

INTERNATIONAL RADIOTELEGRAPH CONFERENCE OF WASHINGTON : 1927

DELEGATION OF THE UNITED STATES OF AMERICA
COMMITTEE ON TARIFFS, WORD COUNT, AND ACCOUNTING

REPORT ON
THE HISTORY OF THE USE OF CODES AND CODE LANGUAGE,
THE INTERNATIONAL TELEGRAPH REGULATIONS
PERTAINING THERETO, AND THE BEARING
OF THIS HISTORY ON THE
CORTINA REPORT

BY
MAJOR WILLIAM F. FRIEDMAN, SIG.-RES.
Technical Adviser, United States Delegation; Chief of Code and Cipher Section,
Office of the Chief Signal Officer, War Department

454



UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON
1928

THE HISTORY OF THE USE OF CODES AND CODE LANGUAGE, INTERNATIONAL TELEGRAPH REGULATIONS PERTAINING THERE TO, AND THE BEARING OF THIS HISTORY ON THE CORTINA REPORT

I. INTRODUCTION

THE CORTINA REPORT

In 1925 the International Telegraph Union held a conference in Paris. Among the many important questions discussed at that time were those relative to the possibility of making certain modifications in the regulations governing the use of code language in international telegrams, in order to bring an end to a situation that was full of difficulties for everybody in general but especially for the telegraph administrations and companies the world over. The conference not being able to come to any decision on this matter, a subcommittee consisting of fifteen delegates was appointed to go into the subject carefully, and to draw up recommendations which were to be submitted to the next international conference, either telegraph or radiotelegraph. The Subcommittee of Study of Code Language held a few preliminary meetings in Paris before that conference adjourned, and in 1926 met at Cortina d'Ampezzo, Italy, where, for almost one month, it studied the subject indicated. It then drew up a report containing its recommendations, this report being termed herein the Cortina report.

This report contains two sets of recommendations; one set, that of the majority, presents the views of, and was subscribed to by fourteen delegates; the other, that of the minority, presents the views of the British delegates. In this paper we are going to be concerned with only a single sentence in the Cortina report, a sentence brief in its wording and substance but extremely large in its import.

The specific sentence of the Cortina report which will be considered in this paper is the first sentence of the majority proposal, on page xiii of that report. It reads as follows:

Making and counting of words. Code words must be formed of a maximum of five letters, chosen at the will of the sender, without any condition.

Following out this recommendation, there was incorporated in the draft of the proposed regulations drawn up in accordance with the majority opinion the following article:

ARTICLE 21, PARAGRAPH 4 (1)

In code language the maximum length of word is fixed at five letters. However, until . . . (a date to be determined later) words may consist of from six to ten letters. Such words are counted as two words.

One unfamiliar with the history of the use and development of code language in telegraphic correspondence can, at first glance, see nothing startling in the recommendation that "code words must be formed of a maximum of five letters, chosen at the will of the sender, without any condition." Yet, this proposal is so revolutionary in its nature, and so marked in its departure from the paths taken by all preceding conferences, that considerable doubt may well be entertained as to whether it will be adopted when the members of the Telegraph Union assemble to act upon it. In order, then, to appreciate the full import of this proposal and to understand all the circumstances which led to its formulation, it is necessary to enter into the history of the use of codes and code language, and the international regulations that have been applied thereto, from the birth of the International Telegraph Union up to the date of the last conference held by it, in 1925.

The difficulties in respect to the use of code language in which the Telegraph Union finds itself to-day, and to which the Cortina committee was intended to suggest a solution, commenced with the earliest days of telegraphy; and despite all efforts of the Union to put an end to them, these difficulties have continued for almost three quarters of a century, always in a more and more aggravated condition. In this paper it will be shown that the fundamental cause for these difficulties and the failure to remove them is that the regulations established by the Telegraph Union in respect to code language have not been, in the words of one of its oldest and most experienced members, "scientific, nor built upon a logical foundation."

DEFINITIONS OF VARIOUS TYPES OF TELEGRAPH LANGUAGES

For purposes of clarity in what is to follow, a few definitions are essential, these being in connection with the various types of language dealt with in telegraphy.

(1) The text of a piece of writing is said to be in *plain language* if it conveys an intelligible meaning in the language in which it is written.

Several implications not apparent on first reading are contained in this definition. In the first place, the text must be composed of *bona-fide* words, these must all belong to the language in question, and each one must have an intrinsic meaning. In the second place, the words must be written separately and must be arranged in a sequence determined by two things: first, the grammatical rules of

the language employed; second, the intelligence or information that is intended to be conveyed by the writing. In the third place, the phrase "an intelligible meaning" implies that there must be persons who understand the language in question, for, unless the language is understood, the mind can comprehend no such thing as a "meaning" in written characters. It does not follow from this last implication that a plain-language text must constitute intelligible text for, and be understood by, any and all persons. The intelligible meaning is, by definition, present in all plain-language texts, but this does not necessarily mean that the meaning or, indeed, even the possible existence of a meaning is obvious to, and comprehended by all persons. The comprehension of the meaning is within the compass of only such persons as are familiar with, or at least have some understanding of, the language in question, and of no others. Thus, a person who understands the English language can easily read plain-language text written, or claimed to be written, in the Spanish language because the latter is composed of alphabetic characters which have practically the same sound values in the English language; but he can grasp the meaning of what he reads (if a meaning is in reality present) only if he understands the Spanish language. Hence, plain language, which, by definition, must have an intelligible meaning, may, in fact, in many cases not be understood and may thus be unintelligible to persons who might be called upon to render a decision on the question of the presence or absence of intelligibility in a given text. We shall see later that the failure to recognize this important implication of the usual definition of plain language led to many serious difficulties in the telegraph industry.

(2) The text of a piece of writing is said to be in *secret language* if it conveys no intelligible meaning in any known language. Secret language may be divided into three fairly distinct classes, based upon differences in their external characteristics: (a) code language, (b) cipher language, and (c) figure language.¹

(3) *Code language* is that which is composed of (a) real words, that is, dictionary words having intrinsic meanings, but associated in sequences not conveying an intelligible meaning in any known language; or of (b) artificial "words", that is, assemblages of vowels and consonants in juxtapositions giving to such assemblages the appearance of being real words.

Obviously, since "words" of the latter type have no intrinsic meaning, their association in sequences cannot convey an intelligible meaning. Code-language telegrams are prepared by means of books termed *codes*, in which appear as separate items plain-language words,

¹ This classification is valid only in connection with the discussion of telegraph languages as set forth in this paper. A classification based upon technical differences in their internal or cryptographic constitution would go much further into detail, but this is here thought to be unnecessary.

phrases, and sentences listed according to some convenient sequence or arrangement (usually alphabetical), each item being assigned an arbitrarily selected code word which can be used to represent it in a message. In drawing up a code telegram by means of such a book one merely substitutes the plain-language words, phrases, or sentences of the plain-language text by the code words indicated in the code, building up the code telegram step by step and in as few code words as possible. It is obvious that correspondents must possess identical copies of the code employed in order to decode, that is, to translate promptly upon their receipt telegrams drawn up by its use. It is equally obvious that the letter constitution of a given code word bears no relation to the plain-language meaning of the word, and is not derived from the plain-language word: one code word of say seven letters may represent a plain-language word of four letters, another code word of seven letters may represent a whole plain-language sentence.

Code language is employed in telegrams for purposes of economy, or secrecy, or both. Economy is made possible by virtue of the condensation which code language affords, since one code word may represent a phrase of several plain-language words, or even a complete sentence. Secrecy is made possible by (a) restricting the distribution of the code book employed, as in the case of private codes compiled by individual firms, or (b) applying to the words of a code-language telegram prepared by means of a public code some system of changing or shifting about the meanings assigned to the code words in the code book.

(4) *Cipher language* is that which is composed of sequences or groups of letters in indiscriminate mixtures of vowels and consonants not at all presenting any appearance of real word formations. This type of language is produced in two principal ways: (a) by applying some system of cryptography to the individual letters of a plain-language text, either by substituting other letters for the original letters according to some cryptographic alphabet, or by rearranging the letters of the plain-language text so that they no longer present the appearance of real words nor yield intelligible text; (b) by applying a system of cryptography to the individual letters of a code-language text, according to either or both of the methods indicated under (a), for the purpose of giving additional secrecy to code telegrams. Cipher language of these two types is often referred to as *cipher language in secret letters*, or as *language in letters having a secret meaning*, in order to distinguish it from that type of language defined in the next paragraph.

(5) *Figure language* is that which is composed of sequences or groups of Arabic numerals. This type of language has two separate origins: (a) in some codes the code groups instead of being in letters,

are in figures, the mechanics of the code and the construction of a code telegram being exactly the same as in codes employing letter groups; (b) cryptographic systems in which the letters of a plain-language text are replaced by figures or groups of figures, according to a cryptographic alphabet. Figure language is often referred to as *cipher language in figures*, but the nomenclature used herein is thought to be shorter and equally descriptive.

(6) *Grouped language* is that which consists of plain-language words in groups composed of the reunions or combinations of two or more whole words, so grouped for the purpose of reducing the number of chargeable words. This type of language affords a saving only when the maximum length of admitted words is greater than the average length of separate words of the plain language employed.¹

(7) *Mixed language* is that which is composed of intermixtures in the same telegram of two or more of the languages defined above. Thus, in a single telegram there may appear parts in plain language, code language, cipher language, figure language, or grouped language, either in unbroken sequences, or in indiscriminate mixtures.

¹ For example, in telegraphic English the average length of word is approximately 5.9 letters; if the maximum length of admitted word is 10 letters, grouped language affords a considerable saving in charges. A unit of grouped language must consist of one or more whole words; fragmentation of words is not permitted. Thus, "HAVENO REPORTFROM HOME" is legitimate but "HAVENOREPO RTFROM HOME" is not accepted. Grouped language is now permitted in telegrams only between the United States, England, and Cuba; between Canada and England; and sometimes between Canada and Norway.

II. HISTORY

THE ORIGIN OF CODES AND THEIR INTRODUCTION INTO TELEGRAPHY

It is perfectly certain that secret language was employed in telegrams almost from the very day of the introduction of electro-magnetic telegraphy into public usage. This is by no means a remarkable circumstance, when it is realized that secret language and cryptographic methods are almost as old as written language itself. Hundreds, perhaps thousands, of years before telegraphy was born men communicated their thoughts by means of messages written in cipher; and as early as the fourteenth century diplomats in Italy were corresponding with each other by means of code books as well as by means of ciphers, and various types of secret language were well known in diplomatic correspondence. Naval communications by means of signaling pennants and flags, using code books for the purpose, were also well known. When in 1794 a system of rapid communication known as *aerial telegraphy*, employing semaphores on a series of high towers visible at considerable distances, was instituted and made rapid progress in Europe, men soon saw that economy as well as secrecy could be achieved by means of codes in which a whole phrase or sentence could be expressed in one group of signals. Thus we find that by 1825 codes employing figure groups were common, and a rather extensive code entitled *Telegraphic Vocabulary for the Line of Semaphoric Telegraphs between Liverpool and Holyhead* was published in London in 1845, a copy being on file in the Library of Congress. In this code there appear words, phrases, and even long sentences, each represented by groups of one to four digits.

This is not the place to enter into a discussion as to who invented the electric telegraph, and indeed, we may agree with the historian Sabine, who said: "The electric telegraph had, properly speaking, no inventor; it grew up . . . little by little." In England the earliest practical trial of electric telegraphy was made in 1837 on the London and Northwestern Railway, and the first public line under the patent of Wheatstone and Cooke was laid from Paddington to Slough on the Great Western Railway in 1843. In America, Joseph Henry in 1835, then a professor at Princeton University, demonstrated electric telegraphy, and in the same year Morse constructed his first rude working model. In 1837 Morse gave the first public exhibition of his telegraph, and in 1844, with the aid of a grant of \$30,000 from Congress, he established the first public telegraph line in America, that between Washington and Baltimore.

It is interesting and important to note that in Morse's original system, the one of 1835, he had not yet conceived the idea of employing the telegraph alphabet (of dots and dashes to represent letters of the alphabet) which now bears his name.¹ What he had was merely an apparatus which employed a set of ten symbols corresponding to the ten digits. The basis of his idea of communicating intelligence by means of these symbols was to be merely a special dictionary or rather a vocabulary in which were listed words serially numbered; in fact, Morse spent a long time in compiling such a vocabulary, and it was employed in his first public exhibition in 1837. But this vocabulary was dropped immediately after the improvement of the system made possible by the introduction of the telegraph alphabet of dots and dashes to represent letters of the alphabet.

In 1845 there was published, by one of Morse's early financial associates, the first code² addressing itself directly to the users of the electro-magnetic telegraph. As its title indicates, the object of this code was to afford secrecy; it contained merely words listed alphabetically, accompanied by groups of numbers in serial order to represent the words. Secrecy was to be achieved by adding to or subtracting from the numbers representing the words to be conveyed, an agreed-upon number; the numbers thus obtained were then to be written down and transmitted.

Although secrecy of communication has always been a factor of some importance in the conduct of commerce (and in these days of increasing competition in foreign trade it is becoming more so), yet rapidity, accuracy, and economy come first. As to economy, the people who quickly saw the advantages of telegraphic communication, the business men, soon realized that a considerable economy in the cost of telegraphic communication could be afforded by the use of codes specially adapted to their requirements. In this country, since the semaphoric telegraphs were not established, and were in all probability little known, the existence of codes of the nature of the *Telegraphic Vocabulary* referred to above was probably unknown, but the principle was so obvious that as early as 1848 American business firms were drawing up improvised codes for their own private usage. In Europe, where such codes were already in use in the semaphoric telegraphs, the introduction of codes for the electro-magnetic telegraph constituted merely a taking over into usage the existent codes from the semaphoric telegraphs.

We have noted that the semaphoric telegraph codes employed figure groups for their code symbols. But it was not very long before an entirely new principle began to be adopted in code communi-

¹ Some historians give the credit for the invention of this extremely important feature of telegraphy to Alfred Vail, Morse's first associate.
² *The Secret Corresponding Vocabulary: Adapted for Use to Morse's Electro-Magnetic Telegraph*, by Francis O. J. Smith, Portland, Maine, 1845.

cation by telegraph. This consisted in the use of real or dictionary words, but without their dictionary meanings, as code symbols. The reasons for this are not hard to understand. From the earliest days of electric telegraphy it has always been the case that telegrams in plain language have cost the users much less than telegrams in secret language, either those composed of sequences of letters in cipher language, or those composed of sequences of numbers in figure language, not because of differences in applied tariffs but by virtue of differences in the number of characters indicated as constituting a chargeable word in each of these languages, as set up by telegraphic regulations. Obviously, secret language occasions more difficulties in transmission and reception than plain language does, and therefore the admitted number of characters constituting a chargeable word in plain-language telegrams has practically always been twice that constituting a chargeable word in secret-language telegrams. Now so far as the mere mechanism of codes is concerned, dictionary words arranged in alphabetical order will serve as code symbols equally as well as figure groups arranged in serial order, and this, of course, soon became patent to the users of codes, who soon saw that there was no reason why they should pay twice as much for their telegrams as was really necessary.¹ Why not use real words as code symbols, instead of groups of figures? Furthermore, taking into consideration only groups of equal length, the number of words available for code purposes is far greater than the number of figure groups, since in the former case there are 26 letters available for permutations and combinations, whereas in the latter case there are only 10 digits. Of course the business man soon saw that the more expressions he could include in his code book, and the longer and more varied these expressions were, the greater was the degree of economy he could obtain by using a code. It is, therefore, not surprising to find that codes employing dictionary words for their code equivalents appeared in America certainly as early as 1860, for in that year a man named Buell published in Buffalo his *Mercantile Cypher for Condensing Telegrams*, in which English dictionary words were employed, and in which we find a fairly complete vocabulary arranged under captions, or most important words in mercantile phrases and expressions. Whether or not similar codes were published before this date it is impossible to say, for no record of them has been encountered, but there is every reason to believe that such codes, perhaps not in printed form, were improvised and used as early as 1850. The Buell code is, however, the earliest of the modern telegraph codes on file in the Library of Congress. Other and similar codes soon followed, most of them compiled and printed by business

¹ For many years, however, most governments continued to employ figure groups for their codes, and even to-day there are some who still use this most expensive form of telegraph correspondence.

houses, for their own use with their agents and correspondents. In 1874 appeared the first edition of one of the public codes destined to have a wide sale and consequently to go through many subsequent editions, the "A B C Code." Other public codes soon followed.

RELATION OF THE SUBMARINE CABLE INDUSTRY TO THE PROBLEM

Before setting forth the various provisions that were adopted from time to time by the successive international telegraph conferences for legalizing, regulating, and facilitating telegraphic correspondence in code language, it will be well to direct the attention of the reader to a very important circumstance which must be noted if one is to understand the causes underlying the difficulties we are going to discuss. It involves a consideration of the rôle played by submarine cables in international communication. It will be recalled that the first attempts to lay a cable across the Atlantic Ocean, those from 1857 to 1865, were unsuccessful; in 1866, however, two were finally laid and placed in operation. These did not last very long, but by 1877 there were four others in operation. There are now many cables spanning the oceans of the world, the great majority of them being owned not by governments as such, but by private companies. Now, for the most part, cables are only intermediate links in communication systems; at either end of any cable there must be thousands of interior stations which serve as feeders and distributors of traffic to and from it. Unless the rules and regulations governing communication are the same throughout the whole system, and are applied with a similar thoroughness, it is obvious that difficulties are bound to come. Let us see, therefore, over how much of this communication system the international convention and regulations were applicable, restricting our attention only to communications between this country and other countries, especially those of Europe.

Now, the United States is not now and never has been a signatory to the international telegraph convention; as a consequence, the rules and regulations drawn up from time to time by the various international conferences that have been held in Europe since 1865 have never been applicable to the internal telegraph systems of this country. The only time that American business men come in contact with the international regulations governing telegraphy is when their telegrams become cablegrams and travel overseas to foreign countries.¹

¹ In the case of radiograms, however, the provisions of the international radiotelegraph convention and regulations are applicable to the correspondence of the American public, since the United States is a signatory to this convention. It is to be noted that the rules relative to languages and word count, as established by international telegraph regulations, are taken over bodily into international radiotelegraph regulations. Hence, while the domestic or internal wire-telegraph business of the American public is not governed by international telegraph regulations (only the overseas business being so governed), all the radiotelegraph business, except so far as concerns point-to-point traffic, of the American public is governed by international telegraph regulations, since in respect to the phases indicated, the telegraph and the radiotelegraph regulations are identical.

It happens, at the same time, that the rules and regulations governing internal telegraph correspondence in this country have never been identical with international rules and regulations; consequently there were bound to be confusion and difficulties, especially since messages from inland destined to some European country, say from Chicago to London, have to be subjected to a word count totally different from that applied in the case of a message from Chicago to New York. For in the former case the cable count, which has always been identical with the international word count, was and is still applied, while in the latter case the domestic or internal telegraph count, different from the international word count, was and is applied. In some respects the domestic word count has always been more liberal in regard to secret language than the international word count, in other respects less liberal. Consequently, code users would naturally have an incentive to compile their codes and prepare their telegrams and cablegrams in such a manner as to gain the benefits of the liberal and avoid the extra costs of the less liberal features of both word-count methods—to the great detriment of the rules and regulations of both systems.

Furthermore, as regards the companies owning and operating cables, international rules and regulations become applicable only if and when these companies become signatories to the convention—and there were, and still are, some cable companies that have not signed. Now the tremendously large investments which cables entail make it essential that they be operated as continuously and as fully to capacity as possible. This means naturally that there must be a striving to obtain business, and under conditions of strenuous competition, practices which under other circumstances would not be tolerated are first admitted under protest, then become accepted without question, and finally are welcomed. Competition between cable companies, therefore, helped still further to break down the rules of the international telegraph regulations, for in all parts of the world, at cable terminals, what one cable company would not accept would be accepted by another; then the first company sooner or later, to survive competition, would have to follow suit.

In Europe, where the telegraph systems have been from very early days completely owned and operated by the respective governments concerned, agreements entered into internationally by a convention, supplemented by regulations, could be much more strictly enforced because of the absence of the severe competition which privately-owned systems and cables had to face. It is not strange, therefore, that infractions of agreements and regulations always occurred first in the privately-owned cable systems, and since the latter could not function alone, but necessarily often had to turn over to government-owned systems cablegrams for delivery to destination, the government systems became involved secondarily in these infractions. One additional important circumstance tending to reduce infractions of regula-

tions in European traffic is important to note. It is that as a general rule the proportion of code communication to plain-language communication in the whole of telegraph traffic is influenced very materially by the tariff applied, which, in turn, is largely determined by the length of the journey covered by the telegrams: Where the distance is great, and the cost of communication is correspondingly high, condensation of text is important, and since the greatest degree of condensation is afforded by the use of code, there we find codes extensively in use. It follows, therefore; that communication between European firms, for example, between two firms located in England, or between a firm located in London and one located in Paris, is not likely to be in code, except in cases where secrecy is required, as in bank correspondences. Therefore, there has never existed, and there still does not exist in European telegraph correspondence an incentive to violate or attempt to violate regulations with a view to reducing costs, such as has always been encountered in extra-European telegraph correspondence. Infractions of rules and regulations always led to attempts on the part of the nations constituting the Telegraph Union to correct conditions and prevent abuses by drawing up new regulations, only to find that the latter were no better than the ones they replaced. Especially as regards the regulations governing the languages permitted in international telegrams do we find change after change, and it is these changes to which we must now hasten to direct our attention.

