REF ID:A64649

CAUTION: THESE RECORDS WILL BE USED FOR OFFICIAL PURPOSES ONLY, DO NOT REMOVE PAPERS NOR REVEAL CONTENTS TO UNAUTHORIZED PERSON(S)

RECORDS CHARGE-OUT

FILE OR SERIAL NUMBER AND SUBJECT	From File of Special Consultant (Friedman) Military Cryptanalysis, Part I Serial 310.21
TO	NAME AND EXTENSION OF PERSON REQUESTING FILE: Mr. William Friedman LI6-8520 ORGANIZATION, BUILDING: 310 2nd, Str, SE,
RETURN TO	Mrs. Christian, AG-24, NSA, Ft. Geo. G. Meade, Md.
INSTRUCTIONS	WHEN TRANSFERRING FILE TO ANOTHER PERSON, COMPLETE SELF-ADDRESSED TRANSFER CO LETTER-SIZE PAPER AND PLACE IN OUT-GOING MAIL SERVICE.

~~CONFIDENTIAL~~

~~Modified Handling Authorized~~

NATIONAL SECURITY AGENCY

MILITARY CRYPTANALYTICS

Part I

By

WILLIAM F. FRIEDMAN

and

LAMBROS D. CALLIMAHOS

National Security Agency
Washington 25, D. C.

April 1956

~~CONFIDENTIAL~~

310.21

NATIONAL SECURITY AGENCY

MILITARY CRYPTANALYTICS
Part I

By
WILLIAM F. FRIEDMAN
and
LAMBROS D. CALLIMAHOS

To Mr. William F. Friedman,
a transcendent cryptologist and venerated teacher, without whose prototypic texts in cryptanalysis this present volume could not have been written, with profound admiration and respect from a devoted pupil striving to follow in his master's footsteps.
Washington, 12 May 1956. — Lambros D. Callimahos

NOTICE: This material contains information affecting the national defense of the United States within the meaning of the espionage laws, Title 18, U.S.C., Secs. 793 and 794, the transmission or revelation of which in any manner to an unauthorized person is prohibited by law.

National Security Agency
Washington 25, D. C.

April 1956

NSATL S-70,021

The Golden Guess
Is Morning-Star to the full round of Truth.
—Tennyson.

Preface

This text represents an extensive expansion and revision, both in scope and content, of the earlier work entitled "Military Cryptanalysis, Part I" by William F. Friedman. This expansion and revision was necessitated by the considerable advancement made in the art since the publication of the previous text.

I wish to express grateful acknowledgment for Mr. Friedman's generous assistance and invaluable collaboration in the preparation of this volume. I also extend particular appreciation to my colleague Robert E. Cefail for his numerous valuable comments and assistance in writing the new material which is contained herein.

—*L. D. C.*

c c c c v	B
c c c c c	A
c c c v c	C
c v v c v	O
c v v c c	N
c c c v v	D
c v c c c	I-J
c c c v v	D
c v v c c	N
c v v c v	O
v c c v c	T
v c v c c	W
v c c c c	R
c v c c c	I-J
v c c v c	T
c c v c c	E
v c c v c	T
c c v v v	H
c v c c c	I-J
v c c c v	S
v c v c c	W
c v v c v	O
v c c c c	R
c v c c v	K

(Only those consonants and vowels marked by dots are the significant Baconian elements, C = a, V = b)

TABLE OF CONTENTS

Military Cryptanalytics, Part I

Monoalphabetic Substitution Systems

~~CONFIDENTIAL~~

[redacted]

CHAPTER I

INTRODUCTORY REMARKS

1. Scope of this text.—*a.* This text constitutes the first of a series of six basic texts [1] on the science of *cryptanalytics* and the art of *cryptanalysis.* Although most of the information contained herein is applicable to cryptograms of various types and sources, special emphasis will be laid upon the principles and methods of solving military [2] cryptograms. Except for an introductory discussion of fundamental principles underlying the science of cryptanalytics, this first text in the series will deal solely with the principles and methods for the analysis of *monoalphabetic substitution ciphers.* Even with this limitation it will be possible to discuss only a few of the many variations of this type that are met in practice; but with a firm grasp upon the general principles few difficulties should be experienced with any modifications or variations that may be encountered.

b. This and the succeeding texts will deal with, among others, some basic types of cryptosystems not because they may be encountered unmodified in military operations but because their study is essential to an understanding of the principles underlying the solution of the modern, very much more complex types of codes, ciphers, and certain encrypted transmission systems that are likely to be employed by the larger governments of today in the conduct of their military affairs in time of war.

c. It is presupposed that the student has no prior background in the field of cryptology; therefore cryptography is presented concurrently with cryptanalysis. It is also presupposed that the reader has had but a minimal mathematical background; a student who has had elementary algebra should encounter no difficulty with the mathematical treatment in the body of the text, and he will be progressively guided into augmenting his mathematical background to fit the needs of cryptanalytics. Basic terminology and preliminary cryptologic considerations are treated in Chapter II; other terms are usually defined upon their first occurrence, or they may be found in the Glossary (Appendix 1). Footnotes, besides amplifying general information, include occasional treatment of mathematical principles that may be beyond a beginner in the field; the student therefore should not spend too much time trying to assimilate all the information contained therein.

d. The cryptograms presented in the examples embrace messages from hypothetical air, ground, and naval traffic; thus, the student will have the opportunity to familiarize himself with the language and phraseology of all three military Services.

---

[1] Each text has its accompanying course in cryptanalysis, so that the student may test his learning and develop his skill in the solution of the types of cryptograms treated in the respective texts. The problems which pertain to this text constitute Appendix 9.

[2] The word "military" is here used in its broadest sense. In this connection see subpar. *d*, below.

2. Mental equipment necessary for cryptanalytic work.—*a.* Captain Parker Hitt, in the first United States Army manual [3] dealing with cryptology, opens the first chapter of his valuable treatise with the following sentence:

"Success in dealing with unknown ciphers is measured by these four things in the order named: perseverance, careful methods of analysis, intuition, luck."

These words are as true today as they were then. There is no royal road to success in the solution of cryptograms. Hitt goes on to say:

"Cipher work will have little permanent attraction for one who expects results at once, without labor, for there is a vast amount of purely routine labor in the preparation of frequency tables, the rearrangement of ciphers for examination, and the trial and fitting of letter to letter before the message begins to appear."

The author deems it advisable to add that the kind of work involved in solving cryptograms is not at all similar to that involved in solving crossword puzzles, for example. The wide vogue the latter have had and continue to have is due to the appeal they make to the quite common interest in mysteries of one sort or another; but in solving a crossword puzzle there is usually no necessity for performing any preliminary labor, and palpable results become evident after the first minute or two of attention. This successful start spurs the crossword "addict" on to complete the solution, which rarely requires more than an hour's time. Furthermore, crossword puzzles are all alike in basic principles and once understood, there is no more to learn. Skill comes largely from the embellishment of one's vocabulary, though, to be sure, constant practice and exercise of the imagination contribute to the ease and rapidity with which solutions are generally reached. In solving cryptograms, however, many principles must be learned, for there are many different systems of varying degrees of complexity. Even some of the simpler varieties require the preparation of tabulations of one sort or another which many people find irksome; moreover, it is only toward the very close of the solution that results in the form of intelligible text become evident. Often, indeed, the student will not even know whether he is on the right track until he has performed a large amount of preliminary "spade work" involving many hours of labor. Thus, without at least a willingness to pursue a fair amount of theoretical study, and a *more than average amount of patience and perseverance*, little skill and experience can be gained in the rather difficult art of cryptanalysis. General Givierge, the author of an excellent treatise on cryptanalysis, remarks in this connection:[4]

"The cryptanalyst's attitude must be that of William the Silent: 'No need to hope in order to undertake, nor to succeed in order to persevere'."

b. As regards Hitt's reference to careful methods of analysis, before one can be said to be a cryptanalyst worthy of the name it is necessary that one should have, firstly, a sound knowledge of the basic principles of cryptanalysis, and secondly, a long, varied, and active *practical* experience in the successful application of those principles. It is not sufficient to have read treatises on this subject. One month's actual practice in solution is worth a whole year's mere reading of theoretical principles. An exceedingly important element of success in solving the more intricate cryptosystems is the possession of the rather unusual mental faculty designated in general terms as the power of inductive and deductive reasoning. Probably this is an inherited rather than an acquired faculty; the best sort of training for its emergence, if latent in the individual, and for its development is the study of the natural sciences, such as chemistry,

[3] Hitt, Capt. Parker, *Manual for the Solution of Military Ciphers.* Army Service Schools Press, Fort Leavenworth, Kansas, 1916. 2d Edition, 1918. (Both out of print.)

[4] Givierge, Général Marcel, *Cours de Cryptographie,* Paris, 1925, p. 301.

physics, biology, geology, and the like. Other sciences such as linguistics, archaeology, and philology are also excellent.

c. Aptitude in mathematics is quite important, more especially in the solution of ciphers and enciphered codes than in codebook reconstruction, which latter is purely and simply a linguistic problem. Although in the early days of the emergence of the science of cryptanalytics little thought was given to the applications of mathematics in this field, many branches of mathematics and, in particular, probability and statistics, have now found cryptologic applications. Those portions of mathematics and those mathematical methods which have cryptologic applications [5] are known collectively as *cryptomathematics*.

d. An active imagination, or perhaps what Hitt and other writers call *intuition*, is essential, but mere imagination uncontrolled by a judicious spirit will be more often a hindrance than a help. In practical cryptanalysis the imaginative or intuitive faculties must, in other words, be guided by good judgment, by practical experience, and by as thorough a knowledge of the general situation or extraneous circumstances that led to the sending of the cryptogram as is possible to obtain. In this respect the many cryptograms exchanged between correspondents whose identities and general affairs, commercial, social, or political, are known are far more readily solved [6] than are isolated cryptograms exchanged between unknown correspondents, dealing with unknown subjects. It is obvious that in the former case there are good data upon which the intuitive powers of the cryptanalyst can be brought to bear, whereas in the latter case no such data are available. Consequently, in the absence of such data, no matter how good the imagination and intuition of the cryptanalyst, these powers are of no particular service to him. Some writers, however, regard the intuitive spirit as valuable from still another viewpoint, as may be noted in the following: [7]

"Intuition, like a flash of lightning, lasts only for a second. It generally comes when one is tormented by a difficult decipherment and when one reviews in his mind the fruitless experiments already tried. Suddenly the light breaks through and one finds after a few minutes what previous days of labor were unable to reveal."

---

[5] It is quite important to stress at this point that in professional cryptologic work the *science* of cryptanalytics is subordinated to the *art* of cryptanalysis, just as in the world of music the technical virtuosity of a great violinist is adjuvant to the expression of music, that is, the virtuosity is a "tool" for the recovery of the complete musical "plain text" conceived by the composer. Since the practice of cryptanalysis *is* an art, mathematical approaches cannot always be expected to yield a solution in cryptology, because art can and must transcend the cold logic of scientific method. By way of example, an experienced Indian guide can usually find his way out of a dense forest more readily than a surveyor equipped with all the refined apparatus and techniques of his profession. Likewise, an experienced cryptanalyst can generally find his way through a cryptosystem more readily than a pure mathematician equipped merely with the techniques of his field no matter how abstruse or refined they may be. A cryptomathematician of repute once stated that "the only effect of [refined mathematical techniques] is frequently to discourage one so much that one does nothing at all and some unmathematical ignoramus then gets the problem out in some very unethical way. This is intensely irritating." See also in this connection the remarks made in subpar. 27*e* in reference to the validity of statistical tests in cryptanalysis.

[6] The application in practical, operational cryptanalysis of "probable words" or "cribs", i. e., plain text assumed or known to be present in a cryptogram, is developed in time of war into a refinement the extent and usefulness of which cannot be appreciated by the uninitiated. Even as great a thinker as Voltaire found the subject of cryptanalysis stretching his credulity to the point that he said:

"Those who boast that they can decipher a letter without knowing its subject matter, and without preliminary aid, are greater charlatans than those who would boast of understanding a language which they have never learned."—*Dictionnaire Philosophique*, under the article "Poste".

[7] Lange et Soudart, ***Traité de Cryptographie***, Libraire Félix Alcan, Paris, 1925, p. 104.

This, too, is true, but unfortunately there is no way in which the intuition may be summoned at will, when it is most needed.[8] There are certain authors who regard as indispensable the possession of a somewhat rare, rather mysterious faculty that they designate by the word "flair", or by the expression "cipher brains". Even so excellent an authority as General Givierge,[9] in referring to this mental faculty, uses the following words:

"Over and above perseverance and this aptitude of mind which some authors consider a special gift, and which they call intuition, or even, in its highest manifestation, clairvoyance, cryptographic studies will continue more and more to demand the qualities of orderliness and memory."

Although the author believes a special aptitude for the work is essential to cryptanalytic success, he is sure there is nothing mysterious about the matter at all. Special aptitude is prerequisite to success in all fields of endeavor. There are, for example, thousands of physicists, hundreds of excellent ones, but only a handful of world-wide fame. Should it be said, then, that a physicist who has achieved very notable success in his field, has done so because he is the fortunate possessor of a *mysterious* faculty? That he is fortunate in possessing a special aptitude for his subject is granted, but that there is anything mysterious about it, partaking of the nature of clairvoyance (if, indeed, the latter is a *reality*) is not granted. While the ultimate nature of any mental process seems to be as complete a mystery today as it has ever been, the author would like to see the superficial veil of mystery removed from a subject that has been shrouded in mystery from even before the Middle Ages down to our own times. (The principal and readily understandable reason for this is that governments have always closely guarded

---

[8] The following extracts are of interest in this connection:

"The fact that the scientific investigator works 50 per cent of his time by non-rational means is, it seems, quite insufficiently recognized. There is without the least doubt an instinct for research, and often the most successful investigators of nature are quite unable to give an account of their reasons for doing such and such an experiment, or for placing side by side two apparently unrelated facts. Again, one of the most salient traits in the character of the successful scientific worker is the capacity for knowing that a point is proved when it would not appear to be proved to an outside intelligence functioning in a purely rational manner; thus the investigator feels that some proposition is true, and proceeds at once to the next set of experiments without waiting and wasting time in the elaboration of the formal proof of the point which heavier minds would need. Questionless such a scientific intuition may and does sometimes lead investigators astray, but it is quite certain that if they did not widely make use of it, they would not get a quarter as far as they do. Experiments confirm each other, and a false step is usually soon discovered. And not only by this partial replacement of reason by intuition does the work of science go on, but also to the born scientific worker—and emphatically they cannot be made—the structure of the method of research is as it were given, he cannot explain it to you, though he may be brought to agree *a posteriori* to a formal logical presentation of the way the method works".—Excerpt from Needham, Joseph, *The Sceptical Biologist*, London, 1929, p. 79.

"The essence of scientific method, quite simply, is to try to see how data arrange themselves into causal configurations. Scientific problems are solved by collecting data and by "thinking about them all the time." We need to look at strange things until, by the appearance of known configurations, they seem familiar, and to look at familiar things until we see novel configurations which make them appear strange. We must look at events until they become luminous. That is scientific method . . . Insight is the touchstone . . . The application of insight as the touchstone of method enables us to evaluate properly the role of imagination in scientific method. The scientific process is akin to the artistic process: it is a process of selecting out those elements of experience which fit together and recombining them in the mind. Much of this kind of research is simply a ceaseless mulling over, and even the physical scientist has considerable need of an armchair . . . Our view of scientific method as a struggle to obtain insight forces the admission that science is half art . . . Insight is the unknown quantity which has eluded students of scientific method".—Excerpts from an article entitled *Insight and Scientific Method*, by Willard Waller, in *The American Journal of Sociology*, Vol. XL, 1934.

[9] *Op cit.*, p. 302.

cryptographic secrets and anything so guarded soon becomes "mysterious".) He would, rather, have the student approach the subject as he might approach any other science that can stand on its own merits with other sciences, because cryptanalytics, like other sciences, has a practical importance in human affairs. It presents to the inquiring mind an interest in its own right as a branch of knowledge; it, too, holds forth many difficulties and disappointments, and these are all the more keenly felt when the nature of these difficulties is not understood by those unfamiliar with the special circumstances that very often are the real factors that led to success in other cases. Finally, just as in the other sciences wherein many men labor long and earnestly for the true satisfaction and pleasure that comes from work well done, so the mental pleasure that the successful cryptanalyst derives from his accomplishments is very often the only reward for much of the drudgery that he must do in his daily work. General Givierge's words in this connection are well worth quoting:[10]

"Some studies will last for years before bearing fruit. In the case of others, cryptanalysts undertaking them never get any result. But, for a cryptanalyst who likes the work, the joy of discoveries effaces the memory of his hours of doubt and impatience."

e. With his usual deft touch, Hitt says of the element of luck, as regards the role it plays in analysis:

"As to luck, there is the old miners' proverb: 'Gold is where you find it.'"

The cryptanalyst is lucky when one of the correspondents whose cryptograms he is studying makes a blunder that gives the necessary clue; or when he finds two cryptograms identical in text but in different keys in the same system; or when he finds two cryptograms identical in text but in different systems, and so on. The element of luck is there, to be sure, *but the cryptanalyst must be on the alert* if he is to profit by these lucky "breaks".

f. If the author were asked to state, in view of the progress in the field since 1916, what elements might be added to the four ingredients Hitt thought essential to cryptanalytic success, he would be inclined to mention the following:

(1) A broad, general education, embodying interests covering as many fields of practical knowledge as possible. This is useful because the cryptanalyst is often called upon to solve messages dealing with the most varied of human activities, and the more he knows about these activities, the easier his task.

(2) Access to a large library of current literature, and wide and direct contacts with sources of collateral information. These often afford clues as to the contents of specific messages. For example, to be able instantly to have at his disposal a newspaper report or a personal report of events described or referred to in a message under investigation goes a long way toward simplifying or facilitating solution. Government cryptanalysts are sometimes fortunately situated in this respect, especially where various agencies work in harmony.

(3) Proper coordination of effort. This includes the organization of cryptanalytic personnel into harmonious, efficient teams of cooperating individuals.

(4) Under mental equipment he would also include the faculty of being able to concentrate on a problem for rather long periods of time, without distraction, nervous irritability, and impatience. The strain under which cryptanalytic studies are necessarily conducted is quite severe and too long-continued application has the effect of draining nervous energy to an unwholesome degree, so that a word or two of caution may not here be out of place. One should continue at work only so long as a peaceful, calm spirit prevails, whether the work is fruitful or not. But just as soon as the mind becomes wearied with the exertion, or just as soon as a feeling

[10] *Op. cit.*, p. 301.

of hopelessness or mental fatigue intervenes, it is better to stop completely and turn to other activities, rest, or play. It is essential to remark that systematization and orderliness of work are aids in reducing nervous tension and irritability. On this account it is better to take the time to prepare the data carefully, rewrite the text if necessary, and so on, rather than work with slipshod, incomplete, or improperly arranged material.

(5) A retentive memory is an important asset to cryptanalytic skill, especially in the solution of codes. The ability to remember individual groups, their approximate locations in other messages, the associations they form with other groups, their peculiarities and similarities, saves much wear and tear of the mental machinery, as well as much time in looking up these groups in indexes.

(6) The assistance of machine aids in cryptanalysis. The importance and value of these aids cannot be overemphasized in their bearing on practical, operational cryptanalysis, especially in the large-scale effort that would be made in time of war on complex, high-grade cryptosystems at a theater headquarters or in the zone of the interior. These aids, under the general category of rapid analytical machines, comprise both punched-card tabulating machinery and certain other general- and special-purpose high-speed electrical and electronic devices. Some of the more compact equipment may be employed by lower echelons within a theater of operations to facilitate the cryptanalysis of medium-grade cryptosystems found in tactical communications.

g. It may be advisable to add a word or two at this point to prepare the student to expect slight mental jars and tensions which will almost inevitably come to him in the conscientious study of this and the subsequent texts. The author is well aware of the complaint of students that authors of texts on cryptanalysis base much of their explanation upon their foreknowledge of the "answer"—which the student does not know while he is attempting to follow the solution with an unbiased mind. They complain, too, that these authors use such expressions as "it is obvious that", "naturally", "of course", "it is evident that", and so on, when the circumstances seem not at all to warrant their use. There is no question that this sort of treatment is apt to discourage the student, especially when the point elucidated becomes clear to *him* only after many hours' labor, whereas, according to the book, the author noted the weak spot at the first moment's inspection. The author can only promise to try to avoid making the steps appear to be much more simple than they really are, and to suppress glaring instances of unjustifiable "jumping at conclusions". At the same time he must indicate that for pedagogical reasons in many cases a message has been consciously "manipulated" so as to allow certain principles to become more obvious in the illustrative examples than they ever are in practical work. During the course of some of the explanations attention will even be directed to cases of unjustified inferences. Furthermore, of the student who is quick in observation and deduction, the author will only ask that he bear in mind that if the elucidation of certain principles seems prolix and occupies more space than necessary, this is occasioned by the author's desire to carry the explanation forward in very short, easily-comprehended, and plainly-described steps, for the benefit of students who are perhaps a bit slower to grasp but who, once they understand, are able to retain and apply principles slowly learned just as well, if not better than the students who learn more quickly.[11]

---

[11] In connection with the use of the word "obvious", the following extract is of interest:

"Now the word 'obvious' is a rather dangerous one. There is an incident, which has become something of a legend in mathematical circles, that illustrates this danger. A certain famous mathematician was lecturing to a group of students and had occasion to use a formula which he wrote down with the remark, 'This statement is obvious.' Then he paused and looked rather hesitantly at the formula. 'Wait a moment,' he said. 'Is it

3. **Validity of results of cryptanalysis.**—Valid or authentic cryptanalytic solutions cannot and do not represent "opinions" of the cryptanalyst. They are valid only so far as they are wholly objective, and are susceptible of demonstration and proof, employing authentic, objective methods. It should hardly be necessary (but an attitude frequently encountered among laymen makes it advisable) to indicate that the results achieved by any serious cryptanalytic studies on authentic material rest upon the same sure foundations and of necessity are reached by the same general steps as the results achieved by any other scientific studies, *viz.*, observation, hypothesis, deduction and induction, and confirmatory experiment. Implied in the latter is the possibility that two or more qualified investigators, each working independently upon the same material, will achieve identical (or practically identical) results—there is one and only one (valid) solution to a cryptogram. Occasionally a "would-be" or pseudo-cryptanalyst offers "solutions" which cannot withstand such tests; a second, unbiased, investigator working independently either cannot *consistently* apply the methods alleged to have been applied by the pseudo-cryptanalyst, or else, if he can apply them at all, the results (plaintext translations) are far different in the two cases. The reason for this is that in such cases it is generally found that the "methods" are not clear-cut, straightforward or mathematical in character. Instead, they often involve the making of judgments on matters too tenuous to measure, weigh, or otherwise subject to careful scrutiny. Often, too, they involve the "correction" of an inordinate number of "errors" which the pseudo-cryptanalyst assumes to be present and which he "corrects" in order to make his "solution" intelligible. And sometimes the pseudo-cryptanalyst offers as a "solution" plain text which is intelligible only to him or which he makes intelligible by expanding what he alleges to be abbreviations, and so on. In all such cases, the conclusion to which the unprejudiced observer is forced to come is that the alleged "solution" obtained by the pseudo-cryptanalyst is purely subjective.[12] In nearly all cases where this has happened (and they occur from time to time) there has been uncovered nothing which can in any way be used to impugn the integrity of the

---

obvious? I think it's obvious.' More hesitation, and then, 'Pardon me, gentlemen, I shall return.' Then he left the room. Thirty-five minutes later he returned; in his hands was a sheaf of papers covered with calculations, on his face a look of quiet satisfaction. 'I was right, gentlemen. It is obvious,' he said, and proceeded with his lecture."—Excerpt from *The Anatomy of Mathematics* by Kershner and Wilcox. New York, 1950.

[12] A mathematician is often unable to grasp the concept behind the expression "subjective solution" as used in the cryptanalytic field, since the idea is foreign to the basic philosophy of mathematics and thus the expression appears to him to represent a contradiction in terms. As an illustration, let us consider a situation in which a would-be cryptanalyst offers a solution to a cryptogram he alleges to be a simple monoalphabetic substitution cipher. His so-called solution, however, requires that he assume the presence of, let us say, approximately 50% garbles (which he claims to have been introduced by cipher clerks' errors, faulty radio reception because of adverse weather conditions, etc.). That is, the "plain text" he offers as the "solution" involves his making helter-skelter many "corrections and emendations", which, one may be sure, will be based on what his subconscious mind expects or desires to find in the cleartext message. Unfortunately, another would-be cryptanalyst working upon the same cryptogram and hypothesis independently might conceivably "degarble" the cryptogram in different spots and produce an entirely dissimilar "plain text" as his "solution". Both "solutions" would be invalid because they are based upon an erroneous hypothesis—the cryptogram actually happens to be a polyalphabetic substitution cipher which when correctly analyzed requires on the part of unbiased observers no assumption of garbles to a degree that strains their credulity. The last phrase is added here because in professional cryptanalytic work it is very often necessary to make a few corrections for errors; but it is rarely the case that the garble rate exceeds more than a few percent of the characters of the cryptogram, say 5 to 10% at the outside. It is to be noted, however, that occasionally the solution to a cryptogram may involve the correction of more than this percentage of errors, but the solution would be regarded as valid only if the errors can be shown to be *systematic* in some significant respect, or can otherwise be explained by *objective* rationalization.

pseudo-cryptanalyst. The worst that can be said of him is that he has become a victim of a special or peculiar form of self-delusion, and that his desire to solve the problem, usually in accord with some previously-formed opinion, or notion, has over-balanced, or undermined, his judgment and good sense.[13]

---

[13] Specific reference can be made to the following typical "case histories":

Donnelly, Ignatius, *The Great Cryptogram.* Chicago, 1888.

Owen, Orville W., *Sir Francis Bacon's Cipher Story.* Detroit, 1895.

Gallup, Elizabeth Wells, *Francis Bacon's Biliteral Cipher.* Detroit, 1900.

Arensberg, Walter Conrad, *The Cryptography of Shakespeare.* Los Angeles, 1922.

The Shakespearean Mystery. Pittsburgh, 1928.

The Baconian Keys. Pittsburgh, 1928.

Margoliouth, D. S., *The Homer of Aristotle.* Oxford, 1923.

Newbold, William Romaine, *The Cipher of Roger Bacon.* Philadelphia, 1928. (For a scholarly and complete demolition of Professor Newbold's work, see an article entitled *Roger Bacon and the Voynich MS,* by John M. Manly, in *Speculum,* Vol. *VI*, No. 3, July 1931.)

Feely, Joseph Martin, *The Shakespearean Cypher.* Rochester, N. Y., 1931.

Deciphering Shakespeare. Rochester, N. Y., 1934.

Roger Bacon's Cypher: the right key found. Rochester, N. Y., 1943.

Wolff, Werner, *Déchiffrement de l'Ecriture Maya.* Paris, 1938.

Strong, Leonell C., *Anthony Askham, the author of the Voynich manuscript,* in *Science,* Vol. 101, June 15, 1945, pp. 608–9.

CHAPTER II

BASIC CRYPTOLOGIC CONSIDERATIONS

4. **Cryptology, communication intelligence, and communication security.**—The need for secrecy in the conduct of important affairs has been recognized from time immemorial. In the case of diplomacy and organized warfare this need is especially important in regard to communications. However, when such communications are transmitted by electrical means, they can be heard and copied by unauthorized persons. The protection resulting from all measures designed to deny to unauthorized persons information of value which may be derived from such communications is called *communication security*. The evaluated information concerning the enemy, derived principally from a study of his electrical communications, is called *communication intelligence*. The collective term including all phases of communication intelligence and communication security is *cryptology*.[1] Or, stated in broad terms, cryptology is that branch of knowledge which treats of hidden, disguised, or secret [2] communications.

5. **Secret communication.**—*a*. Communication may be conducted by any means susceptible of ultimate interpretation by one of the five senses, but those most commonly used are sight and hearing. Aside from the use of simple visual and auditory signals for communication over relatively short distances, the usual method of communication between or among individuals separated from one another by relatively long distances involves, at one stage or another, the act of writing or of speaking over a telephone.

b. Privacy or secrecy in communication by telephone can be obtained by using equipment which affects the electrical currents involved in telephony so that the conversations can be understood only by persons provided with suitable equipment properly arranged for the purpose. The same thing is true in the case of facsimile transmission (i. e., the electrical transmission of pictures, drawings, maps) and television transmission. However, this text will not treat of these aspects [3] of cryptology.

---

[1] From the Greek *kryptos* (hidden) + *logos* (discourse). The prefix "crypto-" in compound words pertains to "cryptologic", "cryptographic", or "cryptanalytic", depending upon the use of the particular word as defined.

[2] In this text the term "secret" will be used in its ordinary sense as given in the dictionary. Whenever the designation is used in the more restricted sense of the security classification as defined in official regulations, it will be capitalized. There are in current use the three classifications CONFIDENTIAL, SECRET, and TOP SECRET, listed in ascending order of degree.

[3] These aspects of cryptology are now known as *ciphony* (from *ci*pher+tele*phony*); *cifax* (from *ci*pher+*fax*simile); and *civision* (from *ci*pher+tele*vision*).

c. Writing may be either *visible* or *invisible*. In the former, the characters are inscribed with ordinary writing materials and can be seen with the naked eye; in the latter, the characters are inscribed by means or methods which make the writing invisible to the naked eye. Invisible writing can be prepared with certain chemicals called *invisible*, *sympathetic*, or *secret inks*, and in order to "develop" such writing, that is, make it visible, special processes must usually be applied. There are also methods of producing writing which is invisible to the naked eye because the characters are of microscopic size, thus requiring special photographic or microscopic apparatus to make such writing visible to the naked eye.

d. Invisible writing and unintelligible visible writing constitute *secret writing*.

6. Plain text and encrypted text.—*a.* Visible writing which is intelligible, that is, conveys a more or less understandable or sensible meaning (in the language in which written) and which is not intended to convey a hidden meaning, is said to be in *plain text*.[4] A message in plain text is termed a *plaintext message*, a *cleartext message*, or a *message in clear*.

b. Visible writing which conveys no intelligible meaning in any recognized language [5] is said to be in *encrypted text* and such writing is termed a *cryptogram*.[6]

7. Cryptography, encrypting, and decrypting.—*a. Cryptography* is that branch of cryptology which treats of various means, methods, and apparatus for converting or transforming plaintext messages into cryptograms and for reconverting the cryptograms into their original plaintext forms by a simple reversal of the steps used in their transformation.

b. To *encrypt* is to convert or transform a plaintext message into a cryptogram by following certain rules, steps, or processes constituting the *key* or *keys* and agreed upon in advance by correspondents, or furnished them by higher authority.

c. To *decrypt* is to reconvert or to transform a cryptogram into the original equivalent plaintext message *by a direct reversal of the encrypting process*, that is, by applying to the cryptogram the key or keys (usually in a reverse order) used in producing the cryptogram.

d. A person skilled in the art of encrypting and decrypting, or one who has a part in devising a cryptographic system is called a *cryptographer;* a clerk who encrypts and decrypts, or who assists in such work, is called a *cryptographic clerk*.

8. Codes, ciphers, and enciphered code.—*a.* Encrypting and decrypting are accomplished by means collectively designated as *codes* and *ciphers*. Such means are used for either or both of two purposes: (1) secrecy, and (2) economy or brevity. Secrecy usually is far more important in military cryptography than economy or brevity. In ciphers or *cipher systems*, cryptograms are produced by applying the cryptographic treatment to *individual letters* of the plaintext messages, whereas, in codes or *code systems*, cryptograms are produced by applying the cryptographic treatment to entire *words, phrases, and sentences* of the plaintext messages. The specialized meanings of the terms *code* and *cipher* are explained in detail later (subpar. 11*d*).

b. A cryptogram produced by means of a cipher system is said to be *in cipher* and is called

---

[4] Visible writing may be intelligible but the meaning it obviously conveys may not be its real meaning, that is, the meaning intended to be conveyed. To quote a simple example of an apparently innocent message containing a secret or hidden meaning, prepared with the intention of escaping censorship, the sentence "Son born today" may mean "Three transports left today." Secret communication methods or artifices of this sort are impractical for field military use but are often encountered in espionage and counter-espionage activities.

[5] There is a certain type of writing which is considered by its authors to be intelligible, but which is either completely unintelligible to the wide variety of readers or else requires considerable mental struggle on their part to make it intelligible. Reference is here made to so-called "modern literature" and "modern verse", products of such writers as E. E. Cummings, Gertrude Stein, James Joyce, et al.

[6] From *kryptos*+*gramma* (that which is written).

a *cipher message*, or sometimes simply a *cipher*. The act or operation of encrypting a cipher message is called *enciphering*, and the enciphered version of the plain text, as well as the act or process itself, is often referred to as the *encipherment*. The cryptographic clerk who performs the process serves as an *encipherer*. The corresponding terms applicable to the decrypting of cipher messages are *deciphering*, *decipherment*, and *decipherer*. A clerk who serves as both an encipherer and decipherer of messages is called a *cipher clerk*.

c. A *cipher device* is a relatively simple mechanical contrivance for encipherment and decipherment, usually "hand-operated" or manipulated by the fingers, as for example a device with concentric rings of alphabets, manually powered; a *cipher machine* is a relatively complex apparatus or mechanism for encipherment and decipherment, usually equipped with a typewriter keyboard and often requiring an external power source.

d. A cryptogram produced by means of a code system is said to be *in code* and is called a *code message*. The text of the cryptogram is referred to as *code text*. This act or operation of encrypting is called *encoding*, and the encoded version of the plain text, as well as the act or process itself, is referred to as the *encodement*. The clerk who performs the process serves as an *encoder*. The corresponding terms applicable to the decrypting of code messages are *decoding*, *decodement*, and *decoder*. A cryptographic clerk who serves as both an encoder and decoder of messages is called a *code clerk*.

e. Sometimes, for special purposes (usually increased security), the code text of a cryptogram undergoes a further step in concealment involving *superencryption*, that is, encipherment of the characters comprising the code text, thus producing what is called an *enciphered-code message*, or *enciphered code*. *Encoded cipher*, that is, the case where the final cryptogram is produced by enciphering the plain text and then encoding the cipher text obtained from the first operation, is also possible, but rare.

9. General system, specific key, and cryptosystem.—*a*. There are a great many different methods of encrypting messages, so that correspondents must first of all be in complete agreement as to which of them will be used in their secret communications, or in different types or classes of such communications. Furthermore, it is to be understood that all the detailed rules, processes, or steps comprising the cryptography agreed upon will be *invariant*, that is, constant or unvarying in their use in a given set of communications. The totality of these basic, invariable rules, processes, or steps to be followed in encrypting a message according to the agreed method constitutes the general cryptographic system or, more briefly, the *general system*.

b. It is usually the case that the general system operates in connection with or under the control of a number, a group of letters, a word, a phrase, or sentence which is used as a *key*, that is, the element which specifically governs the manner in which the general system will be applied in a specific message, or the exact *setting* of a cipher device or a cipher machine at the initial point of encipherment or decipherment of a specific message. This element—usually of a variable nature or changeable at the will of the correspondents, or prearranged for them by higher authority—is called the *specific key*. The specific key may also involve the use of a set of specially prepared tables, a special document, or even a book.

c. The term *cryptosystem*[7] is used when it is desired to designate or refer to all the cryptomaterial (device, machine, instructions for use, key lists, etc.) as a unit to provide a single, complete system and means for secret communication.

---

[7] The term ***cryptosystem*** is used in preference to ***cryptographic system*** so as to permit its use in designating secret communication systems involving means other than ***writing***, such as ciphony and cifax.

