NOTES ON CODE WORDS

By WILLIAM F. FRIEDMAN, Chief of Signal Intelligence Service, War Department, and CHARLES J. MENDELSOHN, College of the City of New York

Introduction

The following notes are believed to be of interest in connection with the present regulations governing the construction of code words, as adopted in 1928 by the International Telegraph Conference of Brussels.[1] Briefly stated, two types of code language are permissible under the present regulations. We shall proceed first to a consideration of words of the first type, called *Category A* words, and conclude with a brief discussion of words of the second type, called *Category B* words, which are, it may as well be stated at once, of far less importance than the Category A words.

That part of the protocol adopted by the 1928 Telegraph Conference dealing with the requirements which words of Category A must fulfill reads as follows:[2]

Category A. Telegrams the text of which contains code words formed of a maximum of 10 letters and in which there is at least one vowel when they have a maximum of 5 letters, at least 2 vowels when they have 6, 7, or 8 letters, and at least 3 vowels when they have 9 or 10 letters. In words of more than 5 letters there must be at least one vowel in the first 5 letters, and at least one vowel in the remainder of the word, it being understood that words of 9 or 10 letters must contain at least a total of 3 vowels. The vowels are a, e, i, o, u, y.

A number of queries suggest themselves in connection with words of this category. In order to understand the origin and significance of these queries, a brief history[3] of code language as employed in international telegrams must first be presented.

It is well known that the basis of all or practically all modern cable and telegraph codes is the 5-letter code word. Through a loophole unforseen by those who participated in formulating the rules drawn up by the International Telegraph Conference of London, in 1903, it became possible to combine two 5-letter code words to form a 10-letter word chargeable as a single word in cablegrams, thus cutting the cost of messages in half. Code compilers and code users were very quick to find and take advantage of this loophole, with the result that

[1] The present regulations went into effect on October 1, 1929. That they are inadequate and that they have failed to accomplish the reforms intended is sufficiently attested to by the fact that proposals for their further modification will constitute one of the most important subjects on the agenda for the forthcoming International Telegraph Conference, which is scheduled to open in September of this year at Madrid.

[2] Code words of this type must conform to other requirements not mentioned herein, but they are of no interest in the present discussion.

[3] For a more detailed history see Friedman, William F., *The history of codes and code language, the international telegraph regulations pertaining thereto, and the bearing of this history on the Cortina report*, Government Printing Office, Washington, 1928.

within a short time 5-letter codes became very widespread and have by this time practically superseded all other types of codes in international cable and radio communications. Thus, while theoretically the 1928 regulations governing Category A code language are applicable to codes based upon words of a length up to 10 letters, in practice they are aimed at codes based upon 5-letter words which, as every code user now knows, are to be combined in pairs to form singly-charged-for 10-letter words. Now let us suppose that in a given code there are some 5-letter words with only one vowel. (With one or two exceptions, this is actually the case in the 5-letter codes constructed up to 1929.) It follows that in a certain number of cases there will be brought together, in the pairing of two 5-letter words to form a single 10-letter word, two words each containing a single vowel, resulting in the formation of a 10-letter word with only two vowels. According to the present regulations, such a word would have to be charged for as two words, thus increasing the cost of messages to an extent more or less dependent upon the number of code words containing only one vowel.[1] The only absolutely sure way of avoiding this source of surcharge is to arrange that *every* word in the code contain at least two vowels. Thus, while the regulations do not specifically state that each 5-letter code word of Category A must contain at least two vowels, it is clear that, in order to insure that the pairing of two 5-letter groups will in *no case* result in the formation of a 10-letter word with only two vowels, *each 5-letter word must contain at least two vowels*. The new codes (those published since October 1, 1929) take this indirectly imposed requirement into account.

Thus far we have been dealing with certain limitations imposed upon codes by the regulations themselves. We come now to the consideration of a very important limitation imposed upon modern codes as a result of practical difficulties inherent in telegraphic communication.

From the earliest days of codes based upon the 5-letter artificial word, it was recognized that some method or device is necessary whereby errors of transmission can be more or less automatically detected and corrected by the recipient of a message; otherwise communication by means of such artificial words becomes impractical. Of the methods elaborated for this purpose, that based upon the principle of including within a given code only such words as differ from one another in at least 2 letters has proved to be the most satisfactory, and has practically superseded all other methods. The 2-letter difference may consist in:

(1) A difference in the identity of 2 letters. For example, if the code contains the word ABABA, it must not contain any word differing from it in only one letter, such as ABABE, ABACA, ABEBA, ACABA, or EBABA. But words

[1] It is incorrect to assume that the number of 10-letter words that will be subject to the double charge is directly dependent upon the actual number of single-vowel, 5-letter words present in the code. Only an actual test made upon many bonafide messages, all prepared with the same code, can satisfactorily determine the number of doubly-charged-for words to be expected for that code, because the words and phrases of any code are employed with greatly varying frequencies.

differing from ABABA in at least two letters, such as ABACE, ABEBE, ACEBA, ECABA, ABECA, etc., are permissible.

(2) A difference in the position of 2 letters. For example, if the code contains the word ABECI, it may contain words such as EBACI, ACEBI, ABICE, BAECI, AEBCI, etc. In the early codes, no attempt was made to eliminate or to suppress one or the other member of a pair of words differing simply in the positions occupied by two adjacent letters. Many of them contained groups, such as ABECI and ABEIC, which are easily converted one into the other by a common type of psychological *lapsus calami* referred to in code work as "transposition." But in the better codes constructed up to about the year 1925, the authors have usually recognized the necessity of avoiding the possibility of errors introduced by a transposition of adjacent letters, and, to a very large extent, have succeeded in eliminating or almost completely suppressing this source of error. Something will be said of this later in these notes. In the more modern codes serious attempts have been made to eliminate errors due to transpositions of both adjacent and alternate letters, and it may be said that the problem is rather a difficult one.