INTERNATIONAL TELEGRAPH CONFERENCES FROM 1865 TO 1879

Although the question of regulation of languages that could be employed in telegrams and the differentiation in their basic telegraphic charges, was one of the most important subjects of discussion from the very beginning of the various conferences held in Europe in the early days, it was not until the third international conference, held at Rome in 1872, that even so much as a definition of what constitutes plain language was incorporated into the regulations of the international convention. For almost a quarter of a century, from 1849 to 1872, code-language words inserted in the text of international telegrams were charged for in practice as if they belonged to plain language. The only language other than plain language defined and provided for until 1872 was cipher language, which could consist of Arabic figures, or of letters having a secret meaning. These characters, with no limitations as to the length of groups, were charged for at the rate of five characters per word.¹

¹ It is interesting to note that in the 1868 Vienna conference, a proposal to make the grouping of cipher language obligatory was rejected, and that one of the important objections to it was that such a requirement would put restrictions upon the use of the telegraph facilities. It is also interesting to note that at this same conference the French delegation proposed that every dispatch composed in ordinary language, but unintelligible, be considered as a secret dispatch; and that one of the principal objections to this proposal was that every dispatch in a foreign language is unintelligible to the average operator who is not a linguist. A Portuguese proposal to create a universal language for international telegraphic communication was also rejected.

Although it has been stated above that only plain and cipher languages were provided for up until 1872, it is worth while calling attention to a rather feeble attempt that was made in the Vienna conference in 1868 to regulate code language. In the regulations drawn up by that conference there appear the following interesting items:

ARTICLE 9 (of the convention)

1. Every dispatch can be written in any one of the languages used in the territories of the contracting states, or in Latin.

6. Dispatches which are not admitted as ordinary dispatches, under the terms of the first paragraph of the present article, are considered as secret dispatches.

ARTICLE 37 (of the convention)

1. The word count for dispatches in figures, or in letters, or considered as secret dispatches is established in the following manner: (the method described is equivalent to counting every five characters as one word).

ARTICLE 5 (of the service regulations)

7. The office of origin is the judge of the application of the last paragraph of Article 9 of the convention, notably as regards the tolerance to be accorded to correspondences which deal with stock market or commercial affairs.

These articles left to the discretion of each administration and each office of origin the decision as to whether those dispatches which, though written in dictionary words, yet conveyed no intelligible meaning to the telegraph offices in correspondence—that is, code-language telegrams—were to be considered as secret dispatches. It is obvious that such an indefinite system was bound to lead to many irregularities and differences in application of the rules. Some offices would naturally be very lenient, others would be very strict. Furthermore, natural competition between companies was bound to break down such a system, for as soon as one company decided to interpret the rule in a very broad manner, so that code-language messages were not considered as secret dispatches, and were thus free from the heavy cost of the latter, other companies had necessarily to follow suit. It is, therefore, not surprising that in the very next conference the indefinite phrase “considered as secret dispatches” was completely modified.

During the 1872 conference considerable thought and discussion was devoted to code language. Since code-language words could be selected from more than fifty plain languages a proposal was made to limit the number of plain languages to three—English, French, and German—and to accept others without any responsibility for errors in transmission. This proposal was, however, rejected. One delegation insisted upon the establishment of a very precise distinction between dispatches written in code language and those composed of figures or secret letters. This delegation stated that code language had come to be accepted as plain language, and that

some provisions for code language should be made. Other delegations insisted that code language ought to be considered as secret language and charged for as such, but the cable companies strenuously objected because of the large proportion of code-language messages they handled for commercial houses. We shall now see how the matter resulted: Code language was charged for at the same rate as plain language.

For at this conference a definition of plain language and a whole new article was incorporated in the convention. It read:

ARTICLE 9 (of the convention)

Dispatches in plain language must offer an intelligible meaning in any one of the languages used in the territories of the contracting States, or in Latin.

Each state designates among the languages used in its territories those which it considers proper for international telegraphic correspondence.

The following are considered as dispatches in secret language:

1. Those which contain a text in figures or in secret letters;
2. Those which include series or groups of figures or letters, the commercial significance of which is not known to the office of origin;
3. Dispatches containing passages in code language, incomprehensible for the communicating offices, or of words not at all corresponding to the languages mentioned in the first paragraph of the present article.

But so far as the charge was concerned, words of code language were counted exactly the same as plain-language words, provided that they were words belonging to one of the languages authorized for international telegraphic use. For Article 37 of the convention read as follows:

In dispatches in secret language, the address, the signature, and that part of the text in ordinary or code language are counted in accordance with the preceding article (plain-language rules).

For the parts of the text composed either in figures or in secret letters, or in a language not admitted under the terms of Article 9, the count of words is . . . (the five-character count here followed).

Thus, if code-language words belonged to a language authorized for plain-language correspondence they were charged for at the same rate as plain-language words; if they did not, they were charged for as cipher words, that is, five characters per word. Since there were at that time, and there still are, over fifty languages authorized for use in plain-language telegrams in international telegraphic correspondence, the limitation as to languages worked no hardships whatsoever upon code users.

Up until the St. Petersburg conference, in 1875, the maximum length of a word was fixed at seven syllables for all telegrams, plain language or code language (in dictionary words of the fifty or more authorized languages). Such a generous measure, coupled with the fact that code-language words were charged for as though they were plain-language words, was bound to lead to abuses. Thus we find

in an old telegraph manual complaints about these abuses and the following "fair specimens of code words in daily use at present":

APOGUMNOUMAI
 APOGUMNOSOMETHA
 EKMISTHOSOISTHEN
 CHINESISKSLUTNINGSDON (21 letters, but only
 6 syllables)

In 1875 the maximum length of words in the European régime¹ was fixed at 15 characters according to the Morse alphabet; and in the extra-European régime, at 10 characters. Still there were no new provisions made for code language, the provisions established in 1872 being retained practically unchanged. Therefore, no curb to abuses was as yet interposed and things went along about as they had gone until then—that is, pretty badly, for reasons which will presently be given in detail.

In 1879, at the London conference, code language was deemed of sufficient importance to warrant the inclusion of the following definition and restrictions into the regulations:

ARTICLE 8

By code language is understood the use of words, which, while presenting an intrinsic meaning, do not at all form intelligible phrases for the corresponding offices.

These words are taken from the vocabularies admitted for international correspondence in code language but of which the composition varies according to whether the régime concerned is the European or the extra-European.

In the European régime telegrams in code language must contain only words belonging to one of the languages mentioned in paragraph 2 of Article 7 (those authorized by each State for correspondence in plain language). Every telegram must contain only words taken from one and the same language.

In the extra-European régime code-language telegrams can contain only words belonging to the German, English, Spanish, French, Italian, Dutch, Portuguese, or Latin languages. Every telegram can contain words taken from all of the aforementioned languages.

Proper names cannot enter into the composition of vocabularies. They are admitted in the text of telegrams in code language only with their plain-language meaning.

The office of origin can demand the production of the vocabulary in order to control the execution of the preceding regulations.

ARTICLE 9

The following are considered as telegrams in cipher language: (a) those which contain a text in figures or in secret letters; (b) those which include either series or groups of figures or letters, the significance of which is not known to the office of origin; or of words or names, or of groups of letters not complying with the conditions for plain language or code language.

¹ The European régime now includes the following countries: All the European states, the Azores, Algeria, Bokhara, the Canary Islands, Gibraltar, Malta, Morocco, Asiatic Russia and the Caucasus, Senegal, Mauretania, Tripoli, Transcaucasia, Tunis, and Asiatic Turkey.

Note the last clause in the last paragraph, that relative to "groups of letters not complying with the conditions for plain language or code language." What was the nature of these words, which, judging by the generosity of the Telegraph Union in respect to the number of languages admitted, we must infer were neither fish, flesh, nor fowl? Here lies an interesting story.

These exceptional, or we may say, outlawed and heavily surtaxed words were of two general classes, and of two different origins. We shall have to deal with them somewhat in detail, for it was these words that are at the bottom of the manifold difficulties in which the Telegraph Union finds itself to-day, and a knowledge of their natures and origins is essential to an understanding of the whole problem.

THE INADEQUACIES OF DICTIONARY WORDS AS CODE WORDS

We have seen that the majority of the early codes employed real or dictionary words for their code arbitraries. Now it often happened in drawing up a code telegram that the sender would not encode every word of the text, but would leave certain words in their unencoded or plain-language form, trusting that the recipient would recognize this fact and would understand that such words were intended to convey their ordinary, or plain-language meaning. But it often happened that the recipient did not recognize such words and proceeded to decode them. Similarly, there would happen cases wherein a word intended to be taken in its code meaning would yield no inconsistency when taken in its dictionary meaning. All this was due to the fact that in these early codes no careful selection of words was made; common words as well as some uncommon ones were included in the code lists, so that when the recipient proceeded to attempt to decode a word which the sender really meant to convey with its ordinary meaning, such a word was not infrequently found to be listed; and, on the other hand, the recipient might fail to decode a word which was intended to be decoded. By more or less rare coincidences, such words, when taken as code words, or not taken, might give meanings not inconsistent with the context of telegrams, and thus introduce errors which, passing unsuspected or unnoticed, might cause much confusion or serious financial loss.¹

¹ In a celebrated case fought through the lower courts and finally settled by the United States Supreme Court in 1894 (*Primrose vs. Western Union Telegraph Co.*, 154 U. S. 1), plaintiff brought suit to recover a heavy financial loss directly brought about by an error in a code telegram. The error consisted in the accidental addition in transmission of a single dot in one word; the code word "bay" was changed into "buy." The word "bay" meant in code "I have bought"; the word "buy" was read by the recipient in its dictionary meaning, and the sense of the telegram was changed from "I have bought 500,000 pounds" to "buy 500,000 pounds." The recipient of the telegram promptly proceeded to buy the indicated quantity (of wool), which later had to be sold at a loss of \$20,000. The most important part of the Supreme Court's ruling in upholding the adverse decision of the lower court reads as follows:

The conclusion is irresistible, that if there was negligence on the part of any of the defendant's servants, a jury would not have been warranted in finding that it was more than ordinary negligence; and that, upon principle and authority, the mistake was one for which the plaintiff, not having had the message repeated according to the terms printed upon the back thereof, and forming part of his contract with the company, could not receive more than the sum which he had paid for sending the single message.

The writer is indebted to Mr. H. F. Taff, General Supt., Eastern Division, Western Union Telegraph Co., for the reference to this citation.

Therefore, it soon became obvious to early code compilers that common words, words that might be used in telegrams in their ordinary, or dictionary sense, ought to be omitted from their code lists, and ought never be used as code arbitraries. In order to contain the number of words necessary for a rather extensive code, this meant that a very large percentage of uncommon, rare, and unusual, dictionary words would have to be introduced into codes for the code arbitraries. Now the fact that telegraph operators are not and can not be expected to be highly skilled linguists was bound to cause difficulties. Not only would they have to know or at least recognize tens of thousands of the most unusual words in their own language, but they would also have to recognize hundreds of thousands of unusual as well as common words in foreign languages. It was indeed too much to expect, and numerous errors were the result.

Code words based upon no scientific principles of construction are extremely susceptible to being mutilated in transmission, reception, or copying, for since they form no intelligible meaning in their sequence in code language, the context of a telegram could not, as in the case of a plain-language message, serve as a guide or control in making errors apparent, or in correcting them, if they should become apparent. Now, real words are of irregular lengths, and composition as regards the placement of vowels and consonants; very often two or more words present such slight differences from each other as to be most easily mutilated, without, however, such mutilation becoming evident on inspection. These differences can be considered under several headings: orthographic, telegraphic, and phonetic.

A. Orthographic errors

Examples of slight orthographic difference between words can be passed over very quickly. Words such as "jeering" and "peering", "morning" and "moaning", "iconical" and "ironical" differ from each other orthographically in but a single letter. Poor or careless handwriting easily leads to errors in such words. Sometimes defective writing can cause two adjacent letters to be mistaken, as, for example, in the case of "juniper" and "jumper."

B. Telegraphic errors

Telegraphic errors are of several types. Under each type thousands of examples could be cited, but we will show only a few.

(1) Errors due to change of a dot into a dash or *vice versa*:

(a) ABBRASSEN—ABBRAUSEN

Here S(•••) is changed into U(••—) by the mere lengthening of the last dot into a dash.

(b) MARROW—NARROW

Here the M(— —) is changed into N(—•) by the mere shortening of the last dash into a dot.

(2) Errors involving a deletion or addition of a dot or dash:

(a) MAZZETTA—TAZZETTA

Here one of the dashes representing M(— —) drops out, giving T(—).

(b) MEDELIJDEN—MEDERIJDEN

Here the last dot of L(• — • •) drops out, giving R(• — •).

(c) CITERONS—CITTRONS

Here the single dot representing E has merely dropped out.

(d) MATGRASS—OATGRASS

Here the addition of a single dash to the signal for M(— —) changes the letter to O(— — —).

(e) MENNER—MENDER

Here the addition of a dot to the signal for N(—•) changes the letter to D(—••).

(3) Errors involving the coalescing of dots into dashes, or the splitting up of dashes into dots:

(a) DENIAL—MENIAL

Here the two dots of the D(—••) coalesce into a dash yielding M(— —).

(b) BATTED—BAITED

Here the dash of the first T(—) splits up into two dots, making I(••).

(4) Errors involving the false grouping or spacing of signals.
• This is a very prolific source of errors of the most insidious type.

(a) $\left\{ \begin{array}{l} \text{AMENDING} = \text{M} \\ \text{ATTENDING} = \text{T T} \end{array} \right.$

(b) $\left\{ \begin{array}{l} \text{AEROLITH} = \text{H} \\ \text{AEROLITES} = \text{E S} \end{array} \right.$

(c) $\left\{ \begin{array}{l} \text{ROSTER} = \text{S T} \\ \text{ROVER} = \text{V} \end{array} \right.$

(d) $\left\{ \begin{array}{l} \underline{\text{ROUGES}} \\ \underline{\text{ROUGH}} \\ \underline{\text{ENOUGH}} \end{array} \right. = \begin{array}{l} \begin{array}{ccc} \text{R} & & \text{E S} \\ \cdot & \cdot & \cdot \cdot \cdot \\ \cdot & \cdot & \cdot \cdot \cdot \end{array} \\ \begin{array}{ccc} \text{R} & & \text{H} \\ \cdot & \cdot & \cdot \cdot \cdot \\ \cdot & \cdot & \cdot \cdot \cdot \end{array} \\ \begin{array}{ccc} & \text{N} & \text{H} \\ \cdot & \cdot & \cdot \cdot \cdot \\ \cdot & \cdot & \cdot \cdot \cdot \end{array} \end{array}$

Under this type may be found thousands of cases wherein the sole telegraphic difference between two words reduces itself to merely a slight difference in the *spacing* of the signals constituting the words. And when, as sometimes happens, two errors in spacing are made in the same word, there results a word hardly bearing any similarity to the original. Note, for example, the following two words:

$\begin{array}{cccccccc} \underline{\text{B}} & \underline{\text{A}} & \underline{\text{N}} & \underline{\text{E}} & \underline{\text{F}} & \underline{\text{U}} & \underline{\text{L}} & \cdot \\ \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \\ \underline{\text{D}} & \underline{\text{U}} & \underline{\text{T}} & \underline{\text{I}} & \underline{\text{F}} & \underline{\text{U}} & \underline{\text{L}} & \cdot \\ \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \end{array}$

The correction of telegraphic errors of these types was made more difficult by the fact that *two different* telegraph alphabets were and still are in use—one in the United States and in Canada, called “American Morse”, the other in all other countries, called “International Morse”.¹ The signals for certain letters in one alphabet apply to different letters in the other, and in a message which is carried over circuits wherein first one alphabet has to be employed, and then the other, the recipient of a mutilated message has to take into consideration errors common in two different alphabets.

In closing this brief treatment of telegraphic errors, it is thought pertinent to add a few remarks in connection with the general manner in which messages were received over cables and over telegraph wires. From days of the first Atlantic cables up to about 1880, the signals arriving at the end of a submarine cable were received and interpreted by means of a small swinging mirror, called the mirror galvanometer. The swinging or deflection of this mirror to the right or left of a neutral or central position corresponded to the dots and dashes of the ordinary Morse system. Two operators were required, one to observe the mirror, interpret the deflections into signals representing letters and words, and call them aloud to the other operator who wrote down what the first operator said. Obviously, in the case of evanescent signals of such a nature, the observer of the mirror had but a single chance—and that only a most fleeting one—to “read” the signals and interpret them; the only recourse in the case of a doubtful reading was to ask for a repetition, since the signals themselves were not automatically recorded as they were received by the galvanometer.

¹ A long and hard fought struggle to replace “American Morse” by “International Morse” was unsuccessful and the matter is now a more or less dead issue. In this respect, radiotelegraphy is in a much better situation, for only one Morse alphabet—the International—is in universal use.

About 1872 the system of evanescent signals of the mirror galvanometer began to give way to the system of recording the signals by means of Thompson’s siphon recorder¹ which made a record of them in the form of a wavy line written by an inking mechanism on a moving tape, in response to the signals received. This system, naturally much improved, is still in use to-day, but two operators are still required. One interprets the signals on the tape, and types them directly upon a typewriter before him; the other checks the “reading” and rolls up the tape as he checks. The record being in a fairly permanent form, doubtful or erroneous readings can be corrected by reference to the tape.

On the other hand, the situation in this respect in the case of land-wire telegraphy was exactly reversed. Originally Morse telegraphy was a system of recorded telegraphy, the dots and dashes being actually written upon a moving tape. This system, however, had but a short life, for even Morse and Vail had noted, from the very first, that the telegraph apparatus made characteristic sounds corresponding to the dots and dashes, and that these sounds could be interpreted by the ear. By 1850 several operators² had become proficient in “sound reading” or sound telegraphy, and although at first the Morse Company, in common with all the other companies, vigorously interdicted the sound system and discouraged its practice to the extent of passing regulations forbidding its use, the new form of receiving soon completely supplanted the old, until the advent of the modern forms of printing telegraph apparatus, which was only of somewhat recent date.

Thus we note that whereas in cable telegraphy progress in the art consisted in supplanting evanescent signals by recorded signals, in land telegraphy progress was in exactly the opposite direction, and a system of recorded signals gave way almost universally to a system of evanescent signals.

Now it is obvious that signals of an evanescent nature require a great deal more concentration and long-sustained mental effort than signals of a more permanent nature require. Then psychological factors begin to play their part in introducing errors. In cable telegraphy, when the mirror galvanometer was in use, continuous observation of the intense shifting light spot very soon led to fatigue of vision, with consequent errors. (In many cases operators had to be retired for partial blindness after only a few years’ work.) In sound telegraphy it is a well known fact that expert telegraphers, in receiving words, do not as a rule listen for and distinguish individual letters. They listen for syllables and entire words as entities. For example, they receive the telegraphic signals representing the word

¹ Patented in England in 1867, practically perfected by 1870, the new instrument was almost immediately introduced into general service by various cable companies.

² See *The Life of James Francis Leonard, the First Practical Sound Reader of the Morse Alphabet*, by John Wilson Townsend, Louisville, Ky., 1909.

"and" not as a set of separate and distinct groupings of dots and dashes comprising single letters, but as a single sequence of signals all run together. Thus, the word "and" is for them not

A N D but is merely a single sequence of dots and dashes:

AND With such a preliminary bias, and in the case of evanescent signals, their own unconscious mistakes are always cropping up. In plain-language reception this causes no serious difficulty because the context of a message serves as a guide, but when code language is involved, the difficulties become more serious. For instance, having received the signals "SATUR", a moment's relaxed attention on the part of the operator results in his thinking the word must be SATURDAY; it may, however, have been SATURNALIA, SATURN, or SATURNINE.

C. Phonetic errors

We come finally to phonetic resemblances between words leading to errors in code communication. We have noted how in the days when the mirror galvanometer was used in cable telegraphy one operator interpreted the signals and called them aloud to another operator who did the writing. Under these circumstances words such as "accept" and "except", "counsel" and "council", "serial" and "cereal", and the like were most easily confused. Furthermore, even if individual letters were called out, the phonetic resemblances between certain letters, such as "b" and "d", "f" and "s", "m" and "n", etc., coupled with the fact that there are hundreds of cases in which either one or the other letter of an easily confused pair of letters would serve equally well in making a *bona-fide* word, such as "abjure" and "adjure", served to increase the difficulties and errors. It is interesting to record what one cable expert¹ of 1880 had to say on this subject:

These (that is, phonetic errors) form the largest percentage of errors made on cables, and are to a great extent due to the defective codes used. Writing or printing a lot of words and calling them a code has caused no end of trouble, delay, and loss to the business community using them. A code to be perfect should be telegraphically checked and the words selected by practical and experienced operators.