10. Cryptanalytics and cryptanalysis.—*a.* In theory any cryptosystem (except one [8]) can be "broken", i. e., solved, if enough time, labor, and skill are devoted to it, and if the volume of traffic in that system is large enough. This can be done even if the general system and the specific key are unknown at the start. In military operations theoretical rules must usually give way to practical considerations. How the theoretical rule in this case is affected by practical considerations will be discussed in Appendix 8, "Principles of cryptosecurity."

b. That branch of cryptology which deals with the principles, methods, and means employed in the solution or *analysis* of cryptosystems is called *cryptanalytics.*

c. The steps and operations performed in applying the principles of cryptanalytics constitute *cryptanalysis.* To *cryptanalyze* a cryptogram is to *solve* it by cryptanalysis.

d. A person skilled in the art of cryptanalysis is called a *cryptanalyst,* and a clerk who assists in such work is called a *cryptanalytic clerk.*

11. Transposition and substitution.—*a.* Technically there are only two distinct types of treatment which may be applied to written plain text to convert it into secret text, yielding two different *classes* of cryptograms. In the first, called *transposition,* the *elements* or *units* of the plain text retain their original identities and merely undergo some change in their relative positions, with the result that the original text becomes unintelligible. In the second, called *substitution,* the elements of the plain text retain their original relative positions but are replaced by other elements with different values or meanings, with the result that the original text becomes unintelligible. Thus, in the case of transposition ciphers, the unintelligibility is brought about merely by a change in the original sequence of the elements or units of the plain text; in the case of substitution ciphers, the unintelligibility is brought about by a change in the elements or units themselves, without a change in their relative order.

b. It is possible to encrypt a message by a substitution method and then to apply a transposition method to the substitution text, or vice versa. Such *combined transposition-substitution* methods do not form a third class of methods. They are occasionally encountered in military cryptography, but the types of combinations that are sufficiently simple to be practicable for field use are very limited.[9]

c. Under each of the two principal classes of cryptograms as outlined above, a further classification can be made based upon the number of characters composing the *textual elements* or *units* undergoing cryptographic treatment. These textual units are composed of (1) individual letters, (2) combinations of letters in regular groupings, (3) combinations of letters in irregular, more or less euphonious groupings called syllables, and (4) complete words, phrases, and sentences. Methods which deal with the first type of units are called *monographic* methods; those which deal with the second type are called *polygraphic* (digraphic, trigraphic, etc.); those which deal with the third type, or syllables, are called *syllabic;* and, finally, those which deal with the fourth type are called *lexical* (of or pertaining to words).

d. It is necessary to indicate that the foregoing classification of cryptographic methods is more or less artificial in nature, and is established for purpose of convenience only. No sharp line of demarcation can be drawn in every case, for occasionally a given system may combine methods of treating single letters, regular or irregular-length groupings of letters, syllables, words, phrases, and complete sentences. When in a single system the cryptographic treatment

[8] The exception is the "one-time" system in which the specific key has no systematic construction and is used only once.

[9] One notable exception is the *ADFGVX system,* used extensively by the Germans in World War I.

is applied to textual units of regular length, usually monographic or digraphic (and seldom longer, or intermixed monographic and digraphic), the system is called a *cipher system*. Likewise, when in a single system the cryptographic treatment is applied to textual units of irregular length, usually syllables, whole words, phrases, and sentences, and is only exceptionally applied to single letters or regular groupings of letters, the system is called a *code system* and generally involves the use of a *code book*.[10]

12. Nature of alphabets.—*a*. One of the simplest kinds of substitution ciphers is that which is known in cryptologic literature as Julius Caesar's Cipher, but which, as a matter of fact, was a favorite long before his day. In this cipher each letter of the text of a message is replaced by the letter standing the third to the right of it in the ordinary alphabet; the letter A is replaced by D, the letter B by E, and so on. The word *cab* becomes converted into FDE, which is cipher.

b. The English language is written by means of 26 simple characters called *letters* which, taken together and considered as a *sequence of symbols*, constitute the alphabet of the language. Not all systems of writing are of this nature. Chinese writing is composed of about 44,000 complex characters, each representing one sense of a word. Whereas English words are composite or polysyllabic and may consist of one to eight or more syllables, Chinese words are all monosyllables and each monosyllable is a word. Written languages of the majority of other civilized peoples of today are, however, alphabetic and polysyllabic in construction, so that the principles discussed here apply to all of them.

c. The letters comprising the English alphabet used today are the results of a long period of evolution, the complete history of which may never fully be known.[11] They are conventional symbols representing *elementary sounds*, and any other simple symbols, so long as the sounds which they represent are agreed upon by those concerned, will serve the purpose equally well. If taught from early childhood that the symbols $, *, and @ represent the sounds "Ay", "Bee", and "See" respectively, the combination @$* would still be pronounced *cab*, and would, of course, have exactly the same meaning as before. Again, let us suppose that two persons have agreed to change the sound values of the letters F, G, and H, and after long practice have become accustomed to pronouncing them as we pronounce the letters A, B, and C, respectively; they would then write the "word" HFG, pronounce it *cab*, and see nothing strange whatever in the matter. But to others no party to their arrangements, HFG constitutes cipher. The combination of sounds called for by this combination of symbols is perfectly intelligible to the two who have adopted the new sound values for those symbols and therefore pronounce HFG as *cab;* but HFG is utterly unpronounceable and wholly unintelligible to others who are reading it according to their own long-established system of sound and symbol equivalents. It would be stated that there is no such word as HFG which would mean merely that the particular combination of sounds represented by this combination of letters has not been adopted by convention to represent a thing or an idea in the English language. Thus, it is seen that, in order for the written words of a language to be pronounceable and intelligible to all who speak that language, it is necessary, first, that the sound values of the letters or symbols be universally understood and agreed upon and, secondly, that the particular combination of sounds denoted by the letters should have been adopted to represent a thing or an idea.

---

[10] A list of single letters, frequent digraphs, trigraphs, syllables, and words is often called a *syllabary;* cryptographic treatment of the units of such syllabaries places them in the category of code systems.

[11] An excellent and most authoritative book on this subject is *The Alphabet: a Key to the History of Mankind* by David Diringer. London, 1949.

d. It is clear also that in order to write a polysyllabic language with facility it is necessary to establish and to maintain by common agreement or convention, equivalency between *two* sets of elements, first, a set of elementary sounds and, second, a set of elementary symbols to represent the sounds. When this is done the result is what is called an *alphabet*, a word derived from the names of the first two letters of the Greek alphabet, "alpha" and "beta".

e. Theoretically, in an ideal alphabet each symbol or letter would represent only one elementary sound, and each elementary sound would invariably be represented by the same symbol. But such an alphabet would be far too difficult for the average person to use. It has been conservatively estimated that a minimum of 100 characters would be necessary for English alone. Attempts toward producing and introducing into usage a practical, scientific alphabet have been made, one being that of the Simplified Spelling Board in 1928, which advocated a revised alphabet of 42 characters. Were such an alphabet adopted into current usage, in books, letters, telegrams, etc., the flexibility of cryptographic systems would be considerably extended and the difficulties set in the path of the enemy cryptanalysts greatly increased. The chances for its adoption in the near future are, however, quite small. Because of the continually changing nature of every living language, it is doubtful whether an initially "perfect alphabet" could, over any long period of time, remain so and serve to indicate with great precision the exact sounds which it was originally designed to represent.

13. Types of alphabets.—*a.* In the study of cryptography the dual nature of the alphabet becomes apparent. It consists of two parts or components, (1) an arbitrarily-arranged sequence of sounds, and (2) an arbitrarily-arranged sequence of symbols.

b. The *normal alphabet* for any language is one in which these two components are the ordinary sequences that have been definitely fixed by long usage or convention. The dual nature of our normal or everyday alphabet is often lost sight of. When we write A, B, C, . . . we really mean:

Sequence of sounds:	"Ay"	"Bee"	"See"	......
Sequence of symbols:	A	B	C	......

Normal alphabets of different languages vary considerably in the number of characters composing them and the arrangement or sequence of the characters. The English, Dutch, and German alphabets each have 26; the French, 25; the Italian, 21; the Spanish, 27 (including the digraphs CH and LL); and the Russian, 31.[12] The Japanese language has a syllabary consisting of 72 syllabic sounds which require 48 characters for their representation.

c. A *cipher alphabet*, or *substitution alphabet* as it is sometimes called, is one in which the elementary speech-sounds are represented by characters other than those representing them in the normal alphabet. These characters may be letters, figures, signs, symbols, or combinations of these.

d. When the plain text of a message is converted into encrypted text by the use of one or more cipher alphabets, the resultant cryptogram constitutes a *substitution cipher*. If only one cipher alphabet is involved, it is called a *monoalphabetic substitution cipher;* if two or more cipher alphabets are involved, it is called a *polyalphabetic substitution cipher*.

e. It is convenient to designate that component of a cipher alphabet constituting the sequence of speech-sounds as the *plain component* and the component constituting the sequence of symbols as the *cipher component*. If omitted in a cipher alphabet, the plain component is understood to be the normal sequence. For brevity and clarity, a letter of the plain text, or of the

[12] In contrast to the foregoing alphabets, it is of interest to note that in the Hawaiian language the alphabet consists of only 12 letters, *viz.*, the five vowels A, E, I, O, U, and the seven consonants H, K, L, M, N, P, W.

plain component of a cipher alphabet, is designated by suffixing a small letter "p" to it: A_p means A of the plain text, or of the plain component of a cipher alphabet. Similarly, a letter of the cipher text, or of the cipher component of a cipher alphabet, will be designated by suffixing a small letter "c" to it: X_c means X of the cipher text, or of the cipher component of a cipher alphabet. The expression $A_p=X_c$ means that A of the plain text, or A of the plain component of a cipher alphabet, is represented by X in the cipher text, or by X in the cipher component of a cipher alphabet.

f. With reference to the arrangement or sequence of letters forming their components, cipher alphabets are of two types:

(1) *Standard cipher alphabets,* in which the sequence of letters in the plain component is the normal, and in the cipher component is the same as the normal, but reversed in direction or shifted from its normal point of coincidence with the plain component.

(2) *Mixed cipher alphabets,* in which the sequence of letters or characters in one or both of the components is no longer the same as the normal in its entirety.

g. Although the basic considerations of the preceding paragraphs place the student in a position to undertake the study of certain fundamental principles of cryptanalysis, this may be a good point at which to pause and to make a few remarks with regard to the role that cryptanalysis plays in the whole chain of more or less complex operations involved in deriving communication intelligence, after which these fundamental cryptanalytic principles will be treated.

CHAPTER III

FUNDAMENTAL CRYPTANALYTIC OPERATIONS

14. The role of cryptanalysis in communication intelligence operations.—*a.* Through the medium of communication intelligence an attempt is made to answer three questions concerning enemy communications: "Who?" "Where?" "What?"—*Who* are their originators and addressees? *Where* are these originators and addressees located? *What* do the messages say?

b. All of the foregoing questions are very important in the military application of communication intelligence. Hence, even though this text deals almost exclusively with the principles and operations involved in deriving the answer to the third question—"What do the messages say?"—a few words on the importance of the first and second questions may be useful. It is a serious mistake to think that one can necessarily and always correctly interpret the mere text of a message without identifying and locating the originator and the addressee or, on many occasions, without having a background against which to interpret the message in order to appreciate its real import or significance.

c. The very first step in the series of activities involved in deriving communication intelligence is the collection of the raw material, that is, the *interception* [1] and copying of the transmissions constituting the messages to be studied and analyzed.

d. Then, with the raw material in hand, studies are made in order to answer the first two questions—"Who?" and "Where?" The answers to these questions are not always obvious in modern military communications, especially in the case of messages exchanged by units in the combat zone, since messages of this sort rarely indicate in plain language *who* the originator and the addressee are or *where* they are located. Consequently, certain apparatus and techniques specifically developed for finding the answers to these questions must be employed. These apparatus and techniques are embraced by that part of communication intelligence theory and practice which is known as *traffic analysis.* This latter subject and interception are treated briefly in Appendix 7, "Communication intelligence operations". (The serious student will derive much practical benefit from a careful reading of this appendix.)

e. The foregoing operations, interception and traffic analysis, along with *cryptanalysis* constitute the first three operations of communication intelligence. But generally there must follow at least one additional operation. If the plain texts recovered through cryptanalysis are

[1] To *intercept* means, in its cryptologic sense, to gain possession of communications which are intended for other recipients, without obtaining the consent of these addressees and without preventing or (ordinarily) delaying the transmission of the communications to them.

in a foreign language, they must usually be translated, and *translation* constitutes this fourth operation. In the course of translating, it may be found that, because of errors in transmission or reception, corrections and emendations must be made in these plain texts; however, although this often requires skill and experience of a high order, it does not constitute another communication intelligence operation, since it is but an auxiliary step to the process of translation.

f. In a large-scale communication intelligence effort these four steps, interception, traffic analysis, cryptanalysis, and translation, must be properly organized and coordinated in order to gain the most benefit from the potentialities of communication intelligence, that is, the production of the maximum quantity of information from the raw traffic. This information must then be *evaluated* by properly trained intelligence specialists, *collated* with intelligence derived from other sources, and, finally, *disseminated* to the commanders who need the intelligence in time to be of *operational* use to them, rather than of mere historical interest. The foregoing operations and especially the first three—interception, traffic analysis, and cryptanalysis—usually complement one another. This, however, is not the place for elaboration on the interrelationships which exist and which when properly integrated make the operations as a whole an efficient, unified complex geared to the fulfillment of its principal goal, namely, the production of timely communication intelligence.

g. With the foregoing general background, the student is prepared to proceed to the technical considerations and principles of cryptanalysis.

15. The four basic operations in cryptanalysis.—*a.* The solution of practically every cryptogram involves four fundamental operations or steps:

(1) The determination of the language employed in the plaintext version.

(2) The determination of the general system of cryptography employed.

(3) The reconstruction of the specific key in the case of a cipher system, or the reconstruction, partial or complete, of the code book, in the case of a code system; or both, in the case of an enciphered code system.

(4) The reconstruction or establishment of the plain text.

b. These operations will be taken up in the order in which they are given above and in which they usually are performed in the solution of cryptograms, although occasionally the second step may precede the first.[2]

---

[2] Although the foregoing four steps represent the classical or ideal approach to cryptanalysis, the art may be reduced to the following:

Procedures in cryptanalysis	*Requirements*
1. Arrangement and rearrangement of data to disclose nonrandom characteristics or manifestations (i. e., in frequency counts, repetitions, patterns, symmetrical phenomena, etc.).	Experience or ingenuity, and time (which latter may be appreciably lowered by the use of machine aids in cryptanalysis).
2. Recognition of the nonrandom characteristics or manifestations when disclosed.	Experience or statistics.
3. Explanation of the nonrandom characteristics when recognized.	Experience or imagination, *and* intelligence.

In all of the foregoing, the element of luck plays a very important part, as it is possible to side-step a large amount of labor and effort, in many cases, if "hunches" or intuition lead the analyst forthwith to the right path. Therefore, the phrase "or luck" should be added to each of the requirements above.

In fact, it all boils down to the simple statement: "Find something significant, and attach some significance thereto."

16. The determination of the language employed.—*a.* There is not much that need be said with respect to this operation except that the determination of the language employed seldom comes into question in the case of studies made of the cryptograms of an organized enemy. By this is meant that during wartime the enemy is of course known, and it follows, therefore, that the language he employs in his messages will almost certainly be his native or mother tongue. Only occasionally nowadays is this rule broken. Formerly it often happened, or it might have indeed been the general rule, that the language used in diplomatic correspondence was not the mother tongue, but French. In isolated instances during World War I the Germans used English when their own language could for one reason or another not be employed. For example, for a year or two before the entry of the United States into that war, during the time America was neutral and the German Government maintained its embassy in Washington, some of the messages exchanged between the Foreign Office in Berlin and the Embassy in Washington were encrypted in English, and a copy of the code used was deposited with the Department of State and our censor. Another instance is found in the case of certain Hindu conspirators who were associated with and partially financed by the German Government in 1915 and 1916; they employed English as the language of their cryptographic messages. Occasionally the cryptograms of enemy agents may be in a language different from that of the enemy. But in general these are, as has been said, isolated instances; as a rule, the language used in cryptograms exchanged between members of large organizations is the mother tongue of the correspondents. Where this is not the case, that is, when cryptograms of unknown origin must be studied, the cryptanalyst looks for any indications on the cryptograms themselves which may lead to a conclusion as to the language employed. Address, signature, and other data, *if in plain text* in the preamble, in the body, or at the end of the cryptogram, all come under careful scrutiny, as well as all extraneous circumstances connected with the manner in which the cryptograms were obtained, the person on whom they were found, or the locale of their origin and destination.

b. In special cases, or under special circumstances a clue to the language employed is found in the nature and composition of the cryptographic text itself. For example, if the letters K and W are entirely absent or appear very rarely in messages, it may indicate that the language is Spanish or Portuguese for these letters are absent in the alphabets of these languages and are used only to spell foreign words or names. The presence of accented letters or letters marked with special signs of one sort or another, peculiar to certain languages, will sometimes indicate the language used. The Japanese Morse telegraph alphabet contains combinations of dots and dashes which are peculiar to that alphabet and thus the interception of messages containing these special Morse combinations at once indicates the language involved. Finally, there are certain peculiarities of alphabetic languages which, in certain types of cryptograms, *viz.*, pure transposition, give clues as to the language used. For example, the frequent digraph CH, in German, leads to the presence, in cryptograms of the type mentioned, of many isolated C's and H's; if this is noted, the cryptogram may be assumed to be in German.

c. In some cases it is perfectly possible to perform certain steps in cryptanalysis *before* the language of the cryptogram has been definitely determined. Frequency studies, for example, may be made and analytic processes performed without this knowledge, and by a cryptanalyst wholly unfamiliar with the language even if it has been identified, or who knows only enough about the language to enable him to recognize valid combinations of letters, syllables, or a few common words in that language. He may, after this, call to his assistance a translator who may not be a cryptanalyst but who can materially aid in making necessary assumptions based upon his special knowledge of the characteristics of the language in question. Thus, cooperation between cryptanalyst and translator results in solution.[8]

17. The determination of the general system.—*a.* Except in the case of the more simple types of cryptograms, the step often referred to as *diagnosis*, that is, ascertaining the general system according to which a given cryptogram has been produced is usually a difficult, if not the most difficult, step in its solution. The reason for this is not hard to find.

b. As will become apparent to the student as he proceeds with his study, *in the final analysis, the solution of every cryptogram involving a form of substitution depends upon its reduction to monoalphabetic terms, if it is not originally in those terms.* This is true not only of ordinary substitution ciphers, but also of combined substitution-transposition ciphers, and of enciphered code. If the cryptogram must be reduced to monoalphabetic terms, the manner of its accomplishment is usually indicated by the cryptogram itself, by external or internal phenomena which become apparent to the cryptanalyst as he studies the cryptogram. If this is impossible, or too difficult, the cryptanalyst must, by one means or another, discover how to accomplish this reduction, by bringing to bear all the special or collateral information he can get from all the sources at his command. If both these possibilities fail him, there is little left but the long, tedious, and often fruitless process of elimination. In the case of transposition ciphers of the more complex type, the discovery of the basic method is often simply a matter of long and tedious elimination of possibilities. For cryptanalysis has unfortunately not yet attained, and may indeed never attain, the precision found today in qualitative analysis in chemistry, for example, where the analytic process is absolutely clear-cut and exact in its dichotomy. A few words in explanation of what is meant may not be amiss. When a chemist seeks to determine the identity of an unknown substance, he applies certain specific reagents to the substance and in a specific sequence. The first reagent tells him definitely into which of two primary classes the unknown substance falls. He then applies a second test with another specific reagent, which tells him again quite definitely into which of two secondary classes the unknown substance falls, and so on, until finally he has reduced the unknown substance to its simplest terms and has found out what it is. In striking contrast to this situation, cryptanalysis affords exceedingly few "reagents" or tests that may be applied to determine positively that a given cipher belongs to one or the other of two systems yielding externally similar results. And this is what makes the analysis of an isolated, complex cryptogram so difficult. Note the limiting adjective "isolated" in the foregoing sentence, for it is used advisedly. It is not often that the general system fails to disclose itself or cannot be discovered by painstaking investigation when there is a great volume of text accumulating from a regular traffic between numerous correspondents in a large organization. *Sooner or later* the system becomes known, either because of blunders and carelessness on the part of the personnel entrusted with the encrypting of the messages, or because the accumulation of text itself makes possible the determination of the general sys-

[8] The writer has seen in print statements that "during the World War . . . . decoded messages in Japanese and Russian without knowing a word of either language." The extent to which such statements are exaggerated will soon become obvious to the student. Of course, there are occasional instances in which a mere clerk with quite limited experience may be able to "solve" a message in an extremely simple system in a language of which he has no knowledge at all; but such a "solution" calls for nothing more arduous than the ability to recognize pronounceable combinations of vowels and consonants—an ability that hardly deserves to be rated as "cryptanalytic" in any real sense. To say that it is possible to solve a cryptogram in a foreign language "without knowing a word of that language" is not quite the same as to say that it is possible to do so with only a slight knowledge of the language; and it may be stated without cavil that the better the cryptanalyst's knowledge of the language, the greater are the chances for his success and, in any case, the easier is his work.

tem by cryptanalytic, including statistical, studies. But in the case of a single or even a few isolated cryptograms concerning which little or no information can be gained by the cryptanalyst, he is often unable, without a knowledge of, or a shrewd guess as to the general system employed, to decompose the heterogeneous text of the cryptogram into homogeneous, monoalphabetic text, which is the ultimate and essential step in analysis. The only knowledge that the cryptanalyst oan bring to his aid in this most difficult step is that gained by long experience and practice in the analysis of many different types of systems. In this respect the practice of cryptanalysis is analogous to the practice of medicine: correct diagnosis is the most important and often the most difficult first step toward success.

c. On account of the complexities surrounding this particular phase of cryptanalysis, and because in any scheme of analysis based upon successive eliminations of alternatives the cryptanalyst can only progress as far as the extent of his own knowledge of *all* the possible alternatives will permit, it is necessary that detailed discussion of the eliminative process be postponed until the student has covered most of the field. For example, the student will perhaps want to know at once how he can distinguish between a cryptogram that is in code or enciphered code from one that is in cipher. It is at this stage of his studies impracticable to give him any helpful indications on his question. In return it may be asked of him why he should expect to be able to do this in the early stages of his studies when often the experienced expert cryptanalyst is baffled on the same score!

d. Much of the labor involved in cryptanalytic work, as referred to in par. 2, is connected with this determination of the general system. The preparation of the text, its rewriting in different forms, sometimes being rewritten in dozens of ways, the recording of letters, the establishment of frequencies of occurrences of letters, comparisons and experiments made with known material of similar character, and so on, constitute much labor that is most often indispensable, but which sometimes turns out to have been wholly unnecessary, or in vain. In one treatise [4] it is stated quite boldly that "this work once done, the determination of the system is often relatively easy." This statement can certainly apply only to the simpler types of cryptosystems; it is entirely misleading as regards the much more frequently encountered complex cryptograms of modern times.

18. The reconstruction of the specific key.—*a.* Nearly all practical cryptographic methods require the use of a specific key to guide, control, or modify the various steps under the general system. Once the latter has been disclosed, discovered, or has otherwise come into the possession of the cryptanalyst, the next step in solution is to determine, if necessary and if possible, the specific key that was employed to encrypt the message or messages under examination. This determination may not be in complete detail; it may go only so far as to lead to a knowledge of the number of alphabets involved in a substitution cipher, or the number of columns involved in a transposition cipher, or that a one-part code has been used, in the case of a code system. But it is often desirable to determine the specific key in as complete a form and with as much detail as possible, for this information will very frequently be useful in the solution of subsequent cryptograms exchanged between the same correspondents, since the nature or source of the specific key in a solved case may be expected to give clues to the specific key in an unsolved case.

b. Frequently, however, the reconstruction of the key is not a prerequisite to, and does not constitute an absolutely necessary preliminary step in, the fourth basic operation, *viz.*, the reconstruction or establishment of the plain text. In many cases, indeed, the two processes are carried along simultaneously, the one assisting the other, until in the final stages both have been com-

[4] Lange et Soudart, *op. cit.*, p. 106.

pleted in their entireties. In still other cases the reconstruction of the specific key may follow the reconstruction of the plain text instead of preceding it and is accomplished purely as a matter of academic interest; or the specific key may, in unusual cases, never be reconstructed.

19. The reconstruction of the plain text.—*a.* Little need be said at this point on this phase of cryptanalysis. The process usually consists, in the case of substitution ciphers, in the establishment of equivalency between specific letters of the cipher text and the plain text, letter by letter, pair by pair, and so on, depending upon the particular type of substitution system involved. In the case of transposition ciphers, the process consists in rearranging the elements of the cipher text, letter by letter, pair by pair, or occasionally word by word, depending upon the particular type of transposition system involved, until the letters or words have been returned to their original plaintext order. In the case of code, the process consists in determining the meaning of each code group and inserting this meaning in the code text to reestablish the original plain text.

b. The foregoing processes do not, as a rule, begin at the beginning of a message and continue letter by letter, or group by group in sequence up to the very end of the message. The establishment of values of cipher letters in substitution methods, or of the positions to which cipher letters should be transferred to form the plain text in the case of transposition methods, comes at very irregular intervals in the process. At first only one or two values scattered here and there throughout the text may appear; these then form the "skeletons" of words, upon which further work, by a continuation of the reconstruction process, is made possible; in the end the complete or nearly complete [5] text is established.

c. In the case of cryptograms in a foreign language, the translation of the solved messages is a final and necessary step, but is not to be considered as a cryptanalytic process. However, it is commonly the case that the translation process will be carried on simultaneously with the cryptanalytic, and will aid the latter, especially when there are lacunae which may be filled in from the context. (See also subpar. 16*c* in this connection.)

20. The utilization of traffic intercepts.[6]—*a.* There are, of course, other operations which are not as basic in nature as those just outlined but which must generally be performed as preliminary steps in *practical* cryptanalytic work (as distinguished from *academic* cryptanalysis). Before a military cryptanalyst can begin the analysis of an enemy cryptosystem, it is necessary for him to study the intercept material that is available to him, isolate the messages that have been encrypted by means of the cryptosystem to be examined, and to arrange the latter in a systematic order for analysis. This work, although apparently very simple, may require a great deal of time and effort.

b. Since, whenever practicable, two or more intercept stations are assigned to copy traffic [7] emanating from the stations of one enemy radio net, it is natural that there should be a certain amount of duplication in the work of these several intercept stations. This is desirable since it provides the cryptanalysts with two or more sets of the same messages, so that when one intercept station fails to receive all the messages completely and correctly, because of radio diffi-

[5] Sometimes in the case of code, the meaning of a small percentage of the code groups occurring in the traffic may be lacking, because there is insufficient text to establish their meaning.

[6] A *traffic intercept* is a copy of a communication gained through interception.

[7] In manual transmission systems, traffic is usually sent in Morse code, consisting of combinations of short signals ("dots") and long signals ("dashes") to make up an "alphabet" for the transmission of the letters, digits, and punctuation symbols of a particular language. It is interesting to note that Samuel F. B. Morse constructed his alphabet in such a manner that, generally speaking, the shorter signals applied to the highest frequency letters in English, while the longer signals were used to represent the lowest frequency letters.

culties, local static, or poor operation, it is possible by studying the other sets to reconstruct accurately the entire traffic of the enemy net.

c. In all intercept activities where operators are used for copying the traffic, one of the most likely errors to be found is caused by the human element in reception. For this reason cryptanalysts and their assistants should be familiar with the international Morse alphabet and the

Ltrs. and Figs.	Morse equivalent	Frequent Errors	Ltrs. and Figs.	Morse equivalent	Frequent Errors
A	.—	i, m, t, et	S	...	h, d, i, r, u
B	—...	d, ts	T	—	a, e, n
C	—.—.	f, k, r, nn	U	..—	a, s, v, it
D	—..	b, s, l, ti	V	...—	h, u, x, st
E	.	t, i	W	.——	a, m, o, r, u, at
F	..—.	r, in	X	—..—	v, k, y, tu
G	——.	m, o, z, me	Y	—.——	x, c, nm
H	....	s, v, b, ii, se	Z	——..	b, g, q, mi
I	..	a, n, s	1	.————	ø, 2
J	.———	w, o, am, eo	2	..———	1, 3
K	—.—	d, o, ta	3	...——	2, 4
L	.—..	r, d, ed	4	....—	3, 5
M	——	a, n, tt	5	.....	4, 6
N	—.	i, m, t, te	6	—....	5, 7
O	———	g, k, w, mt	7	——...	6, 8
P	.——.	j, g, l, w, an	8	———..	7, 9
Q	——.—	o, x, z, ma	9	————.	8, ø
R	.—.	a, f, g, l, n, s, w	ø	—————	9, 1

CHART 1. Most common errors in telegraphic transmission.

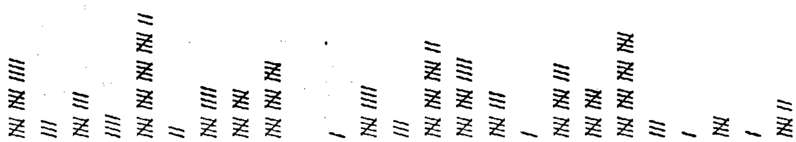
most common errors in wire and radio transmission methods so as to be able to correct garbled groups when they occur. In this connection, Chart 1, above, will be found useful.

d. Besides the message texts themselves, the intercept operator also copies the *call signs* (together with the frequencies on which heard) and the elements of the *preamble* of the messages as transmitted by the enemy. The preamble may have great flexibility among various users, but usually includes a *station serial number* (abbr. "NR") assigned by the radio operator for referencing transmitted traffic, and a *group count* (abbr. "GR") as a check on the number of groups transmitted. In addition, there may also be preamble elements that signify precedence, routing or addressee instructions, the date and time of file, and other items that might facilitate the handling or processing of the traffic.

CHAPTER IV

FREQUENCY DISTRIBUTIONS AND THEIR FUNDAMENTAL USES

21. **The simple or uniliteral frequency distribution.**—*a*. It has long been known to cryptographers and typographers that the letters composing the words of any intelligible written text composed in any language which is alphabetic in construction are employed with greatly varying frequencies. For example, if on cross-section paper a simple tabulation, shown in Fig. 1, called a *uniliteral frequency distribution*, is made of the letters composing the words of the preceding sentence, the variation in frequency is strikingly demonstrated. It is seen that whereas certain letters, such as A, E, I, N, O, R, and T, are employed very frequently, other letters, such as C, G, H, L, P, and S are employed not nearly so frequently, while still other letters, such as F, J, K, Q, V, X, and Z are employed either seldom or not at all.

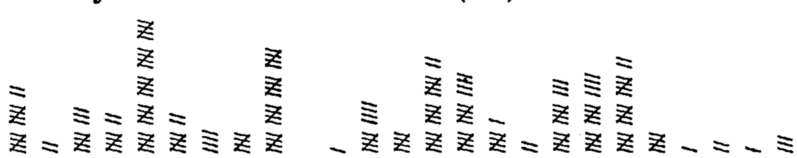


A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
14	3	8	4	22	2	9	10	15	0	1	9	3	17	14	8	1	13	10	20	3	1	5	1	7	0

(Total=200 letters)

FIGURE 1.

b. If a similar tabulation is now made of the letters comprising the words of the second sentence in the preceding subparagraph, the distribution shown in Fig. 2 is obtained. Both sentences have exactly the same number of letters (200).



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
12	2	8	7	25	7	4	5	20	0	1	9	5	17	14	6	2	13	14	17	5	1	2	1	3	0

(Total=200 letters)

FIGURE 2.

c. Although each of these two distributions exhibits great variation in the relative frequencies with which *different* letters are employed in the respective sentences to which they apply, no marked differences are exhibited between the frequencies of the *same* letter in the two distributions. Compare, for example, the frequencies of A, B, C . . . Z in Fig. 1 with those of A, B, C . . . Z in Fig. 2. Aside from one or two exceptions, as in the case of the letter F, these two distributions agree rather strikingly.

d. This agreement, or *similarity,* would be practically complete if the two texts were much longer, for example, five times as long. In fact, when two texts of similar character, each containing more than 1,000 letters, are compared, it would be found that the respective frequencies of the 26 letters composing the two distributions show only very slight differences. This means, in other words, that in normal plain text each letter of the alphabet occurs with a rather *constant* or *characteristic frequency* which it tends to approximate, depending upon the length of the text analyzed. The longer the text (within certain limits), the closer will be the approximation to the characteristic frequencies of letters in the language involved. However, when the amount of text being analyzed has reached a substantial volume (roughly, 1,000 letters), the practical gain in accuracy does not warrant further increase in the amount of text.[1]

e. An experiment along these lines will be convincing. A series of 260 official telegrams [2] passing through the Department of the Army Message Center was examined statistically. The

TABLE 1-A.—*Absolute frequencies of letters appearing in the five sets of Governmental plaintext telegrams, each set containing 10,000 letters, arranged alphabetically*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

[1] See footnote 5, p. 30.

[2] These comprised messages from several official sources in addition to the Department of the Army and were all of an administrative character.

messages were divided into five sets, each totaling 10,000 letters, and the five distributions shown in Table 1–A were obtained.

f. If the five distributions in Table 1–A are summed, the results are as shown in Table 2–A.

TABLE 2–A.—*Absolute frequencies of letters appearing in the combined five sets of messages totaling 50,000 letters, arranged alphabetically*

A-----	3,683	G-----	819	L-----	1,821	Q-----	175	V-----	766
B-----	487	H-----	1,694	M-----	1,237	R-----	3,788	W-----	780
C-----	1,534	I-----	3,676	N-----	3,975	S-----	3,058	X-----	231
D-----	2,122	J-----	82	O-----	3,764	T-----	4,595	Y-----	967
E-----	6,498	K-----	148	P-----	1,335	U-----	1,300	Z-----	49
F-----	1,416								

g. The frequencies noted in Table 2–A above, when reduced to a base of 1,000 letters and then used as a basis for constructing a simple chart that will exhibit the variations in frequency in a striking manner, yield the following distribution which is hereafter designated as the *normal* or *standard uniliteral frequency distribution* for English telegraphic plain text:

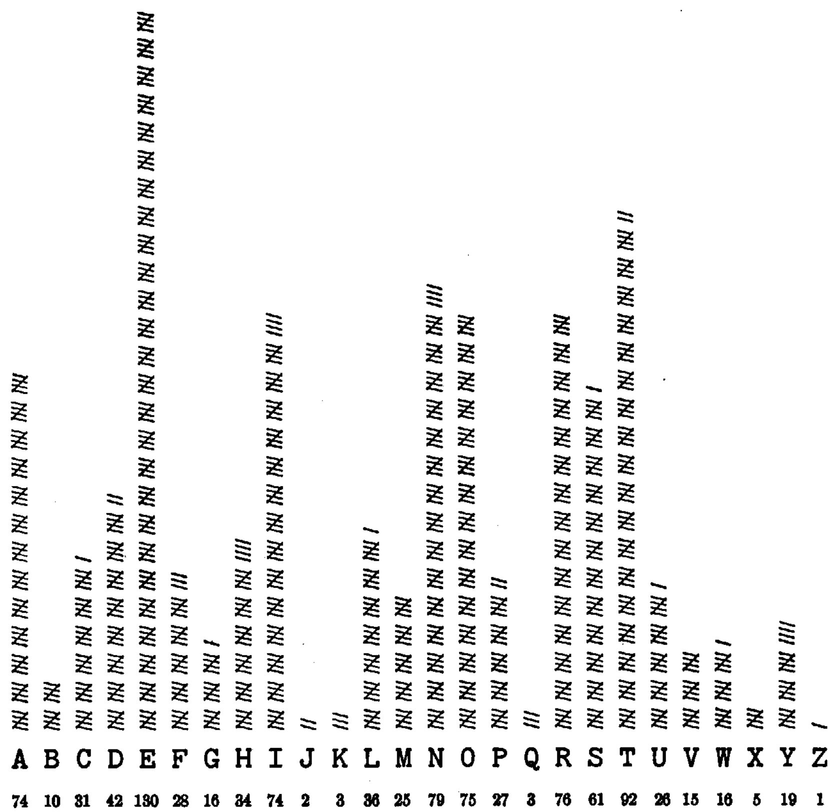


FIGURE 3.