(3) A difference in the identity of one letter and in the position of another. For example, if the code contains the word ABECK, the following would be legitimate words in the same code: ABERC, AERCK, etc.

In short, when at least two homologous letters in a pair of code groups differ in their identities, the code words are said to present a 2-letter difference.

In what follows we shall refer to *classes* and *subclasses* of words. By the designation *class*, we refer to a set of words merely by indicating the *number* of their constituent vowels and consonants. Using the symbol C to represent any of the 20 consonants, and the symbol V to represent any of the 6 vowels, the designation "$3V/2C$" merely indicates that words of this class contain 3 vowels and 2 consonants, without indicating the positions occupied by any of these constituent elements. Under each class, except the first and last, there are different subclasses of words with respect to the exact arrangement or *position* of the vowels and the consonants composing them. In the accompanying Table I there are shown the six classes and the 32 subclasses of words which can be constructed by taking vowels and consonants in groups of five letters.

If we wish to compose 5-letter words imposing no limitations other than that they must show a 1-letter difference, we may place any one of 26 letters in each of the five positions and shall thus have a total of $26 \times 26 \times 26 \times 26 \times 26 =$ 11,881,376 words. We may consider that we are here concerned with a simple multiplication involving five factors the numerical value of each of which is equal to the number of elements available for permutation.

If, however, we now impose the limitation that the words shall show a difference of two letters, the number of words that can be composed will be considerably reduced, because, in the parlance of code compilers, it is necessary to "sacrifice one of the letters." Just what this means will now be explained.

TABLE I

Class No.	Class Description	Subclass No.	Subclass Description
I	5*V*	1	*VVVVV*
II	4*V*/1*C*	2	*VVVVC*
		3	*VVVCV*
		4	*VVCVV*
		5	*VCVVV*
		6	*CVVVV*
III	3*V*/2*C*	7	*VVVCC*
		8	*VVCVC*
		9	*VCVVC*
		10	*CVVVC*
		11	*VVCCV*
		12	*VCVCV*
		13	*CVVCV*
		14	*VCCVV*
		15	*CVCVV*
		16	*CCVVV*
IV	2*V*/3*C*	17	*CCCVV*
		18	*CCVCV*
		19	*CVCCV*
		20	*VCCCV*
		21	*CCVVC*
		22	*CVCVC*
		23	*VCCVC*
		24	*CVVCC*
		25	*VCVCC*
		26	*VVCCC*
V	1*V*/4*C*	27	*VCCCC*
		28	*CVCCC*
		29	*CCVCC*
		30	*CCCVC*
		31	*CCCCV*
VI	5*C*	32	*CCCCC*

Consider for example, code words of two letters. Obviously, with an alphabet of 26 letters a total of 26×26, or 676, 1-letter difference pairs can be constructed. Such pairs will represent *all* the permutations of the 26 letters taken 2 at a time, such as AB, BA, AC, CA, etc. But if a 2-letter difference is desired, then only a total of 26 different pairs can be constructed, whether the pairs be simple doublets, such as AA, BB, · · · , ZZ, or permutations of 2 different letters, such as AB, BC, · · · , ZA. Thus, the formula for 1-letter difference pairs, 26^2, becomes modified to $26^{(2-1)}$ in order to take care of a 2-letter difference. Of the two factors in the multiplication 26×26, one will be reduced to unity, which is what the code compiler really means when he says that "one of the letters must be sacrificed." Similarly, in the case of 3-letter words, the total number of 1-letter difference words is 26^3 or 17,576 words; but if a 2-letter difference is desired the total number becomes reduced to $26^{(3-1)}=26^2=676$ words. For 4-letter groups with a 2-letter difference the number becomes 26^3 words; and for 5-letter groups with a 2-letter difference, 26^4 or 456,976 words. A general formula may be derived from the foregoing:

$$\text{Number of 2-letter difference words} = \lambda^{(n-d+1)},$$

where λ = the number of elements in the alphabet, n = the number of characters per word, d = the differential.

So far we have imposed no restrictions on the *kind* of letter that may occupy any given position in a word: any letter may be a vowel or a consonant as we happen to take it. If we restrict the number of vowels or the number of consonants allowed to a word, the case is altered. When we demand a consonant we are limited to 20 letters instead of 26, and when we demand a vowel, we are limited to 6 letters instead of 26. If we wish to form *all* the words that can be composed containing, say 1 consonant followed by 4 vowels, i.e., words of the form $CVVVV$, we shall have $20\times6\times6\times6\times6=25,920$ words each differing from all the others in at least one letter.

If, in addition to limiting ourselves to words consisting exclusively of one consonant followed by four vowels, we make the further demand that our words shall differ among themselves in at least two letters, we shall further lessen their number. We have already seen that the requirement of a 2-letter difference entails the sacrifice of one factor, or rather the reduction of one factor to unity. Just what this sacrifice will mean when we limit ourselves to a certain number of vowels and a certain number of consonants in each word will become plain as we proceed.

Query I

What is the theoretical maximum number of 5-letter code words that can be constructed with a minimum 2-letter difference, each code word containing a minimum of 2 vowels?

Solution

1. Table II shown below sets forth in brief form all the classes of 5-letter code words and the maximum number of 2-letter difference words in each class. Some explanation as to how the totals for each class of words are derived is added.