All these sorts of errors, inherent in code communication by means of dictionary words, as discussed above, became a source of greater and greater annoyances to all concerned.²

¹ Washburne's *Cable and Telegraph Manual and Error Detector*, New York, 1880.
² One of the curiosities of this period is a 436-page volume containing some 70,000 examples of errors in code words, compiled from actual telegrams. *Guide to the Correction of Errors in Code Telegrams* (4th Ed.) London, 1890.

THE ORIGIN OF THE PRINCIPLE OF THE TWO-LETTER DIFFERENTIAL

In fact, the prevalence of errors of the types enumerated, and the difficulties to which they led both in the way of delays due to the necessity for verifications, and of financial losses caused by undetected errors, soon led to the demand for "safer" words. Code compilers busied themselves examining thousands of potential code words from various dictionaries, and they called into consultation experienced telegraphers to assist them in eliminating telegraphically similar words. In particular, a new principle began to be insisted upon by code experts. It was that code words should differ from each other orthographically by a minimum of *two* letters throughout. This principle, later to become known as the "two-letter differential", and to exercise a profound effect upon the science of code compilation, had for its object that if a telegraph operator made an error involving only a single letter in a word—and fortunately it is true that the average experienced operator will make no more than that, as a rule—the resulting erroneous word would not be a *bona-fide* word in the code, would not be listed, and by its mere absence would thus give indications that an error had been made.

THE ORIGINS OF SPURIOUS DICTIONARY WORDS

Now it was a very difficult task for code compilers to find a number of "safe", two-letter difference, dictionary words sufficient for the needs of large and extensive code books. When in 1880 an American code compiler of long experience compiled a book¹ containing 61,000 carefully selected words, differing from each other by a minimum of two letters, and taken from the standard dictionaries of the eight languages allowed in the extra-European telegraphic traffic (German, English, Spanish, French, Italian, Dutch, Portuguese and Latin), we may be sure that his work was quite exhaustive as well as expensive, for it has been carefully estimated that the limit of the number of grammatical words, differing from each other by a minimum of two letters, none exceeding ten letters in length, and taken from the eight languages mentioned is only 160,000. When from this number thousands of potential words must be discarded because they are telegraphically unsafe, it is seen that this limit becomes considerably reduced.

Some American code compilers, moreover, deemed the inclusion of foreign-language words in their codes to be detrimental. They preferred to restrict their selections to words all belonging to one language, and in order to get around the difficulty of obtaining a sufficient number of suitable words they resorted to an expedient which can best be described by quoting from an actual example. The

¹ *The Merchants' Code, extended and improved*, by John C. Hartfield, New York, 1880.

could be passed over the counter under the pretext of their being real dictionary words, they would then count and be charged for as but single words, whereas, in fact, they are two code groups.

Or, by running together two 5-figure groups, splitting up the 10-figure group into five 2-digit groups, and replacing each 2-digit group by a 2-letter syllable, one would also have a single, euphonious "word" such as CAPAROMATE. Thus, likewise, the scheme would permit of transmitting two code groups at the cost of one.

Or, again, by substituting a single letter for a single digit, according to the following scheme:

1	2	3	4	5	6	7	8	9	0
A	B	C	D	E	F	G	H	I	J

one could replace a 2-group, 10-digit combination of figures, 20534/16719, by a "word" BJECDAFGAI. This amateurish scheme, we may well believe, was adopted only by novices, not by experienced code compilers, for not only were the chances of error great, but also the resultant "words" very frequently bore no resemblance whatever to real words.

B. Syllable-combination systems

Another type of artificial words is that which may be termed the syllable-combination type. Even in the early days there were complicated codes in which two-, three-, and four-letter syllables could be combined to form a single code group. In the United States as early as 1867 there appeared a code for telegraphing stock quotations, in which words such as "CABUHUC", "FEDIXIB", and the like were built up of separate syllables, each syllable having a meaning.¹ Here the saving to code users was in fact greater, for one word might signify three, four, or more phrases or items.

C. Root and terminal systems

Still another source of the artificial words was that wherein such words were built up by means of what are called "roots and terminals." This was really an outgrowth or a development of the type just considered. One great difficulty with code words built up of syllables each having a separate meaning is that an error in a single letter of a syllable may change the whole meaning of the word. There was too much risk to run as regards the all too prevalent inaccuracies in the telegraph service. From groups consisting of two or three letters, code compilers naturally went to combinations of four and five letters and sacrificed condensing power for greater accuracy. The root and terminal system permitted of sending two or more ideas in one word, usually ideas relative to orders or offers,

and the details pertaining to them. For example, the root "APARL" in a certain code meant "We order 1500 at 28 shillings," the terminal "ANFRO" meant "140 lb. jute sacks Duluth Imperial, net c. i. f. London." Putting the two groups together, "APARLANFRO" meant "We order 1500-140 lb. jute sacks Duluth Imperial at 28 shillings, net c. i. f. London." The word "APARLANERE" changed only the destination of the foregoing order from London to Liverpool. Thus, by using a set of roots and another set of terminals one could make thousands of combinations of ideas, each varying from the other in one of a number of details.

Having described these various sorts of artificial and spurious or mutilated words, the question arises: Were such words accepted by the communication agencies as single words on a parity with *bona-fide* dictionary words, and if so, how is this to be explained, seeing that such words were not at all sanctioned? At this point we may merely make reply by saying that whether or not such words would be in reality accepted as legitimate dictionary words would naturally depend upon the linguistic knowledge of the counter clerks in the telegraph offices. But the latter are not and never have been skilled linguists, and it was too much to expect that they would be able to recognize *bona-fide* words from spurious ones among the millions of code words with which they had to deal, words belonging to many languages, all but one of them as a rule being foreign to them. We may be sure that it was not long before such words were accepted. If, at first, there were some disputes, they soon ceased, and natural competition between companies did the rest. What is more, so far as operators were concerned these spurious words were *easier* for them to handle than many of the *bona-fide*, foreign ones, such as "Aardmijten", or "Aangehoogd", and the like. In fact, the telegraph personnel themselves claimed that the artificial words were much to be preferred over the more unusual and difficult real, dictionary words, seeing that for the most part the artificial words were composed of fairly regular alternations of vowels and consonants making them easy to read quickly, remember, and transmit. It was, therefore, not long before such words, appearing first in internal or domestic telegrams within the United States, were carried over into international telegrams, when American business men corresponded with their agents or European representatives and firms abroad. For a long time after their origin, however, these words were not sanctioned by regulations of either the American communication companies, the cable companies, or the European telegraph administrations: they represented abuses and subterfuges of the rules, and their acceptance constituted infractions against legal, authorized procedure.

¹ *Telegraph Cypher for Transmitting Telegrams relating to Foreign News, Stocks, Gold, Cotton, Financial Matters, etc., in a Commercial Form* by Martin K. Thompson, New York, 1867.

THE BIRTH OF THE OFFICIAL VOCABULARY: CONFERENCES OF 1885 TO 1896

Let us return now to a consideration of the measures taken by the International Telegraph Union in its attempts to regulate telegraphic languages and to put a stop to abuses. We have seen that in 1879 the conference restricted the number of languages that could be used for code language in the extra-European régime to eight, with no limitations as to the possibility of mixing languages in the same telegram; in the European régime no restrictions as to the number of legitimate languages was established, but only *one* language could be used in a single telegram. From the point of view of the desire to reduce the difficulties incident to code communication, these two measures present a most striking inconsistency. For, since the admission of several languages in one telegram makes code-language control more difficult than it would be if only one language were permitted in a single telegram, one would expect that the limitations set up by the conference would have sought to prevent this mixing in that régime wherein code language was the most prevalent, viz., the extra-European. But exactly the opposite was the case. In the European régime code language has always played a very minor rôle; in the extra-European régime, always a major rôle. Of these measures M. Crescitz has written:

If one realizes that code language was almost exclusively employed in the extra-European correspondence and almost not at all in the European régime, one can hardly refrain from remarking that the restrictive provisions suggested by the committee (which, as we have seen, were later adopted by the conference) obviously should have been left to rest in the condition of a theoretical measure having only the practical result of uselessly complicating the regulations.¹

During the course of the discussions on the question of languages, which took place in the meetings of one of the committees at this conference, the Eastern Telegraph Co. made a proposal which, for about a quarter of a century thereafter, was the subject of much controversy, almost came to fulfilment, and then suffered an unlamented death. This proposal consisted in confiding to the International Bureau the task of preparing a universal code of 100,000 words which alone would be admissible for extra-European correspondence in code language. Apparently this suggestion, the first allusion to

¹ There will be several occasions to refer to the document from which this quotation is taken. It is entitled *De l'emploi du langage convenu dans la correspondance télégraphique*, written by M. A. Crescitz, vice director of the International Office of the Telegraph Union at Berne, and published in successive monthly issues of the official journal of that bureau, the *Journal télégraphique*, from May 25, 1911, to March 25, 1912. This important and very interesting series of articles will be referred to hereafter as "the Berne Bureau study." It should be added that the document presents the personal opinion of the writer, M. Crescitz, and not the official attitude of the Telegraphic Union.

the compilation of an official vocabulary,¹ arose from a consideration of the second paragraph of the Article 8 which the conference was about to adopt, the paragraph reading:

These words are taken from the vocabularies admitted for international correspondence for code language but of which the composition varies according to whether the régime concerned is European or extra-European.

Since no such vocabularies had as yet been admitted, obviously the inference was that either existing or proposed vocabularies should be submitted for examination and approval, or that the Telegraph Union itself would prepare such vocabularies. But nothing further was done by the conference on this point, and nothing came of the proposal of the Eastern Telegraph Co. at this time.

From 1879 to 1885 the difficulties of code-language communication increased rather than diminished, and at the Berlin conference, in 1885, the Swiss administration resuscitated the Eastern Telegraph Company's proposal to entrust the compilation of a vocabulary for code language to the International Bureau. The following proposal was carefully discussed and drawn up by the committee on tariffs:

ARTICLE 8, PARAGRAPH 2

These words are taken from a vocabulary to be established for that purpose by the labors of the International Bureau and which may contain words of a maximum of ten letters selected from the German, English, Spanish, French, Italian, Dutch, Portuguese, and Latin languages.

The foregoing provisions will not become effective until a date fixed by the International Bureau. Until that date, the provisions of the London regulations will remain in force.

In the plenary session, however, this proposal aroused warm discussions, and was finally rejected, largely through the opposition of the delegates from Great Britain, British East India, Luxemburg, the Netherlands, and France. It is important to note that the principal basis for this opposition was that the article in question would, if adopted, necessitate the *scrapping of existing codes*. The article as finally adopted read as follows:

ARTICLE 8

1. By code language is understood the use of words, which, although each presents an intrinsic meaning, do not at all form expressions intelligible to the offices in correspondence.

¹ This is the first allusion to the compilation of an official vocabulary which was made in any international conference, but the idea must have been current long before this. In the *Journal of the Telegraph*, the official journal of the Western Union Telegraph Company, in the issue of July 18, 1869, there appears the following interesting editorial under the heading "Construction of Cyphers; Necessity of Choosing Cypher Words Dissimilar in Telegraphic Orthography":

No one but a practical telegrapher is competent to make up a safe and reliable practical cypher that will be comparatively free from the danger of errors in transmission over the wires. It might be well for telegraph companies to have made up and published for the benefit of their customers, a vocabulary of cypher words, not easily capable of conversion to other words in transmission, by this slight variation of one or two telegraphic signals. From such a vocabulary business correspondents could select such words, known only to themselves and to such confidential clerks as they deemed proper to intrust with their key, and attach to them such meanings as they might choose to adopt.

2. These words are extracted from vocabularies admitted for international correspondence in code language.

3. Telegrams in code language can contain only words of a maximum of ten characters belonging to the German, English, Spanish, French, Italian, Dutch, Portuguese, and Latin languages. Every telegram can contain words taken from all the above-mentioned languages.

4. Proper names can not enter into the composition of vocabularies. They are admitted in the text of telegrams in code language only with their plain-language significance.

5. The office of origin can demand the production of the vocabulary in order to control the execution of the provisions which precede and to verify the authenticity of the words employed.

Having rejected a proposal to establish an official vocabulary, which would *perhaps* have placed a powerful weapon in the hands of the Union, efficient to enforce its regulations, it seems that the conference clung to the hope of surmounting the difficulties by more strictly applying the provisions of the regulations. It was a vain hope.

The next conference of the Telegraph Union took place in 1890, at Paris. Before taking up the most important question which came before this conference, the one relative to the compilation of an official vocabulary, it is deemed best to give in brief the action taken by the conference on other matters relative to languages.

A proposal of Great Britain, relative to a slight change in Paragraph 2 of Article 8 is of interest. This paragraph, as it appeared in the 1885 regulations, read:

These words are extracted from vocabularies admitted for international correspondence in code language.

The British delegation stated that the use of the word "vocabularies" had "given rise to discussions with senders, the latter pretending that their private lists of words in code language were vocabularies." Here we can get a glimpse into the minute technicalities and unforeseen loopholes in the wording of the regulations that code compilers and code users have always seized upon in their continued attempts to evade the regulations. The British delegation proposed to change this paragraph to read as follows:

These words are taken from conventional dictionaries; the grammatical inflexions of words are, however, permitted.

But this proposal was retired in the course of discussions by the tariff committee. It is unfortunate that the minutes of these discussions do not contain the arguments advanced.

At this time there also took place much discussion of the difficulties encountered in charging for mixed telegrams, that is, those in which code and plain-language words were intermixed, and a new paragraph was adopted and inserted in Article 20, in which the basis of charging for plain-language words under these circumstances was

As to cipher language, it will be recalled that at the St. Petersburg conference in 1875, offices of the extra-European régime were granted the authority to refuse to admit over their lines private telegrams containing letters having a secret meaning. This privilege was continued through the next two conferences, and in the present conference the power to refuse such telegrams was extended to include all offices in both régimes. The fact is mentioned here because it later formed the subject of much argument.

We come now to the principal proposal put before the conference, that relating to the official vocabulary. This time it was Belgium that presented the matter, stating that—

It is indispensable to put an end to the abuses in code-language correspondences.

Certain senders introduce into their telegrams expressions which ought to be charged for as cipher language. Sometimes, each of the consonants which enter into the composition of these words has a code value; the interposed vowels serve only to give these combinations the appearance of words. Other times, the word in code language is preceded or followed by a syllable also having a code meaning. Occasionally, finally, disputes are raised relative to the current usage¹ of the words employed.

An official code, containing in alphabetical order all the words which could be utilized, would cause the major part of these difficulties to disappear.

However, this code ought to include all the words which are at present admissible by virtue of Article 8, paragraph 3. In that manner the vocabularies which have been established, conforming to the international regulations could continue to be utilized. A delay of six months, to date from the putting into effect the new regulations, would be accorded to senders whose nomenclatures would have to be modified.

The public should have the privilege of obtaining, by purchase, copies of the official code.

When the proposal came to be considered, all the members of the tariff committee were agreed in principle that the establishment of such a vocabulary would remove the uncertainties of the charges to be applied and would reduce the number of errors made in transmission. But there were differences of opinion as to how to introduce the vocabulary into practice and the following paragraph was finally adopted in the plenary session:

ARTICLE 8, PARAGRAPH 2

These words are taken from vocabularies admitted for international correspondence in code language, or from an official vocabulary compiled by the International Telegraph Bureau. The use of the official vocabulary will become obligatory after the expiration of a delay of three years following the date of its publication. It will be optional for the correspondences of the extra-European régime.

¹ The expression *l'usage courant*, strangely enough, was the one used. It is worthy of note, for later on we shall see that this expression played an important part in the 1908 conference. It is also necessary to point out that the regulations did not stipulate that the dictionary words employed in code language must be words in current usage, as might be inferred from the remarks of the Belgian delegate.

Here again we note the inconsistency in the policy mentioned once before. Code-language correspondence formed the major portion of the traffic of the extra-European régime, and only a minor portion of that of the European régime; moreover, most of the abuses of code-language correspondence occurred in the extra-European régime. Since the vocabulary had for its aim the suppression of these abuses and the reduction of the difficulties to which they led, it would seem that if an official vocabulary was to be prepared at all, it should have been made obligatory in the extra-European régime and optional in the European. As we have seen, the exact opposite was the case.

The International Bureau then proceeded to compile the official vocabulary and it was published in October 1894. It contained in alphabetic arrangement 256,740 words of five to ten letters, taken from the eight authorized languages. The cost of the vocabulary, for an edition of 15,000 copies, was approximately 123,000 francs, of which 60,000 francs was for compilation and the remainder was for printing and binding.

Shortly after its publication there arose an intense opposition to the vocabulary, the most important of the arguments raised against its adoption being based upon the following features:

- The insufficient number of words;
- The absence of an alphabetic list of terminations;¹
- The weakness of the paper stock used;
- Defects in the numbering system;
- The insufficiency of typographic and telegraphic differentiation;
- The bad choice of words, from the point of view of their meaning;
- The bad choice of words, from the point of view of their pronunciation;
- Typographic errors.

We may be certain that these reasons were *minutiae*, reasons advanced, perhaps, to hide the real ones; for the greatest opposition came both from code compilers, who felt that their territory had been encroached upon by the Telegraph Union, and from those interests which, for one reason or another, did not wish to see an official vocabulary imposed upon commerce. To these interests every excuse which would combat the establishment of the vocabulary was a legitimate one.

At the conference of Budapest, in 1896, the Belgian administration pointed out the inconsistency we have just noted above, and introduced a proposal to make the official vocabulary obligatory for both régimes after January 1, 1908. In agreement with Belgium were the French, Japanese, and Swiss administrations. But among others, Great Britain, British India, New South Wales, Australia, New Zealand, the Netherlands, and the largest of the companies were

¹ A list of all the words alphabetically arranged according to their backward-spelling, useful in correcting errors.

absolutely opposed to the obligatory imposition of any vocabulary. The most important reason advanced was that the failure to include in the vocabulary *all* the perfectly authentic and authorized words contained in existing codes, would cause the scrapping of practically all codes, with a consequent great loss to commerce. But another reason—a real reason scarcely mentioned because no one wished to incur the disfavor of important business interests—was that the imposition of an official vocabulary would curtail all opportunities to employ systems enabling a code user to express two code words at the cost of one, in the manner described herein under the heading of code condenser and root and terminal systems; for, by this time, the use of such systems had become very widespread in international telegraphy.

The spreading use of such systems was due in a large measure to the official recognition accorded by the American companies in 1893 to the legitimacy of artificial words in internal or domestic telegrams. This early sanctioning of artificial code words in the United States is extremely important to note, for it was soon to exercise a profound effect upon international regulations. Let us see how this came about.

In the first place, it must be said that the birthplace of most of the new ideas, schemes, and subterfuges in code compilation was the United States. We have noted that this country is not and never has been a member of the International Telegraph Union, and that the rules governing word count in the internal telegraph traffic of this country have not always coincided with those in international traffic. Now from the early days of American telegraphy up until 1893 the chief difference between American rules and international rules as to word count was that in the former, cipher language was charged for at the rate of one word for each separate character, be it a letter or a figure, whereas in the latter, the count was at the rate of five characters per word;¹ in both, dictionary words when used in code telegrams were charged for the same as in plain-language telegrams. But in the 1893 *Book of Rules* of the Western Union Telegraph Co. there appears for the first time under Rule 4, pertaining to the count of words, a most interesting and curious paragraph.² After stating that dictionary words will be counted and charged for each as one word, the rule reads as follows:

Figures, decimal points and bars of division, and letters (except the pronounceable groups covered by the . . . paragraph below) will be counted—each separately—as one word.

¹ Except for a short period, when three characters constituted a word in both American rules and international rules.

² I am greatly indebted to Mr. J. C. Willever, first vice president of the Western Union Telegraph Co., for the reference and quotation which follow.

All pronounceable groups of letters, when such groups are not combinations of dictionary words, will be counted each group as one word.¹ When such groups are made up of improper combinations of dictionary words, each dictionary word so used will be counted as one word.

Under examples of word count there appear, among others, the following:

Amaurecis.....	1 word ²
Adbantia.....	1 “
Chancin.....	1 “
Interavis.....	1 “

It is, of course, easy to see that code users who found these artificial words safe, practical, and economical for their domestic telegrams, would try to use these words in their cablegrams. If they were acceptable to the American companies, why should they not be to the European telegraph administrations? Once tried, with successful results due to causes already mentioned, the practice grew rapidly until the day came when artificial words had become so deeply rooted in cable telegraphy that any attempts to eradicate them were bound to meet with the most stubborn opposition.