22. Important features of the normal uniliteral frequency distribution.—*a.* When the distribution shown in Fig. 3 is studied in detail, the following features are apparent:

(1) It is quite irregular in appearance. This is because the letters are used with greatly varying frequencies, as discussed in the preceding paragraph. This irregular appearance is often described by saying that the distribution shows marked *crests and troughs*, that is, points of high frequency and low frequency.

(2) The relative positions in which the crests and troughs fall within the distribution, that is, the *spatial relations* of the crests and troughs, are rather definitely fixed and are determined by circumstances which have been explained in subpar. 13*b*.

(3) The relative heights and depths of the crests and troughs within the distribution, that is, the *linear extensions* of the lines marking the respective frequencies, are also rather definitely fixed, as would be found if an equal volume of similar text were analyzed.

(4) The most prominent crests are marked by the vowels A, E, I, O, and the consonants N, R, S, T; the most prominent troughs are marked by the consonants J, K, Q, X, and Z,

(5) The important data are summarized in tabular form in Table 3.

TABLE 3

	Frequency	Percent of total	Percent of total in round numbers
6 Vowels: A E I O U Y	398	39.8	40
20 Consonants:			
5 High Frequency (D N R S T)	350	35.0	35
10 Medium Frequency (B C F G H L M P V W)	238	23.8	24
5 Low Frequency (J K Q X Z)	14	1.4	1
Total	1,000	100.0	100

(6) The frequencies of the letters of the alphabet, reduced to a base of 1,000, are as follows:

A	74	G	16	L	36	Q	3	V	15
B	10	H	34	M	25	R	76	W	16
C	31	I	74	N	79	S	61	X	5
D	42	J	2	O	75	T	92	Y	19
E	130	K	3	P	27	U	26	Z	1
F	28								

(7) The relative order of frequency of the letters is as follows:

E	130	I	74	C	31	Y	19	X	5
T	92	S	61	F	28	G	16	Q	3
N	79	D	42	P	27	W	16	K	3
R	76	L	36	U	26	V	15	J	2
O	75	H	34	M	25	B	10	Z	1
A	74								

(8) The four vowels A, E, I, O (combined frequency 353) and the four consonants N, R, S, T (combined frequency 308) form 661 out of every 1,000 letters of plain text; in other words, *less than one-third of the alphabet is employed in writing two-thirds of normal plain text.*

b. The data given in Fig. 3 and Table 3 represent the relative frequencies found in a large volume of English telegraphic text of a governmental, administrative character.[3] These frequencies will vary somewhat with the nature of the text analyzed. For example, if an equal number of telegrams dealing solely with *commercial* transactions in the *leather industry* were studied statistically, the frequencies would be slightly different because of the repeated occurrence of words peculiar to that industry. Again, if an equal number of telegrams dealing solely with *military* messages of a *tactical* character were studied statistically, the frequencies would differ slightly from those found above for general governmental messages of an administrative character.

c. If ordinary English literary text (such as may be found in any book, newspaper, or printed document) were analyzed, the frequencies of certain letters would be changed to an appreciable degree. This is because, in telegraphic text, words which are not strictly essential for intelligibility (such as the definite and indefinite articles, certain prepositions, conjunctions, and pronouns) are omitted. In addition, certain essential words, such as "stop", "period", "comma", and the like, which are usually indicated in written or printed matter by symbols not easy to transmit telegraphically and which must, therefore, be spelled out in telegrams, occur very frequently. Furthermore, telegraphic text often employs longer and more uncommon words than does ordinary newspaper or book text.

d. As a matter of fact, other tables compiled from Army sources gave slightly different results, depending upon the source of the text. For example, three tables based upon 75,000, 100,000, and 136,257 letters taken from various sources (telegrams, newspapers, magazine articles, books of fiction) gave as the relative order of frequency for the first 10 letters the following:

For 75,000 letters	E T R N I O A S D L
For 100,000 letters	E T R I N O A S D L
For 136,257 letters	E T R N A O I S L D

e. Frequency data applicable purely to English military text were compiled by Hitt,[4] from a study of 10,000 letters taken from orders and reports; these data are given in Table 4, on the next page. Hitt also compiled data for telegraphic text (but does not state what kind of messages); these data are given in Table 5.

---

[3] Just as the individual letters constituting a large volume of plain text have more or less characteristic or fixed frequencies, so it is found that *digraphs* and *trigraphs* (two- and three-letter combinations, respectively) have characteristic frequencies, when a large volume of text is studied statistically. In Table 6 of Appendix 2, "Letter frequency data—English", are shown the relative frequencies of all digraphs appearing in the 260 telegrams referred to in subpar. 21e. This appendix also includes several other kinds of tables and lists of frequency data which will be useful to the student in his work. It is suggested that the student refer to this appendix now, to gain an idea of the data available for his future reference.

Other languages, of course, each have their own individual characteristic plaintext frequencies of single letters, digraphs, trigraphs, etc. A brief summary of the letter frequency data for German, French, Italian, Spanish, Portuguese, and Russian constitutes Appendix 5, "Letter frequency data—foreign languages".

[4] Op. cit., pp. 6–7.

TABLE 4.—*Frequency table for 10,000 letters of nontelegraphic English military text, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

TABLE 5.—*Frequency table for 10,000 letters of telegraphic English military text, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

23. Constancy of the standard or normal uniliteral frequency distribution.—*a.* The relative frequencies disclosed by the statistical study of large volumes of text may be considered to be the standard or *normal* frequencies of the letters of written English. Counts made of smaller volumes of text will tend to approximate these normal frequencies, and, within certain limits,[5]

[5] It is useless to go beyond a certain limit in establishing the normal-frequency distribution for a given language. As a striking instance of this fact, witness the frequency study made by an indefatigable German, Kaeding, who in 1898 made a count of the letters in about 11,000,000 words, totaling about 62,000,000 letters in German text. When reduced to a percentage basis, and when the relative order of frequency was determined, the results he obtained differed very little from the results obtained by Kasiski, a German cryptographer, from a count of only 1,060 letters. See Kaeding, *Haeufigkeitswoerterbuch*, Steglitz, 1898; Kasiski, *Die Geheimschriften und die Dechiffrir-Kunst*, Berlin, 1863.

the smaller the volume, the lower will be the degree of approximation to the normal, until, in the case of a very short message, the normal proportions may not manifest themselves at all. It is advisable that the student fix this fact firmly in mind, for the sooner he realizes the true nature of any data relative to the frequency of occurrence of letters in text, the less often will his labors toward the solution of specific ciphers be thwarted and retarded by too strict an adherence to these generalized principles of frequency.[6] He should constantly bear in mind that such data are merely statistical generalizations, that they will be found to hold strictly true only in large volumes of text, and that they may not even be approximated in short messages.

b. Nevertheless the normal frequency distribution or the "normal expectation" for any alphabetic language is, in the last analysis, the best guide to, and the usual basis for, the solution of cryptograms of a certain type. It is useful, therefore, to reduce the normal, uniliteral frequency distribution to a basis that more or less closely approximates the volume of text which the cryptanalyst most often encounters in individual cryptograms. As regards length of messages, counting only the letters in the body, and excluding address and signature, a study of the 260 telegrams referred to in par. 21 shows that the arithmetical average is 217 letters; the statistical mean, or weighted average,[7] however, is 191 letters. These two results are, however, close enough together to warrant the statement that the *average* length of telegrams is approximately 200 letters. The frequencies given in par. 21 have therefore been reduced to a basis of 200 letters, and the following uniliteral frequency distribution may be taken as showing the most typical distribution to be expected in 200 letters of English telegraphic text:

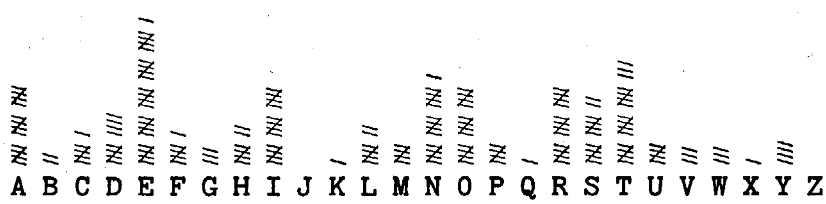


FIGURE 4.

c. The student should take careful note of the appearance of the distribution [8] shown in Fig. 4, for it will be of much assistance to him in the early stages of his study. The manner of setting down the tallies should be followed by him in making his own distributions, indicating every fifth occurrence of a letter by an oblique tally. This procedure almost automatically shows the total number of occurrences for each letter, and yet does not destroy the graphical appearance

[6] A curiosity in this connection is the book "GADSBY" by Ernest Vincent Wright published in Los Angeles, 1939. Written as a tour de force, in this novel of about 50,000 words there is not a single occurrence of the letter "E"!

[7] The arithmetical average is obtained by adding each different length and dividing by the number of different-length messages; the mean is obtained by multiplying each different length by the number of messages of that length, adding all products, and dividing by the total number of messages.

[8] The use of the terms "distribution" and "frequency distribution", instead of "table" and "frequency table," respectively, is considered advisable from the point of view of consistency with the usual statistical nomenclature. When data are given in tabular form, with frequencies indicated by numbers, then they may properly be said to be set out in the form of a *table*. When, however, the same data are distributed in a chart which partakes of the nature of a graph, with the data indicated by horizontal or vertical linear extensions, or by a curve connecting points corresponding to quantities, then it is more proper to call such a graphic representation of the data a *distribution*.

of the distribution, especially if care is taken to use approximately the same amount of space for each set of five tallies. Cross-section paper is very useful for this purpose, since when one is making a frequency distribution on it, he may place each set of five tallies in an individual cell. In making a frequency distribution, each consecutive letter of the sample under study should be recorded as a tally mark; only with this procedure can errors in making a distribution be kept at a minimum. For instance, if the first group of a message is OWQWZ, the first tally mark would be recorded over the "O" in the base of the distribution; the second tally mark, recorded over the "W"; the third tally, over the "Q"; the fourth, over the "W"; and so forth.

d. The word "uniliteral" in the designation "uniliteral frequency distribution" means "single letter," and it is to be inferred that other types of frequency distributions may be encountered. For example, a distribution of pairs of letters, constituting a biliteral frequency distribution, is very often used in the study of certain cryptograms in which it is desired that pairs made by combining successive letters be listed. A biliteral distribution of ABCDEF would take these pairs: AB, BC, CD, DE, EF. The distribution could be made in the form of a large square divided up into 676 cells. When distributions beyond biliteral are required (triliteral, quadriliteral, etc.) they can only be made by listing them in some order, for example, alphabetically based on the 1st, 2d, 3d, . . . letter.

24. The three facts which can be determined from a study of the uniliteral frequency distribution for a cryptogram.—*a.* The following three facts (to be explained subsequently) can usually be determined from an inspection of the uniliteral frequency distribution for a given cipher message of average length, composed of letters:

(1) Whether the cipher belongs to the substitution or the transposition class;

(2) If to the former, whether it is monoalphabetic [9] or nonmonoalphabetic [10] in character;

(3) If monoalphabetic, whether the cipher alphabet is standard (direct or reversed) or mixed.

b. For immediate purposes the first two of the foregoing determinations are quite important and will be discussed in detail in the next two paragraphs; the other determination will be touched upon very briefly, leaving its detailed discussion for subsequent sections of the text.

25. Determining the class to which a cipher belongs.—*a.* The determination of the class to which a cipher belongs is usually a relatively easy matter because of the fundamental difference between transposition and substitution as cryptographic processes. In a transposition cipher the original letters of the plain text have merely been rearranged, without any change whatsoever in their identities, that is, in the conventional values they have in the normal alphabet. Hence, the numbers of vowels (A, E, I, O, U, Y), high-frequency consonants (D, N, R, S, T), medium-frequency consonants (B, C, F, G, H, L, M, P, V, W), and low-frequency consonants (J, K, Q, X, Z) are exactly the same in the cryptogram as they are in the plaintext message. Therefore, the percentages of vowels, high-, medium-, and low-frequency consonants are the same in the transposed text as in the equivalent plain text. In a substitution cipher, on the other hand, the identities of the original letters of the plain text have been changed, that is, the conventional values they have in the normal alphabet have been altered. Consequently, if a count is made

[9] In connection with uniliteral frequency distributions, the term monoalphabetic is considered to embrace the concept of monoalphabetic-monographic-uniliteral systems only, thus excluding *polygraphic* and *multiliteral* systems, both of which, however, usually fall into the monoalphabetic category.

[10] The term nonmonoalphabetic as applied in this instance is considered to embrace all deviations from the characteristic appearance of monoalphabetic distributions. These deviations include the phenomena inherent in polyalphabetic, polygraphic, and multiliteral cryptograms, as well as in *random* text, i. e., text which appears to have been produced by chance or accident, having no discernible patterns or limitations.

of the various letters present in such a cryptogram, it will be found that the number of vowels, high-, medium-, and low-frequency consonants will usually be quite different in the cryptogram from what they are in the original plaintext message. Therefore, the percentages of vowels, high-, medium-, and low-frequency consonants are usually quite different in the substitution text from what they are in the equivalent plain text. From these considerations it follows that if in a specific cryptogram the percentages of vowels, high-, medium-, and low-frequency consonants are approximately the same as would be expected in normal plain text, the cryptogram *probably* belongs to the transposition class; if these percentages are quite different from those to be expected in normal plain text the cryptogram *probably* belongs to the substitution class.

b. In the preceding subparagraph the word "probably" was emphasized by italicizing it, for there can be no certainty in every case of this determination. *Usually* these percentages in a transposition cipher are close to the normal percentages for plain text; *usually*, in a substitution cipher, they are far different from the normal percentages for plain text. But occasionally a cipher message is encountered which is difficult to classify with a reasonable degree of certainty because the message is too short for the general principles of frequency to manifest themselves. It is clear that if in actual messages there were no variation whatever from the normal vowel and consonant percentages given in Table 3, the determination of the class to which a specific cryptogram belongs would be an extremely simple matter. But unfortunately there is always some variation or deviation from the normal. Intuition suggests that as messages decrease in length there may be a greater and greater departure from the normal proportions of vowels, high-, medium-, and low-frequency consonants, until in very short messages the normal proportions may not hold at all. Similarly, as messages increase in length there may be a lesser and lesser departure from the normal proportions, until in messages totalling a thousand or more letters there may be no difference at all between the actual and the theoretical proportions. But intuition is not enough, for in dealing with specific messages of the length of those commonly encountered in practical work the question sometimes arises as to exactly how much deviation (from the normal proportions) may be allowed for in a cryptogram which shows a considerable amount of deviation from the normal and which might still belong to the transposition rather than to the substitution class.

c. Statistical studies have been made on this matter and some graphs have been constructed thereon. These are shown in Charts 2–5 in the form of simple curves, the use of which will now be explained. Each chart contains two curves marking the lower and upper limits, respectively, of the theoretical amount of deviation (from the normal percentages) of vowels or consonants which may be allowable in a cipher believed to belong to the transposition class.

d. In Chart 2, curve V_1 marks the lower limit of the theoretical amount of deviation [11] from the number of vowels theoretically expected to appear [12] in a message of given length; curve V_2 marks the upper limit of the same statistic. Thus, for example, in a message of 100 letters in plain English there should be between 33 and 47 vowels (A E I O U Y). Likewise, in Chart 3, curves H_1 and H_2 mark the lower and upper limits as regards the high-frequency consonants. In a message of 100 letters there should be between 28 and 42 high-frequency consonants (D N R S T). In Chart 4, curves M_1 and M_2 mark the lower and upper limits as regards the medium-frequency consonants. In a message of 100 letters there should be between 17 and 31 medium-frequency

[11] In Charts 2–5, inclusive, the limits of the upper and lower curves have been calculated to include approximately 70 per cent of messages of the various lengths.

[12] The expression "the number of . . . theoretically expected to appear" is often condensed to "the theoretical expectation of . . ." or "the normal expectation of . . ."

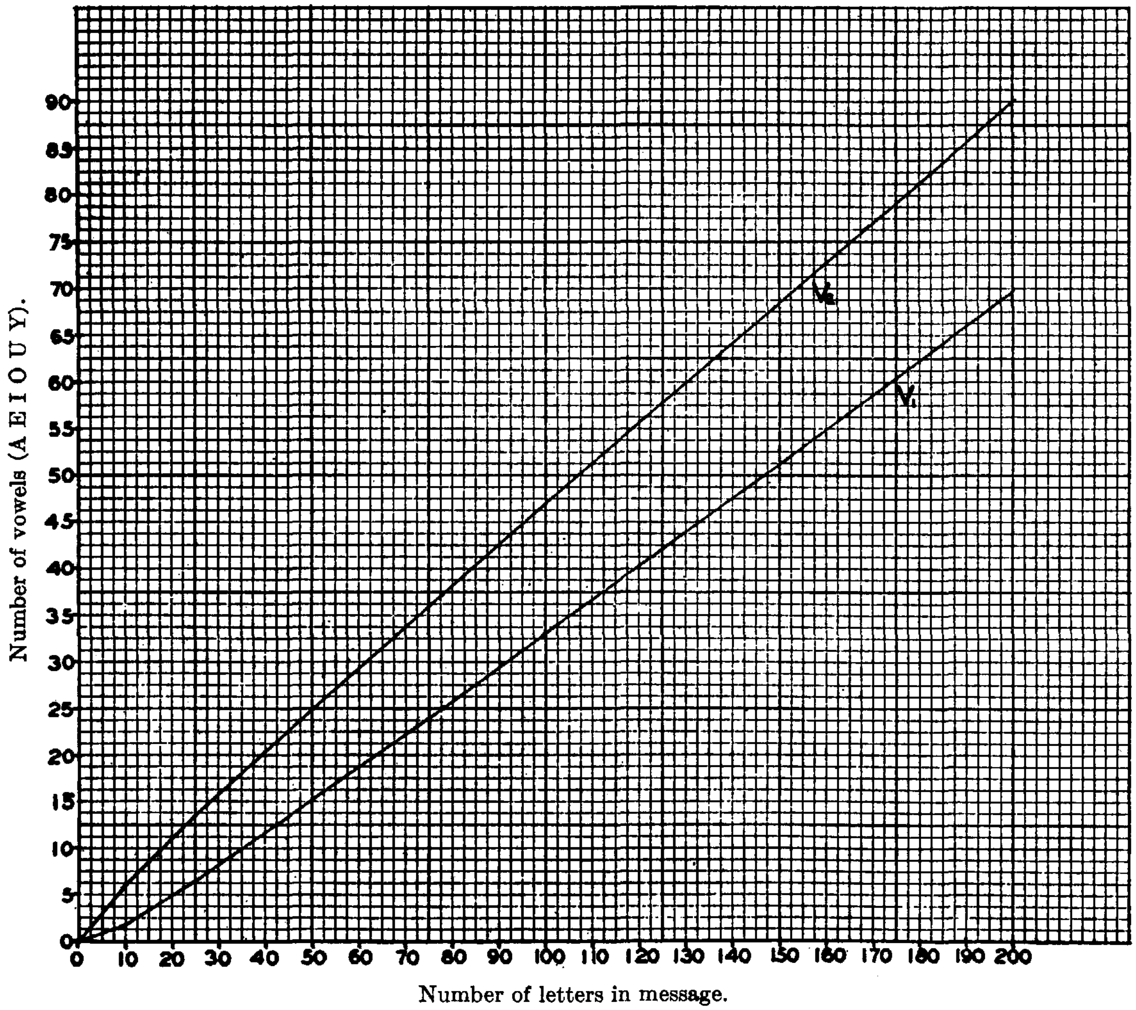


CHART 2. Curves making the lower and upper limits of the theoretical amount of deviation from the number of vowels theoretically expected in messages of various lengths. (See subpar. 25*d*.)

consonants (B C F G H L M P V W). Finally, in Chart 5, curves L_1 and L_2 mark the lower and upper limits as regards the low-frequency consonants. In a message of 100 letters there should be between 0 and 3 low-frequency consonants (J K Q X Z). In using the charts, therefore, one finds the point of intersection of the coordinate (below the chart) corresponding to the length of the message, with the coordinate (to the left of the chart) corresponding to (1) the number of vowels, (2) the number of high-frequency consonants, (3) the number of medium-frequency consonants, and (4) the number of low-frequency consonants actually counted in the message. If all four points of intersection fall within the area delimited by the respective curves, then the numbers of vowels and high-, medium-, and low-frequency consonants correspond with the numbers theoretically expected in a normal plaintext message of the same length; since the message under investigation is not plain text, it follows that the cryptogram may certainly be

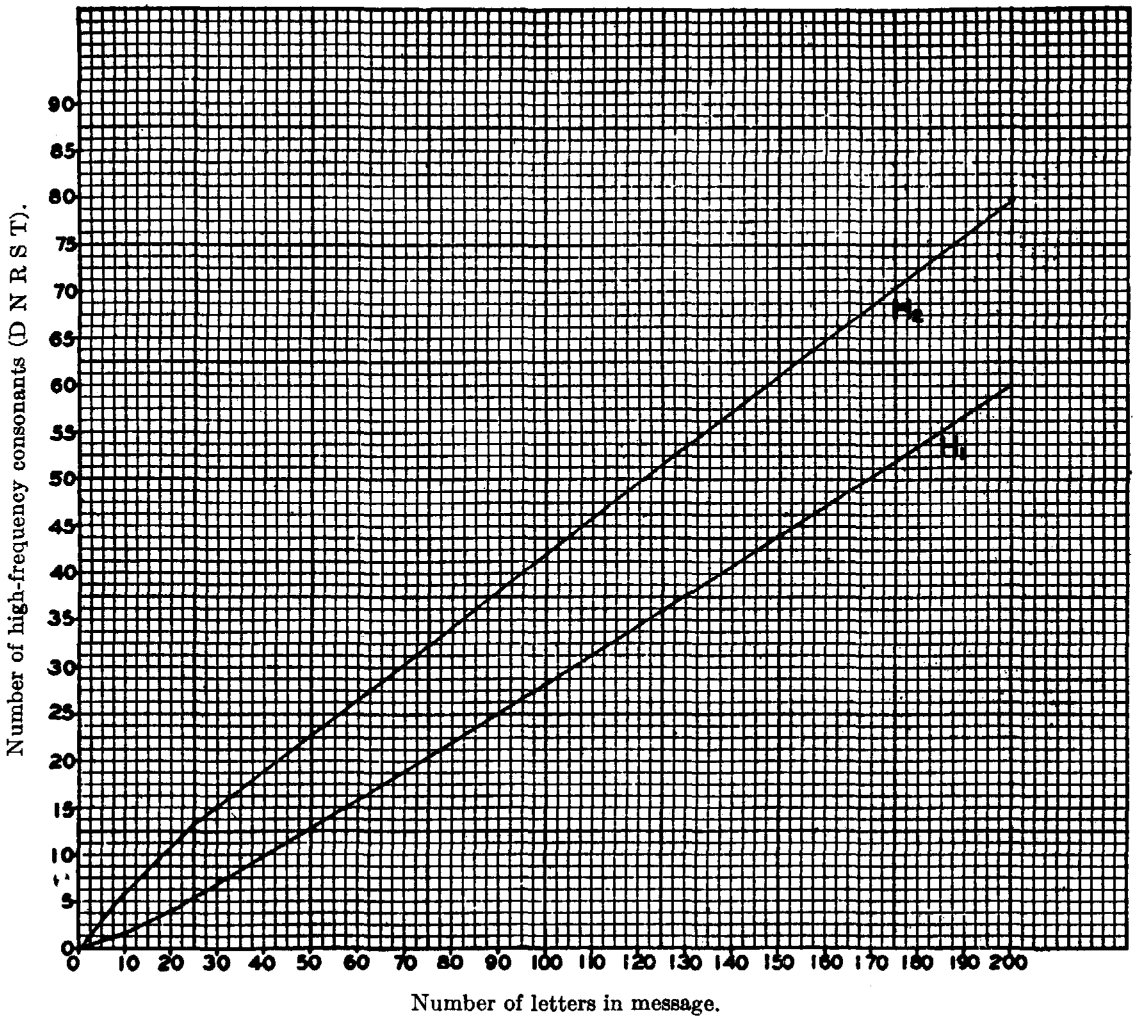


CHART 3. Curves making the lower and upper limits of the theoretical amount of deviation from the number of high-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25*d*.)

classified as a transposition cipher. On the other hand, if one or more of these points of intersection fall outside the area delimited by the respective curves, it follows that the cryptogram is probably a substitution cipher. The distance that the point of intersection falls outside the area delimited by these curves is a more or less rough measure of the improbability of the cryptogram's being a transposition cipher.

e. Sometimes a cryptogram is encountered which is hard to classify with certainty even with the foregoing aids, because it has been consciously prepared with a view to making the classification difficult. This can be done either by selecting peculiar words (as in "trick cryptograms") or by employing a cipher alphabet in which letters of *approximately similar normal frequencies* have been interchanged. For example, E may be replaced by O, T by R, and so on,

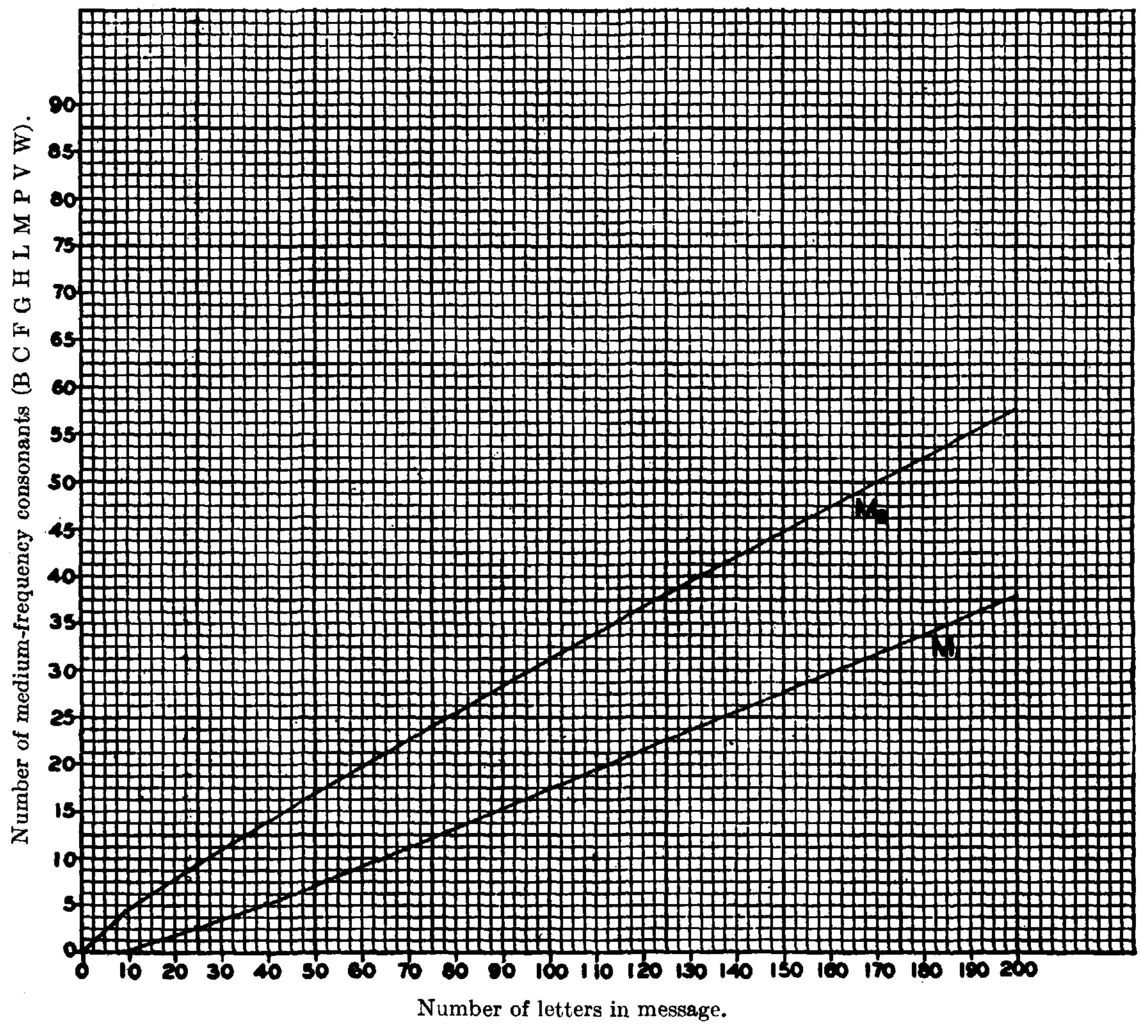


CHART 4. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of medium-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25*d*.)

thus yielding a cryptogram giving external indications of being a transposition cipher but which is really a substitution cipher. If the cryptogram is not too short, a close study will usually disclose what has been done, as well as the futility of so simple a subterfuge.

f. In the majority of cases, in practical work, the determination of the class to which a cipher of average length belongs can be made from a mere inspection of the message, after the cryptanalyst has acquired a familiarity with the normal appearance of transposition and of substitution ciphers. In the former case, his eyes very speedily note many high-frequency letters, such as E, T, N, R, O, A, and I, with the absence of low-frequency letters, such as J, K, Q, X, and Z; in the latter case, his eyes just as quickly note the presence of many low-frequency letters, and a corresponding absence of some of the high-frequency letters.

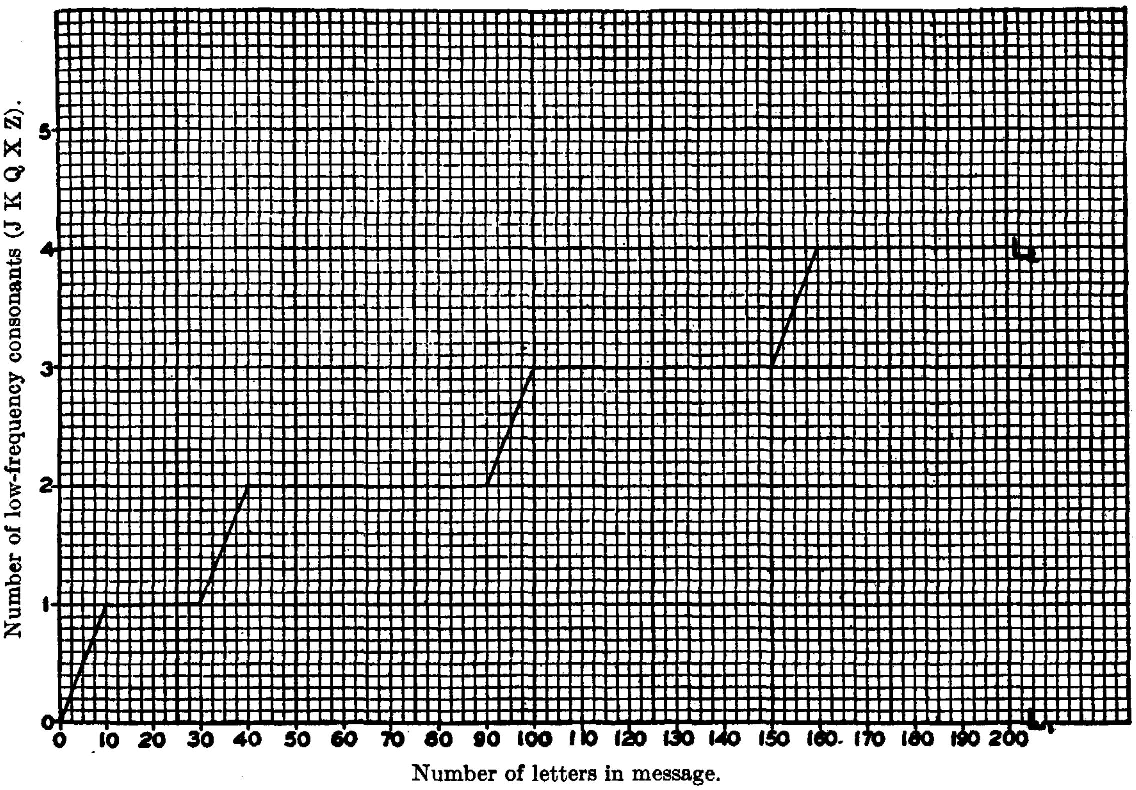


CHART 5. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of low-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25*d*.)

g. Another rather quickly completed test, in the case of the simpler varieties of ciphers, is to look for *repetitions of groups of letters*. As will become apparent very soon, recurrences of syllables, entire words and short phrases constitute a characteristic of all normal plain text. Since a transposition cipher involves a change in the *sequence* of the letters composing a plaintext message, such recurrences are broken up so that the cipher text no longer will show repetitions of more or less lengthy sequences of letters. But if a cipher message does show many repetitions and these are of several letters in length, say over four or five, the conclusion is at once warranted that the cryptogram is most probably a substitution and not a transposition cipher. However, for the beginner in cryptanalysis, it will be advisable to make the uniliteral frequency distribution, and note the frequencies of the vowels and of the high-, medium-, and low-frequency consonants. Then, referring to Charts 2 to 5, he should carefully note whether or not the observed frequencies for these categories of letters fall within the limits of the theoretical frequencies for a normal plaintext message of the same length, and be guided accordingly.

h. It is obvious that the foregoing rule applies only to ciphers composed wholly of letters. If a message is composed entirely of figures, or of arbitrary signs and symbols, or of intermixtures of letters, figures and other symbols, it is immediately apparent that the cryptogram is a substitution cipher.

i. Finally, it should be mentioned that there are certain kinds of cryptograms whose class cannot be determined by the method set forth in subpar. *d* above. These exceptions will be discussed in a subsequent chapter of this text.[13]

26. Determining whether a substitution cipher is monoalphabetic or nonmonoalphabetic.—*a*. It will be remembered that a monoalphabetic substitution cipher is one in which a single cipher alphabet is employed throughout the whole message; that is, a given ciphertext unit invariably represents one and only one particular plaintext unit, this relationship holding throughout the message. On the other hand, a polyalphabetic substitution cipher is one in which two or more cipher alphabets are employed within the same message; that is, a given ciphertext unit may represent two or more different elements in the plain text, according to some rule governing the selection of the equivalent to be used in each case.

b. It is easy to see why and how the appearance of the uniliteral frequency distribution for a substitution cipher may be used to determine whether the cryptogram is monoalphabetic or nonmonoalphabetic in character. The normal distribution presents marked crests and troughs by virtue of two circumstances. First, the elementary sounds which the symbols represent are used with greatly varying frequencies, it being one of the striking characteristics of every alphabetic language that its elementary sounds are used with greatly varying frequencies.[14] In the second place, except for orthographic aberrations peculiar to certain languages (conspicuously, English and French), each such sound is represented by the same symbol. It follows, therefore, that since in a monoalphabetic substitution cipher each different cipher letter (=elementary symbol) represents one and only one plaintext letter (=elementary sound), the uniliteral frequency distribution for such a cipher message must also exhibit the irregular crest-and-trough appearance of the normal distribution, but with this important modification—*the absolute positions of the crests and troughs will not be the same as in the normal.* That is, the letters accompanying the crests and the troughs in the distribution for the cryptogram will be different from those accompanying the crests and the troughs in the normal distribution. But the marked irregularity or "roughness" of the distribution, that is, the presence of accentuated crests and troughs, is in itself an indication that each symbol or cipher letter always represents the same plaintext letter in the cryptogram. Hence the general rule: *A marked crest-and-trough appearance in the uniliteral frequency distribution for a given cryptogram indicates that a single cipher alphabet is involved and constitutes one of the tests for a monoalphabetic substitution cipher.*

c. On the other hand, suppose that in a cryptogram each cipher letter represents several different plaintext letters. Some of them are of high frequency, others of low frequency. The net result of such a situation, so far as the uniliteral frequency distribution for the cryptogram is concerned, is to prevent the appearance of any marked crests and troughs and to tend to reduce the elements of the distribution to a more or less common level. This imparts a "flattened out" appearance to the distribution. For example, in a certain cryptogram of polyalphabetic construction, $K_c=E_p$, G_p and J_p; $R_c=A_p$, D_p, and B_p; $X_c=O_p$, L_p, and F_p. The frequencies of K_c, R_c, and X_c will be approximately equal because the summations of the frequencies of the several plaintext letters which each of these cipher letters represents at different times will be about equal. If this same phenomenon were true of all the letters of the cryptogram, it is clear that the frequencies of the 26 letters, when shown by means of the ordinary uniliteral frequency distribution, would show

[13] Chapter XI.