TABLE II

Based upon an Alphabet of 20 Consonants and 6 vowels

Description		*Maximum number of 2-letter difference words*
(1) Class I—With 5V	$=6\times6\times6\times6\times1=$ 1,296	= 1,296
(2) Class II—With 4V/1C	$=6\times6\times6\times6\times1=$ 1,296;	$1,296\times5=$ 6,480
(3) Class III—With 3V/2C	$=6\times6\times6\times20\times1=$ 4,320;	$4,320\times10=$ 43,200
(4) Class IV—With 2V/3C	$=6\times6\times20\times20\times1=$ 14,400;	$14,400\times10=$ 144,000
(5) Class V—With 1V/4C	$=6\times20\times20\times20\times1=$ 48,000;	$48,000\times5=$ 240,000
(6) Class VI—With 5C	$=20\times20\times20\times20\times1=$ 160,000	= 160,000

(1) In Class I (words composed of 5 vowels) the number of words is the same as if we were forming our words from a 6-letter alphabet. If we allow for a 2-letter difference, which means that we must reduce one of the five factors to unity, we shall, accordingly, have $6\times6\times6\times6\times1=1,296$ words.

(2) Class II consists of words composed of 1 consonant and 4 vowels. There are 5 subclasses according to the position occupied by the consonant. If we take the consonant as the first letter of each word, we can divide our 5 letters into two groups—one of 2 letters, CV, and one of 3 letters, VVV. If we use all the VVV groups, we shall have $6\times6\times6=216$ VVV groups. Each of these must now be combined with as many CV groups as possible to make 5-letter words. Each

of the CV groups with which any one VVV group is combined must differ from the other groups combined with the same VVV group in 2 letters. (We could not, for example, use both BAAAA and BEAAA.) Despite the availability of 20 consonants, since there are only 6 vowels in the alphabet we can form only 6 CV groups differing each from all the other CV groups in 2 letters. We can, for example, combine AAA with BA, CE, DI, FO, GU, HY, but with no more CV groups. Having now associated 6 CV groups with AAA, we can associate 6 others with AAE, say BE, CI, DO, FU, GY, HA, and so on with our remaining VVV groups. Thus we may obtain $6 \times 216 = 1{,}296$ groups of the type $CVVVV$. Since our single consonant can occupy any one of 5 positions in the word, the grand total of $1C/4V$ words will be $5 \times 1{,}296 = 6{,}480$.

(3) It is not possible to increase this number. It is true that we have used only 6 different consonants in all, but the remaining 14 consonants are of no value, for if we use them to replace the single consonant already used in the words we have formed, we shall obtain new words differing in only one letter from those we already have. If, for example, we have BAAAA, CAAAE, DAAAI, FAAAO, GAAAU, HAAAY, we cannot use JAAAA, KAAAE, LAAAI, MAAAO, NAAAU or PAAAY. The case is the same no matter what position in the word the single consonant occupies. And it is manifestly impossible to add to our words by additional vowel variations as those variations have already been exhausted.

(4) It is important to note in the foregoing explanation that the number of permutations of the complete set of 20 consonants suffers a serious reduction as a result of the association of the consonants with a much more limited number of vowels. It is clear, in fact, that if there were but six consonants available instead of 20 the total number of Class II words would still remain the same, 6,480. Fourteen consonants are wholly valueless in composing these words.

(5) Perhaps a concrete example employing a miniature alphabet will be useful in demonstrating this point conclusively. For this purpose, let us take an alphabet composed of three vowels, A, E, I, and of five consonants, B, C, D, F, G, to form 4-letter words with a 2-letter difference. If we require that all four letters be vowels, we shall be able to form, with three vowels, $3 \times 3 \times 3 \times 1 = 27$ 4-letter words with a 2-letter difference. If we set up a further limitation and demand that the first three letters be vowels and the last a consonant ($VVVC$), we can produce no more words than when all four letters were vowels, viz., 27, because, despite the availability of five different consonants, we are unable to make use of them all, as will now be shown.

(6) Let us assume that our $VVVC$ words are divided into two sections, VV and VC. For VV we have $3 \times 3 = 9$ possibilities. For VC we have $3 \times 5 = 15$, if we are forming only 1-letter difference words; but, as we have already seen, in order to obtain a 2-letter difference we must "sacrifice" one factor, that is, reduce it to unity. Now we cannot sacrifice the vowel of the VC section with its factor value of 3, and keep the consonant with its factor value of 5, because we are unable to use the full factor value of the consonant *unless the consonant has*

another letter with a factor value of 5 with which it can be permutatively associated. Since under the conditions of the problem this consonant must be associated with vowels, of which there are only three, our single consonant loses its factor value of 5 and becomes reduced to 3. It is now immaterial whether the factor that must be "sacrificed" in making the final calculation be considered as that pertaining to the vowel or to the consonant: the full number of words is $3\times3\times3\times1=27$. We may actually form a set of *VVVC* words under the conditions given:

AAAB	EAAC	IAAD	AEAC	EEAD	IEAB	AIAD	EIAB	IIAC
AAEC	EAED	IAEB	AEED	EEEB	IEEC	AIEB	EIEC	IIED
AAID	EAIB	IAIC	AEIB	EEIC	IEID	AIIC	EIID	IIIB

We have here our full quota of 27 words, and have not used the consonants F and G at all. We can use them if we will, but only to replace B, C, or D, and that will give us merely *alternative* words, not additional ones. (For example, we might have AIAG or IEAF, but each of these will show only a 1-letter difference from some words we already have.)

(7) In computing the total number of possible words with a 2-letter difference, composed of a mixture of vowels and consonants (always assuming that the consonants outnumber the vowels in the alphabet employed) if the words have one consonant each, the factor value of the set of consonants is exactly the same as that of the set of vowels with which the consonants must be associated; and then, as regards the factor which must be reduced to unity or "sacrificed," it is immaterial which is the one considered to have been sacrificed, that pertaining to the vowels or that pertaining to the consonants. If the words contain more than one consonant each, one of the factors pertaining to the consonant positions must be reduced to unity while all other factors will retain their full value.