This suppression of artificial code words the official vocabulary had as its aim, and we can therefore appreciate the resistance which any endeavor to make its use obligatory was certain to cause. Consequently, we need not be surprised at the vehemence with which the representatives of the large telegraph and cable users at the 1896 telegraph conference opposed the obligatory imposition of the vocabulary. There was no objection, they said, to extending the proposed vocabulary to include more words, but under no circumstances could they agree to its obligatory use.

After much discussion a compromise was reached, a compromise that is best described by our own colloquialisms as “passing the buck,” and “Let George do it.” The conference decided to authorize the extending and improving of the vocabulary, and to leave to a succeeding conference the decision as to the date when the vocabulary would become obligatory. This is the paragraph finally adopted:

ARTICLE 8, PARAGRAPH 5

From a date to be fixed by a next conference all words employed in private telegrams in code language will be taken from the official vocabulary prepared by the International Bureau, duly enlarged.

The comments of the Berne Bureau itself upon this solution are well worth quoting:

When it is recalled that the proposals submitted at the conference and the discussions they evoked, which showed the Congress to be divided into two

¹ It is a most curious circumstance that this rule set no limitation upon the length of such groups. It was fully ten years later that the limit was set at ten letters per word, and the rule made to coincide with the rule adopted by the 1903 London International Telegraph Conference. What an opportunity the code compilers and users overlooked!

² These are all artificial words, but no note to that effect appears in the rule as printed.

irreducibly opposed camps—on the one side, the delegations which demanded the obligatory employ of an official vocabulary in the two régimes, on the other, those which rejected it in the extra-European régime and also in the European—it is not very surprising to see the conference *unanimously* adopt the solution we have just indicated. The reason for it, it seems to us, to tell the truth, was that the solution was not a solution, that the decision taken did not solve the problem, and that the question still remained in its entirety.

The International Bureau again set itself to the extremely difficult, if not impossible, task assigned it. A total of 218 public and private codes, containing over 5,750,000 words, was submitted to the bureau. The work was begun in February 1897, and the first of four large volumes was published in 1900. The other three volumes appeared at successive half-year intervals, the last appearing in May 1901. The four books contained 1,174,864 words; with an appendix, they totaled 1,190,000. The cost for an edition of 7,000 sets totaled over 300,000 francs, of which 154,000 was for compilation, and 121,000 for printing and binding. From 1901 to 1903, when the next conference met in London, there was plenty of time to find fault with the new vocabulary.

THE DEATH OF THE OFFICIAL VOCABULARY: CONFERENCE OF LONDON, 1903

At London the matter of the vocabulary and code language again formed the most important subject of discussion. In view of the remarks made above, it requires no great degree of imagination to understand that the business of adopting the new vocabulary and putting it into effect was to be no simple formality. Events started off with a proposal of Belgium and Japan to make the vocabulary obligatory for all, and to put it into effect at the same time as the new regulations, that is, July 1, 1904; France proposed July 1, 1905. In presenting her proposal, Belgium indicated that when the new vocabulary became effective every code word not included in it would be charged for as cipher, at the rate of five characters per word.

The British delegation then read a long paper, the substance of which is as follows: After a few complimentary remarks stating their appreciation of the excellence of the work of the International Bureau, they had no other alternative than to oppose the adoption of the proposal to make the use of the new vocabulary obligatory. They indicated that many Chambers of Commerce reproached the vocabulary in the following respects: that it did not have a sufficient number of words; that it lacked flexibility; that many private codes would be rendered useless by the obligatory employment of the official vocabulary on account of the accidental omission of admissible words, due to the elimination of all words having less than five letters, and the absence of admissible inflections of words; that the object sought in the obligatory employment of the vocabulary, that is to

say, the verification of the words used in code telegrams, would not, indeed, be possible in practice, because it would be necessary that every office should be provided with the vocabulary and that the clerks should have the time to refer to it; that, far from admitting any restriction whatsoever, they demanded a greater latitude in the employment of code language, in particular the admission of words made by Latin roots and terminals, now in such wide use.

The British delegation felt that these criticisms of the commercial world in general had great weight. After qualifying and summarizing them they went on to say:

Another very serious obstacle to the adoption of the vocabulary arises from the fact that the needs of commerce have caused the compilation of codes of which the words do not comply at all with the present conditions of the regulations, but which have, nevertheless, been in use for a certain time in the commercial world, without appearing to have brought any inconvenience to the telegraphic service.

Many of these codes first appeared in the United States where the telegraph companies do not adhere to the convention, but their usage has greatly extended into the correspondence exchanged with Europe principally as a result of the impotence of the cable companies to enforce the regulations, in view of the pressure of competition.

These codes have to-day become so predominant in this country, as upon the Continent, and their usage has become so deeply rooted, that every effort to prevent their employment, or even to apply a higher rate to the words taken from them, meets with the most lively opposition.

The difficulty of the situation is further increased by the fact that the expressions included in these codes are in certain cases easier to telegraph than many of the words present in the official vocabulary.

For example, a Continental telegrapher would find the transmission of expressions such as "DEMINABAM", "OPORTAVERE", taken from a code used in Liverpool, more easy than the transmission of English words such as "strength", "bowieknife", "awkwardly", and in the same manner an English telegrapher would transmit them more easily than a great number of words of Continental languages.

If, on the one hand, the expressions contained in these irregular codes are thus not improper from the point of view of the telegraph service, they are, in certain respects, better adapted to the special requirements of the commercial world, in that they furnish an elastic and compact system, in which, for example, any root whatsoever can be combined with any termination whatsoever. In this manner the size of the code is considerably diminished (an important consideration for business men and others) and the translation of code language into clear language, or *vice-versa*, is greatly facilitated.

The British administration possesses proofs that the employment of codes of this nature is very widespread among the telegraphing public, which highly approves them, and it seems probable that the development of telegraphy in code language will tend to be in that direction.

If one should succeed in making the public accept the vocabulary as obligatory, the employment of such codes would be excluded, and therein would be a measure which would put great obstacles in the way of the development of code language in telegraphy, with the result that one could not regard it otherwise than as a retrogression, in view of the importance of the rôle played by the telegraph in commercial relations.

As we have said, the codes in question are largely employed in the United States, and anything that would have the effect of excluding them would render more difficult the entrance of the Government of the United States to membership in the Union, a thing which we earnestly desire.

As a consequence, Great Britain proposed that the vocabulary should not be made obligatory, that the provisions of the present regulations be maintained, and that each administration be left free to act with a degree of tolerance toward irregular codes.

Then two delegations remarked that such a tolerance would lead to inequalities in applying the provisions of the regulations, and would thus bring about a situation which would be entirely inconsistent with one of the articles of the convention. When the French delegate asked the British delegate how the Post Office would treat irregular words in code language the reply was in substance as follows: That the British office was not the only one which found itself unable to apply the provisions of the regulations; that it was loyally forcing itself to conform to the regulations, and that, like other offices, it was unsuccessful: the Telegraph Union has found itself powerless in this respect. The British delegation felt that this impotence resulted from the fact that the provisions of the regulations were themselves not scientific, nor built upon a logical foundation, and that, as a result, they were not susceptible of rigorous application.¹ One could pretend that the employment of the vocabulary would give the means of applying these provisions, but it is easy to reply that it is precisely the vocabulary which has placed in full light the defects of the regulations. It is impossible to justify to the public the refusal of such expressions as "deminabam" and at the same time place at their disposal words such as "aangehaard", "aange-lijkt", etc., and, in plain language, to accept words of such languages as Ruthenian, Slavic, Illyriac, and Annamite.

In the face of arguments such as the foregoing, and the powerful opposition made manifest in the conference, all discussion of an obligatory vocabulary was suddenly dropped and new ways and means were sought out of the difficulty.

THE RECOGNITION AND LEGALIZATION OF ARTIFICIAL WORDS

With the official vocabulary a dead issue, and out of the way for good, events now moved with a rapidity that presents a most startling contrast with the staid and deliberate procedure which had theretofore characterized the important legislation enacted by the Telegraph Union from its birth. If, for almost a quarter of a century, extended discussions had been held regarding legislation pertaining to the

¹ It seems hardly necessary to point out that what the British delegate meant by this statement is that the base upon which the regulations in respect to code language were founded is illogical due to the fact that the application of the provisions requires and always has required a far more profound linguistic knowledge than can possibly be expected of telegraph clerks.

adoption of a vocabulary, the Union now proceeded to enact within a period of a few weeks a piece of legislation that had a hundredfold more profound and far-reaching effect than the official vocabulary, either dead or alive, could possibly have had. And if up until now the difficulties which code-language communication had brought to the Telegraph Union seemed serious and burdensome, these difficulties were to become mere petty and trifling annoyances in comparison with the insurmountable ones into which this new legislation was about to entangle all the states belonging to the Union.

We may begin with the proposal of the Belgian delegation concerning code language, which was first presented before the regulations committee, and which read as follows:

1. Code language is that which is composed of euphonious combinations (real or artificial words) not forming phrases that are intelligible to the offices in correspondence.
2. These words must have a phonetic consonance sufficient to render pronunciation possible, by the alternation of consonants and vowels.
3. Words of code language can not be longer than ten characters according to the Morse alphabet.
4. Combinations which are not included within the limits of the two preceding paragraphs are considered as belonging to language in secret letters and are charged for accordingly.

In the ensuing discussions the French delegation, fearing that the conference was on the point of taking radical, perhaps irrevocable, steps along new and untrodden paths, submitted a proposal that the whole matter be turned over to the International Bureau for careful study before action was taken. The German delegation, while recognizing the importance of the present difficulties, was of the opinion that it was preferable to let things be as they were rather than take a premature decision. They remarked that the Belgian delegation, to characterize this code language, proposed the designation "euphonious" combinations. What is the meaning of "euphonious"? they inquired, and pointed out that there was reason to fear that this definition would be insufficient to differentiate code language from cipher language;¹ that this proposal added a new category of words, artificial words.

However, the Belgian proposal gained the support of several other delegations. It is interesting to note that the Italian delegation proposed to add an amendment to require that artificial words be formed in a manner so that there would never be more than three consecutive consonants without being alternated with a vowel. To this the Belgian delegation remarked that this proviso would not be easy to apply. How would one treat the word "mqzoolqk" which

¹ It will be recalled that in 1890 the employment of letters or groups of letters having a secret meaning was not permitted in private telegrams. In the first few sessions of the present conference (1903) it was decided to readmit such correspondence.

could not be pronounced? No; they felt the requirement of pronounceability would much more easily allow the clerks to know how the words ought to be treated.

The French delegation persisted in thinking that such a radical measure as that of Belgium was dangerous, *because the committee itself did not appear to have a sufficient knowledge of the codes to which the new decisions would give birth.*

When the representatives of the various companies were asked for an expression of opinion, the representative of the Great Northern Telegraph Co. agreed with the views of the German and the French delegations and was of the opinion that the Belgian proposal was premature. The representative of the Anglo-American Cable Co. thought that the companies were disposed to accept no more than five characters for one word. The associates of the Eastern Co. expressed their preference that all telegrams in code or secret language be taxed at the rate of five letters per word.

The committee then proceeded to take a vote on the question as to whether the whole matter of code language be submitted to the International Bureau, and this question was voted in the negative, 11 votes against 5, with 2 not voting.

The German delegation then remarked that the Belgian proposal could not be applied to certain languages of which the words, composed of a great number of consonants, are not euphonious. The French delegation, sharing this view, cited an example. "Euphony means 'that which can easily be pronounced.' In the Polish sentence 'Chrzaszcz brzmi w trzcinie,' which is pronounced: 'Krgestch bjmi ve tsiné,' and means 'The maybug is boring into the reed,' are the words euphonious? Assuredly not, and yet these are plain-language words and can not be refused. The qualification that the words must have a euphony *sufficient* for pronounceability is, therefore, not precise enough, and it is necessary to set up another text." As a result, the German delegation proposed to define the possibility of pronunciation by saying that this possibility of pronunciation must exist in only one or the other of the eight authorized languages. Under these conditions, one would admit in code language not only real words, but also artificial words which could be pronounced *according to the usage* of one of these eight languages. Thus it was that a new paragraph 2 of Article 8 came before the committee on regulations in the following terms:

These words, whether real or artificial, are composed of syllables which can be pronounced according to the usage of one of the following languages: German, English, Spanish, French, Dutch, Italian, Portuguese, or Latin.

This wording was approved by the British delegation and by that of the British Indies, the latter stating that they approved, not because they believed this solution to be perfect, not because it was

perhaps, the best possible one,¹ but because it offered a *modus vivendi* for commerce, and regulated a situation full of inconveniences. The associates of the Eastern Co. declared themselves disposed to adhere to the proposal, as did also the Commercial, and the Western Union Cable Companies.

But there were warning voices. The representative of the Great Northern Telegraph Co. pointed out that these proposals merely legalized existing abuses and opened up the door for other abuses. He insisted that in according to the public the liberty of combining any pronounceable syllables whatever up to a limit of ten letters, one could combine in a single word as many as five syllables, each having a secret meaning, that is to say, to unite in a single tax unit as many as five code words. Moreover, he added, the code compilers would see that in order to obtain anything they desired they had only to be insistent enough, and the Telegraph Union would bow to any demands whatsoever.

These views were shared by two other companies, and the French delegation expressed a fear that the authority of the Telegraph Union would be greatly impaired by the fact that it ceased to impose equitable decisions and consented to submit to the unjust demands of a minority.

These voices, however, went unheard, and the Belgian proposal, as amended by Germany, was adopted. The conference in plenary session ratified the decision and the following article was legislated into the regulations:

ARTICLE 8

1. Code language is that which is composed of words not forming intelligible expressions in one or more of the languages authorized for telegraphic correspondence in plain language.
2. The words, whether real or artificial, must be formed of syllables capable of being pronounced according to the usage of one of the following languages: German, English, Spanish, French, Dutch, Italian, Portuguese, or Latin.
3. Code-language words can not be longer than ten characters according to the Morse alphabet.
4. Combinations not complying with the conditions of the two preceding paragraphs are considered as belonging to language in letters having a secret

¹ The only really scientific proposal regarding artificial code language which came before the conference was that presented by the Argentine Government and submitted to the latter by James Nicolson, Superintendent of the Compañía Telefónica del Río de La Plata, Buenos Aires. This proposal concerned the use of artificial words of definite length, composed of regular and systematic alternations of a few specially selected vowels and consonants. Conceived by an able, experienced telegraph engineer, and founded upon thoroughly practical considerations, this scheme, if it had been approved and adopted by the conference, would have profoundly altered the whole future of telegraphic code language, and would most probably have entirely obviated the difficulties that arose after 1903. But the conference unfortunately paid scant attention to Nicolson's excellent ideas. It is interesting to note that at the 1925 Paris conference, as well as at the Cortina conference in 1926, similar proposals, but not nearly so well thought out and scientifically founded, were advanced apparently as new and original contributions. But they were now too late. Seen in retrospect, Nicolson's ideas deserve our full admiration as the concepts of a man who was, perhaps, too far in advance of his times. See *Telegraphic Signals and International Code Vocabularies* (New York, 1897); *Telegraphic Vocabularies Adapted to Telegraphic Signals* (New York, 1902-3); *Consonant-Vowel Vocabulary for Telegrams in Preconcerted Language* (London, 1904); all by James Nicolson.

meaning and are charged for accordingly. However, those formed by the reunion of two or more plain-language words contrary to the usage of the language are not at all admitted.

Let us see what almost immediately followed.

THE BIRTH OF THE FIVE-LETTER CODES

The new regulations were to go into effect on July 1, 1904. In February of that year there was published by an English code compiler a work entitled *Whitelaw's Telegraph Cyphers: 400 Millions of Pronounceable Words*. This was not a code book containing words, phrases, and sentences to which code words were assigned, but merely a list of artificial, pronounceable "words" that could be used for such assignments in private codes. These code groups or "cyphers", as they were termed by the author, were uniformly of five letters each, and here are several examples: FORAB, LUFFA, LOZOJ, FREAN. The subtitle to the work, "400 millions of pronounceable words," is somewhat illusory; before seeing the book one might expect to find 400 million words listed, but actually there were only 20,000 in the list. Where then were the 400 million words? That was easy to answer: since each of the five-letter words was readily pronounceable, and since the maximum length of an admissible artificial code word was, according to paragraph 3 of Article 8 of the regulations, ten letters, then any two of these five-letter words could be run together in one group of ten letters, making a single chargeable word. Thus, two code words could be sent at the cost of a single word. And since any two of these 20,000 cyphers could be joined to form a pair, it was obvious that the total potential number of chargeable words was 20,000 squared, that is, 400 millions, and it was equally obvious (to the uncritical only) that a code employing only these 20,000 groups was equivalent to one of the old type, the dictionary word type, containing 400 millions of words. Of course, the advertised claims of this startling number of code words did not mislead the business men, but the latter saw the real point quickly enough. The immediate possibility of cutting cable costs in half naturally appealed to commercial interests everywhere and the new idea spread like wildfire. Not only did many business houses proceed at once to adopt into their private codes the innovation introduced by *Whitelaw's Cyphers* and the mushroomlike growth of imitations of them, but all code compilers made haste to take advantage of this wonderful boon to their industry, for, as the latter pointed out, no commercial house could any longer afford to retain one of the old-style, dictionary codes, and continue to pay twice as much for its cable correspondence as did the most of its competitors.

As for the general public, and for small business firms whose telegraph correspondence did not justify the compilation of expensive private codes, they too profited. For probably the greatest impetus to the spread of the new code words the world over was given by the appearance in 1906 of "Bentley's Code," a compact phrase code applicable to business affairs in general, selling at a very moderate price. This code gained a wide distribution within an extremely short time, as was to be expected, in view of the economies offered by its employment. The cost of *any* message, even if it were encoded *verbatim* could be cut in half.

There is another very important reason for the rapidity with which these new codes were adopted, and it is that the groups they employed could be transmitted more accurately and could be corrected much more easily if mutilated in transmission, than the old-style, dictionary words. It must be realized that this meant a great deal to firms conducting a regular and voluminous overseas telegraph correspondence, because to such firms rapidity and accuracy in telegraphy are primary considerations; slight or even marked differences in rates and costs are secondary. Unless a cablegram or a telegram is handled expeditiously and *accurately*, they feel it may as well not be sent, and the question of costs, within reasonable limits, of course, does not enter into the calculations.

THE ADVANTAGES OF THE FIVE-LETTER CODES

Now these new code words very quickly demonstrated themselves as being much superior in respect to accuracy to all the old types, either *bona-fide* or counterfeit dictionary words, root and terminal systems, syllable-combination methods, and code-condenser devices. In the first place, the fact that all the code words contained exactly the same number of letters was a great aid to accuracy for all concerned—for code clerks in the offices, as well as for the telegraph personnel; the mere knowledge that each group had to contain a definite and unvarying number of letters, no more and no less, was a very great help. In the second place, the code compilers who devised the new system of code words realized from the very start that if the principle of the two-letter difference between code words was valuable in connection with the old type, dictionary words, it was *absolutely essential* in the new type, since the appearance of these words could hardly serve as a guide in their accurate reception by telegraph, except in a most general way. Consequently, we find that *Whitelaw's Cyphers* embodied the principle of the two-letter differential, as did most of the cyphers of the experienced code compilers who followed his lead. Only in exceptional instances were there published five-letter codes which employed words showing only a single-letter difference, and these codes, we may be sure, were compiled by novices, in firms

that did not appreciate the fact that code compilation is not a playground for amateurs, but had scientific and technical ramifications just as other sciences.

Furthermore, the new codes embodied tables by means of which code groups mutilated or garbled in transmission could be rectified quite readily by the recipient, without having to ask the telegraph service for a verification. Originally these tables consisted of lists in which the code words were arranged alphabetically according to their backward spelling, that is, from the end of the word to the initial letter. It is unnecessary to go further into detail in this respect, for this method of listing the words in terminational order was soon superseded by a much more scientific and practical system involving the use of a chart designated by various names such as "mutilation detector," "permutation table," "error detector," etc. Such a chart, usually on a single sheet of paper, very readily facilitated correction of mutilated code words, and their inclusion into codes soon became a standard practice.