[14] The student who is interested in this phase of the subject may find the following reference of value: *Zipf, G. K., Selected Studies of the Principle of Relative Frequency in Language,* Cambridge, Mass., 1932.

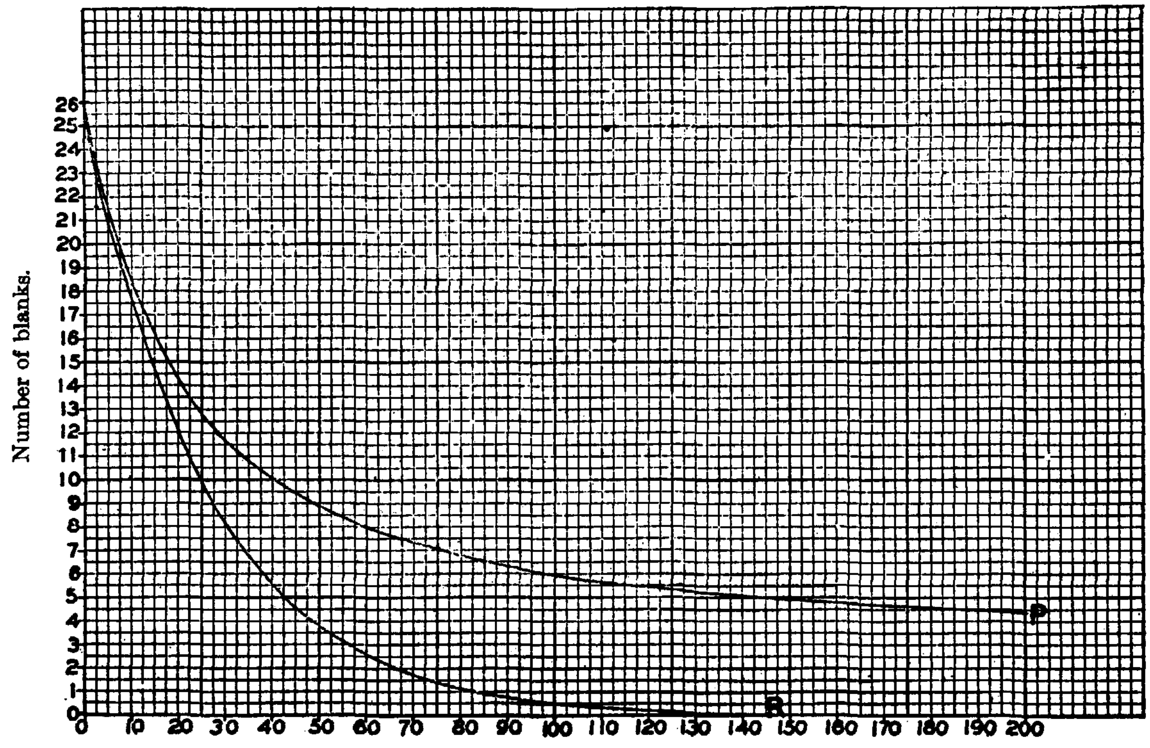


CHART 6. Curves showing the average number of blanks theoretically expected in distributions for plain text (P) and for random text (R) for messages of various lengths. (See subpar. 26*f*.)

no striking differences and the distribution would have the flat appearance of a typical polyalphabetic substitution cipher. Hence, the general rule: *The absence of marked crests and troughs in the uniliteral frequency distribution indicates that a complex form of substitution is involved. The flattened-out appearance of the distribution, then, is one of the criteria for the rejection of a hypothesis of monoalphabetic* [15] *substitution.*

d. The foregoing test based upon the appearance of the frequency distribution is only one of several means of determining whether a substitution cipher is monoalphabetic or nonmonoalphabetic in composition. It can be employed in cases yielding frequency distributions from which definite conclusions can be drawn with more or less certainty by mere ocular examination. In those cases in which the frequency distributions contain insufficient data to permit drawing definite conclusions by such examination, certain statistical tests can be applied. One of these tests, called the ϕ (phi) test, warrants detailed treatment and is discussed in par. 27, below.

e. At this point, however, one additional test will be given because of its simplicity of application. This test, the Λ (lambda) or *blank-expectation test,* may be employed in testing messages up to 200 letters in length, it being assumed that in messages of greater length ocular examination of the frequency distribution offers little or no difficulty. This test concerns the

[15] Cf. footnote 9 on p. 32.

number of blanks in the frequency distribution, that is, the number of letters of the alphabet which are entirely absent from the message. It has been found from statistical studies that rather definite "laws" govern the theoretically expected number of blanks in normal plaintext messages and in frequency distributions for cryptograms of different natures and of various sizes. The results of certain of these studies have been embodied in Chart 6.

f. This chart contains two curves. The one labeled P applies to the average number of blanks theoretically expected in frequency distributions based upon normal plaintext messages of the indicated lengths. The other curve, labeled R, applies to the average number of blanks theoretically expected in frequency distributions based upon *random* assortments of letters; that is, assortments such as would be found by random selection of letters out of a hat containing thousands of letters, all of the 26 letters of the alphabet being present in equal proportions, each letter being replaced after a record of its selection has been made. Such random assortments correspond to polyalphabetic cipher messages in which the number of cipher alphabets is so large that if uniliteral frequency distributions are made of the letters, the distributions are practically identical with those which are obtained by random selections of letters out of a hat.

g. In using this chart, one finds the point of intersection of the vertical line corresponding to the length of the message, with the horizontal line corresponding to the observed number of blanks in the distribution for the message. If this point of intersection falls closer to curve P than it does to curve R, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a plaintext message (or a simple substitution thereof) than it does to a sample of equal length of a more or less "random" assortment of letters (for example, the cipher text of a complex polyalphabetic cipher); therefore, this is evidence that the cryptogram is monoalphabetic. Conversely, if the point of intersection falls closer to curve R than to curve P, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a random text than it does to a plaintext message of the same length; therefore, this is evidence that the cryptogram is nonmonoalphabetic.

27. The ϕ (phi) test for determining monoalphabeticity.—*a.* The student has seen in the preceding paragraph how it is possible to determine by ocular examination whether or not a substitution cipher is monoalphabetic. This tentative determination is based on the presence of a marked crest-and-trough appearance in the uniliteral frequency distribution, and also on the number of blanks in the distribution. However, when the distribution contains a small number of elements, ocular examination and evaluation becomes increasingly difficult and uncertain. In such cases, recourse may be had to a mathematical test, known as the ϕ test, to determine the relative monoalphabeticity or nonmonoalphabeticity of a distribution.

b. Without going into the theory of probability at this time, or into the derivation of the formulas involved, let it suffice for the present to state that with this test the "observed value of ϕ" (symbolized by ϕ_o) for the distribution being tested is compared with the "expected value of ϕ random" (ϕ_r) and the "expected value of ϕ plain" (ϕ_p). The formulas are $\phi_r = .0385N(N-1)$ and, for English military text, $\phi_p = .0667N(N-1)$, where N is the total number of elements in the distribution.[16] The use of these formulas is best illustrated by an example.

[16] The constant .0385 is the decimal equivalent of 1/26, i. e., the reciprocal of the number of elements in the alphabet. The constant .0667 is the sum of the squares of the probabilities of occurrence of the individual letters in English plain text. These constants are treated in detail in *Military Cryptanalytics, Part III.*

c. The following short cryptogram with its accompanying uniliteral frequency distribution is at hand:

O W Q W Z A E D T D Q H H O B A W F T Z W O D E Q

T U W R Q B D Q R O X H Q D A G T B D H P Z R D K

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

ϕ_o for the distribution is calculated by applying the formula $f(f-1)$ to the frequency (f) of each letter and totaling the result; or, expressed in mathematical notation,[17] $\phi_o=\Sigma f(f-1)$. Thus,

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
f	3	3		7	2	1	1	4			1				4	1	6	3		4	1		5	1		3
f(f−1)	6	6		42	2	0	0	12			0				12	0	30	6		12	0		20	0		6

$N=\Sigma f=50$

$\phi_o=\Sigma f(f-1)=154$

For this distribution, $\phi_r=.0385N(N-1)=.0385\times50\times49=94$, and
$\phi_r=.0667N(N-1)=.0667\times50\times49=163$.

Now since ϕ_o, 154, more closely approximates ϕ_p than it does ϕ_r, we have a mathematical corroboration of the hypothesis that the cryptogram is a monoalphabetic substitution cipher. If ϕ_o were nearer to ϕ_r than to ϕ_p, then the assumption would be that the cryptogram is not a monoalphabetic cipher. If ϕ_o were just half way between ϕ_r and ϕ_p, then decision would have to be suspended, since no further statistical proof in the matter is possible with this particular test.[18]

d. Two further examples may be illustrated:

(1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
f(f−1)	0			0	2	6	12	2			0				12	2		0						0		6

$N=25$

$\Sigma f(f-1)=42$

---

[17] The more usual mathematical notation for expressing ϕ_o would be $\sum_{i=A}^{Z} f_i(f_i-1)$, which is read as "the sum of all the terms for all integral values of f from A to Z inclusive." In turn, $\sum_{i=A}^{Z} f_i(f_i-1)$ would be expanded as $f_A(f_A-1)+f_B(f_B-1)+f_C(f_C-1)+\ldots\ldots+f_Z(f_Z-1)$. However, in the interest of simplicity the notation $\Sigma f(f-1)$ is employed; likewise, the notations ϕ_r and ϕ_p are employed in lieu of the more usual $E(\phi_r)$ and $E(\phi_p)$.

[18] Another method of expressing the relative monoalphabeticity of a cryptogram is based upon comparing the *index of coincidence* (abbr. *I. C.*) of the cryptogram under examination with the theoretical I. C. of plain text. The I. C. of a message is defined as the ratio of ϕ_o to ϕ_r; thus, in the example above, the I. C. is $\frac{154}{94}$, which equals 1.64. The theoretical I. C. of English plain text is 1.73, which is the decimal equivalent of $\frac{.0667}{.0385}$, the ratio of the "*plain* constant" to the "*random* constant". The I. C. of random text is 1.00, i. e., $\frac{.0385}{.0385}$.

(2)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
		0				0	0	2	0	0	0	6	0	0		0	2		0	0		0	0	2	6

N=25
$\Sigma f(f-1)=18$

Since both distributions have 25 elements, then for both

$$\phi_r = .0385 \times 25 \times 24 = 23, \text{ and}$$
$$\phi_p = .0667 \times 25 \times 24 = 40.$$

Hence distribution (1) is monoalphabetic, while (2) is not.

e. The student must not assume that statistical tests in cryptanalysis are infallible or absolute in themselves;[19] statistical approaches serve only as a means to the end, in guiding the analyst to the most probably fruitful sources of attack. Since no one test in cryptanalysis gives definite proof of a hypothesis (in fact, not even a battery of tests gives *absolute* proof), all applicable statistical means at the disposal of the cryptanalyst should be used; thus, in examination for monoalphabeticity, the ϕ test, Λ test, and even other tests[20] could profitably be employed. To illustrate this point, if the ϕ test is taken on the distribution of the *plaintext* letters of the phrase

A QUICK BROWN FOX JUMPS OVER THE LAZY DOG

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2				2										12			2			2					

N=33
$\Sigma f(f-1)=20$

$$\phi_r=41;\ \phi_p=70$$

it will be noticed that ϕ_o is less than half of ϕ_r, thus conclusively "proving" that the letters of this phrase could not possibly constitute plain text nor a monoalphabetic encipherment of plain text in *any* language! The student should be able to understand the cause of this cryptologic curiosity.

28. Determining whether a cipher alphabet is standard or mixed.—*a.* Assuming that the uniliteral frequency distribution for a given cryptogram has been made, and that it shows clearly that the cryptogram is a substitution cipher and is monoalphabetic in character, a consideration of the nature of standard cipher alphabets[21] almost makes it obvious how an inspection of the distribution will disclose whether the cipher alphabet involved is a standard cipher alphabet or a mixed cipher alphabet. If the crests and troughs of the distribution occupy positions which correspond to the *relative* positions they occupy in the normal frequency distribution, then the cipher alphabet is a standard cipher alphabet. If this is not the case, then it is highly probable that the cryptogram has been prepared by the use of a mixed cipher alphabet.

b. The difference between the distribution of a direct standard alphabet cipher and one of a reversed standard alphabet cipher is merely a matter of the *direction* in which the sequence of

[19] The following quotation from the Indian mathematician P. C. Mahalanobis, concerning the fallibility of statistics, is particularly appropriate in this connection: "If statistical theory is right, predictions must sometimes come out wrong; on the other hand, if predictions are always right, then the statistical theory must be wrong."—*Sankhyā*, Vol. 10, Part 3, p. 203. Calcutta, 1950.

[20] One of these, the *chi-square test*, will be treated in *Military Cryptanalytics, Part III.*

[21] See par. 12.

crests and troughs progresses—to the right, as is done in normally reading or writing the alphabet (A B C . . . Z), or to the left, that is, in the reversed direction (Z . . . C B A). With a direct standard cipher alphabet the direction in which the crests and troughs of the distribution progress is the normal direction, from left to right; with a reversed standard cipher alphabet this direction is reversed, from right to left.

c. In testing to determine whether a distribution involves encipherment by means of a standard or a mixed alphabet, an attempt is made to locate the more readily-discernible clusters of crests which usually appear in a distribution, such as the distinctive crest-patterns representing the plaintext letters "A . . . E . . . I" and "$\overline{\text{RST}}$." These crest-patterns are searched for, with a quick scanning of the distribution, and then the relative placement with respect to each other is tested to see if it conforms to the expectation for a *direct* standard cipher alphabet, and, if not, then for a *reversed* standard cipher alphabet. During this latter step, which consists of little more than counting in one direction and then (when necessary) in the other, the blank (or nearly-blank) expectation of "$\overline{\text{JK}}_p$" followed by the characteristic curve for "$\overline{\text{LMNOP}}_p$" and the blank "$Q_p$" are also considered.

d. A mechanical test may be applied in doubtful cases arising from lack of material available for study; just what this test involves, and an illustration of its application will be given in the next chapter, using specific examples.

Chapter V

UNILITERAL SUBSTITUTION WITH STANDARD CIPHER ALPHABETS

29. Types of standard cipher alphabets.—*a.* Standard cipher alphabets are of two types:

(1) *Direct standard,* in which the cipher component is the normal sequence but shifted to the right or left of its point of coincidence in the normal alphabet. Example:

```
          ———→
Plain:  ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: QRSTUVWXYZABCDEFGHIJKLMNOP
                    ———→
```

It is obvious that the cipher component can be applied to the plain component at any one of 26 points of coincidence, but since the alphabet that results from one of these applications coincides exactly with the normal alphabet, a series of only 25 (direct standard) cipher alphabets results from the shifting of the cipher component.

(2) *Reversed standard,* in which the cipher component is also the normal sequence but runs in the opposite direction from the normal. Example:

```
          ———→
Plain:  ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: QPONMLKJIHGFEDCBAZYXWVUTSR
                    ←———
```

Here the cipher component can be applied to the plain component at any of 26 points of coincidence, each yielding a different cipher alphabet. There is in this case, therefore, a series of 26 (reversed standard) cipher alphabets.

b. It is often convenient to refer to or designate one of a series of cipher alphabets without ambiguity or circumlocution. The usual method is to indicate a particular alphabet to which reference is made by citing a pair of equivalents in that alphabet, such as, in the example above, $A_p=Q_c$. The *key* for the cipher alphabet just referred to, as well as that preceding it, is $A_p=Q_c$, and it is said that the *key letter* for the cipher alphabet is Q_c.

c. The cipher alphabet in subpar. *a* (2), above, is also a *reciprocal alphabet;* that is, the cipher alphabet contains 13 distinct pairs of equivalents which are reversible. For example, in the alphabet referred to, $A_p=Q_c$ *and* $Q_p=A_c$; $B_p=P_c$ *and* $P_p=B_c$, etc. The reciprocity exists throughout the alphabet and is a result of the method by which it was formed. (Reciprocal alphabets may be produced by juxtaposing *any two components which are identical* but progress in opposite directions.)

30. Procedure in encipherment and decipherment by means of uniliteral substitution.—*a.* When a message is enciphered by means of *monoalphabetic uniliteral substitution*, or *simple substitution* (as it is often called), the individual letters of the message text are replaced by the single-letter equivalents taken from the cipher alphabet selected by prearrangement. Example:

Message: EIGHTEEN PRISONERS CAPTURED
Enciphering alphabet: Direct standard, $A_p = T_c$

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: TUVWXYZABCDEFGHIJKLMNOPQRS

Letter-for-letter encipherment:

EIGHTEEN PRISONERS CAPTURED
XBZAMXXG IKBLHGXKL VTIMNKXW

The cipher text is then regrouped, for transmission, into groups of five.

Cryptogram:

XBZAM XXGIK BLHGX KLVTI MNKXW

b. The procedure in decipherment is merely the reverse of that in encipherment. The cipher alphabet selected by prearrangement is set up with the cipher component arranged in the normal sequence and placed above the plain component for ease in deciphering. The letters of the cryptogram are then replaced by their plaintext equivalents, as shown below.

Cipher: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Plain: HIJKLMNOPQRSTUVWXYZABCDEFG

The message deciphers thus:

Cipher: XBZAM XXGIK BLHGX KLVTI MNKXW
Plain: EIGHT EENPR ISONE RSCAP TURED

The deciphering clerk rewrites the text in word lengths:

EIGHTEEN PRISONERS CAPTURED

c. In subpar. *a*, above, the cryptogram was prepared in final form for transmission by dividing the cryptographic text into groups of five. This is generally the case in military communications involving cipher systems. It promotes accuracy in telegraphic communication since an operator knows he must receive a definite number of characters in each group, no more and no less. Also, it usually makes solution of the messages by unauthorized persons more difficult because the length of the words, phrases, and sentences of the plain text is hidden. If the last group of the cipher text in subpar. 30*a* had not been a complete group of five letters, it might have been completed by adding a sufficient number of meaningless letters (called *nulls*).

31. Principles of solution by construction and analysis of the uniliteral frequency distribution.—*a.* The analysis of monoalphabetic cryptograms prepared by the use of standard cipher alphabets follows almost directly from a consideration of the nature of such alphabets. Since the cipher component of a standard cipher alphabet consists either of the normal sequence merely displaced 1, 2, 3, . . . intervals from the normal point of coincidence, or of the normal sequence proceeding in a reversed-normal direction, it is obvious that the uniliteral frequency distribution for a cryptogram prepared by means of such a cipher alphabet employed monoalphabetically will show crests and troughs whose *relative* positions and frequencies will be exactly the same as in the uniliteral frequency distribution for the plain text of that cryptogram.

The only thing that has happened is that the whole set of crests and troughs of the distribution has been displaced to the right or left of the position it occupies in the distribution for the plain text; or else the successive elements of the whole set progress in the opposite direction. Hence, it follows that the correct determination of the plaintext value of the cipher letter marking *any* crest or trough of the uniliteral frequency distribution, coupled with the correct determination of the relative direction in which the plain component sequence progresses, will result at one stroke in the correct determination of the plaintext values of *all* the remaining 25 letters respectively marking the other crests and troughs in that distribution. The problem thus resolves itself into a matter of selecting that point of attack which will most quickly or most easily lead to the determination of the value of *one* cipher letter. The single word *identification* will hereafter be used for the phrase "determination of the value of a cipher letter"; to *identify* a cipher letter is to find its plaintext value.

b. It is obvious that the easiest point of attack is to assume that the letter marking the crest of greatest frequency in the frequency distribution for the cryptogram represents E_p. Proceeding from this initial point, the identifications of the remaining cipher letters marking the other crests and troughs are tentatively made on the basis that the letters of the cipher component proceed in accordance with the normal alphabetic sequence, either direct or reversed. If the actual frequency of each letter marking a crest or a trough approximates to a fairly close degree the normal or theoretical frequency of the assumed plaintext equivalent, then the initial identification $\theta_c = E_p$ may be *assumed to be correct* and therefore the derived identifications of the other cipher letters also may be assumed to be correct.[1] If the original starting point for assignment of plaintext values is not correct, or if the direction of "reading" the successive crests and troughs of the distribution is not correct, then the frequencies of the other 25 cipher letters will not correspond to or even approximate the normal or theoretical frequencies of their hypothetical plaintext equivalents on the basis of the initial identification. A new initial point, that is, a different cipher equivalent, must then be selected to represent E_p; or else the direction of "reading" the crests and troughs must be reversed. This procedure, that is, the attempt to make the actual frequency relations exhibited by the uniliteral frequency distribution for a given cryptogram conform to the theoretical frequency relations of the normal frequency distribution in an effort to solve the cryptogram, is referred to technically as "fitting the actual uniliteral frequency distribution for a cryptogram to the theoretical uniliteral frequency distribution for normal plain text", or, more briefly, as *"fitting the frequency distribution for the cryptogram to the normal frequency distribution"*, or, still more briefly, *"fitting the distribution to the normal."* In statistical work the expression commonly employed in connection with this process of fitting an actual distribution to a theoretical one is "testing the goodness of fit." The goodness of fit may be stated in various ways, mathematical in character.[2]

c. In fitting the actual distribution to the normal, it is necessary to regard the cipher component (that is, the letters A . . . Z marking the successive crests and troughs of the distribution) as partaking of the nature of a circle, that is, a sequence closing in upon itself, so that no matter with what crest or trough one starts, the spatial and frequency relations of the crests and troughs are constant. This manner of regarding the cipher component as being cyclic in nature is valid *because it is obvious that the relative positions and frequencies of the crests and troughs of any uniliteral*

[1] The Greek letter θ (theta) is used to represent a character or letter without indicating its identity. Thus, instead of the circumlocution "any letter of the plain text" the symbol θ_p is used; and for the expression "any letter of the cipher text", the symbol θ_c is used.

[2] One of these tests for expressing the goodness of fit, the χ (chi) test, will be treated in *Military Cryptanalytics, Part II.*

frequency distribution must remain the same regardless of what letter is employed as the initial point of the distribution. Fig. 5 gives a clear picture of what is meant in this connection, as applied to the normal frequency distribution.

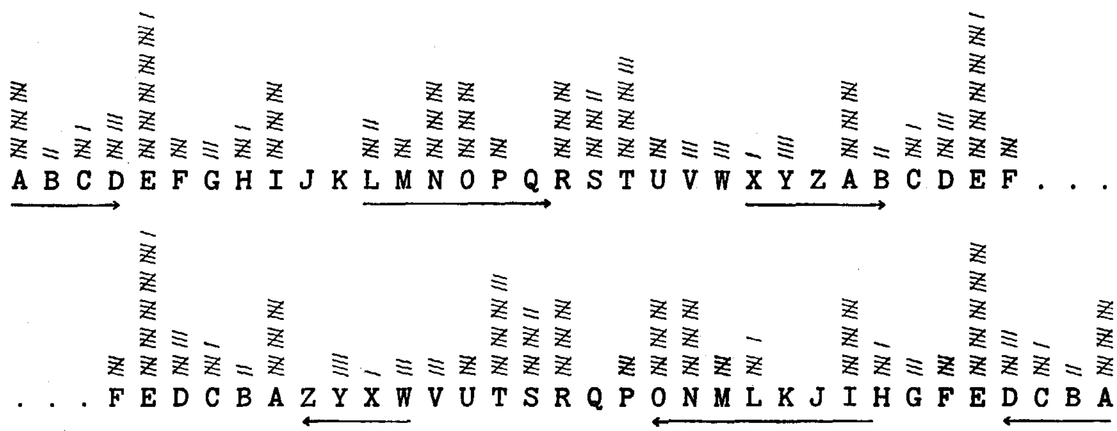


FIGURE 5.

d. In the third sentence of subpar. *b,* the phrase "assumed to be correct" was advisedly employed in describing the results of the attempt to fit the distribution to the normal, because the final test of the goodness of fit in this connection (that is, of the correctness of the assignment of values to the crests and troughs of the distribution) is whether the *consistent* substitution of the plaintext values of the cipher characters in the cryptogram will yield intelligible plain text. If this is not the case, then no matter how close the approximation between actual and theoretical frequencies is, no matter how well the actual frequency distribution fits the normal, the only possible inferences are that (1) either the closeness of the fit is a pure coincidence in this case and that another equally good fit may be obtained from the same data, or else (2) the cryptogram involves something more than simple monoalphabetic substitution by means of a single standard cipher alphabet. For example, suppose a transposition has been applied in addition to the substitution. Then, although an excellent correspondence between the uniliteral frequency distribution and the normal frequency distribution has been obtained, the substitution of the cipher letters by their assumed equivalents will still not yield plain text. However, aside from such cases of double encipherment, instances in which the uniliteral frequency distribution may be easily fitted to the normal frequency distribution and in which at the same time an attempted simple substitution fails to yield intelligible text are rare. It may be said that, in practical operations whenever the uniliteral frequency distribution can be made to fit the normal frequency distribution, substitution of values will result in solution; and, as a corollary, whenever the uniliteral frequency distribution cannot be made to fit the normal frequency distribution, the cryptogram does not represent a case of simple, monoalphabetic substitution by means of a standard alphabet.

32. Theoretical example of solution.—*a.* The foregoing principles will become clearer by noting the encryption and solution of a theoretical example. The following message is to be encrypted.

HOSTILE FORCE ESTIMATED AT ONE REGIMENT INFANTRY AND TWO PLATOONS CAVALRY MOVING SOUTH ON QUINNIMONT PIKE STOP HEAD OF COLUMN NEARING ROAD JUNCTION SEVEN THREE SEVEN COMMA EAST OF GREENACRE SCHOOL FIRED UPON BY OUR PATROLS STOP HAVE DESTROYED BRIDGE OVER INDIAN CREEK.

b. First, solely for purposes of demonstrating certain principles, the uniliteral frequency distribution for this plaintext message is presented in Fig. 6.

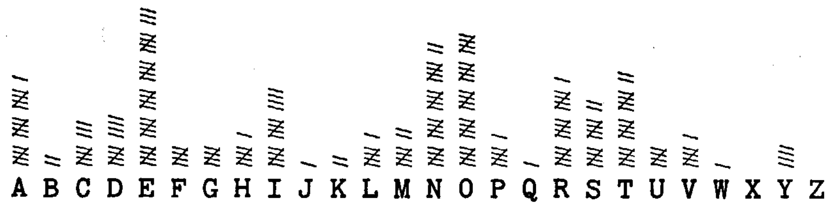


FIGURE 6.

c. Now let the foregoing message be encrypted monoalphabetically by the following standard cipher alphabet, yielding the cryptogram shown below and the frequency distribution shown in Fig. 7.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

Plain	HOSTI	LEFOR	CEEST	IMATE	DATON	EREGI	MENTI	NFANT	RYAND
Cipher	NUYZO	RKLUX	IKKYZ	OSGZK	JGZUT	KXKMO	SKTZO	TLGTZ	XEGTJ
Plain	TWOPL	ATOON	SCAVA	LRYMO	VINGS	OUTHO	NQUIN	NIMON	TPIKE
Cipher	ZCUVR	GZUUT	YIGBG	RXESU	BOTMY	UAZNU	TWAOT	TOSUT	ZVOQK
Plain	STOPH	EADOF	COLUM	NNEAR	INGRO	ADJUN	CTION	SEVEN	THREE
Cipher	YZUVN	KGJUL	IURAS	TTKGX	OTMXU	GJPAT	IZOUT	YKBKT	ZNXKK
Plain	SEVEN	COMMA	EASTO	FGREE	NACRE	SCHOO	LFIRE	DUPON	BYOUR
Cipher	YKBKT	IUSSG	KGYZU	LMXKK	TGIXK	YINUU	RLOXK	JAVUT	HEUAX
Plain	PATRO	LSSTO	PHAVE	DESTR	OYEDB	RIDGE	OVERI	NDIAN	CREEK
Cipher	VGZXU	RYYZU	VNGBK	JKYZX	UEKJH	XOJMK	UBKXO	TJOGT	IXKKQ

Cryptogram

N U Y Z O	R K L U X	I K K Y Z	O S G Z K	J G Z U T	K X K M O
S K T Z O	T L G T Z	X E G T J	Z C U V R	G Z U U T	Y I G B G
R X E S U	B O T M Y	U A Z N U	T W A O T	T O S U T	Z V O Q K
Y Z U V N	K G J U L	I U R A S	T T K G X	O T M X U	G J P A T
I Z O U T	Y K B K T	Z N X K K	Y K B K T	I U S S G	K G Y Z U
L M X K K	T G I X K	Y I N U U	R L O X K	J A V U T	H E U A X
V G Z X U	R Y Y Z U	V N G B K	J K Y Z X	U E K J H	X O J M K
U B K X O	T J O G T	I X K K Q			

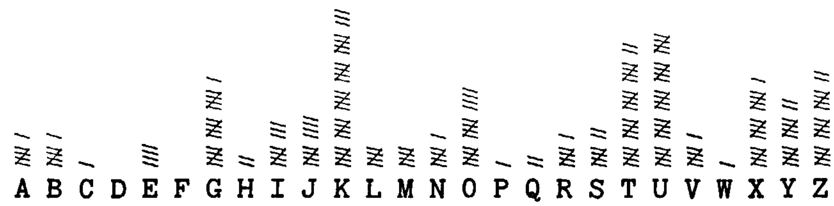


FIGURE 7.

d. Let the student now compare Figs. 6 and 7, which have been superimposed in Fig. 8 for convenience in examination. Crests and troughs are present in both distributions; moreover their relative positions and frequencies have not been changed in the slightest particular.[redacted] Only the absolute position of the sequence as a whole has been displaced six places to the right in Fig. 7, as compared with the absolute position of the sequence in Fig. 6.

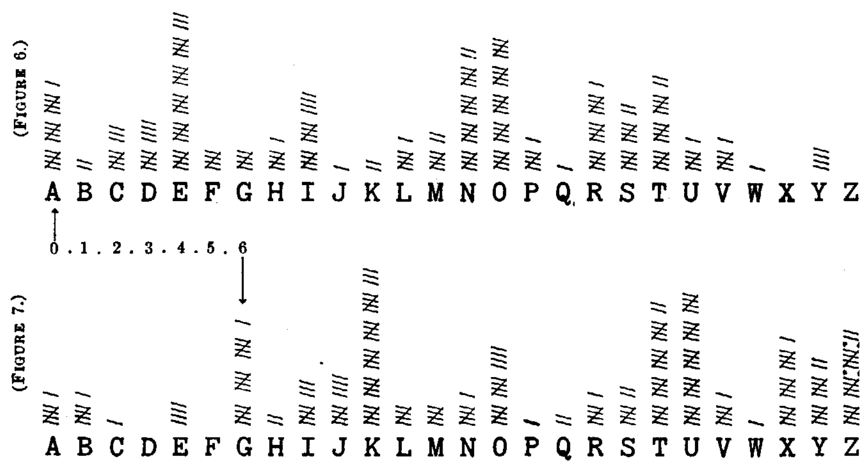


FIGURE 8.

e. If the two distributions are compared in detail the student will clearly understand how easy the solution of the cryptogram would be to one who knew nothing about how it was prepared. For example, the frequency of the highest crest, representing E_p in Fig. 6 is 28; at an interval of four letters before E_p there is another crest representing A_p with frequency 16. Between A and E there is a trough, representing the medium-frequency letters B, C, D. On the other side of E, at an interval of four letters, comes another crest, representing I with frequency 14. Between E and I there is another trough, representing the medium-frequency letters F, G, H. Compare these crests and troughs with their homologous crests and troughs in Fig. 7. In the latter, the letter K marks the highest crest in the distribution with a frequency of 28; four letters before K there is another crest, frequency 16, and four letters on the other side of K there is another crest, frequency 14. Troughs corresponding to B, C, D and F, G, H are seen at H, I, J and L, M, N in Fig. 7. In fact, the two distributions may be made to coincide exactly, by shifting the frequency distribution for the cryptogram six places to the left with respect to the distribution for the equivalent plaintext message, as shown herewith.

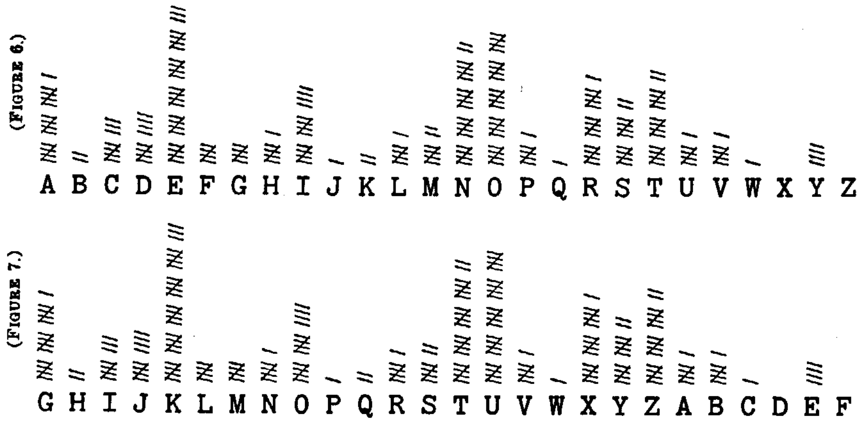


FIGURE 9.

f. Let us suppose now that nothing is known about the process of encryption, and that only the cryptogram and its uniliteral frequency distribution is at hand. It is clear that simply bearing in mind the spatial relations of the crests and troughs in a normal frequency distribution would enable the cryptanalyst to fit the distribution to the normal in this case. He would naturally first assume that $K_c=E_p$, from which it would follow that if a direct standard alphabet is involved, $L_c=F_p$, $M_c=G_p$, and so on, yielding the following (tentative) deciphering alphabet:

Cipher	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain	U V W X Y Z A B C D E F G H I J K L M N O P Q R S T

g. Now comes the final test: If these assumed values are substituted in the cipher text, the plain text immediately appears. Thus:

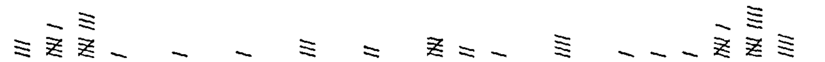
```
N U Y Z O   R K L U X   I K K Y Z   O S G Z K   J G Z U T   etc.
H O S T I   L E F O R   C E E S T   I M A T E   D A T O N   etc.
```

h. It should be clear, therefore, that the initial selection of G_c as the specific key (that is, to represent A_p) in the process of encryption has absolutely no effect upon the relative spatial and frequency relations of the crests and troughs of the frequency distribution for the cryptogram. If Q_c had been selected to represent A_p, these relations would still remain the same, the whole series of crests and troughs being merely displaced further to the right of the positions they occupy when $G_c=A_p$.