(8) From the foregoing this general rule may be stated: In a set of unequal factors which are used as a basis for computing the total number of d-letter difference words that may be constructed, there must be at least d factors of equal and maximum value, and $(d-1)$ of them must be reduced to unity while all other factors retain their full value.

(9) In Class III, the $2C/3V$ class, we can form 10 subclasses, according to the positions occupied by the 2 consonants—*CCVVV, CVCVV, CVVCV, CVVVC, VCCVV, VCVCV, VCVVC, VVCCV, VVCVC, VVVCC*. It is obvious that each of these subclasses will yield the same number of words. Applying the general formula derived above to any of the foregoing ten subclasses, say the subclass *CCVVV*, we have $20\times1\times6\times6\times6=4320$ words for this subclass. Since there are 10 subclasses to each of which the same formula is applicable, we obtain a grand total of $10\times4{,}320=43{,}200$ words of the $2C/3V$ class.

(10) Proceeding with Class IV as with Class III, we find that there are again 10 subclasses, corresponding to the 10 possible positions of the consonants. Taking any subclass, for example, *CCCVV*, and applying the general formula

we have $20\times20\times1\times6\times6=14{,}400$ words. For 10 subclasses we have $10\times14{,}400 =144{,}000$ words.

(11) Similarly Class V ($4C/1V$) will yield $20\times20\times20\times1\times6=48{,}000$ words for each position of the single vowel, or a total of $5\times48{,}000=240{,}000$ words for the entire Class.

(12) Class VI ($5C$) will give $20\times20\times20\times20\times1=160{,}000$ words.

2. We are ready now to study Table II with a view to finding an answer to the question posited: What is the theoretical maximum number of 5-letter code words that can be constructed with a minimum 2-letter difference, each code word containing a minimum of two vowels?

3. Three hypotheses may be considered with respect to how we can obtain from Table II the maximum number of code words conforming to the foregoing specifications. They are:

1st hypothesis: The maximum number may be obtained by taking *all* the words of any one class.

2nd hypothesis: The maximum number may be obtained by taking *all* the words of two or more classes without introducing any "conflicts," i.e., without violating the 2-letter differential in the case of even a single pair of words.

3rd hypothesis: The maximum number may be obtained by taking words from two or more classes, the words being selected in such a way that no conflicts will be introduced.

4. Classes V and VI can immediately be eliminated from consideration, for they do not conform to the requirement that each word contain at least two vowels. There are left for consideration, therefore, only Classes I to IV, inclusive.

5. It is obvious that if two or more of the four classes remaining for consideration can be found to conform to the requirements of the second hypothesis, we shall obtain more words than can be obtained by an adherence to the first hypothesis. Let us begin, therefore, with the class which by itself gives the greatest total number of words, viz., Class IV. Is it possible to combine Class IV words with *all* the words of any of the other three classes, without introducing any conflicts? Let us try to combine Class IV with the class showing the next greatest total number of words, viz., Class III. Let us take subclass 23 of Class IV, of the form $VCCVC$ (see Table I); can we combine words of this subclass with words of subclass 9 of Class III, of the form $VCVVC$, without conflict? To be even more specific, can words of subclass 23, as exemplified in the word ABBAB be used without conflict with words of subclass 9, as exemplified in the word ABAAB? The answer must be in the negative, because as these two words now stand they show only a 1-letter difference. No matter what the specific constitution of any $VCCVC$ word, it is bound to conflict with some $VCVVC$ word, if *all* the words of both subclasses are taken. This is true with respect to all subclasses of Classes III and IV; hence these two classes, each taken in its entirety, cannot be used together without conflict.

6. Can words of Classes II and IV be used together without conflict? Fol-

lowing the same reasoning as in Paragraph 5, it will become apparent that they can. For, let us take any subclass of Class IV, subclass 23 for example, *VCCVC*, and combine it with any subclass of Class II, subclass 5 for example, *VCVVV*; no conflicts can arise because the third and the fifth positions in these subclasses will always be occupied in the one case by two consonants, in the other case, by two vowels. Take another example: combine subclass 3, *VVVCV*, with subclass 25, *VCVCC*; again no conflicts can arise because the second and fifth positions in these two subclasses will always be occupied in the one case by two vowels, in the other, by two consonants. This holds for all the subclasses of Classes II and IV when combined. Hence, the two classes can be used together without conflict, yielding a maximum total of 6,480+144,000=150,480 words.[1]

7. It is obvious that words of Class I cannot be included with words of Classes IV and II without producing conflicts, since words of Classes I and II will show conflicts between themselves.

8. The only other classes that might be employed together in their entireties are Classes I and III, but they will yield a total of only 44,496 words. If, then, the maximum number of words is to be obtained by the association of complete classes, it will be by employing Classes II and IV in complete series, yielding a total of 150,480 words.

9. We are, however, not justified in calling this number the *maximum* number possible unless and until we can dispose of the third hypothesis set forth in paragraph 3: can the maximum number be obtained by taking *some* words from one class and adding *some* words from the other classes, the words being selected in stch a way that no words will conflict with one another? Let us see.

10. We have already associated Classes II and IV. We cannot add to the total number of words by taking any words from Class I. All that any Class I words can do is to *replace* words of Class II, and that will not add to the total. As a matter of fact, it will decrease it. With Class III, however, the case is different. In forming words of Class II we use, or need use, only six different consonants. Just as the words of Class I are formed by associations of one or another of six different vowels in each of the five positions, so the words of Class II use the same six vowels in four of the five positions, and six different consonants in the remaining position. Let us assume that the six different consonants used are BCDFGH.