One more item in connection with the greater accuracy of the artificial five-letter groups must be mentioned. Anybody who has ever played the old-fashioned game called "anagrams" knows how often it is possible to change one word into another by a mere inversion or transposition of the two members of a pair of adjacent or alternate letters. For example, a transposition of the adjacent letters T and I in the word MARTIAL results in forming the word MARITAL; a transposition of the alternate letters O and A in the word CORAL results in CAROL. In code communication such inversions or transpositions are by no means rare; they are due, as a rule, not to mistakes in telegraphy, but to errors in writing, psychological inadvertences, or slips of the pen. If a code contains words that can be altered, by transpositions of this nature, into other words also present in the code, sometimes great errors can result. In codes of the dictionary-word type, there is no way of making sure that one or the other of such words is not included, except by a tremendously laborious examination. But in codes of the artificial-word type, if the permutation table by means of which the code words are made has been scientifically constructed, transpositions of adjacent letters in one code word yielding another code word also present in the same code can be completely avoided. For example, if the word BACED is to be a legitimate word in the code, then the table will automatically prevent the formation of any of the following groups differing from it by a mere transposition of the two members of any adjacent pair of letters: ABCED, BCAED, BAECD, or BACDE. The possibility of accomplishing and incorporating this valuable technical feature without at the same time causing an appreciable reduction in the total number of groups that can be constructed was not realized at first,

but it was soon recognized and adopted by the most competent code compilers.¹

So many were the advantages of these new codes, and so great was the degree of economy they made possible, that in an extremely short time they began to replace all other types of codes used in cabling, the world over. Codes employing dictionary words were discarded by reason of the greater economy and accuracy afforded by the five-letter codes. Root and terminal systems were discarded because of the limitations inherent in them, since a terminal word could never be used as a root, nor could a root be used as a terminal. In the new system any word could be used either as the initial group of a ten-letter word, or as the final group, so that the writing of a telegram suffered no restrictive limitations of verbiage due to limitations of the code system. Syllable-combination codes as well as condenser systems were discarded because of their susceptibility to errors, and because the new codes afforded almost as great economies. Five-letter codes, unknown in 1904, became within the short space of a half decade the most important of all codes used in international telegraph correspondence. Most of the new codes, in fact the great majority of them, were compiled and printed by American and British code makers, and English was the language used in their vocabularies.

Soon code compilers began to vie with each other in the production of new, and *more extensive* codes, for it is obvious that the larger and the more extensive the vocabulary is, the greater the degree of economy effected by the code. But larger codes require more code words to designate their contents, and that is how the troubles which we are now about to consider started. It will be well to state the matter clearly and in detail.

THE DIFFICULTIES INTRODUCED BY THE FIVE-LETTER CODES

As stated above, these codes all embodied the principle of the two-letter differential. Had this principle not been an absolute necessity and an essential adjunct to their successful employment, the situation which soon arose would never have come to pass, because the number of five-letter, easily pronounceable, *single-letter-difference* groups that can be constructed using the 20 consonants and the 6 vowels is far in excess of the number required for even the largest and most comprehensive codes. For example, using five-letter code words of but one form, viz., consonant-vowel-consonant-vowel-consonant (BACED), and without incorporating the two-letter differential, there is available a series of 288,000 groups, of which about 200,000 are easily pronounceable. But if the two-letter differential is incorporated, then the total available number becomes reduced to a mere 14,400, of which but about 11,000 are

¹ Transpositions of alternate letters can be avoided to a large extent, but this is not so important as the suppression of transposition of adjacent letters.

easily pronounceable. This number is, of course, entirely insufficient for the average private or public code, and therefore code words of other types of arrangement of vowels and consonants must be constructed.¹

Now given uniform-length groups of five letters, with no limitations as to the number of vowels or consonants permitted in each group, the total number of different *types* of alternation or arrangement of these vowels and consonants is 32. But only a few of these types may be considered to yield easily pronounceable words; furthermore, in one and the same type of arrangement there will be many cases where the groups are easily pronounceable, but thousands of others where they are utterly unpronounceable. For example, in the type consonant-consonant-vowel-consonant-consonant, the group GRIST is perfectly pronounceable, and GPISR is admittedly unpronounceable. But what should be said of a group such as GWYJP? Obviously that will depend almost entirely on who is to be the judge in the case, who is asked to render the opinion. And the difficulty that became apparent as soon as these code words presented themselves at the telegraph counters for decision was that the judge, the charging clerk, had as his legal guide only that very vague specification: "The words, whether real or artificial, must be formed of syllables capable of being pronounced according to usage in one of the following languages: German, English, Spanish, French, Italian, Dutch, Portuguese, or Latin."

"PRONOUNCEABILITY": WHAT IS IT?

Let us make a searching analysis of the wording of paragraph 2 of Article 8, to see what elements of precision, if any, were contained in it; perhaps we may in that way come to understand why the rule was entirely inadequate to meet the situation that resulted.

First, the rule prescribed, "the words . . . must be formed of syllables," and hence we are led directly to inquire: "What is a syllable?" That this is not so simple a matter as might seem on its face, let us quote from standard dictionaries. This is what *Webster's New International Dictionary* states:

Syllable: 1. An elementary sound, or a combination of such sounds, uttered together with a single effort or impulse of the voice, and constituting a word or a part of a word. In other terms, a syllable is a vowel or a diphthong, either by itself or flanked by one or more consonants, the whole produced by a single impulse or utterance. Certain consonants, as l, m, n, r, may fill the place of a vowel in forming a syllable. Adjoining syllables in a word or phrase need not be marked off by a pause, but only by such an abatement and renewal, or reinforcement, of the stress as to give the feeling of separate impulses.

2. In writing and printing, a part of a word separated from the rest and capable of being pronounced by a single impulse of the voice. It may or may not correspond to a syllable in the spoken language.

¹ See p. 65, footnote, and p. 75.

Funk and Wagnall's Standard Dictionary defines the word as follows:

Syllable: 1. A single or articulated vocal sound; that which is uttered in a single vocal impulse; also, the characters or letters that represent such a sound; a word or part of a word that is capable of separate and complete enunciation by one voice-impulse.

A syllable may be a single vowel or diphthongal sound, or one of these preceded or followed by one or more consonants, or both.

The fact that the word "syllable" has in reality two meanings pertinent to our subject of inquiry is important to note. First, a syllable is a single or articulated vocal sound, and secondly, it is also the character, characters, or letters representing such a sound. Thus we are immediately brought face to face with such very complex technical subjects as orthoëpy, orthography, and phonology. Now if all of the eight languages specified in paragraph 2 of Article 8 were substantially phonetic, that is, if there existed a very close parallelism between the pronunciation or the sound of a syllable and its spelling, the provisions of the paragraph in question would perhaps have presented no serious difficulties. But let us consider only one of the specified languages in this respect, viz., English. That the pronunciation of thousands of words in this language can not be exactly shown by means of our alphabet, that is, by our spelling, is adequately attested to by the lengths to which lexicographers are driven in "respelling" words, and adopting the extremely strange and confusing systems of diacritical markings, such as dots, curves, numbers, and other signs, or of introducing new and unfamiliar letters to designate certain sounds. In other languages, such as German, Italian, or Spanish, the maker of a dictionary has rarely any need to adopt such tactics in order to indicate how a word is to be pronounced, and it is in reality an absurdity that special pronunciation indicators should be necessary for any language, since the original and proper function of spelling is to tell how words are to be pronounced. The English language is indeed unfortunate in that in this respect it shows a greater degree of arbitrariness, irregularity, and inconsistency between pronunciation and spelling than does any other language of civilized peoples. In fact, although in the early periods of English the spelling was substantially phonetic, in modern English the spelling is so far from phonetic that in some cases it scarcely even suggests the pronunciation. Note, for example, such words as "busy", "bosom", "colonel", "bough", "enough", "knight", and so on.¹ How different this situation is in certain others of the eight languages, Italian, for example, wherein practically all words are pronounced as they are spelled. And, since Italian has been mentioned, what idea can the average inhabitant of that land have as to the pronunciation of our word "awkwardly", when their own alphabet entirely lacks the letters K, W, and Y (as well as X)?

¹ Note also "ough", pronounced "coff", and "hiccough", pronounced "hiccup".

So far as the mere question of pronounceability is concerned in its relation with the subject we are considering, a mere deviation from what is usual in pronunciation would be of no moment. What does it matter to a foreign telegraph clerk if, for example, he does not know that the vowel "o" in our word "women" is pronounced like the "i" in "with"? But the utter inconsistency between the pronunciation and spelling in the case of some words, as for example, "wrought", "draught", and "eighths", is, of course, a source of difficulty, because certain of the letters that appear in words in English have a multiplicity of sounds, certain others are often entirely silent and take no part in the pronunciation, while in hundreds of cases the same sound is indicated orthographically in a multiplicity of ways.¹ Yet, technically, all these letters form parts of syllables, and if such syllables are employed in composing artificial code words they impart to such words the most outlandish of appearances, making them seem unpronounceable. If we combine syllables taken from certain *bona-fide* English words we obtain aggregations of letters seemingly difficult or impossible to pronounce, as for example, XYSNIGNUMN, ALLLLAAALA, and AWKMNEPNEU, every syllable of which can be accounted for² according to the provisions of paragraph 2 of Article 8.

It is commonly understood that there can be no syllable without at least one vowel, and, in the vast majority of cases, that is true. But there are certain letters of the English alphabet, as well as in other alphabets, which, although they normally act as consonants, in some cases are technically considered to serve as vowels. This applies particularly to the letters L, M, N, and R. For example, in the common word "BATTLE", the letter L has the value of a vowel, while in the word "BATTling" it has its ordinary consonantal value; in the word "PRISM", the letter M is "syllabic M", whereas in "PRISMATIC" it acts as the ordinary consonantal M; in the word "GIVEN", the N has a syllabic value and forms a syllable by itself, for the word is pronounced "GIV'N", and not "GIV-EN"; and as for the letter R, when it is not followed by a vowel it serves as a vowel itself, as, for example, in the word "PURE", which is perfectly correctly pronounced "PYUAH". It is, therefore, not difficult to understand that code compilers could, with some show of authority, say that an artificial code group like MLNRM is pronounceable, even though it is composed of nothing but letters

¹ For example, one authority has calculated that the pronunciation of the word "scissors" could be indicated in over 6000 different spellings, each having an analogy in some other legitimate word in the English language.

² Syllable "XYS" from the word "XYSMA"; "NIGN" from the word "BENIGN"; "UMN" from the word "COLUMN"; "ALL", a common English word, "LLA" from the word "LLAMA"; "AAL" a red dye (*Standard Dictionary*), and "A" a syllable by itself; "AWK" from the word "AWKWARD"; "MNE" from the word "MNEMONIC"; "PNEI" from the word "PNEUMONIA".

commonly regarded as being consonants. Moreover, *Webster's International Dictionary* says:

Consonant means "sounding with,"—that is, "with " a vowel. But it is not to be supposed that consonants can be uttered only as combined with vowels. Every consonant can be sounded by itself, and such voiced consonants as L, M, N, can be not only indefinitely prolonged, but even musically intoned; or, in other words, they can function as vowels. . . . Similarly, R can function as a vowel. . . . It is thus seen that the line of demarcation between vowel and consonant is not always sharply defined. It is, however, convenient to retain the traditional division of the sounds of the language into vowels and consonants. The respective relations of vowel and consonant to the syllable are a natural consequence of their different characteristics as stated above. Vowels and consonants readily combine into one continuous blend of sound in passing either way, from vowel to consonant or from consonant to vowel.

The *Century Dictionary* says:

Vowel and consonant are relative terms, distinguishing respectively the opener and closer utterances; but there is no absolute division between them. Certain sounds are so open as to be only vowels; certain others, so close as to be only consonants; but there are yet others which have the value now of vowels and now of consonants. Thus, *l* and *n* have frequently vowel-value in English, as in *apple*, *token*; and *r* is in various languages a much-used vowel. Also, the semi-vowels *y* and *w* are not appreciably different from the *i*-vowel (of *pique*) and the *u*-vowel (of *rule*) respectively. A sound, namely, is a vowel if it forms the central or open element of a syllable, being a syllable either alone or in conjunction with the closer sounds (consonants) that accompany it.

When it is realized that this relation of vowels and consonants to pronounceability is a rather flexible one even in the case of *bona-fide* words, what could be expected in the case of artificial words, especially when orthoëpists admit, as noted above, that there is, in reality, no sharp line of demarcation between vowels and consonants and that the traditional division of the sounds of the language into vowels and consonants is retained merely as a convenience?

There is, in addition, another loophole which code compilers could avail themselves of, in regard to syllables. We have seen that the word "syllable" is defined primarily as "an elementary sound, or a combination of such sounds, uttered together with a single effort or impulse of the voice." According to this definition, is the word "SPRY" a single syllable, or is it composed of two syllables? *Webster's Dictionary* says:

Minor superiorities of acoustic prominence are generally not noticed; therefore, such words as SPRY, SKY, RATS, are treated as monosyllabic, though the unsonorous stops P, K, T, really break each of the words into two sound groups.

This can be interpreted, by only a slight straining of terms, as an admission that the combinations SP, or SK, or TS, standing alone and without any vowels, act as syllables, and therefore that an artificial code group such as SPRSK is pronounceable.

Of course, most of the difficulties of pronunciation arise when sequences of consonants appear. But even a sequence of vowels, each of which may be a syllable by itself, can give rise to difficulties. Consider our word QUEUE; does the spelling indicate fully the way in which it is to be pronounced? What of the Dutch word ZEEEENDJE? And what shall we say of a code group like UYEAO—is it pronounceable as a *word*, or does its pronunciation involve merely the sounding of each vowel separately?

After specifying that "the words, whether real or artificial, must be formed of syllables", paragraph 2 of Article 8 goes on to state that these syllables must be "capable of being pronounced." We have seen how the very nature and concept of a syllable involves the whole question of pronounceability, and therefore the added specification in this respect brought no more precision to the paragraph in question than was already inherent in it due to the requirement that the words must be formed of syllables. Perhaps we should take it in connection with the phraseology that followed: "according to the usage of" Again a question arises: What is meant by the phrase "according to usage"? And again we turn to the dictionary: (*Standard*)

Usage: 1. The manner of using or treating a person or thing; treatment; also the act of using; as, mild or harsh *usage*. 2. Customary or habitual practise, or something permitted by it or done in accordance with it; custom or a custom; as, an act permitted by *usage*; ancient *usages*. . . . 4. *Gram. & Rhet.* A use of words or forms considered as enacted or unenacted by reputable authorities.

Presumably, the fourth definition is the one upon which we should lay the most emphasis in connection with our subject. Although the wording "according to usage" did not specify whether the usage applicable in this case was to be of the sanctioned or unsanctioned type, certainly the former was meant. But, as we shall see later, what constitutes "good usage" is not a very simple matter, and, moreover, it is one that is subject to change, just as fashions in dress and so on are subject to change, though perhaps not so violently as in the latter. Furthermore, in the context in which it appears, what does the phrase "according to usage" mean? Does it mean that each syllable used in a code word *must actually appear* as such in a *bona-fide* word, or does it merely mean that the syllable can be considered pronounceable by *analogy*? For instance, should the group of letters OHMSC be considered as a syllable *only* if the same sequence can be found appearing as such in a *bona-fide* word in any one of the eight languages, or should it be considered a syllable if it can really be pronounced according to analogous use of the sounds indicated by these letters in other words, OHM in the word "OHM", and SC in the word "DISC"? Certainly it can not be said that the

rule as it was written indicates which of these two interpretations is the one intended.

Finally, as to the last specification in paragraph 2 of Article 8, relative to usage in any of the eight different languages, it seems hardly necessary to point out that if the complexities we have indicated as being present in only one of the languages mentioned, English, are not to be multiplied by eight in order to gain some measure of the total difficulties, a factor of even two brings us to a realization of what complexities the telegraph charging personnel were expected to resolve.

Enough has been set forth, it is believed, to show that a detailed interpretation of the wording of paragraph 2 of Article 8 involves technical complexities that would tax the knowledge and experience of expert phoneticists, orthoëpists, orthographers, and lexicographers. This being the case, what could have been expected of the poor charging clerks at the telegraph counters, men whose linguistic knowledge is as a rule limited to but one language, but who, if they were to be placed in a position to enforce the provisions of the paragraph in question, should have had a profound linguistic knowledge sufficient to encompass the vagaries of eight different languages? The whole thing, we are forced to admit, was built manifestly upon an illogical, unscientific basis from the beginning, and we must not be surprised, therefore, at the difficulties to which the paragraph in question gave rise.

INCONSISTENCIES IN PRACTICE

Furthermore, to add to the illogic and inconsistency in the situation, whenever a code word did occasionally present in part an unquestioned pronounceability due to the accidental juxtaposition of code letters into a sequence coinciding with a *bona-fide* dictionary word, such a code word was taxed as being *two* words. For example, if the code group OTSER by coincidence preceded the group EPORT, making the ten-letter group OTSEREPORT, such a word would count as two words, one for the code group OTSE, another for the dictionary word REPORT. The reason for this was that according to paragraph 4 of Article 8, as adopted by the London conference, combinations not complying with the provisions in respect to code words were to be considered as cipher language, charged for at the rate of five characters per word. Moreover, there has always existed a prohibition against the alterations, reunions, or combinations of words contrary to the usage of language.

Evidently what the Telegraph Union had in mind when it adopted these rules was to put a stop to the running together of short words in order to reduce the number of chargeable words, as, for example, "AREYOUSURE", "WILLBUYFOR", and so on. But an attempt

was made to interpret the rule to apply to any ten-letter code word whatever, if it is composed of one dictionary word and one code word, or two or more dictionary words, no matter whether such dictionary words happen to be merely accidental or whether they are used in their dictionary sense in the position in which they occur. Of this, one of the best known code compilers complained, in a letter to the British Post Office, in the following terms:

If the interpretation already mentioned is insisted upon, it destroys the efficacy of the rule, as it is simply impossible to provide safe pronounceable words (having written and telegraphic dissimilarities) in sufficient number to meet the needs of the public, without in some cases employing groups of letters which may form a combination of two simple words, in one or other of the eight languages, for instance, "SUJETAROMA". "Sujet" is French and "Aroma" English.

This complaint was certainly a justifiable one, and reference to almost any five-letter code will disclose hundreds of cases which fall under this interpretation purely as a result of coincidence. In 1914, however, an interpretation of the rules was voted upon by the various administrations of the Union, and it was decided that whenever a *bona-fide* dictionary word was *fortuitously* formed by the combining of two code words, and when it was obvious that such a *bona-fide* word was not intended to be conveyed in its dictionary sense, the usual ten-letter count would apply. But, from the point of view of telegraph operation itself, what objection would there possibly be to such *bona-fide* word formations, *whether accidental or intended*, seeing that it would be perfectly possible in almost every case to substitute a less euphonious code word for the real word? For example, in a list found in the Cortina report there appears, among many other cases, this one: UMIRIKILOS. The chances that the code employed does not contain a code equivalent for the word KILOS are so remote as to be hardly worth mentioning; a code word could have been employed, for instance, SKOZH. Nobody will contend that UMIRISKOZH would be easier for the telegraph personnel than UMIRIKILOS, and it is difficult to see wherein complaints directed against the practice of combining a code word with a plain-language word, providing the whole does not exceed ten letters, are fundamentally sound or logical.¹ If a strict interpretation of the rule in question were enforced, all code compilers would have to do what one of them did: He changed or deleted such words as AMADR and ABBIT, because if run together they would read "A MAD RABBIT", chargeable as three words; BABYL and OCKED would read "BABY LOCKED", two words, and so on. This code compiler was fortunate in that he needed but a little over 3,000 five-letter words, and could therefore afford to delete many of the type indicated.

¹ Even the rule prohibiting the reunion of words in plain language has broken down, and over some circuits of the international system a group such as AREYOUSURE counts as but one word.

As has been said, the necessities of the new code language demanded the inclusion of vowel and consonant arrangements of various types, in order to have a sufficient number of code words for the ever expanding codes. If at first 20,000 two-letter difference groups were enough, as Whitelaw thought, this number could be produced without much difficulty, and without bringing any doubts into the matter of their pronounceability. But Bentley, in 1906, needed 32,000 groups, and soon this number was not enough. Presently, 50,000, 60,000 and more groups were considered insufficient. Obviously, such large numbers of groups, embodying the two-letter differences as they had to, could be produced only by considerably stretching the rather vague and, in reality, the definition-defying condition called "pronounceability". To add to the difficulties, while in many cases two five-letter groups might individually be considered as coming within the borderline of pronounceability, yet when they are united into a ten-letter group the ensemble becomes apparently unpronounceable.

It is to be remarked, however, that the question of pronounceability has always been one with which the telegraph services and offices are a great deal more concerned than are the users and compilers of codes. Further on we shall have something more to say on this question.

REACTION OF THE COMMUNICATION AGENCIES TO THE NEW RÉGIME

Let us see now how the telegraph administrations and companies met the situation that began to confront them shortly after the 1903 regulations became effective. We have seen how the telegraphing public took to the new types of codes. If the users of codes and the senders of cablegrams received these new codes with open arms, this was far from the case as regards the telegraph administrations and companies. At first, there arose a stubborn resistance to the acceptance of the new code words, and loud lamentations were heard from the telegraph services. The basis of their opposition and protests was threefold.