33. Practical example of solution by the frequency method.—*a. The case of direct standard alphabet ciphers.* (1) The following cryptogram is to be solved by applying the foregoing principles:

```
N W N V H   C A X X Y   B J C C J   L T R W P   X D A Y X   B R C R X
W B N J B   C X O W N   F C X W B   C X Y Y N   C N A B L   X U R W O
```

(2) From the presence of so many low-frequency letters such as B, W, and X it is at once suspected that this is a substitution cipher. But to illustrate the steps, that must be taken in difficult cases in order to be certain in this respect, a uniliteral frequency distribution is constructed, and then reference is made to Charts 2 to 5 to note whether the actual numbers of vowels, high-, medium-, and low-frequency consonants fall inside or outside the areas delimited by the respective curves.



A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

FIGURE 10*a*.

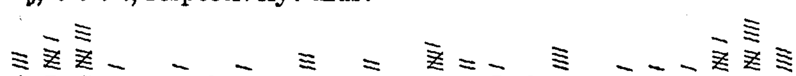
Letters	Frequency	Position with respect to areas delimited by curves
Vowels (AEIOUY)	10	Outside, Chart 1.
High-frequency Consonants (DNRST)	12	Outside, Chart 2.
Medium-frequency Consonants (BCFGHLMPVW)	26	Outside, Chart 3.
Low-frequency Consonants (JKQXZ)	12	Outside, Chart 4.
Total	60	

(3) All four points falling completely outside the areas delimited by the curves applicable to these four classes of letters, the crypogram is clearly a substitution cipher.

(4) The appearance of the frequency distribution, with marked crests and troughs, indicates that the cryptogram is probably monoalphabetic. At this point the ϕ test is applied to the distribution. The observed value of ϕ is found to be 258, while the expected value of ϕ plain and ϕ random are calculated to be 236 and 136, respectively. The fact that the observed value more closely approximates ϕ_p than it does ϕ_r is taken as statistical evidence that the cryptogram is monoalphabetic. Furthermore, reference being made to Chart 6, the point of intersection of the message length (60 letters) and the number of blanks (8) falls directly on curve P; this is additional evidence that the message is probably monoalphabetic.

(5) The next step is to determine whether a standard or a mixed cipher alphabet is involved. This is done by studying the positions and the sequence of crests and troughs in the frequency distribution, and trying to fit the distribution to the normal.

(6) The first assumption to be made is that a direct standard cipher alphabet is involved. The highest crest in the distribution occurs over X_c. Let it be assumed that $X_c=E_p$. Then Y_c, Z_c, A_c, ... $=F_p$, G_p, H_p, ..., respectively: thus:



Cipher.......... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain........... H I J K L M N O P Q R S T U V W X Y Z A B C D E F G

FIGURE 10*b*.

It may be seen quickly that the approximation to the expected frequencies is very poor. There are too many occurrences of J_p, Q_p, U_p, and F_p and too few occurrences of N_p, O_p, R_p, S_p, T_p and A_p. Moreover, if a substitution is attempted on this basis, the following is obtained for the first two cipher groups:

Cipher	N W N V H	C A X X Y
"Plain text"	U D U C O	J H E E F

This is certainly not plain text and it seems clear that X_c is not E_p, if the hypothesis of a direct standard alphabet cipher is correct. A different assumption will have to be made.

(7) Suppose $C_c=E_p$. Going through the same steps as before, again no satisfactory results are obtained. Further trials [3] are made along the same lines, until the assumption $N_c=E_p$ is tested:

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q

FIGURE 10c.

(8) The fit in this case is quite good; possibly there are too few occurrences of A_p, D_p, and R_p. But the final test remains: trial of the substitution alphabet on the cryptogram itself. This is done and the results are as follows:

```
C: N W N V H   C A X X Y   B J C C J   L T R W P   X D A Y X   B R C R X
P: E N E M Y   T R O O P   S A T T A   C K I N G   O U R P O   S I T I O

C: W B N J B   C X O W N   F C X W B   C X Y Y N   C N A B L   X U R W O
P: N S E A S   T O F N E   W T O N S   T O P P E   T E R S C   O L I N F
```

ENEMY TROOPS ATTACKING OUR POSITION EAST OF NEWTON. PETERS COL INF.

(9) It is always advisable to note the specific key. In this case the correspondence between any plaintext letter and its cipher equivalent will indicate the key. Although other conventions are possible, and equally valid, it is usual, however, to indicate the key by noting the cipher equivalent of A_p. In this case $A_p=J_c$.

b. The case of reversed standard alphabet ciphers.—(1) Let the following cryptogram and its frequency distribution be studied.

```
F W F X L   Q S V V U   R J Q Q J   H Z B W D   V P S U V   R B Q B V
W R F J R   Q V E W F   N Q V W R   Q V U U F   Q F S R H   V Y B W E
```

(2) The preliminary steps illustrated above, under subpar. *a* (1) to (4) inclusive, in connection with the test for class and monoalphabeticity, will here be omitted, since they are exactly the same in nature. The result is that the cryptogram is obviously a substitution cipher and is monoalphabetic.

(3) Assuming that it is not known whether a direct or a reversed standard alphabet is involved, attempts are at once made to fit the frequency distribution to the normal direct sequence. If the student will try them he will soon find out that these are unsuccessful. All this takes but a few minutes.

[3] It is unnecessary, of course, to write out all the alphabets and pseudo-decipherments, as shown above, when testing assumptions. This is usually done mentally, using the scanning procedure treated in subpar. 28c.

(4) The next logical assumption is now made, *viz.*, that the cipher alphabet is a reversed standard alphabet. When on this basis F_c is assumed to be E_p, the distribution can readily be fitted to the normal, practically every crest and trough in the actual distribution corresponding to a crest or trough in the expected distribution.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	J	I	H	G	F	E	D	C	B	A	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K

FIGURE 10*d*.

(5) When the substitution is made in the cryptogram, the following is obtained.

Cryptogram	F W F X L	Q S V V U	R J Q Q J	etc.
Plain text	E N E M Y	T R O O P	S A T T A	etc.

(6) The plaintext message is identical with that in subpar. *a*. The specific key in this case is also $A_p = J_c$. If the student will compare the frequency distributions in the two cases, he will note that the relative positions and extents of the crests and troughs are identical; they merely progress in opposite directions.

34. Solution by completing the plain-component sequence.—*a. The case of direct standard alphabet ciphers.* (1) The foregoing method of analysis, involving as it does the construction of a uniliteral frequency distribution, was termed a *solution by the frequency method* because it involves the construction of a frequency distribution and its study. There is, however, another method which is much more rapid, almost wholly mechanical, and which, moreover, does not necessitate the construction or study of any frequency distribution whatever. An understanding of the method follows from a consideration of the method of encipherment of a message by the use of a single, direct standard cipher alphabet.

(2) Note the following encipherment:

Message.................. TWO CRUISERS SUNK

Enciphering Alphabet

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

Encipherment

Plain text	T W O	C R U I S E R S	S U N K
Cryptogram	Z C U	I X A O Y K X Y	Y A T Q

Cryptogram

Z C U I X   A O Y K X   Y Y A T Q

(3) The enciphering alphabet shown above represents a case wherein the sequence of letters of both components of the cipher alphabet is the normal sequence, with the sequence forming the cipher components merely shifted six places to the left (or 20 positions to the right) of the position it occupies in the normal alphabet. If, therefore, two strips of paper bearing the letters

of the normal sequence, equally spaced, are regarded as the two components of the cipher alphabet and are juxtaposed at all of the 25 possible points of coincidence, it is obvious that one of these 25 juxtapositions *must* correspond to the actual juxtaposition shown in the enciphering alphabet directly above.[4] It is equally obvious that if a record were kept of the results obtained by applying the values given at each juxtaposition to the letters of the cryptogram, one of these results would yield the plain text of the cryptogram.

(4) Let the work be systematized and the results set down in an orderly manner for examination. It is obviously unnecessary to juxtapose the two components so that $A_c = A_p$, for on the assumption of a direct standard alphabet, juxtaposing two direct normal components at their normal point of coincidence merely yields plain text. The next possible juxtaposition, therefore, is $A_c = B_p$. Let the juxtaposition of the two sliding strips therefore be $A_c = B_p$, as shown here:

```
Plain-------------------                           ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher------------------ ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
```

The values given by this juxtaposition are substituted for the letters of the cryptogram and the following results are obtained.

Cryptogram	Z C U I X	A O Y K X	Y Y A T Q
1st Test—"Plain text"	A D V J Y	B P Z L Y	Z Z B U R

This certainly is not intelligible text; obviously, the two components were not in the position indicated in this first test. The plain component is therefore slid one interval to the left, making $A_c = C_p$, and a second test is made. Thus

```
Plain-------------------                          ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher------------------ ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
```

Cryptogram	Z C U I X	A O Y K X	Y Y A T Q
2d Test—"Plain text"	B E W K Z	C Q A M Z	A A C V S

Neither does the second test result in disclosing any plain text. But, if the results of the two tests are studied, a phenomenon that at first seems quite puzzling comes to light. Thus, suppose the results of the two tests are superimposed in this fashion.

Cryptogram	Z C U I X	A O Y K X	Y Y A T Q
1st Test—"Plain text"	A D V J Y	B P Z L Y	Z Z B U R
2d Test—"Plain text"	B E W K Z	C Q A M Z	A A C V S

(5) Note what has happened. The net result of the two experiments was merely to continue the normal sequence begun by the cipher letters at the heads of the *columns* of letters. It is obvious that if the normal sequence is completed in each column *the results will be exactly the same as though the whole set of 25 possible tests had actually been performed.* Let the columns therefore be completed, as shown in Fig. 11.

[4] One of the strips should bear the sequence repeated. This permits juxtaposing the two sequences at all 26 possible points of coincidence so as to have a complete cipher alphabet showing at all times.

Z	C	U	I	X	A	O	Y	K	X	Y	Y	A	T	Q
A	D	V	J	Y	B	P	Z	L	Y	Z	Z	B	U	R
B	E	W	K	Z	C	Q	A	M	Z	A	A	C	V	S
C	F	X	L	A	D	R	B	N	A	B	B	D	W	T
D	G	Y	M	B	E	S	C	O	B	C	C	E	X	U
E	H	Z	N	C	F	T	D	P	C	D	D	F	Y	V
F	I	A	O	D	G	U	E	Q	D	E	E	G	Z	W
G	J	B	P	E	H	V	F	R	E	F	F	H	A	X
H	K	C	Q	F	I	W	G	S	F	G	G	I	B	Y
I	L	D	R	G	J	X	H	T	G	H	H	J	C	Z
J	M	E	S	H	K	Y	I	U	H	I	I	K	D	A
K	N	F	T	I	L	Z	J	V	I	J	J	L	E	B
L	O	G	U	J	M	A	K	W	J	K	K	M	F	C
M	P	H	V	K	N	B	L	X	K	L	L	N	G	D
N	Q	I	W	L	O	C	M	Y	L	M	M	O	H	E
O	R	J	X	M	P	D	N	Z	M	N	N	P	I	F
P	S	K	Y	N	Q	E	O	A	N	O	O	Q	J	G
Q	T	L	Z	O	R	F	P	B	O	P	P	R	K	H
R	U	M	A	P	S	G	Q	C	P	Q	Q	S	L	I
S	V	N	B	Q	T	H	R	D	Q	R	R	T	M	J
*T	W	O	C	R	U	I	S	E	R	S	S	U	N	K
U	X	P	D	S	V	J	T	F	S	T	T	V	O	L
V	Y	Q	E	T	W	K	U	G	T	U	U	W	P	M
W	Z	R	F	U	X	L	V	H	U	V	V	X	Q	N
X	A	S	G	V	Y	M	W	I	V	W	W	Y	R	O
Y	B	T	H	W	Z	N	X	J	W	X	X	Z	S	P

FIGURE 11.

An examination of the successive horizontal lines of the diagram discloses *one and only one* line of plain text, that marked by the asterisk and reading T W O C R U I S E R S S U N K.

(6) Since each column in Fig. 11 is nothing but a normal sequence, it is obvious that instead of laboriously writing down these columns of letters every time a cryptogram is to be examined, it would be more convenient to prepare a set of strips each bearing the normal sequence doubled (to permit complete coincidence for an entire alphabet at any setting), and have them available for examining any future cryptograms. In using such a set of sliding strips in order to solve a cryptogram prepared by means of a single direct standard cipher alphabet, or to make a test to determine whether a cryptogram has been so prepared, it is only necessary to "set up" the letters of the cryptogram on the strips, that is, align them in a single row across the strips (by sliding the individual strips up or down). The successive horizontal lines, called *generatrices* (singular, *generatrix*),[5] are then examined in a search for intelligible text. If the cryptogram really belongs to this simple type of cipher, one of the generatrices will exhibit intelligible text all the way across; this text will almost invariably be the plain text of the message. This method of analysis may be termed *a solution by completing the plain-component sequence.* Sometimes it is

[5] Pronounced: *jĕn′ĕr-ā-trī′sēz* and *jĕn′ĕr-ā′trĭks,* respectively.

referred to as "running down" the sequence. The principle upon which the method is based constitutes one of the cryptanalyst's most valuable tools.[6]

b. The case of reversed standard alphabets.—(1) The method described under subpar. *a* may also be applied, in slightly modified form, in the case of a cryptogram enciphered by a single reversed standard alphabet. The basic principles are identical in the two cases, as will now be demonstrated.

(2) Let two sliding components be prepared as before, except that in this case one of the components must be a reversed normal sequence, the other, a direct normal sequence.

(3) Let the two components be juxaposed A to A as shown below, and then let the resultant values be substituted for the letters of the cryptogram. Thus:

Cryptogram

N K S E P   M Y O C P   O O M T W

```
Plain------------------                           ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher----------------- ZYXWVUTSRQPONMLKJIHGFEDCBAZYXWVUTSRQPONMLKJIHGFEDCBA
```

Cryptogram	N K S E P	M Y O C P	O O M T W
1st Test—"Plain text"	N Q I W L	O C M Y L	M M O H E

(4) This does not yield intelligible text, and therefore the reversed component is slid one space forward and a second test is made. Thus:

```
Plain------------------                          ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher----------------- ZYXWVUTSRQPONMLKJIHGFEDCBAZYXWVUTSRQPONMLKJIHGFEDCBA
```

Cryptogram	N K S E P	M Y O C P	O O M T W
2d Test—"Plain text"	O R J X M	P D N Z M	N N P I F

(5) Neither does the second test yield intelligible text. But let the results of the two tests be superimposed. Thus:

Cryptogram	N K S E P	M Y O C P	O O M T W
1st Test—"Plain text"	N Q I W L	O C M Y L	M M O H E
2d Test—"Plain text"	O R J X M	P D N Z M	N N P I F

(6) It is seen that the letters of the "plain text" given by the *second* trial are merely the continuants of the normal sequences initiated by the letters of the "plain text" given by the first trial. If these sequences are "run down"—that is, completed within the columns—the results must obviously be the same as though successive tests exactly similar to the first two were applied to the cryptogram, using one reversed normal and one direct normal component. If the cryptogram has really been prepared by means of a single reversed standard alphabet, one of the generatrices of the diagram that results from completing the sequence *must* yield intelligible text.

(7) Let the diagram be made, or better yet, if the student has already at hand the set of sliding strips referred to in footnote 6, below, let him "set up" the letters given by the *first* trial. Fig. 12 shows the diagram and indicates the plaintext generatrix.

[6] A set of heavy paper strips, suitable for use in completing the plain-component sequence, has been prepared for use as a training aid in connection with the courses in Military Cryptanalytics.

N	K	S	E	P	M	Y	O	C	P	O	O	M	T	W
N	Q	I	W	L	O	C	M	Y	L	M	M	O	H	E
O	R	J	X	M	P	D	N	Z	M	N	N	P	I	F
P	S	K	Y	N	Q	E	O	A	N	O	O	Q	J	G
Q	T	L	Z	O	R	F	P	B	O	P	P	R	K	H
R	U	M	A	P	S	G	Q	C	P	Q	Q	S	L	I
S	V	N	B	Q	T	H	R	D	Q	R	R	T	M	J
*T	W	O	C	R	U	I	S	E	R	S	S	U	N	K
U	X	P	D	S	V	J	T	F	S	T	T	V	O	L
V	Y	Q	E	T	W	K	U	G	T	U	U	W	P	M
W	Z	R	F	U	X	L	V	H	U	V	V	X	Q	N
X	A	S	G	V	Y	M	W	I	V	W	W	Y	R	O
Y	B	T	H	W	Z	N	X	J	W	X	X	Z	S	P
Z	C	U	I	X	A	O	Y	K	X	Y	Y	A	T	Q
A	D	V	J	Y	B	P	Z	L	Y	Z	Z	B	U	R
B	E	W	K	Z	C	Q	A	M	Z	A	A	C	V	S
C	F	X	L	A	D	R	B	N	A	B	B	D	W	T
D	G	Y	M	B	E	S	C	O	B	C	C	E	X	U
E	H	Z	N	C	F	T	D	P	C	D	D	F	Y	V
F	I	A	O	D	G	U	E	Q	D	E	E	G	Z	W
G	J	B	P	E	H	V	F	R	E	F	F	H	A	X
H	K	C	Q	F	I	W	G	S	F	G	G	I	B	Y
I	L	D	R	G	J	X	H	T	G	H	H	J	C	Z
J	M	E	S	H	K	Y	I	U	H	I	I	K	D	A
K	N	F	T	I	L	Z	J	V	I	J	J	L	E	B
L	O	G	U	J	M	A	K	W	J	K	K	M	F	C
M	P	H	V	K	N	B	L	X	K	L	L	N	G	D

FIGURE 12.

(8) The only difference in procedure between this case and the preceding one (where the cipher alphabet was a direct standard alphabet) is that the letters of the cipher text are first "deciphered" by means of *any* reversed standard alphabet and then the columns are "run down", according to the normal A B C . . . Z sequence. For reasons which will become apparent very soon, the first step in this method is technically termed *converting the cipher letters into their plain-component equivalents;* the second step is the same as before, *viz., completing the plain-component sequence.*

35. Special remarks on the method of solution by completing the plain-component sequence.—*a.* The terms employed to denote the steps in the solution set forth in subpar. 34*b* (8), *viz.*, "converting the cipher letters into their plain-component equivalents" and "completing the plain-component sequence", accurately describe the process. Their meaning will become more clear as the student progresses with the work. It may be said that whenever the components of a cipher alphabet are *known* sequences, no matter how they are composed, the difficulty and time required to solve any cryptogram involving the use of those components is considerably reduced. *In some cases this knowledge facilitates, and in other cases is the only thing that makes possible, the solution of a very short cryptogram that might otherwise defy solution.* Later on an example will be given to illustrate what is meant in this regard.

b. The student should take note, however, of two qualifying expressions that were employed in a preceding paragraph to describe the results of the application of the method. It was stated that "one of the generatrices will exhibit intelligible text *all the way across;* this text will *almost invariably* be the plain text." Will there ever be a case in which more than one generatrix will yield intelligible text through its extent? That obviously depends almost entirely on the number of letters that are aligned to form a generatrix. If a generatrix contains but a very few letters, only five, for example, it may happen as a result of pure chance that there will be two or more generatrices showing what might be "intelligible text." Note in Fig. 12, for example, that there are several cases in which 3-letter and 4-letter English words (LAD, COB, MESH, MAPS, etc.) appear on generatrices that are not correct, these words being formed by pure chance. But there is not a single case, in this diagram, of a 5-letter or longer word appearing fortuitously, because obviously the longer the word the smaller the probability of its appearance purely by chance; and the probability that two generatrices of 15 letters each will both yield intelligible text along their entire length is exceedingly remote, so remote, in fact, that in practical cryptology such a case may be considered nonexistent.[7]

c. The student should observe that in reality there is no difference whatsoever in principle between the two methods presented in subpars. *a* and *b* of par. 34. In the former the preliminary step of converting the cipher letters into their plain-component equivalents is apparently not present but in reality it is there. The reason for its apparent absence is that in that case the plain component of the cipher alphabet is identical in all respects with the cipher component, so that the cipher letters require no conversion, or, rather, they are identical with the equivalents that would result if they were converted on the basis $A_c = A_p$. In fact, if the solution process had been arbitrarily initiated by converting the cipher letters into their plain-component equivalents at the setting $A_c = O_p$, for example, and the cipher component slid one interval to the right thereafter, the results of the first and second tests of par. 34*a* would be as follows:

Cryptogram	Z	C	U	I	X	A	O	Y	K	X	Y	Y	A	T	Q
1st Test—"Plain text"	N	Q	I	W	L	O	C	M	Y	L	M	M	O	H	E
2d Test—"Plain text"	O	R	J	X	M	P	D	N	Z	M	N	N	P	I	F

Thus, the foregoing diagram duplicates in every particular the diagram resulting from the first two tests under par. 34*b:* a first line of cipher letters, a second line of letters derived from them but showing externally no relationship with the first line, and a third line derived immediately from the second line by continuing the direct normal sequence. This point is brought to attention only for the purpose of showing that a single, broad principle is the basis of the general method of solution by completing the plain-component sequence, and once the student has this firmly in mind he will have no difficulty whatsoever in realizing when the principle is applicable, what a powerful cryptanalytic tool it can be, and what results he may expect from its application in specific instances.

d. In the two foregoing examples of the application of the principle, the components were normal sequences; but it should be clear to the student, if he has grasped what has been said in the preceding subparagraph, that these components may be mixed sequences which, if known (that is, if the sequence of letters comprising the sequences is known to the cryptanalyst), can be handled just as readily as can components that are normal sequences.

---

[7] A person with patience and an inclination toward the curiosities of the science might construct a text of 15 or more letters which would yield two "intelligible" texts on the plain-component completion diagram.

e. It is entirely immaterial at what points the plain and the cipher components are juxtaposed in the preliminary step of converting the cipher letters into their plain-component equivalents. For example, in the case of the reversed alphabet cipher solved in subpar. 34*b*, the two components were arbitrarily juxtaposed to give the value $A_p=A_c$, but they might have been juxtaposed at any of the other 25 possible points of coincidence without in any way affecting the final result, *viz.*, the production of one plaintext generatrix in the completion diagram.

36. Value of mechanical solution as a short cut.—*a*. It is evident that *the very first step the student should take in his attempts to solve an unknown cryptogram that is obviously a substitution cipher is to try the mechanical method of solution by completing the plain-component sequence, using the normal alphabet, first direct, then reversed.* This takes only a very few minutes and is conclusive in its results. It saves the labor and trouble of constructing a frequency distribution in case the cipher is of this simple type. Later on it will be seen how certain variations of this simple type may also be solved by the application of this method. Thus, a very easy short cut to solution is afforded, which even the experienced cryptanalyst never overlooks in his first attack on an unknown cipher.

b. It is important now to note that *if neither of the two foregoing attempts is successful in bringing plain text to light and the cryptogram is quite obviously monoalphabetic in character, the cryptanalyst is warranted in assuming that the cryptogram involves a mixed cipher alphabet.*[8]

37. Basic reason for the low degree of cryptosecurity afforded by monoalphabetic cryptograms involving standard cipher alphabets.—The student has seen that the solution of monoalphabetic cryptograms involving standard cipher alphabets is a very easy matter. Two methods of analysis were described, one involving the construction of a frequency distribution, the other not requiring this kind of tabulation, being almost mechanical in nature and correspondingly rapid. In the first of these two methods it was necessary to make a correct assumption as to the value of but one of the 26 letters of the cipher alphabet and the values of the remaining 25 letters at once became known; in the second method it was not necessary to assume a value for even a single cipher letter. The student should understand what constitutes the basis of this situation, *viz.*, the fact that the two components of the cipher alphabet are composed of *known sequences*. What if one or both of these components are for the cryptanalyst *unknown sequences?* In other words, what difficulties will confront the cryptanalyst if the cipher component of the cipher alphabet is a mixed sequence? Will such an alphabet be solvable as a whole at one stroke, or will it be necessary to solve its values individually? Since the determination of the value of one cipher letter in this case gives no direct clues to the value of any other letter, it would seem that the solution of such a cipher should involve considerably more analysis and experiment than has the solution of either of the two types of ciphers so far examined. The steps to be taken in the cryptanalysis of a mixed-alphabet cipher will be discussed in the next chapter.

[8] There is but one other possibility, already referred to under subpar. 31*d*, which involves the case where transposition and monoalphabetic substitution processes have been applied in successive steps. This is unusual, however, and will be discussed in a subsequent text.

Chapter VI

UNILITERAL SUBSTITUTION WITH MIXED CIPHER ALPHABETS

38. Literal keys and numerical keys.—*a.* As has been previously mentioned, most cryptosystems involve the use of a specific key to control the steps followed in encrypting or decrypting a specific message (see subpar. 9*b*). Such a key may be in literal form or in numerical form.

b. It is convenient to designate a key which is composed of letters as a *literal key.* As already mentioned, a literal key may consist of a single letter, a single word, a phrase, a sentence, a whole paragraph, or even a book; and, of course, it may consist merely of a sequence of letters chosen at random.

c. Certain cryptosystems involve the use of a *numerical key,* which may consist of a relatively long sequence of numbers difficult or impossible for the average cipher clerk to memorize. Several simple methods for deriving such sequences from words, phrases, or sentences have been devised, and a numerical key produced by any of these methods is called a *derived numerical key* (as opposed to a key consisting of randomly-selected numbers). One of the commonly-used methods consists of assigning numerical values to the letters of a selected literal key in accordance with their *relative* positions in the ordinary alphabet, as exemplified in the following subparagraph.

d. Let the prearranged *key word* be the word LOGISTICS. Since C, the penultimate letter of the key word, appears in the normal alphabet before any other letter of the key word, it is assigned the number 1:

```
L O G I S T I C S
              1
```

The next letter of the normal alphabet that occurs in the key word is G, which is assigned the number 2. The letter I, which occurs twice in the key word, is assigned the number 3 for its first occurrence (from left to right) and the number 4 for its second occurrence; and so on. The final result is:

```
L O G I S T I C S
5 6 2 3 7 9 4 1 8
```

This method of assigning the numbers is very flexible and varies with different uses to which numerical keys are put. It may, of course, be applied to phrases or to sentences, so that a very long numerical key, ordinarily impossible to remember, may be thus derived at will from an easily-remembered *key text*.

e. As far as the cryptanalyst is concerned, the derivation of a numerical key from a specific literal key is of interest to him because this knowledge may assist in subsequent solutions of cryptograms prepared according to the same basic system, or in identifying the source from which the literal key was selected—perhaps an ordinary book, a magazine, etc. However, it should be pointed out that in some instances the cryptanalyst may be unaware that a literal key has in fact been used as the basis for deriving a numerical key.

39. Types of mixed cipher alphabets.—*a*. It will be recalled that in a mixed cipher alphabet the sequence of letters or characters in one of the components (usually the cipher component) does not correspond to the normal sequence. There are various methods of composing the sequence of letters or elements of this mixed component, and those which are based upon a scheme that is systematic in its nature are very useful because they make possible the derivation of one or more mixed sequences from any easily-remembered word or phrase, and thus do not necessitate the carrying of written memoranda. Alphabets involving a systematic method of mixing are called *systematically-mixed cipher alphabets*.

b. One of the simplest types of systematically-mixed cipher alphabets is the *keyword-mixed alphabet*. The cipher component consists of a key word or phrase (with repeated letters, if present, omitted after their first occurrence),[1] followed by the letters of the alphabet in their normal sequence (with letters already occurring in the key omitted, of course). Example, with GOVERNMENT as the key word:

```
 Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: GOVERNMTABCDFHIJKLPQSUWXYZ
```

c. It is possible to disarrange the sequence constituting the cipher component even more thoroughly by applying a simple method of transposition to the keyword-mixed sequence. Two common methods are illustrated below, using the key word TELEPHONY.

(1) *Simple columnar transposition:*

```
T E L P H O N Y
A B C D F G I J
K M Q R S U V W
X Z
```

[1] Mixed alphabets formed by including all repeated letters of the key word or key phrase in the cipher component were common in Edgar Allan Poe's day but are impractical because they are ambiguous, making decipherment difficult; an example:

(a) Alphabet for enciphering---------------------------------

```
Plain:  ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: NOWISTHETIMEFORALLGOODMENT
```

(b) Inverse form of (a), for deciphering-----------------------

```
Cipher: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Plain:  P  VHMSGD  QKAB  OEF  C
            L   J  RWYN    I
            X         T    Z
                      U
```

The average cipher clerk would have considerable difficulty in decrypting a cipher group such as TOOET, each letter of which has three or more equivalents, and from which the plaintext fragments (N)INTH., .. FT THI(S), IT THI ... , etc. can be formed on decipherment.

Mixed sequence (formed by transcribing the successive columns from left to right):

TAKXEBMZLCQPDRHFSOGUNIVYJW

(2) *Numerically-keyed columnar transposition:*

```
7-1-3-6-2-5-4-8
T E L P H O N Y
A B C D F G I J
K M Q R S U V W
X Z
```

Mixed sequence (formed by transcribing the columns in a sequence determined by the numerical key derived from the key word itself):

EBMZHFSLCQNIVOGUPDRTAKXYJW

d. The last two systematically-mixed sequences are examples of *transposition-mixed sequences.* Almost any method of transposition may be used to produce such sequences.

e. Another simple method of forming a mixed sequence is the *decimation method.* In this method, letters in the normal alphabet, or in a keyword-mixed sequence, are "counted off" according to any selected interval. As each letter is decimated—that is, eliminated from the basic sequence by counting off—it is entered in a separate list to form the new mixed sequence. For example, to form a mixed sequence by this method from a keyword-mixed sequence based on the key phrase SING A SONG OF SIXPENCE with 7 the interval selected, proceed as follows: Keyword-mixed (or basic) sequence:

SINGAOFXPECBDHJKLMQRTUVWYZ

When the letters are counted off by 7's from left to right, F will be the first letter arrived at, H the second, T the third:

```
S I N G A O F X P E C B D H J K L M Q R T U V W Y Z
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
```

These letters are entered in a separate list (F first, H second, T third, and so on) and eliminated from the keyword-mixed sequence. When the end of the keyword-mixed sequence is reached, return to the beginning, skipping the letters already eliminated:

```
S I N G A O F X P E C B D H J K L M Q R T U V W Y Z
                                          1 2 3 4 5
6 7 1 2 3 4   5 6 7 1 2 3     4 5 6 7
```

The decimation-mixed sequence:

FHTIEMZPQNDWCVBSLXAGOKYJRU

f. Practical considerations, of course, set a limit to the complexities that may be introduced in constructing systematically-mixed alphabets. Beyond a certain point there is no object in further mixing. The greatest amount of mixing by systematic processes will give no more security than that resulting from mixing the alphabet by random selection, such as by putting the 26 letters in a box, thoroughly shaking them up, and then drawing the letters out one at a time. Whenever the laws of chance operate in the construction of a mixed alphabet, the probability of producing a thorough disarrangement of letters is very great. *Random-mixed alphabets* give more cryptographic security than do the less complicated systematically-mixed alphabets, because they afford no clues to positions of letters, given the position of a few of them. Their chief disadvantage is that they must be reduced to writing, since they cannot readily be remembered, nor can they be reproduced at will from an easily-remembered key word.

40. Additional remarks on cipher alphabets.—*a.* Cipher alphabets may be classified on the basis of their arrangement as *enciphering* or *deciphering alphabets.* An enciphering alphabet is one in which the sequence of letters in the plain component coincides with the normal sequence and is arranged in that manner for convenience in encipherment. In a deciphering alphabet the sequence of letters in the cipher component coincides with the normal, for convenience in deciphering. For example, (1), below, shows a mixed cipher alphabet arranged as an enciphering alphabet; (2) shows the corresponding deciphering alphabet. An enciphering alphabet and its corresponding deciphering alphabet present an *inverse* relationship to each other.

Enciphering Alphabet

(1) Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: JKQVXZWESTRNUIOLGAPHCMYBDF

Deciphering Alphabet

(2) Cipher: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Plain: RXUYHZQTNABPVLOSCKIJMDGEWF

b. As has been previously mentioned,[2] a series of related reciprocal alphabets may be produced by juxtaposing at all possible points of coincidence two components which are identical but progress in opposite directions. This holds regardless of whether the components are composed of an even or an odd number of elements. The following reciprocal alphabet is one of such a series of 26 alphabets:

Plain: HYDRAULICBEFGJKMNOPQSTVWXZ
Cipher: GFEBCILUARDYHZXWVTSQPONMKJ

A single or isolated reciprocal alphabet may be produced in one of two ways:

(1) By constructing a complete reciprocal alphabet by arbitrary or random assignments of values in pairs. That is, if A_p is made the equivalent of K_c, then K_p is made the equivalent of A_c; if B_p is made R_c, then R_p is made B_c, and so on. If the two components thus constructed are slid against each other no additional reciprocal alphabets will be produced.

(2) By juxtaposing a sequence comprising an *even* number of elements against the same sequence shifted exactly half way to the right (or left), as seen below:

HYDRAULICBEFGJKMNOPQSTVWXZ
HYDRAULICBEFGJKMNOPQSTVWXZHYDRAULICBEFGJKMNOPQSTVWXZ

[2] Subpar. 29*c*.

41. Preliminary steps in the analysis of a monoalphabetic, mixed-alphabet cryptogram.—*a.* The student is now ready to resume his cryptanalytic studies. Note the following cryptogram:

```
SFDZF IOGHL PZFGZ DYSPF HBZDS GVHTF UPLVD FGYVJ VFVHT GADZZ AITYD ZYFZJ
ZTGPT VTZBD VFHTZ DFXSB GIDZY VTXOI YVTEF VMGZZ THLLV XZDFM HTZAI TYDZY
BDVFH TZDFK ZDZZJ SXISG ZYGAV FSLGZ DTHHT CDZRS VTYZD OZFFH TZAIT YDZYG
AVDGZ ZTKHI TYZYS DZGHU ZFZTG UPGDI XWGHX ASRUZ DFUID EGHTV EAGXX
```

b. A casual inspection of the text discloses the presence of several long repetitions as well as of many letters of normally low frequency, such as F, G, V, X, and Z; on the other hand, letters of normally high frequency, such as the vowels, and the consonants N and R, are relatively scarce. The cryptogram is obviously a substitution cipher and the usual mechanical tests for determining whether it is possibly of the monoalphabetic, standard-alphabet type are applied. The results being negative, a uniliteral frequency distribution is immediately constructed, as shown in Fig. 13, and the ϕ test is applied to it.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	4	1	23	3	19	19	15	10	3	2	5	2	0	3	5	0	2	10	22	5	16	1	8	14	35

$\phi_p=3668$ $\phi_r=2117$ $\phi_o=3862$

FIGURE 13.

c. The fact that the frequency distribution shows very marked crests and troughs indicates that the cryptogram is very probably monoalphabetic, and the results of the ϕ test further support this hypothesis. The fact that the cryptogram has already been tested by the method of completing the plain-component sequence and found not to be of the monoalphabetic, standard-alphabet type, indicates with a high degree of probability that it involves a mixed cipher alphabet. A few moments might be devoted to making a careful inspection of the distribution to insure that it cannot be made to fit the normal; the object of this would be to rule out the possibility that the text resulting from substitution by a standard cipher alphabet had not subsequently been transposed. But this inspection in this case is hardly necessary, in view of the presence of long repetitions in the message.[3] (See subpar. 25*g*.)

d. One might, of course, attempt to solve the cryptogram by applying the simple principles of frequency. One might, in other words, assume that Z_c (the letter of greatest frequency) represents E_p, D_c (the letter of next greatest frequency) represents T_p, and so on. If the message

[3] This possible step is mentioned here for the purpose of making it clear that the plain-component sequence completion method cannot solve a case in which transposition has followed or preceded monoalphabetic substitution with standard alphabets. Cases of this kind will be discussed in a later text. It is sufficient to indicate at this point that the frequency distribution for such a combined substitution-transposition cipher would present the characteristics of a standard alphabet cipher and yet the method of completing the plain-component sequence would fail to bring out any plain text.

were long enough this simple procedure might more or less quickly give the solution. But the message is relatively short and many difficulties would be encountered. Much time and effort would be expended unnecessarily, because it is hardly to be expected that in a message of only 235 letters the relative order of frequency of the various cipher letters should exactly coincide with, or even closely approximate the relative order of frequency of letters of normal plain text found in a count of 50,000 letters. *It is to be emphasized that the beginner must repress the natural tendency to place too much confidence in the generalized principles of frequency and to rely too much upon them.* It is far better to bring into effective use certain other data concerning normal plain text, such as digraphic and trigraphic frequencies.