11. In the two positions in the words of Class III where consonants are used, *all* the consonants are employed. As a consequence we can add a certain number of words from Class III without conflicting with any of the words of Class II. Thus, we shall have in Class II, say, the word AEIOB. We can now take AEIJK, AEIKL, AEILM, AEIMN, etc., from Class III, and still preserve the 2-letter differential between the words of Class II and those of Class III.

12. There are fourteen consonants not employed in words of Class II. These

[1] Credit for the association of Class IV with Class II to produce the number 150,480 is due to Dr. L. H. Canfield, of the College of the City of New York.

14 consonants will form $6\times6\times6\times14=3,024$ combinations, which when multiplied by the ten possible positions of the two consonants in each word will yield 30,240 Class III words. All these words can be used with the complete series of words of Class II without interfering with the 2-letter differential.

13. We have, however, in addition to the complete set of words of Class II, used, in obtaining our 150,480 words, a complete set of words of Class IV. Before concluding that the 30,240 words just obtained can be added to our total, we must see whether they or any of them conflict with the Class IV words.

14. Both Class III and Class IV, at those places in their words where they employ consonants, use, for the complete set of words, the full number of twenty consonants. The only difference between a word of Class III and one of Class IV is that at one point where a word of Class III has a vowel a word of Class IV will have a consonant. It is therefore impossible to add words of Class III to a complete set of words of Class IV. We can *substitute* a word of Class III for one of Class IV, but this is an even exchange and will not increase our word total. If, then, we increase our word total by adding, to the words of Class II, non-conflicting words from Class III, we lose one word from Class IV for each word so added, and at the end we are exactly where we were before.

15. One possibility remains: it might be possible to omit *some* of the words of Class II or Class IV and substitute for each word so omitted more than one word from Class I or Class III. But if we replace any word of Class II by a word of Class I, we lose five words in Class II for each one that is taken for such replacement in Class I. And if we replace any word of Class II by a word of Class III, we shall gain ten words in Class III for every five so replaced in Class II—but we shall lose ten words of Class IV for every ten words taken from Class III. That is, for every ten words gained, 15 are lost as a result of the substitution.

16. It has been shown that we can obtain 150,480 words by using all the words of Class II and all those of Class IV. This total cannot be increased by adding words of Class I or Class III. Neither can it be increased by dropping some of the words of Class II or Class IV and adding a greater number of Class I or Class III. That is, it cannot be increased at all, and is the maximum number obtainable.

Query II

The 2-letter difference code words of modern 5-letter codes are usually compiled by reference to tables known under various designations, such as "Permutation Table," "Mutilation Chart," "Error Detector Chart," etc. An example of a typical table is shown in Table III below. We shall refer to such a table as a code-word construction table.

The code words afforded by the foregoing table are formed from the table by combining three elements: (1) a pair of letters from section 1, (2) a single letter from section 2, and (3) a pair of letters from section 3; and the combination must be made according to the rule that the initial pair and the middle letter must lie in the same vertical line (extended), the middle letter and the final pair must lie in the same horizontal line (extended).

Section 1

ac	ad	ae	af	ag	ah	ai	aj	ak	al	am	an	ao	ap	aq	ar	as	at	au	av	aw	ax	ay
bd	be	bf	bg	bh	bi	bj	bk	bl	bm	ba	bo	bp	bq	br	bs	bt	bu	bv	bw	bx	by	bz
ce	cf	cg	ch	ci	cj	ck	cl	cm	ca	cb	cp	cq	cr	cs	ct	cu	cv	cw	cx	cy	cz	cn
df	dg	dh	di	dj	dk	dl	dm	da	db	dc	dq	dr	ds	dt	du	dv	dw	dx	dy	dz	dn	do
eg	eh	ei	ej	ek	el	em	ea	eb	ec	ed	er	es	et	eu	ev	ew	ex	ey	ez	en	eo	ep
fh	fi	fj	fk	fl	fm	fa	fb	fc	fd	fe	fs	ft	fu	fv	fw	fx	fy	fz	fn	fo	fp	fq
gi	gj	gk	gl	gm	ga	gb	gc	gd	ge	gf	gt	gu	gv	gw	gx	gy	gz	gn	go	gp	gq	gr
hj	hk	hl	hm	ha	hb	hc	hd	he	hf	hg	hu	hv	hw	hx	hy	hz	hn	ho	hp	hq	hr	hs
ik	il	im	ia	ib	ic	id	ie	if	ig	ih	iv	iw	ix	iy	iz	in	io	ip	iq	ir	is	it
jl	jm	ja	jb	jc	jd	je	jf	jg	jh	ji	jw	jx	jy	jz	jn	jo	jp	jq	jr	js	jt	ju
km	ka	kb	kc	kd	ke	kf	kg	kh	ki	kj	kx	ky	kz	kn	ko	kp	kq	kr	ks	kt	ku	kv
la	lb	lc	ld	le	lf	lg	lh	li	lj	lk	ly	lz	ln	lo	lp	lq	lr	ls	lt	lu	lv	lw
mb	mc	md	me	mf	mg	mh	mi	mj	mk	ml	mz	mn	mo	mp	mq	mr	ms	mt	mu	mv	mw	mx
nq	nr	ns	nt	nu	nv	nw	nx	ny	nz	nn	nb	nc	nd	ne	nf	ng	ni	nh	nj	nk	nl	nm
or	os	ot	ou	ov	ow	ox	oy	oz	on	oo	oc	od	oe	of	og	oh	oj	oi	ok	ol	om	oa
ps	pt	pu	pv	pw	px	py	pz	pn	po	pp	pd	pe	pf	pg	ph	pi	pk	pj	pl	pm	pa	pb
qt	qu	qv	qw	qx	qy	qz	qn	qo	qp	qq	qe	qf	qg	qh	qi	qj	ql	qk	qm	qa	qb	qc
ru	rv	rw	rx	ry	rz	rn	ro	rp	rq	rr	rf	rg	rh	ri	rj	rk	rm	rl	ra	rb	rc	rd
sv	sw	sx	sy	sz	sn	so	sp	sq	sr	ss	sg	sh	si	sj	sk	sl	sa	sm	sb	sc	sd	se
tw	tx	ty	tz	tn	to	tp	tq	tr	ts	tt	th	ti	tj	tk	tl	tm	tb	ta	tc	td	te	tf
ux	uy	uz	un	uo	up	uq	ur	us	ut	uu	ui	uj	uk	ul	um	ua	uc	ub	ud	ue	uf	ug
vy	vz	vn	vo	vp	vq	vr	vs	vt	vu	vv	vj	vk	vl	vm	va	vb	vd	vc	ve	vf	vg	vh
wz	wn	wo	wp	wq	wr	ws	wt	wu	wv	ww	wk	wl	wm	wa	wb	wc	we	wd	wf	wg	wh	wi
xn	xo	xp	xq	xr	xs	xt	xu	xv	xw	xx	xl	xm	xa	xb	xc	xd	xf	xe	xg	xh	xi	xj
yo	yp	yq	yr	ys	yt	yu	yv	yw	yx	yy	ym	ya	yb	yc	yd	ye	yg	yf	yh	yi	yj	yk
zp	zq	zr	zs	zt	zu	zv	zw	zx	zy	zz	za	zb	zc	zd	ze	zf	zh	zg	zi	zj	zk	zl