In the first place, they felt that the new system constituted merely a subterfuge and an evasion of the regulations, made possible by a lack of precision in the wording of Article 8. This contention was justifiable. For it was perfectly clear that what the 1903 conference had in mind when it adopted this article was to permit the use of individual, singly-chargeable, pronounceable combinations of letters which, by virtue of alternations of vowels and consonants, *would very closely resemble bona-fide, dictionary words*. But as soon as the new regulations became effective, all the members of the Telegraph Union, to their utter astonishment and chagrin, found that instead of legalizing, as they thought, merely the usage of words similar to those cited to the conference by the proponents of artificial code language,

euphonious combinations of vowels and consonants such as "DEMI-NABAM" and "OPORTAVERE", they had in reality legalized the usage of not only these euphonious and unobjectionable words, but also such outlandish assemblages of vowels and consonants as EYYHOGORGT, BEEUHDWEWF, and ZAQRUARSC. And when the compilers of the codes, and the senders of telegrams containing these difficult words, were reproached with infractions of the regulations, they merely referred to the letter of the rule itself—with the intent and spirit of the rule *they* were not concerned, they averred. Were not their words formed of *syllables* capable of being pronounced according to usage in any of the eight languages? Certainly, they were; for each doubtful syllable they could cite a word in usage in *one or the other* of the specified languages. The rule, they pointed out did not say that the words thus formed must resemble real words; nor did it specify that all the syllables in each word had to be pronounceable in only one and the same language of the eight permitted. The fact that in a given word there might be three or four syllables each of which might be pronounceable in a different language did not make these words illegitimate, they declared.

Arguments such as these, we must admit, were logical and incontrovertible. But this did not help the telegraph services, for whom the new words increased the seriousness of the difficulties and problems of code communication, instead of lessening them. This was the primary basis of their opposition and protests.

A secondary basis for their distaste for the new condition of affairs was the loss of time and the engendering of disputes which the acceptance of and charging for code telegrams occasioned at the telegraph counters, on account of the doubtful pronounceability of certain words. For in each case, such words would be accorded not the code count, at the rate of ten letters per word, but the cipher count, five letters per word. Hence, in order to convince the counter clerk on the score of pronounceability the senders of telegrams would, as expressively described by a telegraph official, "manage by means of facial contortions to pronounce words containing up to four or five consecutive consonants," and thus engage in acrimonious and time-consuming disputations. In addition, there was the irritating and entangling business of collecting supplementary charges, either from the recipient of a telegram, or the sender, or the office of origin, and of keeping the international accounts relative thereto in good order.

Finally, there was, but only for a very short period, the matter of reduced revenues, due to the fact that the charge for telegrams employing the new code language was only half of what it really should be, since each ten-letter word was in fact two code words. But the fears entertained on this score must soon have dispelled themselves when the total traffic handled by practically all adminis-

trations and companies increased very markedly, due to the lowering in cost to the public, not as a result of a decrease in unit rates, but as a result of the new code groups, which permitted two words to be transmitted at the cost of one.

THE LISBON CONFERENCE OF 1908

In 1908 the Telegraph Union held a conference in Lisbon and again code language formed one of the most important subjects of discussion. In the words of one of the French delegates, "it was obvious that the solution adopted by the conference of London had not attained the object desired, and that the language euphonic had become the language cacophonous." A proposal of the Belgian delegation to tax telegrams according to the *number of letters they contained*, at the rate of five letters per word, was seriously considered and rejected as being impractical. The Russian delegation proposed that the provisions of the regulations relative to artificial words be made to read: "Artificial code words must be formed of syllables whose pronunciation is easy." It is hardly necessary to say that this proposal was not seriously entertained. The British delegation proposed to replace the paragraph in question by the following:

The words, whether real or artificial, must be formed according to the ordinary usage of one of the languages German, English, etc., in a manner capable of being pronounced according to the ordinary usage of the language in question.

The motive, as declared by the British delegation was to modify the wording of the regulation in a manner so as to specify that the rule prescribing conformity with the usage of one of the eight languages should be based on the *ordinary* usage and that it should be applicable to the whole of the word in its ensemble and not solely to each individual syllable. But the proposal as given above was likewise rejected. The subject was then referred to a subcommittee for study.

This subcommittee recognized that in certain codes the artificial words were formed in a manner that was not intended by the London conference, and that the surest means of excluding combinations which did not correspond with these intentions, and to prevent abuses, would be to adopt a more precise and a more strict wording of Article 8 of the regulations. But in order to avoid the inconveniences which would be brought to commerce if that were done, the subcommittee recommended another means which, in appealing to the interests of code compilers, would assure their full cooperation, as well as the commercial public, by way of applying in advance the provisions of the regulations in a spirit more in conformity with that which brought them about. This recommendation was to accord to the public the facility of obtaining approval, by the Telegraph Union, of the codes they wanted to use, in order to have the assurance that

they conformed to the provisions of the regulations. In case a code did not receive this approval, each administration would remain the judge, just as was then the case, of the admissibility of each of the words it contained; at the same time, in using an approved code the public would have the guaranty that none of the words of such a code would be rejected by any administration whatever as not conforming to the regulations. The subcommittee further recommended that the work of examining codes be entrusted to the German, French, and British administrations acting in concert. With the idea that these proposals, if adopted, made a complete change in the wording of Article 8 then in effect unnecessary, the following new paragraph 2 of Article 8 was drawn up by the subcommittee and submitted to the next plenary session of the conference:

ARTICLE 8

2. The words, whether real or artificial, must be formed of syllables capable of being pronounced according to the *current* usage of one of the following languages: German, English, Spanish, French, Dutch, Italian, Portuguese, or Latin. Artificial words must not contain the accented letters *ã, à, â, é, ê, ñ, õ, ü.*¹

Codes designed for correspondence in code language can be submitted to the telegraph administrations designated for this purpose, with a view of permitting those interested to obtain the assurance that the words contained in these codes comply with the conditions of the present paragraph.

It will be noted that the *sole* change in the first part of this paragraph over the wording adopted by the London conference in 1903 was the insertion of the word "current" before the word "usage".

"CURRENT USAGE": WHAT IS IT? COULD IT SOLVE THE DIFFICULTIES?

When this slight change in the wording of paragraph 2 came up for discussion in the tariff committee, the Russian delegation declared that they did not see the necessity of modifying the regulations by substituting the words "current usage" for the much shorter "usage", and that they did not understand the import of this change. The British delegate remarked that the best way of answering the question raised was to cite an example, which was as follows:

A code maker submitted to the British administration a word in French, which appeared, moreover, in the vocabulary of the International Bureau.

This word is "MDOLL", which, according to the Larousse dictionary, means: "a very high, cocked hat, with wide brims, made of palm leaves and ornamented with ostrich feathers, used among the desert Algerians."

This word, although taken from one of the eight languages submitted for code language, is not in current use in French; code makers ought not be able to invoke an exceptional usage of words of this type as authorizing the employment, in a general way, of the letters MD as a base for forming artificial words.

¹ This last sentence was added by the conference at the second reading of the article.

It is not surprising that this explanation did not satisfy the Russian delegate. For, in truth, if mere "usage" is a debatable issue, "current usage" is an even more debatable one. What is "current usage"? The following is what the *Standard Dictionary* has to say on this score:

usage is infinitely various, so that the whole of it can not possibly be recorded. It varies with the time and place, with the culture and social status of the individual, with the speech-habit of the community.

change is the law of life for a spoken language, and usage is therefore never absolutely binding. We call it binding in proportion as it has endured for a long time and has been accepted by everybody. Thus arise the so-called laws of language—brief, summary statements of accepted usage. Whatever runs counter to those laws must be regarded, provisionally at least, as incorrect or bad; and it is the plain duty of the lexicographer to state the fact as it is. But all things are in a flux, and that which is bad to-day may become good to-morrow. A fashion may spring up in defiance of previous usage, of logic, even of common sense; if it prevails it is a part of the language, and must be recognized as such by the lexicographer. But when may a usage be said to have prevailed? After a decade, or a century, or five centuries? How large a following must it have to give it entire respectability? To such questions no precise answer is possible. The dictionary is a record of fashion, and fashion varies with the time and place.

To these words we may simply add that, even brushing aside uncertainties as to whether this or that word should be considered as being in "current usage" or not, as in the case of the word "MDOLL", there are in the English language alone hundreds of words in daily use to-day which, if employed as sources for the selection of syllables to form artificial code words, yield difficult and impossible situations as regards pronounceability. For example, let us take two words: "psychology" and "phthisis", both in current usage among intelligent English-speaking people. If we take the syllables "psych" and "phthi" as code words, and run them together into a ten-letter code word we get PHTHPSYCH or PSYCHPHTHI. Are these pronounceable? It all depends upon who is asked. If an illiterate person is asked, one who has never heard either the word "psychology" or the word "phthisis" pronounced, he would be at an utter loss to vocalize the code words formed from them, as indicated above. Even an educated English-speaking person might well hesitate before such an assortment of consonants. What shall we say of the telegraph personnel in foreign countries in this respect?

Thus it can not be seen how even a very strict interpretation of what is to be meant by "current usage" would ameliorate conditions. In the introductory remarks to a list of artificial code words compiled by one of the most scrupulously conscientious code compilers, there appears the following statement:

I have taken the official vocabulary and *Standard Dictionaries* as my guide and can guarantee that every group of five letters not only consists of syllables in *current use* in the eight languages, but they can be verified by the official vocabulary.

The search for euphonious syllables has been a long, tedious and trying piece of work. Many a one has been found only after much research. Take for instance "ryb" found in "Eurybates"; "vus" found in "Corvus"; "ams" found in "broughams"; "yb" found in "Molybdena".

Yet, in his compilation of only 21,323 five-letter words will be found many cases wherein the reputed "pronounceability" would be of little assistance to telegraph personnel. For example, EQGUG-UXJIK and JIUPHIQHUG are two ten-letter words that would be formed by combining five-letter ones found in his list.

In the 1912 Berne Bureau series of articles on code language there appear some remarks on the work of the Lisbon conference in regard to this matter of "current usage," remarks which are well worth quoting and comment. The writer of this series of articles, M. Crescitz, said:

The Conference of Lisbon resolutely discarded all proposals having as their object the modification of the régime applied to code language. As to those which, in fully maintaining the régime in effect, tending to reinforce the regulations, it adopted them only in part.

It is thus that in paragraph 2 of Article 8 the conference replaced "The words must be formed of syllables capable of being pronounced according to the usage" by "The words must be formed of syllables capable of being pronounced according to the *current usage*."

This modification offers the advantage of more clearly indicating the intentions of the Union as far as pronounceability is concerned. It was adopted in order to suppress bizarre syllables which can be found exceptionally in some real words, but which are not in current usage.

In that connection the word MDOLL was cited as an example. We would like to remark that even with the old text this word should have been refused as a code word.

Since the word MDOLL figures in a French dictionary (*Larousse Dictionary*), it must be admitted that, from the point of view of telegraphy, it belongs to the French language, even though the dictionary in question cites it as a word of the Arabic language.

As a consequence, until the London conference this word could legitimately have been employed as a code word, since it is a real word and the regulations antedating that conference admitted all real words of an authorized language, without imposing any condition as to pronounceability.

But, since then, the situation has completely changed: the London conference, in admitting artificial words, imposed on all code words, *real words* as well as *artificial*, the obligation of being pronounceable, according to usage, or at least the obligation of being formed of syllables capable of being pronounced according to usage.

Hence, since July 1, 1904, if a *real word* is composed of syllables not capable of being pronounced according to usage, such a word ought to be refused as a code word; in spite of its being a real word.

Such is our opinion in regard to the word MDOLL.

Admitting the force and logic of this argument, does it help to relieve the situation in any respect? Has it not been shown that even in words in everyday use in English there are syllables which,

when employed as bases for the construction of artificial code words, lead to difficult if not impossible situations?

Again, a bit further on, the quoted remarks continue as follows:

The deliberations of the London conference left no doubt whatever as to the intent of the Telegraphic Union: it wished to authorize the use of artificial expressions resembling real words and capable of being telegraphed with as much facility as the latter. But, if the regulatory provisions are insufficient to make the public respect these intentions, definite even as they were legitimate, it is that the provisions are drawn up in the regulations in a very much less explicit manner.

Under these conditions it appeared logical and useful to give them preciseness and to complete them. Such, it seems to us, was the purpose of the proposition.

One could even contend that the proposal was superfluous, for the pronounceability of words goes without saying.

The regulations speak of *artificial words* and not of *groups of letters*. But one of the characteristics of a *word* is precisely that of pronounceability: by definition, a *word* must be capable of being pronounced according to certain rules; pronounceability distinguishes the *word* from the *group of letters*.

Thus the proposition reduces itself to a precise text and puts its true meaning in a clear light:

Likewise, in commenting upon these remarks, granting the force and logic of this argument, does it help matters? If "pronounceability" were a condition susceptible of physical measurement, if it were a thing that could be weighed, a thing or a condition to which definite, unvarying, standard constants or measures of precision could be applied, the weight of the argument would be greatly increased. But is such a thing possible? In the light of all that has been here pointed out, it is believed that further comment is unnecessary.

It becomes obvious, then, that the addition of the word "current" to the word "usage" in paragraph 2 of Article 8 by the Lisbon conference could not possibly have served to ameliorate the difficulties of code language acceptance, charging, and transmission. In order to enforce the provisions of the paragraph in question, telegraph personnel would have to be intimately familiar with all words in "current usage"—whatever that is—not only in one language, but in eight. And we have also seen that even syllables taken from words in current usage to form the base for code words can give rise to assemblages of letters difficult or impossible to pronounce. It is not surprising, therefore, that no improvement whatsoever, in respect to this matter of pronounceability was noted after the 1908 regulations became effective, and that conditions went from bad to worse; as code users and code compilers, not satisfied with 65,000 or 75,000 groups, extended themselves to include 100,000 or more in a single book.

Code words of the form shown in the following list (a few of a larger collection to be found on page 61 of the Cortina report, taken from actual telegrams handled by offices of the Netherland telegraph administration on three days in December 1925) became widely prevalent:

YIDENVOHFK
LIYFKBVEVA
BYMGVRASMB
LNAGMYMJMY
YLGMPFAHVY
UYSPHGEJMB
GOPBLJOURI
IWCCAEBWML
VYENVZMOOA
IAMKHZVESK

NHAZCBLYOJ
IJWREOPGTS
IPMYOONPSH
EAPWUYPKTS
AHGUPINJBL
VYLDZZMORE
WRIETYPKTS
OXUZHVGUIJ
XUOXMPWUCY
ETC., ETC.

Beside each code word in the list mentioned, the Netherland administration has indicated, under a heading labeled "Unpronounceability of consecutive consonants," the sequence of letters which it considers difficult. While in very many cases it must be admitted that the sequences indicated are certainly unpronounceable, yet in some cases, though the sequences are formidable in appearance *when lifted out of their relations with the remaining letters*, the words in which they occur are not, in reality, wholly unpronounceable, if one takes into consideration what has been pointed out herein on the subject of the syllable. For example, take the case of the very first word in the list: YIDENVOHFK, of which the sequence HFK is indicated as unpronounceable. Is it legitimate to lift the HFK out of its relation with the vowel element in the group VOHFK? If it is, then it is just as legitimate to do the same in the case of good English words such as "first", "strength", "eighth", and say that the sequences of consonants "rst", "ngth", "ghth" are unpronounceable. The fact that one might not be able to cite offhand a word in current usage in one of the eight languages to give as a precedent for the sequence HFK is of no moment, and, in any case, the Netherland administration has not made any complaints based upon this factor. In reality, the word VOHFK can be legitimately considered as forming a syllable in strict accordance with the accepted definition of the term: "an elementary sound, or a combination of such sounds, uttered together with a single effort or impulse of the voice."

Take the second example: LIYFKBVEVA, in which the sequence FKBV is indicated as being unpronounceable. If it is legitimate to lift the sequence FKBV out of its relation with the vowels that precede and follow it, then it is just as legitimate to lift analogous sequences out of the following good English words taken out of this evening's newspaper after a most casual search, and say that they are unpronounceable:

BEDSPREAD.....	DSPR
THANKSGIVING.....	NKSG
ANTHRAX.....	NTHR
TOUCHSTONE.....	CHST

All this is cited not with any pretense of showing that the code words objected to by many administrations are in reality easy to pronounce, or are pronounceable if one tries hard enough; it is cited merely to show that in many cases the code words which are considered difficult to pronounce conform to the provisions of that paragraph in international regulations which was intended to prevent their construction and use. The provisions, we must conclude, were faultily drawn up; they were not worded with a precision sufficient to accomplish the purpose the conference had in mind when, in 1903, it decided to legitimize artificial code words; and when, in 1908, it failed, as we have seen, to add any real precision to the paragraph in question, no improvement should have, or could have been expected to result in international code language communication.¹

THE WORK OF THE CODE CONTROL COMMITTEE

With the coming of the World War, the conference which was to be held in 1915 had to be postponed. Possibly, if it had been held, some remedy could have been found and applied before it became too late. Vague as the London definition of pronounceability was, even as amended at Lisbon, it might have been practicable to enforce as rigidly as possible whatever elements of precision the rules did contain, by mutual agreement among the telegraph services, aided by the regulations relative to the examination of codes by the code control committee established, as we have noted, by the Lisbon conference. But it is to be feared that not much could have been expected as a result of the latter facility, for, in my opinion, in its essence this facility is not as important as it might seem on its face, for two reasons.

In the first place, to examine a code *after* it has been compiled and printed, merely to determine whether or not the code words satisfy given conditions, is to reverse the normal order of things. The amount of work and the cost of publishing a code of any size is very considerable. After it has been printed, little can be done to remedy its defects, except to bring out an entirely new edition. Code compilers therefore would assume a considerable risk in submitting their codes after all this outlay of money and labor has been made. Suppose the code is entirely disapproved, or that a certificate of the second class is the best that could be obtained?

In the second place, it is obviously unjust to require the users of a code for secret or confidential private correspondence to file a copy with an agency over which they have absolutely no control. To do this vitiates the entire secrecy of the code. Of course, what the conference could have done, and still attain the object desired, would

¹ Not so, however, was the case with respect to code language in internal telegraph communications within the United States. See page 71.

have been to provide merely that the list of code words which the compiler intends to use be submitted for examination, preferably in manuscript form, or at least before the work of putting together the code words and the phrases which they are to represent has been initiated. Thus, if the words proved to be objectionable, or if only a certain number of them were considered to infringe the rules, changes could be made before printing. And, of course, the secrecy of the code would in no way be impaired by such an examination of the code words alone.

The code control committee, authorized and organized in 1908, faithfully pursued its work for over three years before the first reports of their study of such codes as had been submitted were made to the Berne Bureau. Quicker results could hardly have been expected, for the examination of *all* the code words of a large code is a particularly difficult and tedious task even if done only casually.¹ What then could be hoped for if it is remembered that each code word had to be carefully scrutinized to see whether it was pronounceable? One can well picture the quandaries in which the committee found itself time and again. Most discouragingly, the labor seemed to be entirely in vain. For, submission of codes was not compulsory; the compilation of larger and larger codes went on; a strict and uniform enforcement of the provisions of the regulations was impossible for reasons that, it is hoped, have been made clear; competition between companies continued—and code language, far from showing any improvement, continued to go from bad to worse.

THE PARIS CONFERENCE, 1925, AND CORTINA

As already stated, the conference which was to take place in 1915 had to be postponed on account of the World War. During the four years from 1914 to 1918 an almost world-wide censorship over communications served practically to suppress international commercial communication in code language. But almost on the day the armistice was signed censorship was lifted by several governments—those which realized what a striving for business would immediately ensue. International communication in code language immediately resumed its course, and increased by leaps and bounds. More codes were compiled and printed in the next five or six years than had been prepared in the twenty years that preceded. Code compilers did not even trouble themselves to submit their new codes to the code control committee, for it was evident that almost any code word carrying even the faintest possibility of pronounceability would do. The situation, of course, became an intolerable one for the communication companies and administrations.

¹ By 1921 approximately 108 public codes and 15 private ones had been examined. A list will be found in the United States Department of Commerce publication, *International Communications and the International Telegraph Convention*, 1923. (Miscellaneous Series No. 121.)

At last, in 1925, the Telegraph Union held a conference in Paris, and it was now faced with the condition where it had to admit that code language was hardly distinguishable from cipher language in secret letters.

Pursuing the path that had been suggested to and rejected by the 1903 conference, the Paris conference decided to submit the whole question of code language to a special committee. This was the committee whose recommendations formed the opening remarks of this report, and, as stated, their principal recommendation, aimed at reducing the difficulties occasioned by code language correspondence, was as follows:

Making and counting of words: Code words must be formed of a maximum of five letters, chosen at the will of the sender, without any condition.

III. THE CORTINA PROPOSALS

THE ELIMINATION OF PRONOUNCEABILITY

It is believed that the Cortina majority proposal is logical and is, moreover, one which the Telegraph Union will, if not now, then ultimately have to adopt.

In the first place, if adopted, the question of pronounceability and the illogical and difficult situations arising directly therefrom will disappear forever from the discussions of future conferences. In my opinion, no loss worth mentioning will result to anybody by reason of the suppression of this illogical and impracticable restriction: on the contrary, its suppression would be to the great advantage of all concerned. Let us look into this matter of pronounceability a bit in detail, from the point of view of its advantages in telegraph technique of to-day.