42. Preparation of the work sheet.—*a.* The details to be considered in this paragraph may at first appear to be superfluous, but long experience has proved that systematization of the work and preparation of the data in the most utilizable, condensed form is most advisable, even if this seems to take considerable time. In the first place, if it merely serves to avoid interruptions and irritations occasioned by failure to have the data in an instantly available form, it will pay by saving mental wear and tear. In the second place, especially in the case of complicated cryptograms, painstaking care in these details, while it may not always bring about success, is often the factor that is of greatest assistance in ultimate solution. The detailed preparation of the data may be irksome to the student, and he may be tempted to avoid as much of it as possible, but, unfortunately, in the early stages of solving a cryptogram he does not know (nor, for that matter, does the expert always know) just which data are essential and which may be neglected. Even though not all of the data may turn out to have been necessary, as a general rule, time is saved in the end if all the usual data are prepared as a regular preliminary to the solution of most cryptograms.

b. First, the cryptogram is recopied in the form of a *work sheet.* This sheet should be of a good quality of paper so as to withstand considerable erasure. If the cryptogram is to be copied by hand, cross-section paper of ¼-inch squares is extremely useful, because each letter may be written in an individual cell. The writing should be in ink, and plain, carefully-made roman capital letters should be used in all cases.[4] If the cryptogram is to be copied on a typewriter, the ribbon employed should be impregnated with an ink that will not smear or smudge under the hand.

c. The arrangement of the characters of the cryptogram on the work sheet is a matter of considerable importance. If the cryptogram as first obtained is in groups of regular length (usually five characters to a group) and if the uniliteral frequency distribution shows the cryp-

---

[4] It is advisable to use, for this purpose, the system of standardized manual printing adopted by Service communications personnel. The use of this system, appended below, assures that work sheets are completely legible, not only to the person preparing them, but to others as well.

A	B	C	D	E	F	G	H	I	J	K	L
M	N	O	P	Q	R	S	T	U	V	W	X
Y	Z	1	2	3	4	5	6	7	8	9	Ø

togram to be monoalphabetic, the characters should be copied without regard to this grouping. It is advisable to allow one space between letters (this is especially true for work sheets prepared on the typewriter), and to write a constant number of letters per line, approximately 25. At least two spaces, preferably three spaces, should be left between horizontal lines, to allow room for multiple assumptions. Care should be taken to avoid crowding the letters in any case, for this is not only confusing to the eye but also mentally irritating when later it is found that not enough space has been left for making various sorts of marks or indications. If the cryptogram is originally in what appears to be word lengths (and this is the case, as a rule, only with the cryptograms of amateurs), naturally it should be copied on the work sheet in the original groupings.[5] If further study of a cryptogram shows that some special grouping is required, it is often best to recopy it on a fresh work sheet rather than to attempt to indicate the new grouping on the old work sheet.

d. In order to be able to locate or refer to specific letters or groups of letters with speed, certainty, and without possibility of confusion, it is advisable to use coordinates applied to the lines and columns of the text as it appears on the work sheet. To minimize possibility of confusion, it is best to apply letters to the horizontal lines of the text, numbers to the vertical columns. In referring to a letter, the horizontal line in which the letter is located is usually given first. Thus, referring to the work sheet shown below, coordinates A17 designate the letter Y, the 17th letter in the first line. The letter I is usually omitted from the series of line indicators so as to avoid confusion with the figure 1. If lines are limited to 25 letters each, then each set of 100 letters of the text is automatically blocked off by remembering that 4 lines constitute 100 letters.

e. Above each character of the cipher text may be some indication of the frequency of that character in the whole cryptogram. This indication may be the actual number of times the character occurs, or, if colored pencils are used, the cipher letters may be divided up into three categories or groups—high-frequency, medium-frequency, and low-frequency. It is perhaps simpler, if clerical help is available, to indicate the actual frequencies. This saves constant reference to the frequency tables, which interrupts the train of thought, and saves considerable time in the end, since it enables the student better to visualize *frequency-patterns* of words. In any case, it is recommended that the frequencies of the letters comprising the repetitions be inscribed over their respective letters; likewise, the frequencies of the first 10 and last 10 letters should also be inscribed, as these positions often lend themselves readily to attack.[6]

f. After the special frequency distribution, explained in par. 43 below, has been constructed, repetitions of digraphs and trigraphs should be underscored. In so doing, the student should be particularly watchful for trigraphic repetitions which can be further extended into tetragraphs and polygraphs of greater length. If a repetition continues from one line to the next, put an arrow at the end of the underscore to signal this fact. Reversible digraphs and trigraphs should also be indicated by an underscore with an arrow pointing in both directions. Anything which strikes the eye as being peculiar, unusual, or significant as regards the distribution or recurrence of the characters should be noted. All these marks should, if convenient, be made with ink

[5] In some cryptosystems, certain low-frequency letters are employed as *word separators* to indicate the end of a word; if the meaning of these letters is discovered, it is tantamount to having the cryptogram in word lengths and thus the work sheet is made accordingly. See also in this connection the treatment on word separators in Chapter VII.

[6] See Appendix 4 in this connection.

so as not to cause smudging. The work sheet will now appear as shown below (not all the repetitions are underscored):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	10	19	23	35	19	10	3	19	15	5	5	35	19	19	35	23	14	10	5	19	15	4	35	23	10
A	S	F	D	Z	F	I	O	G	H	L	P	Z	F	G	Z	D	Y	S	P	F	H	B	Z	D	S
	19	16	15	22	19	5	5	5	16	23	19	19	14	16	3	16	19	16	15	22	19	8	23	35	35
B	G	V	H	T	F	U	P	L	V	D	F	G	Y	V	J	V	F	V	H	T	G	A	D	Z	Z
	8	10	22	14	23	35	14	19	35	3	35	22	19	5	22	16	22	35	4	23	16	19	15	22	35
C	A	I	T	Y	D	Z	Y	F	Z	J	Z	T	G	P	T	V	T	Z	B	D	V	F	H	T	Z
	23	19	8	10	4	19	10	23	35	14	16	22	8	3	10	14	16	22	3	19	16	2	19	35	35
D	D	F	X	S	B	G	I	D	Z	Y	V	T	X	O	I	Y	V	T	E	F	V	M	G	Z	Z
	22	15	5	5	16	8	35	23	19	2	15	22	35	8	10	22	14	23	35	14	4	23	16	19	15
E	T	H	L	L	V	X	Z	D	F	M	H	T	Z	A	I	T	Y	D	Z	Y	B	D	V	F	H
	22	35	23	19	2	35	23	35	35	3	10	8	10	10	19	35	14	19	8	16	19	10	5	19	35
F	T	Z	D	F	K	Z	D	Z	Z	J	S	X	I	S	G	Z	Y	G	A	V	F	S	L	G	Z
	23	22	15	15	22	1	23	35	2	10	16	22	14	35	23	3	35	19	19	15	22	35	8	10	22
G	D	T	H	H	T	C	D	Z	R	S	V	T	Y	Z	D	O	Z	F	F	H	T	Z	A	I	T
	14	23	35	14	19	8	16	23	19	35	35	22	2	15	10	22	14	35	14	10	23	35	19	15	5
H	Y	D	Z	Y	G	A	V	D	G	Z	Z	T	K	H	I	T	Y	Z	Y	S	D	Z	G	H	U
	35	19	35	22	19	5	5	19	23	10	8	1	19	15	8	8	10	2	5	35	23	19	5	10	23
J	Z	F	Z	T	G	U	P	G	D	I	X	W	G	H	X	A	S	R	U	Z	D	F	U	I	D
	3	19	15	22	16	3	8	19	8	8															
K	E	G	H	T	V	E	A	G	X	X															

43. Triliteral frequency distributions.—*a.* In what has gone before, a type of frequency distribution known as a uniliteral frequency distribution was used. This, of course, shows only the number of times each individual letter occurs. In order to apply the normal digraphic and trigraphic frequency data (given in Appendix 2) to the solution of a cryptogram of the type now being studied, it is obvious that the data with respect to digraphs and trigraphs occurring in the cryptogram should be compiled and should be compared with the data for normal plain text. In order to accomplish this in suitable manner, it is advisable to construct a more comprehensive form of distribution termed a *triliteral frequency distribution.*[7]

b. Given a cryptogram of 50 or more letters and the task of determining what trigraphs are present in the cryptogram, there are three ways in which the data may be arranged or assembled. One may require that the data show (1) each letter with its two succeeding letters; (2) each letter with its two preceding letters; (3) each letter with one preceding letter and one succeeding letter.

[7] It is felt wise here to distinguish between two closely related terms. A triliteral distribution of A B C D E F would consider the groups A B C, B C D, C D E, D E F; a trigraphic distribution would consider only the trigraphs A B C and D E F. (See also subpar. 23*d*.)

c. A distribution of the first of the three foregoing types may be designated as a "triliteral frequency distribution showing two suffixes"; the second type may be designated as a "triliteral frequency distribution showing two prefixes"; the third type may be designated as a "triliteral frequency distribution showing one prefix and one suffix." Quadriliteral and pentaliteral frequency distributions may occasionally be found useful.

d. Which of these three arrangements is to be employed at a specific time depends largely upon what the data are intended to show. For present purposes, in connection with the solution of a monoalphabetic substitution cipher employing a mixed alphabet, possibly the third arrangement, that showing one prefix and one suffix, is most satisfactory.

e. It is convenient to use ¼-inch cross-section paper for the construction of a triliteral frequency distribution in the form of a distribution showing crests and troughs, such as that in Fig. 14. In that figure the prefix to each letter to be recorded is inserted in the left half of the cell directly above the cipher letter being recorded; the suffix to each letter is inserted in the right half of the cell directly above the letter being recorded; and in each case the prefix and the suffix to the letter being recorded occupy the same cell, the prefix being directly to the left of the suffix. The number in parentheses gives the total frequency for each letter.

f. The triliteral frequency distribution is now to be examined with a view to ascertaining what digraphs and trigraphs occur two or more times in the cryptogram. Consider the pair of columns containing the prefixes and suffixes to D_c in the distribution, as shown in Fig. 14. This pair of columns shows that the following digraphs appear in the cryptogram:

Digraphs based on prefixes (arranged as one reads up the column)

FD, ZD, ZD, VD, AD, YD, BD,
ZD, ID, ZD, YD, BD, ZD, ZD,
ZD, CD, ZD, YD, VD, SD, GD,
ZD, ID

Digraphs based on suffixes (arranged as one reads up the column)

DZ, DY, DS, DF, DZ, DZ, DV,
DF, DZ, DF, DZ, DV, DF, DZ,
DT, DZ, DO, DZ, DG, DZ, DI,
DF, DE

The nature of the triliteral frequency distribution is such that in finding what digraphs are present in the cryptogram it is immaterial whether the prefixes or the suffixes to the cipher letters are studied, *so long as one is consistent in the study*. For example, in the foregoing list of digraphs based on the prefixes to D_c, the digraphs FD, ZD, ZD, VD, etc., are found; if now, the student will refer to the suffixes of F_c, Z_c, V_c, etc., he will find the very same digraphs indicated. This being the case, the question may be raised as to what value there is in listing both the prefixes and the suffixes to the cipher letters. The answer is that by so doing the trigraphs are indicated at the same time. For example, in the case of D_c, the following trigraphs are indicated:

FDZ, ZDY, ZDS, VDF, ADZ, YDZ, BDV, ZDF, IDZ, ZDF, YDZ, BDV, ZDF,
ZDZ, ZDT, CDZ, ZDO, YDZ, VDG, SDZ, GDI, ZDF, IDE.

g. The *repeated* digraphs and trigraphs can now be found quite readily. Thus, in the case of D_c, examining the list of digraphs based on suffixes, the following repetitions are noted:

DZ appears 9 times; DF appears 5 times; DV appears 2 times

Examining the trigraphs with D_c as central letter, the following repetitions are noted:

ZDF appears 4 times; YDZ appears 3 times; BDV appears 2 times

CONDENSED TABLE OF REPETITIONS

Digraphs			*Trigraphs*		*Longer Polygraphs*
DZ–9	TZ–5	VF–4	DZY–4	FHT–3	HTZAITYDZY–2
ZD–9	TY–5	VT–4	HTZ–4	TYD–3	BDVFHTZDF–2
HT–8	FH–4	ZF–4	ITY–4	YDZ–3	ZAITYDZY–3
ZY–6	GH–4	ZT–4	ZDF–4	ZAI–3	FHTZ–3
DF–5	IT–4	ZZ–4	AIT–3		
GZ 5					

																									UD
																									FT
																									UF
																									DG
																									YY
																									ZT
																									GZ
																									DY
																									TA
																									OF
																									YD
																									DR
			IE																						GD
			ZF																HV						GY
			GI																ZG						ZJ
			SZ																IY						DZ
			VG		DU	AX													ZK						KD
			YZ		ZZ	EH													IY						TD
			ZO		FH	WH													HZ						DY
			CZ		ZF	PD													VY		TE				TA
			ZT		VS	TU	GT												HC		AD				XD
			ZZ		DK	ZH	GX												DH		ST			ZS	ZT
			ZF		VH	DZ	GU												HZ		AF			TZ	GZ
			BV		DM	YA	KI												IY		DF			ZG	DY
			YZ		EV	LZ	FT												HZ		LX			TD	TD
			ZF		DX	YA	HT	UD										AR	ZH		FM			TZ	TB
			IZ		VH	SZ	TH	DX										YD	VE		YT			ZG	JT
EG			ZF		YZ	MZ	FT	HT										RV	VX		YT		X–	ZB	FJ
XS			BV		VV	BI	MT	AT										FL	HZ		DF		GX	TD	DY
GV			YZ		DG	TP	TL	XS										IG	VZ		TT		HA	IV	ZA
ZI			AZ		TU	TA	FT	AT			SG				UG			JX	PV	FI	FH		IW	ZV	DZ
GV	YD		VF		PH	FY	VT	OY			LV				GT			XB	ZG	RZ	JF		SI	ZF	BD
ZI	SG		ZS	VA	ZG	SV	VT	GD	ZS		HL			DZ	UL			DG	IY	GP	YJ		VZ	TD	GD
ZI	ZD		ZY	DG	ZI	FZ	FB	AT	ZZ	TH	PV	FH		XI	SF		SU	YP	HG	HZ	LD		TO	GV	PF
GD	HZ	TD	FZ	TF	SD	OH	GL	FO	VV	FZ	HP	VG		IG	LZ		ZS	–F	HF	FP	GH	XG	FS	DS	DF
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
(8)	(4)	(1)	(23)	(3)	(19)	(19)	(15)	(10)	(3)	(2)	(5)	(2)	(0)	(3)	(5)	(0)	(2)	(10)	(22)	(5)	(16)	(1)	(8)	(14)	(35)

FIGURE 14.

h. It is unnecessary, of course, to go through the detailed procedure set forth in the preceding subparagraphs in order to find all the repeated digraphs and trigraphs. The repeated trigraphs with D_c as central letter can be found merely from an inspection of the prefixes and suffixes opposite D_c in the distribution. It is necessary only to find those cases in which two or more prefixes are identical at the same time that the suffixes are identical. For example, the distribution shows at once that in four cases the prefix to D_c is Z_c at the same time that the suffix to this letter is F_c. Hence, the trigraph ZDF appears four times. The repeated trigraphs may all be found in this manner.

i. The most frequently repeated digraphs and trigraphs are then assembled in what is termed a *condensed table of repetitions,* so as to bring this information prominently before the eye. As a rule, in messages of average length, digraphs which occur less than four or five times, and trigraphs which occur less than three or four times may be omitted from the condensed table as being relatively of no importance in the study of repetitions. In the condensed table the frequencies of the individual letters forming the most important digraphs, trigraphs, etc., should be indicated.

44. Classifying the cipher letters into vowels and consonants.—*a.* Before proceeding to a detailed analysis of the repeated digraphs and trigraphs, a very important step can be taken which will be of assistance not only in the analysis of the repetitions but also in the final solution of the cryptogram. This step concerns the classification of the high-frequency cipher letters into two groups—(1) those which most probably represent vowels, and (2) those which most probably represent consonants. For if the cryptanalyst can quickly ascertain the equivalents of the four vowels, A, E, I, and O, and of only the four consonants, N, R, S, and T, he will then have the values of approximately two-thirds of all the cipher letters that occur in the cryptogram; the values of the remaining letters can almost be filled in automatically.

b. The basis for the classification will be found to rest upon a comparatively simple phenomenon: the associational or combinatory behavior of vowels is, in general, quite different from that of consonants. If an examination be made of Table 7–B in Appendix 2, showing the relative order of frequency of the 18 digraphs composing 25 per cent of English telegraphic text, it will be seen that the letter E enters into the composition of 9 of the 18 digraphs; that is, in exactly half of all the cases the letter E is one of the two letters forming the digraph. The digraphs containing E are as follows:

```
ED  EN  ER  ES
    NE  RE  SE  TE  VE
```

The remaining nine digraphs are as follows:

```
AN  ND  OR  ST
IN  NT      TH
ON          TO
```

c. None of the 18 digraphs is a combination of vowels. Note now that of the 9 combinations with E, 7 are with the consonants N, R, S, and T, one is with D, one is with V, and *none is with any vowel.* In other words, E_p combines most readily with consonants but not with other vowels, or even with itself. Using the terms often employed in the chemical analogy, E shows a great "affinity" for the consonants N, R, S, T, but not for the vowels. Therefore, if the letters of highest frequency occurring in a given cryptogram are listed, together with the number of times each of them combines with the assumed cipher equivalent of E_p, those which show considerable combining power or affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of N, R, S, T_p; those which do not show any affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of A, I, O, U_p. Applying these principles to the problem in hand, and examining the triliteral frequency distribution, it is quite certain that $Z_c=E_p$ not only because Z_c is the letter of highest frequency, but also because it combines with *several* other high-frequency letters, such as D_c, F_c, G_c, etc. The nine letters of next highest frequency are:

23	22	19	19	16	15	14	10	10
D	T	F	G	V	H	Y	S	I

Let the combinations these letters form with Z_c be indicated in the following manner:

Number of times Z_c occurs as prefix	//// ~~////~~	////	////	/			/ ~~////~~		
Cipher letter	D(23)	T(22)	F(19)	G(19)	V(16)	H(15)	Y(14)	S(10)	I(10)
Number of times Z_c occurs as suffix	~~////~~ ////	~~////~~	//	~~////~~			//		

d. Consider D_c. It occurs 23 times in the message and 18 of those times it is combined with Z_c, 9 times in the form Z_cD_c $(=E\theta_p)$, and 9 times in the form D_cZ_c $(=\theta E_p)$. It is clear that D_c must be a consonant. In the same way, consider T_c, which shows 9 combinations with Z_c, 4 in the form Z_cT_c $(=E\theta_p)$ and 5 in the form T_cZ_c $(=\theta E_p)$. The letter T_c appears to represent a consonant, as do also the letters F_c, G_c, and Y_c. On the other hand, consider V_c, occurring in all 16 times but never in combination with Z_c; it appears to represent a vowel, as do also the letters H_c, S_c, and I_c. So far, then, the following classification would seem logical:

Vowels	Consonants
$Z_c(=E_p)$, V_c, H_c, S_c, I_c	D_c, T_c, F_c, G_c, Y_c

45. Further analysis of the letters representing vowels and consonants.—*a.* O_p is usually the vowel of second highest frequency. Is it possible to determine which of the letters V, H, S, I_c is the cipher equivalent of O_p? Let reference be made again to Table 6 in Appendix 2, where it is seen that the 10 most frequently occurring diphthongs are:

Diphthong	IO	OU	EA	EI	AI	IE	AU	EO	AY	UE
Frequency	41	37	35	27	17	13	13	12	12	11

If V, H, S, I_c are really the cipher equivalents of A, I, O, U_p (not respectively), perhaps it is possible to determine which is which *by examining the combinations they make among themselves and with* Z_c $(=E_p)$. Let the combinations of V, H, S, I, and Z that occur in the message be listed. There are only the following:

ZZ_c—4 VH_c—2 HH_c—1 HI_c—1 IS_c—1 SV_c—1

ZZ_c is of course EE_p. Note the doublet HH_c; if H_c is a vowel, then the chances are excellent that $H_c=O_p$ because the doublets AA_p, II_p, UU_p, are practically non-existent, whereas the double vowel combination OO_p is of next highest frequency to the double vowel combination EE_p. If $H_c=O_p$, then V_c must be I_p because the digraph VH_c occurring two times in the message could hardly be AO_p, or UO_p, whereas the diphthong IO_p is the one of high frequency in English. So far then, the tentative (because so far unverified) results of the analysis are as follows:

$$Z_c=E_p \quad H_c=O_p \quad V_c=I_p$$

This leaves only two letters, I_c and S_c (already classified as vowels) to be separated into A_p and U_p. Note the digraphs:

$$HI_c=O\theta_p \quad IS_c=\theta\theta_p \quad SV_c=\theta I_p$$

Only two alternatives are open:

(1) either $I_c=A_p$ and $S_c=U_p$,
(2) or $I_c=U_p$ and $S_c=A_p$.

If the first alternative is selected, then

$$HI_c=OA_p \quad SV_c=UI_p \quad IS_c=AU_p$$

If the second alternative is selected, then

$$HI_c=OU_p \quad SV_c=AI_p \quad IS_c=UA_p$$

The eye finds it difficult to choose between these alternatives; but suppose the frequency values of the plaintext diphthongs as given in Table 6 of Appendix 2 are added for each of these alternatives, giving the following:

$HI_c=OA_p$, frequency value= 7	$HI_c=OU_p$, frequency value=37
$SV_c=UI_p$, frequency value= 5	$SV_c=AI_p$, frequency value=17
$IS_c=AU_p$, frequency value=13	$IS_c=UA_p$, frequency value= 5
Total................25	Total................59

Mathematically, the second alternative appears to be more probable than the first.[8] Let it be assumed to be correct and the following (still tentative) values are now at hand:

$$Z_c=E_p \quad H_c=O_p \quad V_c=I_p \quad S_c=A_p \quad I_c=U_p$$

[8] A more accurate guide for choosing between the alternative groups of digraphs could be obtained through a consideration of the *logarithmic weights* of their assigned probabilities, rather than their plaintext *frequency* values. These weights are given in Appendix 2, along with an explanation of the method for their derivation; a detailed treatment of their application is presented in *Military Cryptanalytics, Part II.*

b. Attention is now directed to the letters classified as consonants: How far is it possible to ascertain their values? The letter D_c, from considerations of frequency alone, would seem to be T_p, but its frequency, 23, is not considerably greater than that for T_c. It is not much greater than that for F_c or G_c, with a frequency of 19 each. But perhaps it is possible to ascertain not the value of one letter alone but of two letters at one stroke. To do this one may make use of a tetragraph of considerable importance in English, *viz.*, $TION_p$. For if the analysis pertaining to the vowels is correct, and if $VH_c=IO_p$, then an examination of the letters immediately before and after the digraph VH_c in the cipher text might disclose both T_p and N_p. Reference to the text gives the following:

$$\begin{array}{cc} GVHT_c & FVHT_c \\ \theta IO\theta_p & \theta IO\theta_p \end{array}$$

The letter T_c follows VH_c in both cases and very probably indicates that $T_c=N_p$; but as to whether G_c or F_c equals T_p cannot be decided. However, two conclusions are clear: first, the letter D_c is neither T_p nor N_p, from which it follows that it must be either R_p or S_p; second, the letters G_c and F_c must be either T_p and S_p, respectively, or S_p and T_p, respectively, because the only tetragraphs usually found (in English) containing the diphthong IO_p as central letters are $SION_p$ and $TION_p$. This in turn means that as regards D_c, the latter cannot be *either* R_p or S_p; it *must* be R_p, a conclusion which is corroborated by the fact that ZD_c ($=ER_p$) and DZ_c ($=RE_p$) occur 9 times each. Thus far, then, the identifications, when inserted in an *enciphering* alphabet, are as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																			F	G						

46. Substituting deduced values in the cryptogram.—*a*. Thus far the analysis has been almost purely hypothetical, for as yet not a single one of the values deduced from the foregoing analysis has been tried out in the cryptogram. It is high time that this be done, because the final test of the validity of the hypotheses, assumptions, and identifications made in any cryptographic study is, after all, only this: do these hypotheses, assumptions, and identifications ultimately yield verifiable, intelligible plain text when *consistently* applied to the cipher text?

b. At the present stage in the process, since there are at hand the assumed values of but 9 out of the 25 letters that appear, it is obvious that a continuous "reading" of the cryptogram can certainly not be expected from a mere insertion of the values of the 9 letters. However, the substitution of these values should do two things. First, it should immediately disclose the fragments, outlines, or "skeletons" of "good" words in the text; and second, it should disclose no places in the text where "impossible" sequences of letters are established. By the first is meant that the partially deciphered text should show the outlines or skeletons of words such as may be expected to be found in the communication; this will become quite clear in the next subparagraph. By the second is meant that sequences, such as "AOOEN" or "TNRSENO" or the like, obviously not possible or extremely unusual in normal English text, must not result from the substitution of the tentative identifications resulting from the analysis. The appearance of several such extremely unusual or impossible sequences would at once signify that one or more of the assumed values is incorrect.

c. Here are the results of substituting the nine values which have been deduced by the reasoning based on a classification of the high-frequency letters into vowels and consonants and the study of the members of the two groups:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	10 S A	19 F T S	23 D R	35 Z E	19 F T S	10 I U	3 O	19 G S T	15 H O	5 L	5 P	35 Z E	19 F T S	19 G S T	35 Z E	23 D R	14 Y	10 S A	5 P	19 F T S	15 H O	4 B	35 Z E	23 D R	10 S A
B	19 G S T	16 V I	15 H O	22 T N	19 F T S	5 U	5 P	5 L	16 V I	23 D R	19 F T S	19 G S T	14 Y	16 V I	3 J	16 V I	19 F T S	16 V I	15 H O	22 T N	19 G S T	8 A	23 D R	35 Z E	35 Z E
C	8 A	10 I U	22 T N	14 Y	23 D R	35 Z E	14 Y	19 F T S	35 Z E	3 J	35 Z E	22 T N	19 G S T	5 P	22 T N	16 V I	22 T N	35 Z E	4 B	23 D R	16 V I	19 F T S	15 H O	22 T N	35 Z E
D	23 D R	19 F T S	8 X	10 S A	4 B	19 G S T	10 I U	23 D R	35 Z E	14 Y	16 V I	22 T N	8 X	3 O	10 I U	14 Y	16 V I	22 T N	3 E	19 F T S	16 V I	2 M	19 G S T	35 Z E	35 Z E
E	22 T N	15 H O	5 L	5 L	16 V I	8 X	35 Z E	23 D R	19 F T S	2 M	15 H O	22 T N	35 Z E	8 A	10 I U	22 T N	14 Y	23 D R	35 Z E	14 Y	4 B	23 D R	16 V I	19 F T S	15 H O
F	22 T N	35 Z E	23 D R	19 F T S	2 K	35 Z E	23 D R	35 Z E	35 Z E	3 J	10 S A	8 X	10 I U	10 S A	19 G S T	35 Z E	14 Y	19 G S T	8 A	16 V I	19 F T S	10 S A	5 L	19 G S T	35 Z E
G	23 D R	22 T N	15 H O	15 H O	22 T N	1 C	23 D R	35 Z E	2 R	10 S A	16 V I	22 T N	14 Y	35 Z E	23 D R	3 O	35 Z E	19 F T S	19 F T S	15 H O	22 T N	35 Z E	8 A	10 I U	22 T N
H	14 Y	23 D R	35 Z E	14 Y	19 G S T	8 A	16 V I	23 D R	19 G S T	35 Z E	35 Z E	22 T N	2 K	15 H O	10 I U	22 T N	14 Y	35 Z E	14 Y	10 S A	23 D R	35 Z E	19 G S T	15 H O	5 U
J	35 Z E	19 F T S	35 Z E	22 T N	19 G S T	5 U	5 P	19 G S T	23 D R	10 I U	8 X	1 W	24 G S T	15 H O	8 X	8 A	10 S A	2 R	5 U	35 Z E	23 D R	19 F T S	5 U	10 I U	23 D R
K	3 E	19 G S T	15 H O	22 T N	16 V I	3 E	8 A	19 G S T	8 X	8 X															

CONFIDENTIAL

d. No impossible sequences are brought to light, and, moreover, several long words, nearly complete, stand out in the text. Note the following portions:

```
          A21
          H B Z D S G V H T F
    (1)   O ? E R A S I O N T
                    T       S
          C15
          T V T Z B D V F H T Z D F
    (2)   N I N E ? R I T O N E R T
                        S         S
          F22
          S L G Z D T H H T
    (3)   A ? S E R N O O N
              T
```

The words are obviously OPERATIONS, NINE PRISONERS, and AFTERNOON. The value G_c is clearly T_p; that of F_c is S_p; and the following additional values are certain:

$$B_c = P_p \qquad L_c = F_p$$

47. Completing the solution.—*a.* Each time an additional value is obtained, substitution is at once made throughout the cryptogram. This leads to the determination of further values, in an ever-widening circle, until all the identifications are firmly and finally established, and the message is completely solved. In this case the decipherment is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	S	F	D	Z	F	I	O	G	H	L	P	Z	F	G	Z	D	Y	S	P	F	H	B	Z	D	S
	A	S	R	E	S	U	L	T	O	F	Y	E	S	T	E	R	D	A	Y	S	O	P	E	R	A
B	G	V	H	T	F	U	P	L	V	D	F	G	Y	V	J	V	F	V	H	T	G	A	D	Z	Z
	T	I	O	N	S	B	Y	F	I	R	S	T	D	I	V	I	S	I	O	N	T	H	R	E	E
C	A	I	T	Y	D	Z	Y	F	Z	J	Z	T	G	P	T	V	T	Z	B	D	V	F	H	T	Z
	H	U	N	D	R	E	D	S	E	V	E	N	T	Y	N	I	N	E	P	R	I	S	O	N	E
D	D	F	X	S	B	G	I	D	Z	Y	V	T	X	O	I	Y	V	T	E	F	V	M	G	Z	Z
	R	S	C	A	P	T	U	R	E	D	I	N	C	L	U	D	I	N	G	S	I	X	T	E	E
E	T	H	L	L	V	X	Z	D	F	M	H	T	Z	A	I	T	Y	D	Z	Y	B	D	V	F	H
	N	O	F	F	I	C	E	R	S	X	O	N	E	H	U	N	D	R	E	D	P	R	I	S	O
F	T	Z	D	F	K	Z	D	Z	Z	J	S	X	I	S	G	Z	Y	G	A	V	F	S	L	G	Z
	N	E	R	S	W	E	R	E	E	V	A	C	U	A	T	E	D	T	H	I	S	A	F	T	E
G	D	T	H	H	T	C	D	Z	R	S	V	T	Y	Z	D	O	Z	F	F	H	T	Z	A	I	T
	R	N	O	O	N	Q	R	E	M	A	I	N	D	E	R	L	E	S	S	O	N	E	H	U	N
H	Y	D	Z	Y	G	A	V	D	G	Z	Z	T	K	H	I	T	Y	Z	Y	S	D	Z	G	H	U
	D	R	E	D	T	H	I	R	T	E	E	N	W	O	U	N	D	E	D	A	R	E	T	O	B
J	Z	F	Z	T	G	U	P	G	D	I	X	W	G	H	X	A	S	R	U	Z	D	F	U	I	D
	E	S	E	N	T	B	Y	T	R	U	C	K	T	O	C	H	A	M	B	E	R	S	B	U	R
K	E	G	H	T	V	E	A	G	X	X															
	G	T	O	N	I	G	H	T	(X	X)															

CONFIDENTIAL

Message: AS RESULT OF YESTERDAYS OPERATIONS BY FIRST DIVISION THREE HUNDRED SEVENTY NINE PRISONERS CAPTURED INCLUDING SIXTEEN OFFICERS. ONE HUNDRED PRISONERS WERE EVACUATED THIS AFTERNOON, REMAINDER LESS ONE HUNDRED THIRTEEN WOUNDED ARE TO BE SENT BY TRUCK TO CHAMBERSBURG TONIGHT.

b. The solution should, as a rule, not be considered complete until an attempt has been made to discover all the elements underlying the general system and the specific key to a message. In this case, there is no need to delve further into the general system, for it is merely one of uniliteral substitution with a mixed cipher alphabet (with the convention that Q_p may be used to represent a comma and X_p may be used for a period). It is necessary or advisable, however, to reconstruct the cipher alphabet because this may give clues that later may become valuable.

c. Cipher alphabets should, as a rule, be reconstructed by the cryptanalyst in the form of *enciphering* alphabets because they will then usually be in the form in which the encipherer used them. This is important for two reasons. First, if the sequence in the cipher component gives evidence of system in its construction or if it yields clues pointing toward its derivation from a key word or a key phrase, this may often corroborate the identifications already made and may lead directly to additional identifications. A word or two of explanation is advisable here. For example, refer to the skeletonized enciphering alphabet given at the end of subpar. 45*b*:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																			F	G						

Suppose the crypanalyst, looking at the sequence DGFI or DFGI in the cipher component, suspects the presence of a keyword-mixed alphabet. Then DFGI is certainly a more plausible sequence than DGFI. Examining the skeleton cipher component more carefully, he notes that S . . . Z would allow for insertion of three of the missing letters UWXY since the letters T and V occur later, probably in the key word itself; further, he notes that the key word probably begins under F_p and ends in TH, making it probable that the TH is followed by AB, AC, or BC. This means that if $P_p=A_c$, Q_p=either B_c or C_c; but if $P_p=B_c$, then $Q_p=C_c$. Referring to the frequency distribution, he notes that C_c (with one occurrence) would make an excellent Q_p; however, either A_c (8 occurrences, or 3.4%) or B_c (4 occurrences, or 1.7%) might represent P_p in this single, isolated message. A trial of these values would materially hasten solution because it is often the case in cryptanalysis that if the value of a very low-frequency letter can be surely established it will yield clues to other values very quickly. Thus, if Q_p is definitely identified it almost invariably will identify U_p, and will give clues to the letter following the U_p, since it must be a vowel. For the foregoing reason an attempt should always be made in the early stages of the analysis to determine, if possible, the basis of construction or derivation of the cipher alphabet; as a rule this can be done only by means of the enciphering alphabet, and not the deciphering alphabet. For example, the skeletonized *deciphering* alphabet corresponding to the enciphering alphabet directly above is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain				R		T	S	O	U										A	N		I				E
						S	T																			

Here no evidences of a keyword-mixed alphabet are seen at all. However, if the enciphering alphabet has been examined and shows no evidences of systematic construction, the deciphering alphabet should then be examined with this in view, because occasionally it is the deciphering alphabet which shows the presence of a key or keying element, or which has been systematically derived from a word or phrase. The second reason why it is important to try to discover the basis of construction or derivation of the cipher alphabet is that it affords clues to the general type of key words or keying elements employed by the enemy. This is a psychological factor, of course, and may be of assistance in subsequent studies of his traffic. It merely gives a clue to the general type of thinking indulged in by certain of his cryptographers.

d. In the case of the foregoing solution, the complete enciphering alphabet is found to be as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	

Obviously, the letter Q, which is the only letter not appearing in the cryptogram, should follow P in the cipher component. Note now that the latter is based upon the key word LEAVENWORTH, and that this particular cipher alphabet has been composed by shifting the mixed sequence based upon this key word five intervals to the right so that the key for the message is $A_p=S_c$.[9] Note also that the deciphering alphabet fails to give any evidence of keyword construction based upon the word LEAVENWORTH.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	H	P	Q	R	G	S	T	O	U	V	W	F	X	J	L	Y	Z	M	A	N	B	I	K	C	D	E

e. If neither the enciphering nor the deciphering alphabet exhibits characteristics which give indication of derivation from a key word by some form of mixing or disarrangement, the use of such a key word for this purpose is nevertheless not finally excluded as a possibility. For the reconstruction of such mixed alphabets the cryptanalyst must use ingenuity and a knowledge of the more common methods of suppressing the appearance of key words in the mixed alphabets. Several of these methods are given detailed treatment in par. 51 below.

f. It is very important in practical cryptanalytic work to prepare a technical summary of the solution of a system.[10] Step-by-step commentaries should accompany an initial solution, especially those steps leading to the first plaintext entries; the steps taken should be jotted down *as they are made*, and at the end they should be combined into a complete résumé of the analysis. The résumé should be brief and concise, yet comprehensive enough that at any future time the solution may be reconstructed following the exact manner in which it was originally accomplished. Assumptions of words, etc., should be referred to with worksheet line- and column indicators, and should be couched in the proper cryptologic language or symbols. A short exposition of the mechanics of the general system, enciphering alphabets, enciphering diagrams, etc., as well as all key words (together with their derivation) and specific keys should be included. On the work sheet there should be a letter-for-letter decryptment *under* the cipher text [11]; the

[9] It is usual practice to employ as the specific key the equivalent of either A_p, or the equivalent of the first letter of the plain component when this component is a mixed sequence.