d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a
f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b
g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c
h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d
i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e
j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i
n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j
o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k
p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l
q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m
r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n
s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y

Section 2

TABLE III

An Example of a Code Word Construction Table.

Reproduced by permission of the Code Compiling Co., Inc., New York, publis Universal Trade Code, 1921. All the cells in Section 1 and Section 3, many of which a the original table, have been filled in with the proper pairs of letters by the present meet the special requirements of this topic.

bb	cc	dd	ee	ff	gg	hh	ii	jj	kk	ll	mm	nz	on	po	qp	rq	sr	ts	ut	vu
ba	cb	dc	ed	fe	gf	hg	ih	ji	kj	lk	ml	ny	oz	pn	qo	rp	sq	tr	us	vt
bm	ca	db	ec	fd	ge	hf	ig	jh	ki	lj	mk	nx	oy	pz	qn	ro	sp	tq	ur	vs
bl	cm	da	eb	fc	gd	he	if	jg	kh	li	mj	nw	ox	py	qz	rn	so	tp	uq	vr
bk	cl	dm	ea	fb	gc	hd	ie	jf	kg	lh	mi	nv	ow	px	qy	rz	sn	to	up	vq
bj	ck	dl	em	fa	gb	hc	id	je	kf	lg	mh	nu	ov	pw	qx	ry	sz	tn	uo	vp
bi	cj	dk	el	fm	ga	hb	ic	jd	ke	lf	mg	nt	ou	pv	qw	rx	sy	tz	un	vo
bh	ci	dj	ek	fl	gm	ha	ib	jc	kd	le	mf	ns	ot	pu	qv	rw	sx	ty	uz	vn
bg	ch	di	ej	fk	gl	hm	ia	jb	kc	ld	me	nr	os	pt	qu	rv	sw	tx	uy	vz
bf	cg	dh	ei	fj	gk	hl	im	ja	kb	lc	md	nq	or	ps	qt	ru	sv	tw	ux	vy
be	cf	dg	eh	fi	gj	hk	il	jm	ka	lb	mc	np	oq	pr	qs	rt	su	tv	uw	vx
bd	ce	df	eg	fh	gi	hj	ik	jl	km	la	mb	no	op	pq	qr	rs	st	tu	uv	vv
bc	cd	de	ef	fg	gh	hi	ij	jk	kl	lm	ma	nn	oo	pp	qq	rr	ss	tt	uu	vv
bn	co	dp	eq	fr	gs	ht	iu	jv	kw	lx	my	na	ob	pc	qd	re	sf	tg	uh	vi
bz	cn	do	ep	fq	gr	hs	it	ju	kv	lw	mx	nm	oa	pb	qc	rd	se	tf	ug	vl
by	cz	dn	eo	fp	gq	hr	is	jt	ku	lv	mw	nl	om	pa	qb	rc	sd	te	uf	v
bx	cy	dz	en	fo	gp	hq	ir	js	kt	lu	mv	nk	ol	pm	qa	rb	sc	td	ue	v
bw	cx	dy	ez	fn	go	hp	iq	jr	ks	lt	mu	nj	ok	pl	qm	ra	sb	tc	ud	v
bv	cw	dx	ey	fz	gn	ho	ip	jq	kr	ls	mt	nh	oi	pj	qk	rl	sm	ta	ub	v
bu	cv	dw	ex	fy	gz	hn	io	jp	kq	lr	ms	ni	oj	pk	ql	rm	sa	tb	uc	v
bt	cu	dv	ew	fx	gy	hz	in	jo	kp	lq	mr	ng	oh	pi	qj	rk	sl	tm	ua	v
bs	ct	du	ev	fw	gx	hy	iz	jn	ko	lp	mq	nf	og	ph	qi	rj	sk	tl	um	v
br	cs	dt	eu	fv	gw	hx	iy	jz	kn	lo	mp	ne	of	pg	qh	ri	sj	tk	ul	
bq	cr	ds	et	fu	gv	hw	ix	jy	kz	ln	mo	nd	oe	pf	qg	rh	si	tj	uk	
bp	cq	dr	es	ft	gu	hv	iw	jx	ky	lz	mn	nc	od	pe	qf	rg	sh	ti	uj	
bo	cp	dq	er	fs	gt	hu	iv	jw	kx	ly	mz	nb	oc	pd	qe	rf	sg	th	ui	

Section 3

Given a specific, completely-filled code-word construction table for constructing 5-letter words with a 2-letter difference, this question now arises: How many words are there of each of the six classes discussed under Query I?