In the first place, it is necessary to remark that the matter of pronounceability of five-letter code words has always been of a great deal less importance to the users of such code words than to the telegraph personnel entrusted with their transmission. The writer of a code telegram consisting of five-letter groups can not trust to the sound of the groups at all. He must see and write down each letter as it appears in the code. Of course, if the groups were strictly euphonious, with no confusing letters such as C and S, I and Y, or K and X, whose sounds are often interchangeable, pronounceability would be of some assistance, but even then, it would not be as much as might be imagined. But such not being the case, the writers of code telegrams can not trust to pronounceability; they must see and note down the words by their individual letters, not by the sound of their ensemble.

In the second place, certainly, the suppression of the pronounceability requirement for the present-day ten-letter code words would work little, if any, hardship upon the telegraph services. The truth is that if pronounceability was an aid to correct and expeditious handling of code language in days gone by—and nobody can doubt that this was the case—to-day it is of no aid worth considering and retaining, especially in view of the other difficulties to which it gives rise by way of disputes with customers.

In the old days, before 1903, when code language did consist for the most part of dictionary words and really euphonious artificial words of the type discussed in this paper, a transmitting operator could take in at a single glance a whole word, or even a group of words, easily remember them, and transmit them without having to refer

two or more times to the telegraph blank. This facilitated speed in transmission, if not accuracy. At the receiving station, the operator was also aided in rapid recognition of signals, when the latter assembled themselves into letters representing a dictionary word or a euphonious artificial word. Errors, it is true, were frequent.

But not long after the new five-letter codes were introduced, the aid which pronounceability had theretofore afforded began to decrease, for reasons which, it is hoped, have been made apparent in the technical survey herein presented. To-day, everybody admits, pronounceability, as far as ten-letter code words are concerned, is a nonexistent, purely theoretical condition.

Furthermore, whatever value pronounceability may have had, in preceding periods, as a technical aid in the correct reception of code telegrams over land wires, its value in this respect has greatly decreased due to the very extensive replacement of auditory or Morse-sounder systems by automatic printing-telegraph systems the world over. In this country and in Canada approximately 85 to 90 per cent of telegraph traffic over land wires is now handled by means of printing-telegraph apparatus, and the time is not far off when practically all traffic, except to the smallest and most remote offices, in practically all countries will be so handled. What does "pronounceability" mean to such machines? It is true that the keyboard operators of such machines can manipulate the keys, either in direct transmission or in preparing a perforated tape, somewhat faster when plain language is to be sent than when code language is to be sent, but when it comes to the present-day code language of the ten-letter type, with its difficult and really outlandish combinations of letters, no keyboard operator would dare send except visually from the telegraph form itself. As a matter of fact, the operators of these keyboards must be proficient touch typists, and it is doubtful whether a candidate for such a position who is not a touch typist would even be considered by any company. Touch-typing being therefore the rule, and all this being the case in the present-day practice on land wires, the question of pronounceability has become of less and less importance to all administrations and companies.

In the preceding paragraph it was stated that a keyboard operator can manipulate his keyboard somewhat faster when sending plain language than when sending code language of the present-day type. It is pertinent to inquire: "How much faster?"

If one asks the average, experienced telegrapher: "How long does it take to transmit this present-day code language as compared with plain language?" the answer will undoubtedly be: "From at least two to three times as long." If he is then asked to state the basis upon which he makes this comparison, the answer will be, as a rule, "personal experience." No actual data based upon tests are offered.

The Cortina committee, feeling that personal observation could not be taken as a guide in determining this question, and appreciating the necessity for statistical data, drew up a *questionnaire* addressed to all telegraph services, in which the following tests (among many others) were requested to be made:

- What is the time required to transmit or to perforate
- (1) 5000 ten-letter, imperfectly pronounceable words taken from actual telegrams in code language?
 - (2) 5000 words taken from actual texts of international telegrams written in plain language in the mother tongue?

The averages based upon the data furnished the committee on these two tests are very interesting. They show that for the Baudot type of automatic telegraph equipment, which is very widely used in the world to-day, and considering the data from 14 of the largest administrations, it takes 1.56 more time to transmit code *words* of the type indicated than to transmit plain-language *words* of the mother tongue. But when the calculations are based, not upon *words*, but upon *letters*, it is found that it takes exactly the same amount of time to transmit this type of code language as it does to transmit plain language in the mother tongue. This was, I believe, a most startling discovery and occasioned considerable surprise among the members of the Cortina committee.

The same test, made by certain administrations with another type of automatic equipment, the Hughes, showed that a *letter* of code language takes 1.21 time units as against 1.00 for plain language. But this increase ought to be discounted to some extent, because the Baudot apparatus is much more extensively used than the Hughes in international services, according to the statement of the chairman of the Cortina committee (page 211, Cortina). Tests made by nine cable and radio companies, using apparatus of various types, averaged 1.25 for code *letters* as against 1.00 for plain-language *letters*.

Suppose we take the mean of these three averages, and say that a *letter* of the present-day code language requires 1.15 more time than a *letter* of plain language, that is 15 per cent more time. This is nowhere near the "at least two or three times as long" that telegraphers usually estimate the comparison to be. Where does the discrepancy originate? It is my opinion that it can be explained in this manner. When an operator is transmitting or perforating plain-language words in his mother tongue, the telegram is intelligible to him; he proceeds with the work in hand with a great deal less mental effort and a great deal less concentration than he is forced to apply when he is dealing with code-language words, which are unintelligible to him, and which at the same time are often made up of letters infrequently used in plain language. Moreover, the average length of plain-language words for all languages, as established by the Cortina

committee from statistical studies, is only 6.5 letters as against a uniform average of 10 letters for a code-language word. Thus, the transmission of a code-language telegram *seems* to the operator to take a *much* longer time than does a plain-language telegram, and we must recognize that statements by telegraphers relative to the much speedier transmission of plain language than code language are based more upon psychological, subjective impressions than upon actual, objective, physical, or technical realities. Either that is the case, or the statistics furnished the Cortina committee are faulty. It is difficult to believe that the latter could be true, and hence we are driven to the conclusion that pronounceability to-day is not so important a factor in telegraph operating as is generally thought to be the case.

THE FIVE-LETTER MAXIMUM FOR CODE WORDS

It will be said, of course, that the problems of code-language communication for the telegraph agencies are by no means at an end when the mere initial transmission has been completed, and this is true. For, between the transmitting and receiving operators the repetition of entire telegrams, or partial repetitions in order to correct errors and to eliminate doubtful receptions require a considerable additional length of time in the case of code-language telegrams, as compared with plain-language telegrams. In fact, this augmentation of time has been presented in the Cortina report by the German delegation (page 25) as being in the neighborhood of 30 per cent for ten-letter code-language words. Furthermore, there is an additional augmentation of almost 5 per cent for service messages asking for repetitions; and of 1 per cent for investigations required by the latter, making a total of 6 per cent. In fact, when all increases in time required for correct code-language transmission are taken into consideration, the German delegation finds (page 26) that if the transmission of a plain-language word represents seven letters, then the transmission of a ten-letter code word represents in reality 15.16 letters, so far as the telegraph service is concerned. Thus, according to their calculations a ten-letter code word to-day is 2.17 times as expensive to the German telegraph administration as a plain-language German word of seven letters.

The German delegation then went on to calculate this ratio on the assumption that code words would be transmitted in groups of only five letters, and found (page 27) that if a plain-language word in German is taken as representing seven letters, a five-letter code word of the present-day type, made simply by splitting the ten-letter groups into their two original, or natural parts, represents 6.93 letters. In order to make a just comparison and to find the time required for the two groups resulting when ten-letter groups are split into their

natural components, we should double this 6.93, making 13.86 letters. If, on the basis of ten-letter groups, the comparison is 7 letters for plain language to 15.16 letters for code language, the new ratio, on the basis of five-letter groups is 7 to 13.86. This means that a reduction of 9 per cent in total cost to the telegraph services for the transmission of code language can be effected if only the present ten-letter code words are separated into their original two component groups of five letters each.

In this reduction of difficulty in transmission by splitting ten-letter groups the code-language customers would also benefit quite materially. For according to the Cortina statistics the number of errors committed in transmitting ten-letter groups is very considerably reduced when the transmission is made in five-letter groups. Exactly what this reduction amounts to in percentage was not agreed upon by the Cortina committee on account of difficulties in interpreting the data; but all the members of the committee were in full accord in recognizing that a very material decrease did result therefrom. This, of course, would lessen the difficulties for the recipients of code telegrams when it comes to the decoding process.

Furthermore, in the mere preparation of code versions of telegrams, or in the writing down of code text, clerks would find it easier to handle five-letter groups, taken just as they naturally appear in the code book itself, and fewer errors would be committed at the initial or final stages of the business. Especially when it comes to typing code words upon telegraph forms the number of errors would be reduced, if five-letter groups were the rule.

ADVANTAGES FOR CODE USERS AND CODE COMPILERS

If the pronounceability restriction, which to-day is almost a purely theoretical one for all practical purposes, is suppressed, codes employing much better code equivalents could be compiled. Let us see upon what grounds this statement rests.

According to the mathematical laws of permutation,¹ using the 26 letters of the alphabet taken in groups of five, it is possible to make 26^5 , or 11,881,376 groups differing from each other by one or more letters. We have seen, however, that the two-letter difference is an absolute necessity for code communication. Therefore, if the groups must differ from each other by a minimum of two letters, the total number of permutations becomes reduced to 26^4 , or 456,976.

¹ The mathematical formula applicable is as follows:

$$n = 26^{(r-t)}$$

where n = the total number of permutations
 r = the number of letters per group
 t = the differential

When r is 5, and t is 2, the formula becomes $n = 26^{(5-2)}$, that is $n = 26^3$.

Now 456,976 is a really tremendous number of code words to include in one book, and there is reason to believe that the public does not show a preference for codes in a direct relation to their size or extensiveness. One code compiler of long experience has summed this up when he stated (page 124, Cortina):

My personal opinion is that large and bulky codes are not those preferred by the public. The advantages resulting from the increase in the number of expressions is more than compensated for by the additional difficulty in handling the code and the difficulty in finding what is wanted. Let us recall that 100,000 does not have the effect of two times 50,000. If a code containing 50,000 of the best expressions is increased to 100,000 one will note that the telegrams are nevertheless formed almost entirely of the first 50,000 expressions. An increase in the number of expressions beyond a certain point would only make for a slight decrease in their preciseness, and would hardly constitute an appreciable improvement. However, in some special businesses, where tables of descriptive matter, dimensions, etc., of a very great extent are in use, a very great number of code words could be advantageously employed.

With these remarks I am in full accord, and it is my belief that code compilers in general are in substantial agreement on this matter. They can all point to one actual example, in fact, where the remarks made are fully applicable.

What then will code compilers do with the large number of groups made available by discarding pronounceability? They can make much better selections of groups. They can introduce a complete or a modified form of *three-letter differential*, which would make code communication much safer than at present.

The total number of five-letter groups with a three-letter difference is $26^{(5-3+1)}$, or 17,576. This number is, of course, too small for the average private or public code; it might do in exceptional cases. But it is possible to employ a modified form of a three-letter differential, in the following manner. If for each three-letter difference group there were one group differing from it by only two letters, then the total number of both three- and two-letter difference groups becomes $2[26^{(5-3+1)}]$, that is, 35,152. If for each three-letter difference group there were two groups differing from it and from one another by only two letters, then the total number of both three- and two-letter difference groups becomes $3[26^{(5-3+1)}]$, that is, 52,728. In order to have a code of 100,000 groups, based upon this principle, one would require that for each three-letter difference group there be six two-letter difference groups, since $6[26^{(5-3+1)}]=105,456$. Thus, every set of seven groups in such a code would present a three-letter difference from every other set; within each set there would be only a two-letter difference. Such groups could, in the great majority of cases, be corrected, if mutilated in transmission, by reference to garble tables without having to request repetition by the telegraph service.

Again, it might be more desirable to take advantage of the possibilities afforded, if pronounceability be eliminated as a requirement, by suppressing a certain number of the letters of the alphabet, those which are especially liable to confusion orthographically or telegraphically. Such letters as B and D, G and Q, S and H, etc., are very often mistaken for each other in telegraphy; the letter E, consisting of a single dot, is often lost. By eliminating eight of the worst letters, there would still be possible a total of $18^{(6-2+1)}$, or 104,976 two-letter difference groups. Some persons might even consider the further elimination of difficult letters more advantageous than the retention of the two-letter difference. For them a reduction of the alphabet to but ten letters would give 100,000 groups. Various other sorts of systems and schemes falling outside the scope of this paper would be rendered available for the improvement of codes, providing only that this now practically useless restriction is removed.

ADVANTAGES FOR THE COMMUNICATION AGENCIES

If, from the point of view of the code users, a considerable amelioration of conditions would follow from the suppression of the pronounceability requirement and the establishment of five-letter groups as the rule in code language, from the point of view of the telegraph operating agencies there would result an even greater improvement in the situation.

In the first place, as already pointed out, the removal of the greatest cause for disputes and irritating altercations with the users of code language would immediately follow the elimination of the pronounceability restriction, permitting counter clerks to handle and charge for telegrams in a very much more expeditious manner than is now the case throughout the world.

Secondly, as pointed out above, the mere splitting up of the present ten-letter groups into their natural component parts, and thus setting up five-letter groups as a rule, would result in a marked reduction in the number of service telegrams requesting repetition which at present are transmitted gratuitously by the telegraph agencies. In some cases this reduction amounts perhaps to as much as 50 per cent.

In this connection it is pertinent to note that wherever the cost of transmission does not enter into the question, five-letter code words are transmitted as such, without combination in pairs. This is the case in the naval communication services of the large governments. The United States Navy and the British Navy, for example, long ago discarded the practice of combining five-letter groups into ten-letter groups in their radio services. As a result, errors in transmission are certainly fewer in number in these services. Anybody who has ever had experience in code work can testify to the fact that a five-letter group, even if composed entirely of consonants,

can be readily taken in at one glance and remembered long enough to be transcribed correctly. But this is generally impossible in the case of ten-letter groups, and any attempt to transcribe such a group from memory after but one glance very frequently results in errors. When code clerks and telegraph operators are pressed for time, or the occasion is one of urgency requiring speed, chances are taken on this transcribing of ten-letter words, hoping that no errors are introduced. But such hopes are most often ill-founded and the result is that telegraph and cable companies are very frequently called upon to repeat messages, or parts of them. Sometimes the errors are committed by the originator of the message, but very often they are committed by the telegraph personnel, and in the measure of their frequency, the extra cost of the repetitions made necessary must be borne by the telegraph services. Anything that will promote a reduction in this source of unremunerative work ought to be embraced. The five-letter maximum for code words will undoubtedly reduce the number of errors and repetitions and therefore it is rightly and especially favored by the communication agencies.

Finally, there would be one more source of benefit to the operating agencies in case the five-letter rule is adopted. At present, with the ten-letter rule in effect, many code users feel that they must, in order to avoid the raising of a doubt in the mind of the recipient, make every code word a complete, ten-letter group. Now it often happens that in the interior of a message a proper name must be mentioned, or that a word not present in the code vocabulary must be indicated. If it happens at the same time that the code word of five letters immediately preceding is not the second half of a ten-letter group, but stands by itself as a five-letter group, the writer of the telegram will complete this half word by adding a code word which is either nonsignificant in its meaning, and is specially included in some codes for this particular purpose, or is not essential to the text of the telegram itself. This, of course, costs the sender of the message nothing, and in his opinion eliminates any doubt on the part of the recipient as to whether the second half of a word has been erroneously omitted in the transmission. But, for the operating agencies it adds just that much more work, without any increase in receipts. This same factor applies to the final group of 50 per cent of all telegrams, since there are only two possibilities in this place: either a message has an even number of five-letter groups or an odd number. But if the five-letter rule is adopted there would be no reason whatever for the senders of telegrams to add these extra groups which cost them nothing, make more work for the telegraph agencies, and, in a certain percentage of cases, when transmission errors happen to strike these particular groups, make trouble for the recipient of their telegrams. What the saving to the operating

agencies would be from these two sources is difficult to estimate, but I should judge that a saving of two to three per cent of the total number of words handled would not be excessive.

Thus far we have considered only the technical advantages that would accrue to the public as well as to the telegraph services if the five-letter rule were adopted and the pronounceability requirement suppressed. Let us see what technical disadvantages would be brought to each of these two interested parties.

DISADVANTAGES FOR THE COMMUNICATION AGENCIES

As for the operating agencies, splitting of ten-letter groups into two groups of five letters each would introduce an additional signaling element in transmission, that coming from the insertion of the space between the two groups. According to the Cortina statistics the insertion of this space adds, in the case of the Baudot apparatus, 7 per cent to the time required for transmission; in the case of the Hughes apparatus, it adds nothing at all; and according to the figures furnished by the cable and radio companies, the increase in transmission time due to the insertion of the space is more than offset by the greater rapidity in perforating or in operating the transmitting keyboard, so that there is in reality a reduction in time of 2 per cent. Evidently the conditions under which this particular test was made by the various administrations and companies must have been widely different, so that it is difficult to render a judgment on the matter. In this connection it is thought that the following remarks, appearing as a footnote to the figures furnished by the Western Union Telegraph Co. are pertinent:

These questions as formulated do not appear, to the Western Union, to have any value in respect to the question of codes. The conditions under which the various administrations and companies work are so different, and the circumstances under which the tests were made in order to reply to these questions must have been so dissimilar that they make the value of the replies absolutely without consequence. This is especially the case so far as the comparison of the transmission of groups of five and of ten letters is concerned. Be that as it may, the operators for whom a great percentage of work consists of code groups of ten letters, are accustomed to transmit such words and have adopted for the transmission a cadence based upon ten characters; consequently, if they are required to make a test on a base of five characters they must adopt another kind of cadence which completely differs from that to which they have been accustomed, and since they have not had sufficient practice with groups of five letters so as to become habituated to this type of transmission, it is absolutely useless to desire to establish a comparison between the base of ten letters to which they are accustomed and that of five letters to which they are not accustomed.¹

The Western Union figures showed that the time required for operating the keyboard of the Kleinschmidt apparatus in order to perforate the tape, was reduced by 2 per cent as a result of splitting

¹ Cortina, p. 89, note 31.

the groups. The time required for transmission, that is, to run the tape through the transmitter, was, however, increased by 5 per cent, obviously due to the space between the two groups of five letters. This separation into the two steps required, as done by the Western Union, was unfortunately not followed by all the administrations and companies making the tests; moreover, the results would differ according to the various types of apparatus used. In any case, it may be concluded that if there is an increase in time required for perforation or for transmission, or for keyboard-transmitter operation, occasioned by the splitting of ten-letter groups into five-letter ones, this increase is not very important, and would be more than offset by the reduction in service messages and the costs pertaining thereto.

DISADVANTAGES FOR THE CODE USERS AND CODE COMPILERS

As to the technical disadvantages that would be brought to code users as a result of the requirement that code language be written in five-letter groups, it is obvious that for those users who now employ five-letter codes—and they are in an overwhelming majority—no hardships whatsoever would be worked upon them, since they would merely have to write the code words as they naturally appear in their codes. In fact, in a way, they would gain, for, so far as the technique of code correspondence by means of these codes is concerned, the work done by the sender, who combines pairs of five-letter groups into ten-letter ones, and the work done by the receiver, who has to separate the ten-letter groups into their original five-letter ones, is utterly unnecessary and represents just so much waste of time, labor, and money.

But what about those who still employ codes using the old style dictionary words, or any of the other types of code words we have enumerated in this paper? If the Cortina majority recommendation is adopted and only five-letter words are admitted in code language, the users of dictionary-word codes would be forced to give up these uneconomical and truly out-of-date codes for the much more economical, modern, five-letter codes. But the users of these old codes form as has been stated, only a very small minority of the code-using public. Should the great majority be made to suffer because a small minority persists in clinging to old-fashioned, expensive methods of communication, especially if, by forcing a change, the minority would in reality soon be the gainers, too? It might be said, of course, that in the United States a large number of firms still use dictionary-word codes, but if the matter be carefully studied it will be found that these codes are used only for internal correspondence, because the rules applying to our domestic telegraph correspondence in code language are now different from those applying to similar telegrams in international correspondence.

CODE LANGUAGE REGULATIONS IN THE UNITED STATES SINCE 1909

For, from 1903 to 1909, the rules for internal telegrams with respect to code-language words, whether real or artificial, were the same as in international regulations,¹ but when the Lisbon conference in 1908 failed to adopt any real measures toward bringing about a reform in a situation rapidly becoming intolerable, the two American companies, the Western Union and Postal Telegraph, decided that they had had enough of pronounceability. Finding no other way out of the dilemma, they seized the bull by the horns, and despite loud complaints, amended their rules with respect to artificial words in the following manner:

All groups of letters, when such groups are not dictionary words, and are not combinations of dictionary words, will be counted at the rate of five letters or fraction of five letters to a word.

