 3... OF A WORD (PER WORD) MAGNIFIED