Solution

1. The initial step in the solution to this query consists in determining by actual count the number of words of the $5V$ class in the table under investigation. Let us assume that in a certain table this was found to be 236 words.

2. Words of the class $4V/1C$ comprise five subclasses, according to the position of the consonant, (see Table I). Each of these subclasses will yield 1,296 2-letter difference words. Any word of the class $VVVVV$ will eliminate one word from each of the foregoing subclasses of $4V/1C$ words. For example, AEIOU would eliminate AEIOC, AEICU, AECOU, ACIOU, and CEIOU. If then, we use 236 five-vowel words, we shall have remaining $1,296-236=1,060$ words in each of the five subclasses comprising the class $4V/1C$, making a total of $5\times 1,060=5,300$ words.

3. Words of the class $3V/2C$ comprise ten subclasses, according to the positions of the two consonants (see Table I). Each of these subclasses will yield 4,320 2-letter difference words. Certain subclasses of class $4V/1C$ will conflict with certain subclasses of class $3V/2C$. For example, words of subclass 2, such as AEIOR, will conflict with words of subclass 7, such as AEIQR, as well as with words of subclass 8, such as AEQOR, of subclass 9, such as AQIOR, and of subclass 10, such as QEIOR. Thus, each word of subclass 2 will eliminate *four* words from class $3V/2C$ words. Since there are 1,060 words of subclass 2 ($VVVVC$) and 4,320 words in each of the subclasses 7, 8, 9 and 10, (see Table I) we shall have remaining in each of these four last-named subclasses the difference between 4,320 and 1,060, which is 3,260 words. Applying the same reasoning to all the subclasses of class II, we draw up the following table:

TABLE IV

Subclass	2—$VVVVC$ conflicts with	Subclass	3—$VVVCV$ conflicts with
"	7—$VVVCC$, with	"	7—$VVVCC$, with
"	8—$VVCVC$, with	"	11—$VVCCV$, with
"	9—$VCVVC$, and with	"	12—$VCVCV$, and with
"	10—$CVVVC$	"	13—$CVVCV$
Sublcass	4—$VVCVV$ conflicts with	Subclass	5—$VCVVV$ conflicts with
"	8—$VVCVC$, with	"	9—$VCVVC$, with
"	11—$VVCCV$, with	"	12—$VCVCV$, with
"	14—$VCCVV$, and with	"	14—$VCCVV$, and with
"	15—$CVCVV$	"	16—$CCVVV$

Subclass	6—$CVVVV$ conflicts with
"	10—$CVVVC$, with
"	13—$CVVCV$, with
"	15—$CVCVV$, and with
"	16—$CCVVV$

4. Each of the ten subclasses 7 to 16, inclusive, is here duplicated in conflicts; but we need, of course, to eliminate words only once, in order to dispose of the conflicts with retained words of subclasses 2 to 6, inclusive. Therefore, we shall have remaining in each of subclasses 7 to 16, inclusive, 4,320 − 1,060 or 3,260 words, and hence the total number remaining, after deduction for conflicts, for the whole of class III is 3,260×10 or 32,600 words.

5. If we continue this process with the remaining classes of words, the following final results are obtained, shown in condensed mathematical form:

TABLE V

Class	*Description*	*Calculation*		*No. of Words*
I	5*V*	(as found by actual count in a specific table)	=	236
II	4*V*/1*C* =	1,296 − 236 = 1,060; 1,060 × 5	=	5,300
III	3*V*/2*C* =	4,320 − 1,060 = 3,260; 3,260 × 10	=	32,600
IV	2*V*/3*C* =	14,400 − 3,260 = 11,140; 11,140 × 10	=	111,400
V	1*V*/4*C* =	48,000 − 11,140 = 36,860; 36,860 × 5	=	184,300
VI	5*C*	= 160,000 − 36,860 = 123,140		123,140
			Grand total	456,976

6. It will be noted that the grand total in the foregoing calculation checks, since $26^4 = 456{,}976$. In similar calculations applicable to any other code-word construction table, while the individual totals may vary (according to the number of 5*V* words formed by the table), the grand total for a completely filled construction table must obviously be the same as that obtained above, viz., 456,976 words.

Query III

Experience has shown that if a code contain two such code words as, for example, ABCDE and BACDE, confusion may arise from the accidental transposition (in writing or telegraphing) of the letters A and B. It has accordingly been found advisable to construct codes so that no two code words differing from each other merely in the transposition of two adjacent letters will be included in the same code. A code-word construction table affording code words which will show no transpositions of adjacent letters can, however, be made when the number of different letters, λ, used in its construction is odd. The English alphabet contains 26 letters. To drop one letter in order to make λ odd would reduce the total number of words available. We may, however, add an extra character to the alphabet, giving $(\lambda+1)$ characters, construct a table without transpositions, and then eliminate all words containing the extra character. This will leave only words containing the λ letters, and these will contain no transpositions. We shall, however, lose a certain number of words that can be made from λ letters. How many shall we lose?