[10] For an illustration of a technical report, see par. 10 of Appendix 7.

[11] It is desirable to standardize work sheets where possible, since it lessens the chance of notations being misread by a cryptanalyst looking over the work of another. The particular reason for printing the plaintext recoveries *under* the cipher text is that this procedure permits the frequencies and other notations to be placed over the cipher letters.

final plaintext version should be in word lengths, with any errors or garbles corrected. Nulls or indicators showing sentence separation, change of key, etc., may be enclosed in parentheses. *All* work sheets and notes should be kept together with the solution.

48. General remarks on the foregoing solution.—*a.* The example solved above is admittedly a more or less artificial illustration of the steps in analysis, made so in order to demonstrate general principles. It was easy to solve because the frequencies of the various cipher letters corresponded quite well with the normal or expected frequencies. However, all cryptograms of the same monoalphabetic nature can be solved along the same general lines, after a certain amount of experimentation, depending upon the length of the cryptogram, and the skill and experience of the cryptanalyst.[12]

b. It is no cause for discouragement if the student's initial attempts to solve a cryptogram of this type require much more time and effort than were apparently required in solving the foregoing purely illustrative example. It is indeed rarely the case that *every* assumption made by the cryptanalyst proves in the end to have been correct; more often it is the case that a good many of his initial assumptions are incorrect, and that he loses much time in casting out the erroneous ones. The speed and facility with which this elimination process is conducted is in many cases all that distinguishes the expert from the novice.

c. Nor will the student always find that the initial classification into vowels and consonants can be accomplished as easily and quickly as was apparently the case in the illustrative example. The principles indicated are very general in their nature and applicability, and there are, in addition, some other principles that may be brought to bear in case of difficulty. Of these, perhaps the most useful are the following:

(1) In normal English it is unusual to find more than two consonants in succession, each of high frequency. If in a cryptogram a succession of three or four letters of high-frequency appear in succession, it is practically certain that at least one of these represents a vowel.[13]

(2) Successions of three vowels are rather unusual in English.[14] Practically the only time this happens is when a word ends in two vowels and the next word begins with a vowel.[15]

(3) When two letters already classified as vowel-equivalents are separated by a sequence

---

[12] The use of simple substitution in modern military operations is exceedingly rare because of the ease of solution. However, such cases have occurred, and one rather illuminating instance may be cited. In an important communication on 5 August 1918, General Kress von Kressenstein used a single mixed alphabet, and the intercepted radio message was solved at American GHQ very speedily. A day later another message, but in a very much more difficult cipher system, was intercepted and solved. When translated, it read as follows:

"GHQ Kress:

The cipher prepared by General von Kress was at once solved here. Its further use and employment is forbidden.

Chief Signal Officer, Berlin."

[13] Sequences of as many as eight consonants are not impossible, however, as in STRENGTHS THROUGH.

[14] Note that the word RADIOED, past tense of the verb RADIO, is in use.

[15] A sequence of seven vowels is not impossible, however, as in THE WAY YOU EARN

of six or more letters, it is either the case that one of the supposed vowel-equivalents is incorrect, or else that one or more of the intermediate letters is a vowel-equivalent.[16]

(4) Reference to Table 7–B of Appendix 2 discloses the following:

Distribution of first 18 digraphs forming 25 per cent of English text

Number of consonant-consonant digraphs	4
Number of consonant-vowel digraphs	6
Number of vowel-consonant digraphs	8
Number of vowel-vowel digraphs	0

Distribution of first 53 digraphs forming 50 per cent of English text

Number of consonant-consonant digraphs	8
Number of consonant-vowel digraphs	23
Number of vowel-consonant digraphs	18
Number of vowel-vowel digraphs	4

The latter tabulation shows that of the first 53 digraphs which form 50 per cent of English text, 41 of them, that is, over 75 per cent, are combinations of a vowel with a consonant. In short, in normal English the vowels and the high-frequency consonants are in the long run distributed fairly evenly and regularly throughout the text.

(5) As a rule, repetitions of trigraphs in the cipher text are composed of high-frequency letters forming high-frequency combinations. The latter practically always contain at least one vowel; in fact, if reference is made to Table 10–A of Appendix 2 it will be noted that 36 of the 56 trigraphs having a frequency of 100 or more contain one vowel, 17 of them contain two vowels, and only three of them contain no vowel. In the case of tetragraphic repetitions, Table 11–A of Appendix 2 shows that no tetragraph listed therein fails to contain at least one vowel; 27 of them contain one vowel, 25 contain two vowels, and 2 contain three vowels.

(6) Quite frequently when two known vowel-equivalents are separated by six or more letters none of which seems to be of sufficiently high frequency to represent one of the vowels A E I O, the chances are good that the cipher-equivalent of the vowel U or Y is present.

d. Another method for the determination of vowels which is of especial importance in a difficult case of monoalphabetic substitution, is that known as the *consonant-line method.* The fact that there is a very strong tendency in English for low-frequency consonants to be flanked on one or both sides by vowels is exploited in this method. If a distribution is made of the contacts of the low-frequency ciphertext letters in a monoalphabetic cryptogram, one or more vowel-equivalents should be identifiable by its high occurrence on both sides of the "consonant-line"

[16] Some cryptanalysts place a good deal of emphasis upon this principle as a method of locating the remaining vowels after the first two or three have been located. They recommend that the latter be marked throughout the text and then all sequences of five or more letters showing no marks be studied attentively. Certain letters which occur in several such sequences are sure to be vowels. An arithmetical aid in the study is as follows: Take a letter thought to be a good possibility as the cipher equivalent of a vowel (hereafter termed a *possible vowel-equivalent*) and find the length of each interval from the possible vowel-equivalent to the next *known* (fairly surely determined) vowel-equivalent. Multiply the interval by the number of times this interval is found. Add the products and divide by the total number of intervals considered. This will give the *mean* interval for that possible vowel-equivalent. Do the same for all the other possible vowel-equivalents. The one for which the mean is the greatest is most probably a vowel-equivalent. Mark this letter throughout the text and repeat the process for locating additional vowel-equivalents, if any remain to be located. One convention used for marking vowel-equivalents is to place a *red* dot over these letters; a *blue* dot is reserved for consonant equivalents, when so identified.

diagram. As an example, the consonant-line diagram for the distribution in Fig. 14 is given below. (The letters above the horizontal line are the lowest-frequency cipher letters, i. e., in this case, those letters with a frequency of 4 or less. The letters to the left of the vertical line are those which occurred as prefixes of the low-frequency cipher letters, while the letters to the right of the line are the suffixes of those letters.)

```
        4 1 3 3 2 2 3 2 1
        B C E J K M O R W
        -----------------
                Y |
              D D | D D D
              S S | S S
                  | G G G G G
          Z Z Z Z | Z Z Z Z
                H | H H
            T T T |
            V V V | V
                  | A
              F F | F
              X X |
                I | I
                  | U
```

From this diagram it is easy to see that Z_c in all likelihood is a plaintext vowel-equivalent, and that D_c and S_c are probable vowel-equivalents; furthermore, H_c, V_c, F_c, and I_c are possible vowel-equivalents. (Actually, Z_c, S_c, H_c, V_c, and I_c are vowel-equivalents.)

e. To recapitulate the general principles, vowels may then be distinguished from consonants in that they are usually represented by:

(1) high-frequency letters;
(2) high-frequency letters which do not readily contact each other;
(3) high-frequency letters which have a great variety of contact;
(4) high-frequency letters which have an affinity for low-frequency letters (i. e., low-frequency plaintext consonants).

f. In the foregoing example the amount of experimentation or "cutting and fitting" was practically nil. (This is not true of real cases as a rule.) Where such experimentation is necessary, the underscoring of all repetitions of several letters is very essential, as it calls attention to peculiarities of structure that often yield clues.

g. After a few basic assumptions of values have been made, if short words or skeletons of words do not become manifest, it is necessary to make further assumptions for unidentified letters. This is accomplished most often by assuming a word.[17] Now there are two places in

[17] This process does not involve anything more mysterious than ordinary, logical reasoning; there is nothing of the subnormal or supernormal about it. If cryptanalytic success seems to require processes akin to those of medieval magic, if "hocus-pocus" is much to the fore, the student should begin to look for items that the claimant of such success has carefully hidden from view for the mystification of the uninitiated. If the student were to adopt as his personal motto for all his cryptanalytic ventures the quotation (from Tennyson's poem *Columbus*) appearing on the back of the title page of this text, he will frequently find short cuts to his destination and will not too often be led astray!

every message which lend themselves more readily to successful attack by the assumption of words than do any other places—the very beginning and the very end of the message. The reason is quite obvious, for although words may begin or end with almost any letter of the alphabet, they usually begin and end with but a few very common digraphs and trigraphs. Very often the association of letters in peculiar combinations will enable the student to note where one word ends and the next begins. For example, suppose E, N, S, and T have been definitely identified, and a sequence like the following is found in a cryptogram:

. . . E N T S N E . . .

Obviously the break between two words should fall either after the S of E N T S or after the T of E N T, so that two possibilities are offered: . . . E N T S / N E . . ., or . . . E N T / S N E . . .. Since in English there are very few words with the initial trigraph S N E, it is most likely that the proper division is . . . E N T S / N E . . .. Of course, when several word divisions have been found, the solution is more readily achieved because of the greater ease with which assumptions of additional new values may be made.

h. Although a considerable amount of detailed treatment has been devoted to vowel-consonant analysis, it is felt advisable again to caution the student against the natural tendency to accept without question the results of any one cryptanalytic technique exclusively, even one such as vowel-consonant analysis which seems quite scientific in character.

49. The "probable-word" method; its value and applicability.—*a.* In practically all cryptanalytic studies, short cuts can often be made by assuming the presence of certain words in the message under study. Some writers attach so much value to this kind of an "attack from the rear" that they practically elevate it to the position of a method and call it the "intuitive method" or the "probable-word method." It is, of course, merely a refinement of what in everyday language is called "assuming" or "guessing" a word in the message. The value of making a "good guess" can hardly be overestimated, and the cryptanalyst should never feel that he is accomplishing a solution by an illegitimate subterfuge when he has made a fortunate guess leading to solution. A correct assumption as to plain text will often save hours or days of labor, and sometimes there is no alternative but to try to "guess a word", for occasionally a system is encountered the solution of which is absolutely dependent upon this artifice.

b. The expression "good guess" is used advisedly. For it is "good" in two respects. First, the cryptanalyst must use care in making his assumptions as to plaintext words. In this he must be guided by extraneous circumstances leading to the assumption of *probable* words—not just any words that come to his mind. Therefore he must use his imagination but he must nevertheless carefully control it by the exercise of *good* judgment. Second, only if the "guess" is correct and leads to solution, or at least puts him on the road to solution, is it a good guess. But, while realizing the usefulness and the time- and labor-saving features of a solution by assuming a probable word, the cryptanalyst should exercise discretion in regard to how long he may continue in his efforts with this method. Sometimes he may actually waste time by adhering to the method too long, if straightforward, methodical analysis will yield results more quickly.

c. Obviously, the "probable-word" method has much more applicability when working upon material the general nature of which is known, than when working upon more or less isolated communications exchanged between correspondents concerning whom or whose activities nothing is known. For in the latter case there is little or nothing that the imagination can seize upon as a background or basis for the assumptions.[18] However, in the case of military cryptanalysis in time of active operations there is, indeed, so great a probability that certain words and expressions are present in certain cryptograms that those words and expressions ("clichés") are often referred to as "cribs" (as defined in Webster's New Collegiate Dictionary: ". . . a plagiarism; hence, a translation, etc., to aid a student in reciting."). The cryptanalyst is quite sure they are present in the cryptogram under examination—what he must do is to "fit the crib to the text", that is, locate it in the cipher text.

d. Very frequently, the choice of probable words is aided or limited by the number and positions of repeated letters. These repetitions may be *patent*—that is, externally visible in the cryptographic text as it originally stands—or they may be *latent*—that is, externally invisible but susceptible of being made patent as a result of the analysis. For example, in a monoalphabetic substitution cipher, such as that discussed in the preceding paragraph, the repeated letters are directly exhibited in the cryptogram; later the student will encounter many cases in which the repetitions are latent, but are made patent by the analytical process. When the repetitions are patent, then the *pattern* or *formula* to which the repeated letters conform is of direct use in assuming plaintext words; and when the text is in word lengths, the pattern is obviously of even greater assistance. Suppose the cryptanalyst is dealing with military text, in which case he may expect such words as DIVISION, BATTALION, etc., to be present in the text. The positions of the repeated letter I in DIVISION, of the reversible digraph AT, TA in BATTALION, and so on, constitute for the experienced cryptanalyst telltale indications of the presence of these words, even when the text is not divided up into its original word lengths.

e. The important aid that a study of word patterns can afford in cryptanalysis warrants the use of definite terminology and the establishment of certain data having a bearing thereon. The phenomenon herein under discussion, namely, that many words are of such construction as regards the number and positions of repeated letters as to make them readily identifiable, will be termed *idiomorphism* (from the Greek "idios"=one's own, individual, peculiar+"morphe"=form). Words which show this phenomenon will be termed *idiomorphic*. It will be useful to deal with the idiomorphisms symbolically and systematically as described below.

f. The most usual practice in designating idiomorphic patterns and classifying them into systematic lists is to assign a literal nomenclature to that portion of a word (or sequence of plaintext letters) which contains the distinctive pattern, beginning with the first letter which is repeated in the pattern and ending with the last letter which is repeated in the pattern. Thus, the word DIVISION would be termed an idiomorph of the *abaca* class (based on the sequence IVISI contained therein), and the word BATTALION as an idiomorph of the *abba* class (based on the sequence ATTA). In Appendix 3 will be found a compendium of the more frequent military words in English, arranged according to word lengths in alphabetical order and in rhyming order;

[18] General Givierge in his *Cours de Cryptographie* (p. 121) says: "However, expert cryptanalysts often employ such details as are cited above [in connection with assuming the presence of 'probable words'], and the experience of the years 1914 to 1918, to cite only those, proves that in practice one often has at his disposal elements of this nature, permitting assumptions much more audacious than those which served for the analysis of the last example. The reader would therefore be wrong in imagining that such fortuitous elements are encountered only in cryptographic works where the author deciphers a document that he himself enciphered. Cryptographic correspondence, if it is extensive, and if sufficiently numerous working data are at hand, often furnishes elements so complete that an author would not dare use all of them in solving a problem for fear of being accused of obvious exaggeration."

in addition, there will be found in this appendix a listing of idiomorphs arranged first according to pattern and then according to the first letter of the idiomorphic sequence.[19]

50. Solution of additional cryptograms produced by the same components.—*a.* To return, after a rather long digression, to the cryptogram solved in pars. 44–47, once the components of a cipher alphabet have been reconstructed, subsequent messages which have been enciphered by means of the same components may be solved very readily, and without recourse to the principles of frequency, or application of the probable-word method. It has been seen that the illustrative cryptogram treated in pars. 41–47 was enciphered by juxtaposing the cipher component against the normal sequence so that $A_p=S_c$. It is obvious that the cipher component may be set against the plain component at any one of 26 different points of coincidence, each yielding a different cipher alphabet. After the components have been reconstructed, however, they become *known* sequences and the method of converting the cipher letters into their plain-component equivalents and then completing the plain-component sequence[20] begun by each equivalent can be applied to solve any cryptogram which has been enciphered by these components.

b. An example will serve to make the process clear. Suppose the following message, passing between the same two stations as before, was intercepted shortly after the first message had been solved:

IYEWK CERNW OFOSE LFOOH EAZXX

It is assumed that the same components were used, but with a different key letter. First the initial two groups are converted into their plain-component equivalents by setting the cipher component against the plain component at any arbitrary point of coincidence. The initial letter of the former may as well be set against A of the latter, with the following result:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z

Cryptogram	I Y E W K	C E R N W	. . .
Equivalents	P Y B F R	L B H E F	. . .

The plain component sequence initiated by each of these conversion equivalents is now completed, with the results shown in Fig 15. Note the plaintext generatrix, CLOSEYOURS, which manifests itself without further analysis. The rest of the message may be read either by continuing the same process, or, what is even more simple, the key letter of the message may now be determined quite readily and the message deciphered by its means.

---

[19] When dealing with cryptograms in which the word lengths are determined or specifically shown, it might be convenient to indicate their lengths and their patterns in a slightly modified form, such as is illustrated below:

3/aba: DID, EVE, EYE, etc.
abb: ADD, ALL, ILL, OFF, etc.
4/abac: ARAB, AWAY, etc.
abbc: ALLY, BEEN, etc.
abca: AREA, BOMB, DEAD, etc.
abcb: ANON, CEDE, etc.
etc. etc.

[20] It must be noted that if the plain component is a *mixed* sequence, then it is this mixed sequence which must be used to complete the columns.

I	Y	E	W	K	C	E	R	N	W
P	Y	B	F	R	L	B	H	E	F
Q	Z	C	G	S	M	C	I	F	G
R	A	D	H	T	N	D	J	G	H
S	B	E	I	U	O	E	K	H	I
T	C	F	J	V	P	F	L	I	J
U	D	G	K	W	Q	G	M	J	K
V	E	H	L	X	R	H	N	K	L
W	F	I	M	Y	S	I	O	L	M
X	G	J	N	Z	T	J	P	M	N
Y	H	K	O	A	U	K	Q	N	O
Z	I	L	P	B	V	L	R	O	P
A	J	M	Q	C	W	M	S	P	Q
B	K	N	R	D	X	N	T	Q	R
*C	L	O	S	E	Y	O	U	R	S
D	M	P	T	F	Z	P	V	S	T
E	N	Q	U	G	A	Q	W	T	U
F	O	R	V	H	B	R	X	U	V
G	P	S	W	I	C	S	Y	V	W
H	Q	T	X	J	D	T	Z	W	X
I	R	U	Y	K	E	U	A	X	Y
J	S	V	Z	L	F	V	B	Y	Z
K	T	W	A	M	G	W	C	Z	A
L	U	X	B	N	H	X	D	A	B
M	V	Y	C	O	I	Y	E	B	C
N	W	Z	D	P	J	Z	F	C	D
O	X	A	E	Q	K	A	G	D	E

FIGURE 15.

c. In order that the student may understand without question just what is involved in the latter step, that is, discovering the key letter after the first two or three groups have been deciphered by the conversion-completion process, the foregoing example will be used. It was noted that the first cipher group was finally deciphered as follows:

Cipher........... I Y E W K
Plain............ C L O S E

Now set the cipher component against the normal sequence so that $C_p = I_c$. Thus:

Plain............ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher........... F G I J K M P Q S U X Y Z L E A V N W O R T H B C D

It is seen here that when $C_p = I_c$ then $A_p = F_c$. This is the key for the entire message. The decipherment may be completed by direct reference to the cipher alphabet. Thus:

Cipher........... I Y E W K   C E R N W   O F O S E   L F O O H   E A Z X X
Plain............ C L O S E   Y O U R S   T A T I O   N A T T W   O P M(X X)

Message: CLOSE YOUR STATION AT TWO PM

d. The student should make sure that he understands the fundamental principles involved in this quick solution, for they are among the most important principles in cryptanalytics. How useful they are will become clear as he progresses into more and more complex cryptanalytic studies.

e. It must be kept in mind that there are *four* ways that two basic sequences may be used to form a cipher alphabet, subject to the instructions guiding the cryptographer in the use of his cryptosystem; this fact must be considered when additional cryptograms appear in a particular cryptosystem for which the primary components have been recovered. Assuming that the sequences just recovered are labelled "A" and "B", then the following contingencies might arise in the encryption of subsequent messages:

(1) "A" direct for the plain component, and "B" direct for the cipher component (as in the original recovery);

(2) "A" direct for the plain, and "B" reversed for the cipher;

(3) "B" direct for the plain, and "A" direct for the cipher; and

(4) "B" direct for the plain, and "A" reversed for the cipher.

51. Recovery of key words.—*a.* Concurrent with the solution of a cryptogram, there should be a simultaneous effort in the reconstruction of cipher alphabets and recovery of key words. Much labor can thus be saved as recovery of the keys early in the stages of solution may transform the process of cryptanalysis into one of decipherment.

b. A mixed cipher alphabet falls into one of five categories, according to the composition of its components, *viz.*,

(1) the plain component is the normal sequence and the cipher component is mixed;

(2) the cipher component is the normal sequence and the plain component is mixed;

(3) both components are the same mixed sequence;

(4) both components are the same mixed sequence, but running in reverse; or

(5) the components are different mixed sequences.

c. Let us examine several types of mixed sequences, using the key word HYDRAULIC as an example. The ordinary keyword-mixed sequence produced from this key word is:

(1) H Y D R A U L I C B E F G J K M N O P Q S T V W X Z

The two principal transposition-mixed types based on this key word are derived from the diagram:

```
H Y D R A U L I C
B E F G J K M N O
P Q S T V W X Z      and read:
```

(2) *Simple columnar*

H B P Y E Q D F S R G T A J V U K W L M X I N Z C O and

(3) *Numerically-keyed columnar*

A J V C O D F S H B P I N Z L M X R G T U K W Y E Q

Other types may arise from various types of route transpositions such as the following, using the foregoing diagram:

(4) *Alternate vertical*

H B P Q E Y D F S T G R A J V W K U L M X Z N I C O

(5) *Alternate diagonal*

H Y B P E D R F Q S G A U J T V K L I M W X N C O Z

(6) *Simple diagonal*

P B Q H E S Y F T D G V R J W A K X U M Z L N I O C

(7) *Alternate horizontal*

H Y D R A U L I C O N M K J G F E B P Q S T V W X Z

(8) *Spiral counterclockwise*

O C I L U A R D Y H B P Q S T V W X Z N M K J G F E

Still other types are possible from the foregoing diagram which do not follow a simple, clear-cut route, such as the following:

(9) H Y E B P Q S T G F D R A U K J V W X Z N M L I C O
(10) C P I O Q B L N S E H U M Z T F Y A K X V G D R J W

Any transposition system may be employed to produce a systematically-mixed sequence; practicability of method is the only determining factor. It must be remembered that the greatest amount of systematic mixing will produce a sequence inherently no more secure than a random-mixed alphabet.

d. The student would do well to construct both enciphering and deciphering versions of cipher alphabets recovered, as has been previously mentioned. For example, in the following case

Plain: J Q N M F H L E B R S K G Y Z O T I C D U V A W P X
Cipher: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

no semblance of a key is apparent; but in the inverse form

Plain: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher: W I S T H E M F R A L G D C P Y B J K Q U V X Z N O

the key phrase "NOW IS THE TIME FOR ALL GOOD MEN TO COME TO THE AID OF THEIR PARTY" is quite clear. In other types of mixed sequences, first the one form is attacked, and then if negative results are obtained the inverse form is treated.

e. Let us consider the following cipher alphabet:

P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: D W Z M S O C R Y A T X B E F U G Q H I V J K L N P

The section $\begin{smallmatrix} V & W & X \\ J & K & L \end{smallmatrix}$ seems to comprise superimposed parts of the non-keyword portions of mixed sequences. Adding Y Z to the plain component, we get $\begin{smallmatrix} V & W & X & Y & Z \\ J & K & L & N & P \end{smallmatrix}$ which is certainly consistent as far as alphabetical progression goes, and indicates that the letters M and O are present in the key word of the cipher component. Continuing in this vein, the section

M	N	O	Q	S	T	V	W	X	Y	Z
B	E	F	G	H	I	J	K	L	N	P

is rapidly established by correlating both sequences. It is obvious that the plain component key word begins right after the Z, and that the cipher component key word probably just precedes the B. Going to the right, $\begin{smallmatrix} Z & R & H \\ P & Q & R \end{smallmatrix}$ suggests key words like RHOMBOID, RHEUMATISM etc. These trials are quickly repudiated; therefore we go on to $\begin{smallmatrix} Z & R & E \\ P & Q & S \end{smallmatrix}$ which is acceptable. $\begin{smallmatrix} Z & R & E & K \\ P & Q & S & T \end{smallmatrix}$ is found wanting, but $\begin{smallmatrix} Z & R & E & P \\ P & Q & S & U \end{smallmatrix}$ is very satisfactory, and this is soon expanded to $\begin{smallmatrix} Z & R & E & P & U & B & L & I & C \\ P & Q & S & U & V & W & X & Y & Z \end{smallmatrix}$, and in a moment or two we recover the complete cipher alphabet:

P:	R	E	P	U	B	L	I	C	A	N	D	F	G	H	J	K	M	O	Q	S	T	V	W	X	Y	Z
C:	Q	S	U	V	W	X	Y	Z	D	E	M	O	C	R	A	T	B	F	G	H	I	J	K	L	N	P

f. In the example below the student will observe that the alphabets are reciprocal: this is an indication that identical sequences at a shift of 13 have been employed, or that a mixed sequence is running against itself in reverse. In this case the $\begin{smallmatrix} W & X & Y & Z \\ Z & Y & X & W \end{smallmatrix}$ points to the latter hypothesis.

P:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
C:	H	O	J	F	T	D	N	A	K	C	I	M	L	G	B	S	U	V	P	E	Q	R	Z	Y	X	W

Starting with the $\begin{smallmatrix} V & W & X & Y & Z & R \\ R & Z & Y & X & W & V \end{smallmatrix}$ cluster, we see that the key word begins with the letter R; therefore the next letter should be a vowel. $\begin{smallmatrix} Z & R & A \\ W & V & H \end{smallmatrix}$ is not acceptable, but $\begin{smallmatrix} Z & R & E \\ W & V & T \end{smallmatrix}$ is fine, showing that the letter U appears in the key word. Continuing the same line of reasoning as in the preceding example, and with a little further experimentation, the final alphabet is discovered to be

P:	R	E	P	U	B	L	I	C	A	N	D	F	G	H	J	K	M	O	Q	S	T	V	W	X	Y	Z
C:	V	T	S	Q	O	M	K	J	H	G	F	D	N	A	C	I	L	B	U	P	E	R	Z	Y	X	W

g. In the next example, all efforts to derive key words on the basis of keyword-mixed sequences are fruitless. The conclusion is therefore drawn that this is a case of a transposition.

P:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
C:	A	C	S	E	J	Y	I	G	W	L	F	V	M	H	X	N	K	Z	P	B	Q	R	D	U	T	O

Considering the mechanics of the cryptography involved, and assuming for the time being that Z is at the bottom of the matrix and not in the key word, we start with the letters to the left of Z in the cipher component (or if this fails, with the letters to the right of Z), obtaining the column
N
K which is not incompatible if N is in the key word on the top row. If we place Y to the left of Z
Z

and build up *its* column, we get

```
E N
J K
Y Z
```

which is excellent. This is expanded into

```
I M E N
G H J K
W X Y Z
```

which quickly becomes

```
7 1 8 4 3 5 2 6 9
P A R L I M E N T
B C D F G H J K O
Q S U V W X Y Z
```

This last example was very easy because none of the letters V W X Y Z appeared in the key word; but other cases should hardly prove more difficult.

h. Two additional methods that have been encountered for deriving mixed sequences may be mentioned. One is a slight modification of the system in the preceding subparagraph, when the key word contains repeated letters:

```
1 8 7 3 4 9 5 2 6
C O M . I T . E .
A B D F G H J K L
N P Q R S U V W X
Y Z
```

which produces the mixed sequence:

C A N Y E K W F R I G S J V L X M D Q O B P Z T H U

The other method is an interrupted-key columnar transposition system:[21]

```
5 1 3 4 2 6
V A L . E Y
B C)
D F G H I)
J K M)
N O P Q)
R)
S T U W X Z)
```

which produces the mixed sequence:

A C F K O T E I X L G M P U H Q W V B D J N R S Y Z

The first example will succumb to the treatment outlined in subpar. *g*, whereas the second method is vulnerable owing to the presence of the fragments D J N, F K O, and G M P in the sequence

[21] It is to be noted that in this particular case the numerical key serves two purposes: (1) determining the cut-off point (and therefore the number of letters) in each *row* of the diagram, after the appearance of the key word; and (2) determining the order of transcription of the *columns*.

which may be anagrammed. Note the fair-sized fragment B D J N R S, composed of an ascending sequence of letters; this is an outward manifestation of the interrupted-key columnar method.

i. There are still other methods used for the production of mixed sequences, but space does not permit giving further examples. However, the student should by this time be able to devise methods of attack for any special cases that may present themselves, based upon the cryptanalytically exploitable weaknesses or peculiarities inherent in the system of cryptography involved.

CHAPTER VII

MULTILITERAL SUBSTITUTION WITH SINGLE-EQUIVALENT CIPHER ALPHABETS

52. General types of multiliteral cipher alphabets.—*a.* Monoalphabetic substitution methods in general may be classified into uniliteral and multiliteral systems. In the former there is a strict "one-to-one" correspondence between the length of the units of the plain and those of the cipher text; that is, each letter of the plain text is replaced by a single character in the cipher text. In the latter this correspondence is no longer $1_p:1_c$ but may be $1_p:2_c$, where each letter of the plain text is replaced by a combination of two characters in the cipher text; or $1_p:3_c$, where a three-character combination in the cipher text represents a single letter of the plain text, and so on. A cipher in which the correspondence is of the $1_p:1_c$ type is termed *uniliteral* in character; one in which it is of the $1_p:2_c$ type, *biliteral;* $1_p:3_c$, *triliteral,* and so on. Ciphers in which one plaintext letter is represented by cipher characters of two or more elements are classed as multiliteral.[1]

b. Biliteral alphabets are usually composed of a set of 25 or 26 combinations of a limited number of characters taken in pairs. An example of such an alphabet is the following:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M
Cipher	WW	WH	WI	WT	WE	HW	HH	HI	HT	HT	HE	IW	IH
Plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	II	IT	IE	TW	TH	TI	TT	TE	EW	EH	EI	ET	EE

This alphabet is derived from the *cipher square* or *matrix* shown in Fig. 16. The cipher equivalent of each plaintext element is made up of two letters from outside the cipher matrix, one letter being the letter beside the row, the other being the letter above the column in which the plaintext letter is located. In other words, the letters at the side and top of the matrix have been used to designate, according to a coordinate system, the cell occupied by each letter within the matrix. The letters (or figures) at the side and top of the matrix are termed *row and column coordinates,* respectively, or row and column *indicators.*

[1] The terms uniliteral and multiliteral, although originally applied only to cipher text composed of letters, are used here in their broader sense to embrace cipher text in letters, digits, and even other symbols. In more precise terminology, these terms would probably be *monosymbolic* and *polysymbolic,* respectively, but the terms uniliteral and multiliteral are too well established in literature to be changed at this late time.

θ_p^1 \ θ_p^2	W	H	I	T	E
W	A	B	C	D	E
H	F	G	H	IJ	K
I	L	M	N	O	P
T	Q	R	S	T	U
E	V	W	X	Y	Z

FIGURE 16.

c. If a message is enciphered by means of the foregoing biliteral alphabet, the cryptogram is still monoalphabetic in character. A frequency distribution based upon pairs of letters will obviously have all the characteristics of a simple, uniliteral distribution for a monoalphabetic substitution cipher.

d. The cipher alphabets shown thus far in this text have involved only letters, but alphabets in which the cipher component consists of figures, or groups of figures, are not uncommon in military cryptography.[2] Since there are but 10 digits it is obvious that, in order to represent an alphabet of more than 10 characters by means of figure ciphers, combinations of at least two digits are necessary. The simplest kind of such an alphabet is that in which A_p=01, B_p=02, . . . Z_p=26; that is, one in which the plaintext letters have as their equivalents two-digit numbers indicating their positions in the normal alphabet.

e. Instead of a simple alphabet of the preceding type, it is possible to use a diagram of the type shown in Fig. 17. In this cipher the letter A_p is represented by the *dinome*[3] **11**, B_p by the dinome 12, etc. Furthermore, this matrix includes provision for the encipherment of some of the frequently-used punctuation marks in addition to the 26 letters.

	1	2	3	4	5	6	7	8	9	Ø
1	A	B	C	D	E	F	G	H	I	J
2	K	L	M	N	O	P	Q	R	S	T
3	U	V	W	X	Y	Z	.	,	:	;

FIGURE 17.

f. Other types of biliteral cipher alphabets are illustrated in the examples below:

	5	6	7	8	9	Ø
1	A	B	C	D	E	F
2	G	H	IJ	K	L	M
3	N	O	P	Q	R	S
4	T	UV	W	X	Y	Z

FIGURE 18.

	1	2	3	4	5	6	7	8	9
1	A	B	C	D	E	F	G	H	I
2	J	K	L	M	N	O	P	Q	R
3	S	T	U	V	W	X	Y	Z	*

FIGURE 19.

[2] Although, as an extension of this idea, cipher components employing signs and symbols are possible, such alphabets are not suitable for modern cryptography because they can be neither telegraphed nor telephoned with any degree of accuracy, speed, or facility.

[3] A pair of digits is called a *dinome;* similarly, a *trinome* is a set of three digits; a *tetranome,* a set of four digits; etc. Although a single digit would properly be termed a mononome, for the sake of euphony it is shortened into the term *monome.*

	M	U	N	I	C	H
B	A	7	E	5	R	M
E	G	1	N	Y	B	2
R	C	3	D	4	F	6
L	H	8	I	9	J	Ø
I	K	L	O	P	Q	S
N	T	U	V	W	X	Z

FIGURE 20.