This paragraph was followed by some explanatory matter well worth quoting:

It will be noted that December 1 and thereafter no distinction will be made between pronounceable and unpronounceable artificial groups of letters, but that all will alike be counted at the rate of five letters or fraction of five letters to a word. When such groups are made up of combinations of dictionary words, each dictionary word used will be counted as one word.

Receiving clerks and other employees who accept messages for transmission are notified that all rules of the company for the counting of messages shall be strictly enforced. They must carefully scrutinize all messages for the detection of all attempted evasions, such as deliberately misspelled or improperly combined or formed words, and that all must be checked strictly in accordance with the amended rule. No exceptions will be made in any case, and any infractions of these instructions which may be observed by intermediate or delivery offices should be reported for correction of the service.

This new rule was made effective on December 1, 1909, simultaneously by the two companies. The result was that by 1910 they were no longer troubled with difficulties concerning the matter of pronounceability, and the Gordian knot was cut at one stroke.

Since under these rules nothing can any longer be gained by combining pairs of five-letter groups, there is no great incentive for large business firms to compile new codes of the five-letter type for their internal or domestic correspondence; such firms as had gone to the expense of compiling codes using dictionary words find it convenient to keep these codes in service for their occasional internal code telegrams. *But for their external or overseas telegrams these same firms, almost without exception, employ public five-letter codes, or have had such codes compiled for them by code-compiling agencies, or by their own personnel, because in these telegrams they can send two code words at the cost of only one.*

¹ See page 32 and note 1.

CODE LANGUAGE REGULATIONS AND CRYPTOGRAPHIC APPARATUS

There is one further point of great importance in connection with secret language that has as yet not been mentioned in this paper. This is the age of mechanical and electrical operation of most facilities; the machine is more and more replacing the hand and the brain. The time and labor required to encode and decode long messages by means of code books is enormous. For those who are more interested in secrecy and rapidity of communication than in economy, the possibility of using already existing cryptographic machinery would be a great boon. But such machinery can not regularly produce pronounceable groups in converting a plain-language text into a secret-language text; only cipher language can be produced.¹ The present discrimination against cipher language which makes the cost of international telegrams in that type of language twice that of telegrams in what is to-day euphemistically designated as code language, has served to prevent the use of existing cryptographic apparatus, and to check the development of improved apparatus.

One of the most serious practical objections to the use of radio by many commercial firms is that the secrecy of this method of communication is more dependent upon the observance of laws than upon the mere physical impracticability of the unauthorized interception of communications, as is the case in wire or cable telegraphy. For some firms, such as press and news associations, and commercial houses conducting a voluminous correspondence, it is practically impossible to encode by means of code books the mass of material they transmit, so that radio transmission in plain language does not seem desirable. But if the cost of cipher language were the same as code language, or better, if the present unjustifiable distinction between the two languages as regards cost were eliminated, then automatic cryptographic apparatus could be employed by these firms, and their messages could then be transmitted by radio. Even a simple type of cryptographic machinery would do for this purpose, and the additional cost of enciphering and deciphering would be very little. There is even reason to believe that certain firms which at the present time prefer to use wire and cable telegraphy exclusively and must, for purposes of secrecy, as is the case with banks and brokerage houses, use code, would find the greater rapidity of cryptographic apparatus

¹ For the technical cryptographic distinction between code and cipher languages, refer to the definitions on pages 2-3. Cryptographic apparatus as a rule can produce only substitution ciphers, though apparatus to produce transposition ciphers is not inherently impossible.

advantageous, and would discard code books.¹ But all this is dependent, of course, upon the placing of cipher language on a parity with code language as regards the cost of transmission.

THE CORTINA MINORITY PROPOSALS

If the recommendations of the Cortina majority report are not adopted, if the present unenforceable provisions as to pronounceability are not eliminated, what then? The recommendations of the minority, those of the British delegation? The latter recommendations contemplate keeping in force, until December 31, 1931, the present provisions as to the ten-letter words and pronounceability, and to require that a code word contain at least one vowel in the first five letters and at least one in the succeeding letters; also that in no case must a word contain more than five consecutive vowels or consonants. No specification is made as to what letters will be considered as falling in each of the two categories. In any case, the latter requirement would cause trouble for many words of the artificial type, and for a few of the dictionary type. For example, in one code selected at random, the combination of ERRGD and SCHNE yields ERRGDSCHNE, in which seven consonants appear consecutively; there are hundreds of cases in which six would appear consecutively. As to dictionary words, the words "Angstthau" and "Angstzweet", taken from the official vocabulary may be cited.

After January 1, 1932, the British proposal would establish the following:

- (a) Conserve the limit of ten letters per word.
- (b) Abandon the rule concerning pronounceability.
- (c) Add the following new rules:
 1. A code-language word must contain at least two vowels in the first five letters and at least two vowels in the second group of letters. (Exceptionally, words of six, seven, or eight letters can include two vowels in the first five letters and one vowel in the succeeding letters.
 2. It is not permitted to employ in succession more than four consonants or more than four vowels in the same word.

One of the principal objections of the British delegation to the proposals of the majority at Cortina was that the setting up of five letters as a maximum for length of code words would force the discarding of dictionary-word codes, at considerable expense to their users. Of the British proposals cited above it may be stated that, if adopted, they would force the scrapping of practically *all* codes, dictionary-word codes, as well as five-letter-word codes. Note the

¹ The inherent and fundamental difference between code systems and cipher systems makes it certain that the latter will never replace the former in commercial telegraph correspondence. For in cipher systems, the final text is usually just as long as the original plain-language text, whereas in code systems, because of the condensation and abbreviation effected, the final text is usually much shorter than the original plain-language text, with a consequent saving in cost.

following cases found after only a very casual search through several of the most widespread and most important codes of to-day:

<i>Dictionary-word type</i>	<i>5-letter-word type</i>
ABSCHLACHTEN	BLUZBSTOWJ
ABSCHRAUBEN	MAIYOAAIEQ
ABSCHWEIFEN	KWUNGBLYXK
BLENDTRITT	CLEZNCHULF
BLENKLECHT	BLENDCHOND
BOCKSPIEL	EBUMTCHRUY
BRANDS	CROUYEYAIT
BRANDSCHIFF	BLAGMBRAKS
CHAFFWEED	WAYOYYOILT
DRESSMAKER	KHISJDHUF8
etc., etc.	etc., etc.

In the ten examples of the five-letter type there are two that were taken, not from codes compiled after 1920, but from one compiled before 1908. In an actual examination of the fifteen most important public codes, only *one* was encountered which would not run afoul of the rules contemplated by the British proposal, and that single example has now been superseded by a new edition by the same compiler, in which this is no longer the case.

It would be interesting to know how long it took the reader to determine, in each of the twenty examples cited above, wherein the word violated the proposed rules. This would give some idea of what the receiving clerks would have to face, in the acceptance of telegrams. *Each word would have to be examined carefully* to see that it contained the proper number of vowels, and that it did not contain more than the admitted number of consonants or vowels in succession. The loss in time thus occasioned would be a factor of considerable importance to the operating agencies, and of considerable annoyance to the public. Nor would a strict adherence to the rules solve the technical troubles of perforation or transmission, for words such as ZYYZZZZYYZ, XYUPBCSOIH, and HSAEIOZUYI would conform to the rules. What then are the advantages of such a system?

ANOTHER PROPOSAL

In some of the sessions at Cortina there were discussions concerning the possibility of modifying the present code-language regulations so as to require merely that there be at least one vowel among the first five letters, and at least one among the second five of a ten-letter group. Furthermore, this proposal seems even now to be viewed favorably by certain of the administrations. In truth, the efficacy of such a modification, if adopted, seems most questionable. In the first place, even a most casual inspection of any of the large

codes now in use, those having about 100,000 code words, will soon show that the proposed requirement has already been fulfilled; in fact, it will be found that the average number of vowels per five-letter word is somewhere between two and three,¹ and groups containing only one vowel are rather the exception. Has the presence of two vowels per group aided materially in reducing the technical difficulties of code-language communication? What would be gained by *lowering* the average, especially if the ten-letter maximum were retained? Would the one-vowel requirement reduce the number of bizarre groups that could be formed? If we are to be guided by the mathematics set forth at Cortina² by some of those present, the total number of two-letter difference groups containing at least one vowel is 69,934. But this number is most certainly erroneous, and is far short of the correct number. While it is difficult to present a mathematical formula to cover this particular case, we can arrive at an approximation which will be suitable for our purposes. We have seen that the maximum number of five-letter code groups with a two-letter difference and no limitations as to their vowel-consonant composition is $26^{(5-2+1)}$ or 456,976. We have also seen that the total number of possible forms or types of groups as regards their vowel-consonant composition is 32. All these forms are shown herewith, where the letter C represents any consonant, the letter V, any vowel. Of these 32 different types of groups the one and *only one* that would be ruled out by a regulation requiring at least one vowel per five-letter group would be the last one, CCCCC. All the other types have at least one vowel in their composition.

Now the maximum theoretical number of groups of the form CCCCC, with a two-letter difference, and employing all the letters of the alphabet except A, E, I, O, U, and Y, is $20^{(5-2+1)}$, which is 20^4 , or 160,000. The actual number that would be obtained from any specific permutation table is a function of the arrangement of the letters within the table, so that in most cases the total actual number produced would fall below 160,000. But even taking the maximum theoretical number, and finding the difference between 456,976 for

¹ According to statistics compiled by the Great Northern Telegraph Co., the average number of vowels per ten-letter code word is 4.322 (Cortina, p. 268). But in making pronounceability possible, it is the distribution of the vowels, not their number, which is the more important factor. Note the English word "strengths," with eight consonants and only one vowel.

² See Cortina, page 268.

1 VVVVV	17 CVVVV
2 VVVVC	18 CVVVC
3 VVVCV	19 CVVCV
4 VVCCV	20 CVVCC
5 VVCVV	21 CVCVV
6 VVCVC	22 CVCVC
7 VVCCV	23 CVCCV
8 VVCCC	24 CVCCC
9 VCVVV	25 CCVVV
10 VCVVC	26 CCVVC
11 VCVCV	27 CCVCV
12 VCVCV	28 CCVCC
13 VCCVV	29 CCCVV
14 VCCVC	30 CCCVC
15 VCCCV	31 CCCC
16 VCCCC	32 CCCCC

groups of all forms and 160,000 for groups of only the form CCCCC, we obtain 296,976 five-letter groups, each containing at least one vowel, and all showing a two-letter difference from one another. This is considerably higher than 69,934, the number stated in the Cortina report. Even if code compilers found it inconvenient or unnecessary to make use of all the 296,000 groups, even 200,000 would practically be double the present number of groups employed in codes and since the ten-letter maximum would still be retained, as proposed in connection with the proposition under discussion, the adoption of this proposal would serve only to *increase* the present technical difficulties. Of course, if the five-letter maximum were established, the requirement of one vowel per group would reduce the total number of groups available, but *in itself* the latter requirement would result in no worth-while amelioration in conditions, would serve only as an irritating—because useless—restriction upon code compilers, and would entail much loss of time at telegraph charging counters, since each group would have to be examined to make sure of the presence of at least one vowel. Even then, groups such as EPBQZHJXQY and ZBVWQEKVWP would conform to the provisions of the regulations, whereas even at present they could be ruled out or double-checked. What then are the advantages of such a proposal? This conclusion is then irresistible: to place any limitations whatsoever upon the constitution of code words, other than that they must not exceed five letters in length, and must consist of letters capable of transmission according to the Morse alphabet, is in reality impractical, illogical, and in its final analysis, if there is an incentive on the part of the telegraphing public to circumvent them, unenforceable.

THE QUESTION OF NEW RATES TO BE APPLIED

This paper has dealt only with technical matters connected with code language in international telegraphy; all matters connected with the question of rates have been avoided as falling outside the scope and aims of the paper. There is, however, one phase of the latter subject which can here legitimately be considered because it appears to the writer to be of vital and fundamental importance in connection with any discussion of code language. The *basis* upon which rates are to be fixed is, of course, of first consideration, and in that connection the question arises: Should the code user pay more for code language *because* this type of language allows him to transmit a greater quantity of information or intelligence than would be the case if code language were not permitted? There seem to be many persons who believe that the code user ought to pay more for code language merely because a *special service* is rendered him. Throughout the Cortina deliberations, in fact, there ran this current

of thought, and it has been one of long standing. But the justice of the contention may well be questioned.

In the first place, *after the additional expense which the handling of code occasions for the agencies has been taken into account*, whether a group of letters expresses just one word or a whole sentence should really be of no material interest to the agencies, because as far as the operators, the apparatus used, or the electric impulses and channels conveying these characters are concerned, what the agencies are being paid for are signaling units. It is difficult to see what the amount of intelligence conveyed by these units has to do with the case.

In M. Crescitz's study of code language already referred to, published by the International Bureau in 1911-12 in the *Journal télégraphique*, it is stated that all questions of rates should be evolved by taking into account these two elements: The work done; the service rendered.

After discussing the first element, to which we take no exception, M. Crescitz says, with respect to the second element, that since each code word generally represents an entire phrase, or at least three or four words, it would be logical to apply a special rate higher than that for plain language. Then he goes on to say that examples wherein this idea of setting up rates according to the service rendered is applied, can be cited and that one is not embarrassed to find many examples. However, he cites only two. The first relates to the railroad service where the charges are based, he says, largely upon the value of the merchandise transported. For equal weights a costly object, a piece of furniture, for example, pays three or four times as much as ordinary or heavy merchandise. The second example cited relates to the postal service, wherein the cost of transporting letters is greater than that for printed matter. Then he remarks that in the telegraph service likewise there is a charge according to the service rendered. The regulations provide a reduction of 50 per cent for press telegrams, not because transmission of these telegrams occasions less work than ordinary telegrams, he says, but because of other considerations.

Let us now examine these examples. Take the railroad service, wherein differential commodity rates are common, and an expensive, small article may be more costly to ship than a very much larger, heavy article. The increased cost in shipping is brought about by factors other than the *service rendered* the consignor or the consignee; the increased cost is due to the different type of handling required. For a small, easily damaged article must be handled very much more carefully than an equally heavy metal object, or an equal weight of bricks, for example. Again, a small package may be much more easily lost and therefore requires more careful accounting and checking.

The type of car used, whether open or closed, what other articles can be placed in the same car, and so on, enters into the matter. In short, the writer believes that the *service rendered* in this example is not a factor in the sense that M. Crescitz's discussion would seem to imply.

With regard to the second example taken from the postal service, it is true that the rate for printed matter is less than that for first-class matter, but here again there are considerations other than those of service rendered, similar in nature to those M. Crescitz himself refers to in connection with the reduced rates for press telegrams. Furthermore, to take this same example, does it cost any more to transport a letter containing just a single sheet of paper upon which perhaps a single sentence is written, as compared with another letter of equal weight which contains several sheets on which may be written a short story? Again, to take this time an example from the telephone service, is the greater charge for a person-to-person toll call than of a station-to-station call due to the greater value which the telephone company estimates that the communicants place upon the fact that they have been put into direct contact? Is it not due solely to the greater labor, personnel, and plant facilities involved in establishing such a connection? And, is it not pertinent to ask, after the connection has been established does the company charge according to the amount of intelligence transmitted?

The moment one admits that in the telegraph service a charge ought to be made according to the amount of information transmitted, the limits to which such an admission might be carried are hard to foresee. People in general exhibit wide differences in ability to compose telegrams; one person requires 50 words to convey an idea which another can express equally well by employing only 10 words. Why not charge accordingly? Thus, if the general principle of charging according to the amount of information transmitted is once accepted, then one could conceive of a day when counter clerks would be empowered to examine every plain-language telegram and every code telegram, measure the amount of information conveyed, and charge accordingly, whereupon the whole matter is seen to have been reduced somewhat to an absurdity.¹

It seems clear that the writer of the Berne articles confuses two very different things: "service rendered" and "quality of service rendered." The latter is wholly different in nature from the former, both as regards origin and as regards the value placed upon it by the person to whom the service is rendered. And, as regards quality of service rendered, telegrams of the same class, those in code as well as those in clear, are supposed to be accorded similar treatment in all respects, accuracy, speed, and so on. In the deferred-rate services

¹ On the basis of service rendered, why should not a person who gains \$1,000 as a result of a telegraphic order to his broker pay more for his telegram than another person who loses a similar amount on a like transaction, or one who sends simply a birthday greeting?

the lower cost is due to the fact that an important item in *quality of service rendered*—delay in transmission—enters into the question. It is factors like these that legitimately may be and usually are taken into consideration under quality of service rendered.

Certain administrations may say, of course, that the use of code cuts down their revenue because fewer words are transmitted by code than would otherwise be the case. But is the revenue really diminished? When, after 1903, the evasion of the new regulations permitted two code words to be sent at the price of one, was the revenue cut in half? On the contrary, an era of greater use of the telegraph was the immediate result with greater increase in revenue. Let us quote from an authoritative book on submarine telegraphy:¹

From the point of view of the general public, the economy effected by the use of code is often even a more important consideration than its secrecy. A single code word, charged for only at a slightly higher rate than one ordinary word, may be made to convey the sense of a good many. The telegraph cable thus becomes available for business and other purposes by many people who could not otherwise afford it, and the number of messages which pass over it daily have enormously increased in consequence. And with this increase in the number of them, there has not been the corresponding decrease in their length which might have been anticipated. The public has simply become educated to the more liberal use of the telegraph, and has availed itself of its facilities in the measure and in the spirit in which they have been granted to it. The increase of the total volume of traffic, and of business leading to still greater traffic in the future, has more than compensated the companies for the economies effected by its code-using customers.

The fact is, but for the code system, the existing number of cables would, in many cases, be quite inadequate for the demands of the present traffic. This remark applies most conspicuously to the case of the North Atlantic, and will be readily understood when it is stated that, whereas prior to the universal recognition and adoption of code transmission, the average length of telegrams used to be 35 words, it is now only eleven. In other words, but for code, the companies might, by now, be asked to transmit more than three times as many words as they are transmitting within the same time. More probably the proportion would not be so great in practice, for reasons already given. But even an addition of only half as many again would be embarrassing to the operators—and indeed, to all concerned, excepting telegraph engineers and contractors, who would, in consequence, have extra cables to lay.

Finally, the weakness of the arguments on this score are directly attested to by the attitude of all of the American communication companies who state that they do not care in the slightest how much more can be expressed by a code message than by a plain message. In fact, one might believe that they wish to see more people use code, as the consequent cheapening of communication is an incentive to increased use of telegraphy, and all companies seem to be desirous of extending their services and of having more people avail themselves of their facilities. To sum up this argument, the writer believes that the question of how much more meaning

¹ *Submarine Telegraphs*, Charles Bright, London, 1898. The quotation given, although written many years ago, is just as true to-day as it was then.

a code message can convey than a plain-language message ought certainly to be omitted from consideration in making calculations having for their aim the establishment of rates.

FACSIMILE TRANSMISSION: WILL IT SOLVE THE DIFFICULTIES?

From various quarters one hears predictions that this matter of code-language regulation will solve itself when facsimile processes of transmission have been fully developed and introduced into daily use. Certainly the day is not far distant when this may be the case as regards radioelectric methods, either by wire or wireless. Then the sender of a message may indeed write what he wishes to convey in any way he may choose; the basis of charging will be the extent of the physical surface covered by his message, and whatever errors are occasioned in his message will be his own. But can we foresee the day when *all* telegraphic traffic will be so conducted? Is not this method likely to be restricted for a considerable period to communication between only the largest fixed stations? What about the thousands of small offices scattered throughout the civilized world, and what about the thousands of ships on the seas?

Furthermore, it is not likely that radiotelegraphy will supplant submarine cable telegraphy for a long time to come. Each of these two grand methods has its place in world communications, and they are indeed to-day interwoven in a complicated, all-embracing network. Now the high-frequency electrical oscillations involved in the transmission of facsimiles, either by wire or by wireless methods, are of such an order that their use on ocean cables is at the present day a practical impossibility. It was only a comparatively short time ago that a frequency of but two or three hundred cycles per second of alternating current was made possible over the new permalloy submarine cables and hailed as a great achievement. This frequency falls far short of that required for practical facsimile transmission, and probably it will be a long time before frequencies of the order required for this method will be realized over ocean cables. This being the case, what bearing has facsimile transmission upon the present-day problems of code language over ocean cables? When it is considered that by far the larger part of the communication between the continents of the world is to-day—and will be for some little time to come—conducted through the intermediacy of ocean cables, the idea that facsimile transmission methods will solve the problems of code language seems to be one of more theoretical than practical interest at the present moment; and despite the undoubted advances that will be made within the next five or ten years along the lines of facsimile transmission, the need for a solution to the present difficulties of code language in international communication, in order to reduce the inconveniences, errors, and delays incident to the present long-established, and practical methods, is most pressing.

○