Solution

1. We may experiment with the miniature $(\lambda+1)$ table below, where the extra character, added to the alphabet to make the λ letter alphabet a $(\lambda+1)$ letter alphabet, is represented by the asterisk:

TABLE VI

	Section I														
	AA	AB	AC	AD	AE	AF	A*								
	BB	BC	BD	BE	BF	B*	BA								
1st and	CC	CD	CE	CF	C*	CA	CB								
2nd	DD	DE	DF	D*	DA	DB	DC				(λ=6)				
letters	EE	EF	E*	EA	EB	EC	ED								
	FF	F*	FA	FB	FC	FD	FE								
	**	*A	*B	*C	*D	*E	*F								
	A	B	C	D	E	F	*	AA	BB	CC	DD	EE	FF	**	
	B	C	D	E	F	*	A	BA	CB	DC	ED	FE	*F	A*	
3rd	C	D	E	F	*	A	B	CA	DB	EC	FD	*E	AF	B*	Section
letter	D	E	F	*	A	B	C	DA	EB	FC	*D	AE	BF	C*	III
	E	F	*	A	B	C	D	EA	FB	*C	AD	BE	CF	D*	
	F	*	A	B	C	D	E	FA	*B	AC	BD	CE	DF	E*	
	*	A	B	C	D	E	F	*A	AB	BC	CD	DE	EF	F*	
	Section II							4th and 5th letters							

2. Examination of this miniature table of $(\lambda+1)$ characters, where λ is 6, will show that it can yield words containing only the original λ letters[1] as follows:

(1) One set of words having $\lambda^2+\lambda(\lambda-1)(\lambda-1)$ words
(2) $(\lambda-1)$ sets of words each having $\lambda(\lambda-1)+(\lambda-1)(\lambda-1)(\lambda-1)$ words
(3) One set of words having $\lambda(\lambda-1)(\lambda-1)$ words

Adding all words we have:

$$\lambda^4-\lambda^3+\lambda^2-\lambda+1 \text{ or } \lambda^4-(\lambda^2+1)(\lambda-1)$$

3. Now λ letters arranged in a construction table will give λ^4 2-letter difference words. Hence, by adding the extra character to the table and eliminating words in which the extra character appears, we shall lose $(\lambda^2+1)(\lambda-1)$ words.

4. Accordingly, if $\lambda=26$, we shall lose from the complete table for 26 letters $(26^2+1)25=677\times25=16{,}925$ words. This leaves $456{,}976-16{,}925=440{,}051$ words.

5. It may be interesting to know how many words will be eliminated by the process described, from the complete table based on $(\lambda+1)$ letters. This table will yield $(\lambda+1)^4$ words, which equals $\lambda^4+4\lambda^3+6\lambda^2+4\lambda+1$ words. As already shown, if we omit words containing the extra character, we shall have $\lambda^4-\lambda^3+\lambda^2-\lambda+1$ words. Hence

[1] One special case must be considered. The added letter will appear twice in each column of Section I and twice in each row of Section III, except that there will be one column and one row respectively where it will appear only once. Should the arrangement of the code-word construction table as a whole be such that this one column and one row are not associated in forming words, one additional word will be lost. The deduction of the words from the total for λ letters will then be $(\lambda^2+1)(\lambda-1)+1$ and the deduction from the total for $\lambda+1$ letters will be $5\lambda^3+5\lambda^2+5\lambda+1$.

$$(\lambda + 1)^4 - (\lambda^4 - \lambda^3 + \lambda^2 - \lambda + 1) = 5\lambda^3 + 5\lambda^2 + 5\lambda$$

words will be eliminated.

For 27 characters the net total would therefore be:

$$27^4 - [5(27^3) + 5(27^2) + 5(27)] = 531,441 - (87,880 + 3380 + 130)$$
$$= 531,441 - 91,390 = 440,051.$$

6. It may further be interesting to see what is gained by this process over the simpler method of constructing a table with $(\lambda - 1)$ letters. This table would give $(\lambda-1)^4 = \lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1$ words. The table containing $\lambda + 1$ letters gives, as we have seen, when words containing the extra letters are rejected, $\lambda^4 - (\lambda^2+1)(\lambda-1)$ words. Subtracting $\lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1$ from the latter quantity, we have $3\lambda^3 - 5\lambda^2 + 3\lambda$ as the difference between the respective numbers of words yielded by the two tables. If $\lambda = 26$, a table of $(\lambda - 1)$ or 25 letters will give 390,625 words, while a table of $(\lambda+1)$ or 27 letters, omitting words containing the extra letter, will give 440,051, as has been shown.

Note

So far as words of Category B are concerned, since no limitations are placed upon their composition by the present regulations, the total number of code words with a 2-letter difference available for code compilers and code users is 26^4 or 456,976 words. If nontransposability of adjacent letters referred to in the preceding section is taken into consideration in the elaboration of the construction table, this total becomes reduced to either 440,051 words or 390,625 words, depending upon the method selected.

CYCLOIDAL CURVES

By SOLOMON BILINSKY, Washington University

When one plane curve rolls upon another, every point fixed relative to the rolling curve and in its plane describes a new curve. In the particular instance where the fixed curve is a straight line and the rolling curve a circle every point on the circle describes an ordinary cycloid. Other particular instances are quite as well known. It is the purpose of this discussion, however, to treat the general case, and to demonstrate the simplicity and straightforwardness with which vector analysis provides a means for the study of such curves, which will in the general case be designated as cycloidal curves.[1]

We suppose that the curve Γ_1 rolls upon the fixed curve Γ and that the point P, fixed relative to Γ_1, describes a curve Γ_2. As Γ_1 rolls along Γ the tangents and

---

[1] A different treatment of this subject may be found in A. Mannheim, *Géométrie Descriptive*, 2nd edition (1886), pp. 175–180.