	A	B	C	D	E	F	G	H	I
A	A	D	G	J	M	P	S	V	Y
B	B	E	H	K	N	Q	T	W	Z
C	C	F	I	L	O	R	U	X	1
D	2	3	4	5	6	7	8	9	Ø

FIGURE 21.

g. It is to be noted that in alphabets of the foregoing types, the row indicators may be distinct from the column indicators (e. g., Fig. 18), or they may not (e. g., Fig. 19); of course, when there is any duplication beween the row and column indicators, it is necessary to agree beforehand upon which indicator will be given as the first half of the equivalent for a letter, in order to avoid ambiguity. (In all of the systems described in this and subsequent sections of this text, the row indicator will always form the first part of an equivalent.) When letters are used as row and column indicators they may form a key word (e. g., Fig. 20), or they may not (e. g., Fig. 21); the key words, if formed, may be identical (e. g., Fig. 16) or different (e. g., Fig. 20). Furthermore, the plaintext letters may be arranged within the matrix as a mixed sequence (e. g., Fig. 20), either systematically- or random-mixed; and the matrix may contain, in addition to the letters of the alphabet, punctuation symbols (Fig. 17), numbers (Figs. 20, 21), etc., permitting their encipherment as such, instead of having to be spelled out. When the digits are included within a matrix they are usually inscribed in sequence (such as in Fig. 21), or in some systematic fashion (such as in Fig. 20, where A is followed by "1," B by "2," . . ., J by "Ø".

h. When letters are used as row and column indicators, they may be selected so as to result in producing cipher text that resembles artificial words; that is, words composed of alternate vowels and consonants. For example, if in Fig. 16 the row indicators consisted of the vowels A E I O U in this sequence from the top down, and the column indicators consisted of the consonants B C D F G in this sequence from left to right, the word RAIDS would be enciphered as OCABE FAFOD, which very closely resembles code of the type formerly called artificial code language. Such a system may be called a *false*, or *pseudo-code* system.[4]

i. As a weak type of subterfuge, ciphers which are essentially biliteral may involve a third character appended to the basic two-character cipher unit; this is done to "camouflage" the biliteral nature of the cipher text. This third character may be produced through the use of a cipher matrix of the type illustrated in Fig. 22 (wherein A_p=611, B_p=612 etc.); or the third character may be a "sum-checking" digit which is the *noncarrying sum* (i. e., the sum modulo 10)[5] of the preceding two digits, such as in the trinomes 25<u>7</u>, 83<u>1</u>, and 66<u>2</u>; or it may involve "self-summing" groups, such as the trinomes 254, 83Ø, and 669, all of which sum to a constant "1"; or it may merely be a randomly-selected character (inserted solely for the purpose of leading the cryptanalyst astray).

[4] Prior to 1934, international telegraph regulations required code words of five letters to contain at least one vowel and code words of ten letters to contain at least three vowels. The International Telegraph Conference held in Madrid in 1932 amended these regulations to permit the use of 5-letter code groups containing any combination of letters. These unrestricted code groups were authorized for use after 1 January 1934.

[5] The term *modulo* (abbreviated *mod*) pertains to a cyclic scale or basis of arithmetic; thus, in the *modulus* of 7, the numbers 8 and 15 are equivalent to 1, and 9 and 16 are equivalent to 2, etc.; or expressed differently, 8 mod 7 is 1, 9 mod 7 is 2. In cryptology, many operations are expressed mod 10 and mod 26.

	1	2	3	4	5
61	A	B	C	D	E
72	F	G	H	IJ	K
83	L	M	N	O	P
94	Q	R	S	T	U
05	V	W	X	Y	Z

FIGURE 22.

j. Another possibility that lends itself to certain multiliteral ciphers is the use of a *word spacer* or *word separator*. This word separator might be represented by a value in the matrix; i. e., the separator is *enciphered* (for instance, the dinome "39" in Fig. 19 might stand for a word separator). The word separator might instead be a *single* element not otherwise used in the cryptosystem; i. e., *unenciphered*, thus not giving rise to any possible ambiguity. Thus, in Fig. 19 the digit Ø and in Fig. 21 the letter J might be used as word separators, since no confusion would arise in decrypting.

k. The alphabets yielded by the matrices of Figs. 16–22 may also be termed *bipartite*, because the cipher units of these alphabets may be divided into two separate *parts* whose functions are clearly defined, *viz.*, row indicators and column indicators. As will be discussed later, this bipartite nature of most biliteral alphabets produced from cipher matrices constitutes one of the weaknesses of these alphabets which make them recognizable as such to a cryptanalyst. However, it is possible to employ a cipher matrix in a manner which will produce a biliteral alphabet *not* bipartite in character. For example, using the matrix of Fig. 23 one could produce the following biliteral cipher alphabet in which the equivalent for any letter in the matrix is the sum

	1	2	3	4	5
09	H	Y	D	R	A
15	U	L	IJ	C	B
21	E	F	G	K	M
27	N	O	P	Q	S
33	T	V	W	X	Z

FIGURE 23.

of the two coordinates which indicate its cell in the matrix:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M
Cipher	14	20	19	12	22	23	24	10	18	18	25	17	26

Plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	28	29	30	31	13	32	34	16	35	36	37	11	38

The cipher units of this alphabet are, of course, biliteral; but they are not bipartite. Note the equivalent of A_p, that is 14—if divided, it yields the digits 1 and 4 which have no meaning *per se*: plaintext letters whose cipher equivalents begin with 1 may be found in *two* different rows of the matrix, and those whose equivalents end in 4 appear in *three* different columns.

53. The Baconian and Trithemian ciphers.—*a.* An interesting example in which the cipher equivalents are five-letter groups and yet the resulting cipher is strictly monoalphabetic in character is found in the cipher system invented by Sir Francis Bacon (1561–1626) over 300 years ago. Despite its antiquity the system possesses certain features of merit which are well worth noting.[6] Bacon proposes the following 24-element cipher alphabet, composed of arrangements of five elements, each of which may be chosen from one of two categories:[7]

A=aaaaa	I–J=abaaa	R=baaaa
B=aaaab	K=abaab	S=baaab
C=aaaba	L=ababa	T=baaba
D=aaabb	M=ababb	U–V=baabb
E=aabaa	N=abbaa	W=babaa
F=aabab	O=abbab	X=babab
G=aabba	P=abbba	Y=babba
H=aabbb	Q=abbbb	Z=babbb

If this were all there were to Bacon's invention it would be hardly worth bringing to attention. But what he pointed out, with great clarity and simple examples, was how such an alphabet might be used to convey a secret message by enfolding it in an innocent, external message which might easily evade the strictest kind of censorship. As a very crude example, suppose that a message is written in capital and lower-case letters, any capital letter standing for an "a" element of the cipher alphabet, and any small letter, for a "b" element. Then the external sentence "All is well with me today" can be made to contain the secret message "Help."

Thus:

A L l i s	W E l L W	I t H m E	T o d a Y
a a b b b	a a b a a	a b a b a	a b b b a
H	E	L	P

Instead of employing a device so obvious as capital and small letters, suppose that an "A" element be indicated by a very slight shading, or a very slightly heavier stroke. Then a secret message might easily be thus enfolded within an external message of exactly opposite meaning. The number of possible variations of this basic scheme is very high. The fact that the characters of the cryptographic text are hidden in some manner or other has, however, no effect upon the strict monoalphabeticity of the scheme.

---

[6] For a true picture of this cipher, the explanation of which is often distorted beyond recognition even by cryptographers, see Bacon's own description of it as contained in his *De Augmentis Scientiarum* (*The Advancement of Learning*), as translated by any first class editor, such as Gilbert Watts (1640) or Ellis, Spedding, and Heath (1857, 1870). The student is cautioned, however, not to accept as true any alleged "decipherments" obtained by the application of Bacon's cipher to literary works of the 16th century. These readings are purely subjective.

[7] Bacon's alphabet was called by him a "biliteral alphabet" because it employs permutations of two letters. But from the cryptanalytic standpoint the significant point is that each plaintext letter is represented by a 5-character equivalent. Hence, present terminology requires that this alphabet be referred to as a *quinqueliteral alphabet.* Although the quinqueliteral alphabet affords 32 permutations, Bacon used only 24 of them, because in the 16th century the letters I and J, U and V were used interchangeably. Note the regularity of construction of Bacon's biliteral alphabet, a feature which easily permits its reconstruction from memory.

b. Another historical multiliteral cipher, sometimes attributed to the abbot Trithemius, born Johann von Heydenberg (1462–1516), is that incorporating a *triliteral alphabet*. Trithemius was said to have invented this alphabet for use in a fashion similar to Bacon's alphabet; i. e., as a means of disguise or cover for a secret text. This alphabet, modified to include the 26 letters of the present-day English alphabet, is shown in Fig. 23, below; it consists of all the permutations (with repetitions allowed) of three things taken three at a time, i. e., 3^3 or 27 in all.

A=111	D=121	G=131	J=211	M=221	P=231	S=311	V=321	Y=331
B=112	E=122	H=132	K=212	N=222	Q=232	T=312	W=322	Z=332
C=113	F=123	I=133	L=213	O=223	R=233	U=313	X=323	*=333

FIGURE 23.

The cipher text of course does not have to be restricted to digits; any groupings of three things taken three at a time will do.

54. Analysis of multiliteral, monoalphabetic substitution ciphers.—*a*. Biliteral ciphers and those of the other multiliteral (triliteral, quadriliteral, . . .) types are often readily detected externally by the fact that the cryptographic text is usually composed of but a very limited number of different characters. They are handled in exactly the same manner as are uniliteral, monoalphabetic substitution ciphers. So long as the same character, or combination of characters, is always used to represent the same plaintext letter, and so long as a given letter of the plain text is always represented by the same character or combination of characters, the substitution is strictly monoalphabetic and can be handled in the simple manner described in the preceding chapter of this text.

b. In the case of biliteral ciphers in which the row and column indicators are not identical, and the direction of reading the cipher pairs is chosen at will for each succeeding cipher pair, an analysis of the *contacts* of the letters comprising the cipher pairs will disclose that there are *two* distinct families of letters, and a cipher pair will never consist of two letters of the same family. With this fact discovered, the cipher may be quickly reduced to uniliteral terms and solved in the manner previously mentioned.

c. If a multiliteral cipher includes provision for the encipherment of a word separator, the cipher equivalent of this word separator may be readily identified because it will have the *highest* frequency of any cipher unit.[8] On the other hand, if the word separator is a *single* character (see subpar. 52*j*, on the use of the digit Ø and the letter J), this character may be identified throughout the encrypted text by its positional appearance spaced "wordlength-wise" in the cipher text, and by the fact that it never contacts itself. If this single character is used as a null indiscriminately throughout the cipher text, instead of as a word separator, the analysis is a bit more complicated but not as great as might be thought.

d. As a general rule, it is advisable to reduce multiliteral cipher text to uniliteral equivalents, especially if a triliteral frequency distribution is to be made. If not more than 36 different

[8] For English, since the average word length is 5.2 letters, the word separator will have a percentage frequency of 16%. The letters of the alphabet will now take on new percentage frequencies as follows:

A	6.2	F	2.3	K	0.25	O	6.3	S	5.1	W	1.3
B	0.84	G	1.3	L	3.0	P	2.3	T	7.7	X	0.41
C	2.6	H	2.9	M	2.1	Q	0.25	U	2.2	Y	1.6
D	3.5	I	6.2	N	6.6	R	6.4	V	1.3	Z	0.08
E	11.0	J	0.16								

combinations are present in a cryptogram, the extra values over 26 may be represented by digits for the purpose of this reduction. If, however, more than 36 different combinations are found in the encrypted text, it is usually not worth the trouble to attempt any uniliteral reduction, and the cipher text can be attacked in its multiliteral groupings.

e. As one of the first steps in the solution of any multiliteral cipher in letters which appears to involve the use of a cipher matrix, it is generally advisable to anagram the letters comprising the row and column indicators in an attempt to disclose any key words for these indicators. When the anagramming process *does* disclose such a key word or words, the next step is to make a skeleton reconstruction matrix which is a duplicate of the original enciphering matrix in that the indicators are arranged in the same order as on the original. Then, as plain text is recovered in the cryptogram by any of the methods outlined in the previous chapter of this text, the recovered plaintext letters should be inserted in the proper cells of the reconstruction matrix, so that any systematic arrangement of the plaintext letters, if present in the original, may be disclosed prior to recovery of the complete plain text. Furthermore, it may in some instances be found worthwhile, immediately after successfully uncovering the key words used as indicators, to make a frequency distribution of the particular cryptogram in the form of tally marks within the properly arranged frame of the reconstruction matrix, because a few moments' study of the *locations* of the crests and troughs in the distribution made in that form may, if the letters of the underlying plain component have been arranged in the normal sequence or in a keyword-mixed sequence (especially if it is related to the key words for the indicators), provide a basis for the recovery of this sequence at one stroke, without recourse to analysis of the cipher text.

55. Historically interesting examples.—*a.* Two examples of multiliteral ciphers of historical interest will be cited as illustrations. During the campaign for the presidential election of 1876 (Hayes vs. Tilden) many cipher messages were exchanged between the Tilden managers and their agents in several states where the voting was hotly contested. Two years later the New York Tribune [9] exposed many irregularities in the campaign by publishing the decipherments of many of these messages. These decipherments were achieved by two investigators employed by the Tribune, and the plain text of the messages seems to show that illegal attempts and measures to carry the election for Tilden were made by his managers. Here is one of the messages:

```
                                        JACKSONVILLE, Nov. 16 (1876).
GEO. F. RANEY, Tallahassee.
    P p y y e m n s n y y y p i m a s h n s y y s s i t e p a a e n s h n
s p e n s s h n s m m p i y y s n p p y e a a p i e i s s y e s h a i n s
s s p e e i y y s h n y n s s s y e p i a a n y i t n s s h y y s p y y p
i n s y y s s i t e m e i p i m m e i s s e i y y e i s s i t e i e p y y
p e e i a a s s i m a a y e s p n s y y i a n s s s e i s s m m p p n s p
i n s s n p i n s i m i m y y i t e m y y s s p e y y m m n s y y s s i t
s p y y p e e p p p m a a a y y p i i t
L'Engle goes up tomorrow.
                                                           DANIEL.
```

Examination of the message discloses that only ten different letters are used. It is probable, therefore, that what one has here is a cipher which employs a multiliteral alphabet. First

[9] New York Tribune, Extra No. 44, *The Cipher Dispatches*, New York, 1879.

assuming that the alphabet is one in which combinations of two letters represent single letters of the plain text, the message is rewritten in pairs and substitution of arbitrary letters for the pairs is made, as seen below:

PP	YY	EM	NS	NY	YY	PI	MA	SH	NS	YY	SS	etc.
A	B	C	D	E	B	F	G	H	D	B	I	etc.

A triliteral frequency distribution is then made and analysis of the message along the lines illustrated in the preceding chapter of this text yields solution, as follows:

Jacksonville, Nov. 16.

GEO. F. RANEY, Tallahassee:
Have Marble and Coyle telegraph for influential men from Delaware and Virginia. Indications of weakening here. Press advantage and watch Board. L'Engle goes up tomorrow.

DANIEL.

b. The other example, using numbers, is as follows:

Jacksonville, Nov. 17.

S. PASCO and E. M. L'ENGLE:

84	55	84	25	93	34	82	31	31	75	93	82	77	33	55	42
93	20	93	66	77	66	33	84	66	31	31	93	20	82	33	66
52	48	44	55	42	82	48	89	42	93	31	82	66	75	31	93

DANIEL.

There were, of course, several messages of like nature, and examination disclosed that only 26 different dinomes in all were used. Solution of these ciphers followed very easily, the decipherment of the one given above being as follows:

Jacksonville, Nov. 17.

S. PASCO and E. M. L'ENGLE:
Cocke will be ignored, Eagan called in. Authority reliable.

DANIEL.

c. The Tribune experts gave the following alphabets as the result of their decipherments:

AA=O	EN=Y	IT=D	NS=E	PP=H	SS=N
AI=U	EP=C	MA=B	NY=M	SH=L	YE=F
EI=I	IA=K	MM=G	PE=T	SN=P	YI=X
EM=V	IM=S	NN=J	PI=R	SP=W	YY=A
20=D	33=N	44=H	62=X	77=G	89=Y
25=K	34=W	48=T	66=A	82=I	93=E
27=S	39=P	52=U	68=F	84=C	96=M
31=L	42=R	55=O	75=B	87=V	99=J

They did not attempt to correlate these alphabets, or at least they say nothing about a possible relationship. The author [W. F. F.] has, however, reconstructed the square upon which these alphabets are based, and it is given below (Fig. 24).

1st Letter or Number \ 2d Letter or Number	H 1	I 2	S 3	P 4	A 5	Y 6	M 7	E 8	N 9	T 0
H 1										
I 2					K		S			D
S 3	L		N	W					P	
P 4		R		H				T		
A 5		U			O					
Y 6		X				A		F		
M 7					B		G			
E 8		I		C			V		Y	
N 9			E			M			J	
T 0										

FIGURE 24.

It is amusing to note that the conspirators selected as their key a phrase quite in keeping with their attempted illegalities—HIS PAYMENT—for bribery seems to have played a considerable part in that campaign. The blank cells in the matrix probably contained proper names, numbers, etc.

56. The international (Baudot) teleprinter code.—*a*. Modern printing telegraph systems,[10] or *teleprinter systems* as they are more often called, make use of a five-unit code[11] or alphabet which is similar to the Baconian alphabet treated in par. 53. The teleprinter alphabet is composed of all the possible permutations (with repetitions allowed) of five elements, each of which may be chosen from one of two categories, making it possible to obtain 32 different permutations, 26 of which are assigned to the letters of the alphabet, leaving 1 for an "idle condition" and 5 for certain printer operations called *functions*, such as "space," "figure shift," "letter shift," etc.

[10] Such systems are characterized by the transmission and *reception printing* of messages by electrical means, incorporating two electrically-connected instruments resembling typewriters. When a key of the keyboard on the transmitting instrument is depressed, an electrical signal is transmitted to the receiving instrument, causing the corresponding character to be printed therein. Usually the message is printed at the local as well as the distant station. The system has been adapted to radio as well as wire and overseas cable transmission.

[11] The five-unit code was first applied to teleprinter systems by Jean Maurice Emile Baudot (1845–1903), and is commonly known as the Baudot Code. It is worthwhile to point out that Baudot apparently constructed his alphabet to correspond with normal frequencies of characters (with certain exceptions), since the most frequent ones are represented by permutations requiring the least electrical energy on the basis of "marking" and "spacing." In this respect Baudot "took a leaf out of Morse's note-book." Seven-unit codes are also in existence; the characters in these alphabets are always composed of 3 mark impulses, so that the adding or dropping of an impulse will at once be recognized as an error.

b. During electrical transmission, the two distinct elements of which each character is composed take the form of (1) a timed interval of electrical current and (2) a timed interval of no current, which are commonly referred to as "mark" impulses and "space" impulses, respectively, and these impulses are transmitted serially. In certain operations, a paper tape is prepared of the traffic to be transmitted, or a paper tape may be prepared of the incoming traffic at the receiving end; in such tapes, the elements of the Baudot characters take the form of punched holes ("mark" impulses) and imperforate positions ("space" impulses).

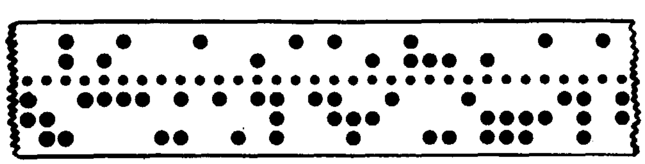
c. The teleprinter code in international use is given in Chart 7, below, wherein the mark and space impulses (known collectively as *bauds*) are illustrated as the holes (shown as black dots) and "no-holes" of a teleprinter tape. The letter equivalents ("lower case") are self-explanatory.

UPPER CASE — WEATHER SYMBOLS	↑	⊕	○	⁄	3	→	↘	↓	8	↙	←	↖	•	●	9	Ø	I	4	🔔	5	7	⌽	2	/	6	+	−	⟨	≡			
UPPER CASE — COMMUNICATIONS	−	?	:	$	3	!	&	£	8	'	(	)	.	,	9	Ø	I	4	🔔	5	7	;	2	/	6	"	≀	⟨	≡			
LOWER CASE	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	BLANK	C.R.	L.F.	SPACE	LTR. SHIFT	FIG. SHIFT
1	●	●		●	●	●				●	●						●		●		●		●	●	●	●					●	●
2	●		●				●		●	●	●	●				●	●	●			●	●	●						●		●	●
3			●			●		●	●		●		●	●		●	●		●		●	●		●	●					●	●	
4		●	●	●		●	●			●	●		●	●	●			●				●		●				●			●	●
5		●					●	●				●	●		●	●	●			●		●	●	●	●	●					●	●

CHART 7. International teleprinter code.

The *figure shift* is used to change the meaning of a particular character to an "upper case" equivalent, and when it is desired to return to lower case, the *letter shift* is used; in regular teleprinter usage, the "communications" set of upper-case equivalents are the ones recorded on the typed copy by the teleprinter, whereas the "weather symbols" are the upper-case equivalents which are printed in teleprinter systems designed for the sending and receiving of weather information. The *space* is used to separate words; the *carriage return* (C. R.) effects the return of the teleprinter carriage to the right and the *line feed* (L. F.) rolls the platen to the next line for printing (cf. the corresponding functions of an ordinary typewriter). In addition, when the upper-case equivalent of "S" is used, a bell rings in the receiving teleprinter as a signal to call the operator to his machine, or to indicate that traffic is about to be sent.

d. In Fig. 25 is shown a portion of a teleprinter tape containing the beginning of the phrase "Now is the time for all good men . . ."



NOW IS THE TIME FOR ALL GOOD MEN

FIGURE 25.

The small holes, one of which appears in every position of the tape between the second and third levels, are sprocket holes used for advancing the tape through the teleprinter unit. Tapes may be of two kinds: (a) tapes in which the holes are fully perforated, called "chad tape" or "fully-chadded tape"; or (b) tapes in which the holes are cut as little round flaps or lids (i. e., the punchings are left attached to the body of the tape), called "chadless tape." This latter tape was developed so as to permit an easily readable typed record on a perforated tape without increasing the width of the standard tape or changing punching dimensions.

e. It is to be emphasized that messages are *not* made secure from unauthorized reading merely by sending them by means of an ordinary teleprinter system—the teleprinter alphabet is internationally known, just as the English, Russian, etc. alphabets are. In order to provide security for a teleprinter message, it is just as necessary to apply thereto some sort of cryptographic treatment as it is to any other kind of message. The cryptosystems used for teleprinter encryption may involve either, or both, of the two classes of cryptographic treatment, *viz.*, substitution and transposition. A substitution treatment might involve changing certain of the mark impulses of the characters comprising a message to space impulses, and vice versa, according to a prearranged system; a transposition treatment might involve changing the *order* of the 5 impulses in the Baudot equivalents for the characters comprising a message; and so on. The cryptographic treatment can be accomplished by a special cipher attachment (called an "appliqué unit") to a teleprinter; thus no modification of the teleprinter itself would be necessary. There are, of course, self-contained cipher teleprinters designed as such for engineering or cryptographic reasons, or both.

f. In the analysis of encrypted teleprinter systems, recourse is had to special tables[12] of the frequencies of single Baudot characters, digraphs, trigraphs, etc., as they appear in teleprinter traffic. It is important to note that in teleprinter traffic, as in any other type of traffic involving the use of a word separator, this character has the highest frequency of any plaintext element. Furthermore, one of the highest-frequency plaintext digraphs, in addition to those wherein the word separator constitutes one of the elements, will be the combination "carriage-return/line-feed", since this combination of characters is used in the normal procedure of typing each line of text on the teleprinter.

---

[12] In such tables, as is common in cryptanalytic practice, the mark impulses are designated by a plus symbol (+), and the space impulses are designated by a minus symbol (−). In addition, it is usual in such tables to denote the character representing the carriage return by the digit "3," the line feed by "4," the figure shift by "5." the blank by "7," the letter shift by "8," and the space by "9."

CHAPTER VIII

MULTILITERAL SUBSTITUTION WITH VARIANTS

57. Purpose of providing variants in monoalphabetic substitution.—*a.* It has been seen that the individual letters composing ordinary intelligible plain text are used with varying frequencies; some, such as (in English) E, T, R, I, and N, are used much more often than others, such as J, K, Q, X, and Z. In fact, each letter has a *characteristic frequency* which affords definite clues in the solution of simple monoalphabetic ciphers, such as those discussed in the preceding chapters of this text. In addition, the associations which individual letters form in combining to make up words, and the peculiarities which certain of them manifest in plain text, afford further direct clues by means of which ordinary monoalphabetic substitution encipherments of such plain text may be more or less speedily solved. This has led cryptographers to devise methods for disguising, suppressing, or eliminating the foregoing characteristics manifested in cryptograms produced by the simpler methods of monoalphabetic substitution. One category of such methods, the one to be discussed in this chapter, is that in which the letters of the plain component of a cipher alphabet are assigned two or more cipher equivalents, which are called *variant values* (or, more simply, *variants*).

b. Basically, systems involving variants are multiliteral[1] and, in such systems, because of the large number of equivalents made available by the combinations and permutations of a limited number of elements, each letter of the plain text may be represented by several multiliteral cipher equivalents which may be selected at random. For example, if 3-letter combinations are employed as the multiliteral equivalents, there are available 26^3 or 17,576 such equivalents for the 26 letters of the plain text; they may be assigned in equal numbers of different equivalents for the 26 letters, in which case each letter would be representable by 676 different 3-letter equivalents; or they may be assigned on some other basis, for example, proportionately to the

[1] *Uniliteral* substitution with variants is also possible, but not very practical. Note the following cipher alphabet, illustrated by Captain Roger Baudouin in his excellent treatise, *Eléments de Cryptographie*, p. 101 (Paris, 1939):

Plain:	A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T	U	V	X	Z
Cipher:	L	G	O	R	F	Q	A	H	C	M	B	T	I	D	N	P	U	S	Y	E	W	J
					K							X					Z					
					V																	

Baudouin proposed that J_p and Y_p be replaced by I_p; K_p by C_p or Q_p; and W_p by VV_p—thus four cipher letters would be available as variants for the high-frequency plaintext letters in French. (Cf. the variant scheme in Edgar Allan Poe's day, in footnote 1 on p. 62, in which the decipherment may be ambiguous.)

relative frequencies of plaintext letters. For this reason this type of system may be more completely described as a *monoalphabetic, multiliteral substitution with a multiple-equivalent cipher alphabet.*[2] Some authors term such a system "simple substitution with multiple equivalents"; others term it "monoalphabetic substitution with variants", or *multiliteral substitution with variants*. For the sake of brevity and precise terminology, the latter designation will be employed in this text, it being understood without further restatement that only such systems as are monoalphabetic will be discussed.

c. The primary object of substitution with variants is, as has been mentioned above, to provide several values which may be employed at random in a simple substitution of cipher equivalents for the plaintext letters.

d. A word or two concerning the underlying theory of (monoalphabetic) multiliteral substitution with variants may not be amiss. Whereas in simple or single-equivalent substitution it has been seen that

(1) the same letter of the plain text is invariably represented by but one and always the same character or cipher unit of the cryptogram, and

(2) the same character or cipher unit of the cryptogram invariably represents one and always the same letter of the plain text,

in multiliteral substitution with variants it will be seen that

(1) the same letter of the plain text may be represented by one or more different cipher units of the cryptogram, but

(2) the same cipher unit of the cryptogram nevertheless invariably represents one and always the same letter of the plain text.

58. Simple types of cipher alphabets with variants.—*a*. The matrices shown below represent some of the simpler means for accomplishing monoalphabetic substitution with variants. The systems incorporating these matrices are extensions of the basic ideas of multiliteral substitution treated in par. 52. The variant equivalents for any plaintext letter may be chosen at will; thus, in Fig. 26, E_p=10, 15, 60, or 65; in Fig. 27, E_p=AU_c, AZ_c, FU_c, FZ_c, LU_c, or LZ_c; etc.

		6	7	8	9	0
		1	2	3	4	5
6	1	A	B	C	D	E
7	2	F	G	H	IJ	K
8	3	L	M	N	O	P
9	4	Q	R	S	T	U
0	5	V	W	X	Y	Z

FIGURE 26.

			V	W	X	Y	Z
			Q	R	S	T	U
L	F	A	A	B	C	D	E
M	G	B	F	G	H	IJ	K
N	H	C	L	M	N	O	P
O	I	D	Q	R	S	T	U
P	K	E	V	W	X	Y	Z

FIGURE 27.

				A	E	I	O	U
T	N	H	B	A	B	C	D	E
V	P	J	C	F	G	H	IJ	K
W	Q	K	D	L	M	N	O	P
X	R	L	F	Q	R	S	T	U
Z	S	M	G	V	W	X	Y	Z

FIGURE 28.

[2] Cf. the title of the preceding chapter, "Multiliteral substitution with *single*-equivalent cipher alphabets."

	V Q L F A	W R M G B	X S N H C	Y T O I D	Z U P K E
V Q L F A	A	B	C	D	E
W R M G B	F	G	H	IJ	K
X S N H C	L	M	N	O	P
Y T O I D	Q	R	S	T	U
Z U P K E	V	W	X	Y	Z

FIGURE 29.

	O M J F A	N K G B	L H C	I D	E
O M J F A	E	N	A	L	U
N K G B	T	R	S	F	W
L H C	O	IJ	H	Y	X
I D	D	C	M	V	K
E	P	G	B	Q	Z

FIGURE 30.

	Z W S N	X T O	Y U P	V Q	R
M J F A	E	N	A	L	U
K G B	T	R	S	F	W
L H C	O	IJ	H	Y	X
I D	D	C	M	V	K
E	P	G	B	Q	Z

FIGURE 31.

	1	2	3	4	5	6	7	8	9	Ø
7 4 1	A	B	C	D	E	F	G	H	I	J
8 5 2	K	L	M	N	O	P	Q	R	S	T
9 6 3	U	V	W	X	Y	Z	.	,	:	;

FIGURE 32.

	1	2	3	4	5	6	7	8	9
7 4 1	A	B	C	D	E	F	G	H	I
8 5 2	J	K	L	M	N	O	P	Q	R
9 6 3	S	T	U	V	W	X	Y	Z	*

FIGURE 33.

	1	2	3	4	5	6	7	8	9
5 1	A	B	C	D	E	F	G	H	I
6 2	J	K	L	M	N	O	P	Q	R
7 3	S	T	U	V	W	X	Y	Z	1
8 4	2	3	4	5	6	7	8	9	Ø

FIGURE 34.

	1	2	3	4	5	6	7	8	9
Ø 8 5 1	T	E	R	M	I	N	A	L	S
9 6 2	B	C	D	F	G	H	J	K	O
7 3	P	Q	U	V	W	X	Y	Z	1
4	2	3	4	5	6	7	8	9	Ø

FIGURE 35.

b. It is to be noted that encipherment by means of the matrices in Figs. 27, 28, and 31 is *commutative;* i. e., the coordinates may be read in either row-column or column-row order without cryptographic ambiguity, since there is no duplication between the row and column coordinates. The remaining matrices above are *noncommutative;* therefore a convention must be agreed upon as to the order of reading the coordinates. It should also be noted that in Figs. 30 and 31 the letters in the square have been inscribed in such a manner that, coupled with the particular arrangement of the row and column coordinates, the number of variants available for each plaintext letter is roughly proportional to the frequencies of the letters in plain text. A similar idea is found in Fig. 35, wherein the top row of the rectangle contains a word composed of high-frequency letters, and the coordinates are arranged in a manner roughly corresponding to the frequencies of plaintext letters. The matrix in Fig. 28 is a modification of the pseudo-code system described in par. 52*h*, with the added feature of variants.

c. Other simple ideas for producing variant systems are matrices such as the following:

A	B	C	D	E	F	G	H	IJ	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	01	02	03	04	05	06	07
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	26	27	28	29	30	31	32	33	34
68	69	70	71	72	73	74	75	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
87	88	89	90	91	92	93	94	95	96	97	98	99	00	76	77	78	79	80	81	82	83	84	85	86

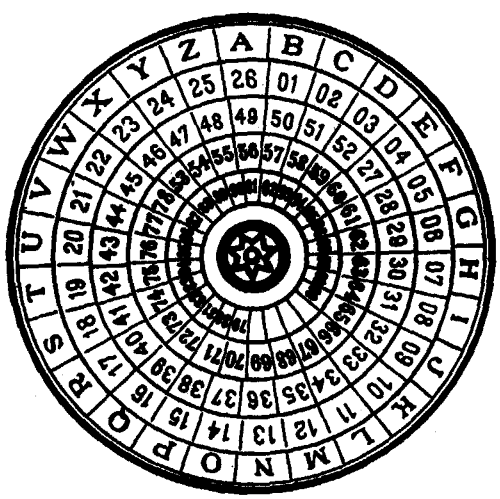
FIGURE 36.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
14	15	16	17	18	19	20	21	22	23	24	25	26	01	02	03	04	05	06	07	08	09	10	11	12	13
27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	53	54	55	56	57
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	//////	//////	//////	//////	79	80

FIGURE 37.

In these two matrices there has been a regular inscription of the dinomes in the rows. Furthermore, in Fig. 36 the dinomes 01, 26, 51, and 76 (i. e., the lowest number in each of the four sequences) give the key word (TRIP) for that matrix; and in Fig. 37, the dinomes 01, 27, 53, and 79 denote the key word (NAVY) for that matrix. The security of systems involving such matrices would of course be greatly improved if the dinomes were assigned in a random manner; but then the easy mnemonic feature of the four sequences and the key word would be lost.

d. An interesting adaptation in a disc form of the type of matrix illustrated in Fig. 37 is the following device reputedly once used by the Mexican Army:



The device consisted of five concentric discs, the outer disc bearing the 26 letters of the alphabet, and the other four bearing the sequences 01–26, 27–52, 53–78, and 79–00. The rotatable discs made it possible to change the keys at frequent intervals, without the necessity of writing out a new matrix each time.

59. More complicated types of cipher alphabets with variants.—*a.* Matrices such as those in Figs. 38, 39, and 40 below are termed *frequential matrices,* since the number of cipher values available for any given plaintext letter closely approximates its relative plaintext frequency.

b. In the fragmentary matrix illustrated in Fig. 38, the number of occurrences of a particular letter within the matrix is proportional to its frequency in plain text; the letters are inscribed in a random manner, in order to enhance further the security of the system. In Fig. 39, we have a modification of the idea set forth in Fig. 38, except that the size of the matrix has been reduced from 26 x 26 to 10 x 10; in this case, the letters (with appropriate number of repetitions) have been inscribed in a simple diagonal route (lower left to upper right) within the square, and the coordinates have been scrambled, for greater security. In Fig. 40, there is illustrated a type of

	A	B	C	D	E		V	W	X	Y	Z
A	T	G	A	U	R		I	E	C	A	P
B	S	L	I	E	Y		F	R	N	S	T
C	C	N	D	O	M	. . .	E	L	T	I	H
D	R	A	P	T	F		O	Y	S	O	V
E	N	T	X	N	E		C	E	R	E	D
		.							.		
		.							.		
		.							.		
V	N	O	A	T	E		A	L	E	Z	H
W	I	H	R	O	Q		E	T	R	B	T
X	O	I	E	T	A	. . .	C	N	P	E	S
Y	F	T	L	O	S		A	M	T	I	U
Z	I	S	N	D	R		I	E	D	O	N

(676-cell matrix)

FIGURE 38.

	6	8	9	1	5	4	3	7	2	0
7	A	A	A	C	D	E	E	I	L	N
1	A	A	C	D	E	E	H	K	N	O
3	A	B	D	E	E	H	J	N	O	R
8	A	D	E	E	H	I	N	O	R	S
9	C	E	E	G	I	N	O	R	S	T
2	E	E	F	I	M	O	Q	S	T	T
0	E	F	I	M	O	P	R	T	T	U
5	F	I	L	N	P	R	S	T	U	X
6	I	L	N	P	R	S	T	U	W	Y
4	L	N	O	R	S	T	T	V	Y	Z

FIGURE 39.

	0	1	2	3	4	5	6	7	8	9
0	E	N	T	R	U	C	K	I	N	G
1	Q	U	A	R	A	N	T	I	N	E
2	U	N	E	X	P	E	N	D	E	D
3	I	M	P	O	S	S	I	B	L	E
4	V	I	C	T	O	R	I	O	U	S
5	A	D	J	U	D	I	C	A	T	E
6	L	A	B	O	R	A	T	O	R	Y
7	E	I	G	H	T	E	E	N	T	H
8	N	A	T	U	R	A	L	I	Z	E
9	T	W	E	N	T	Y	F	I	V	E

FIGURE 40.

cipher square which is known in cryptologic literature as the *Grandpré cipher;* in this square there are inscribed ten 10-letter words containing all the letters of the alphabet in their approximate plaintext frequencies. These ten words are further linked together by a 10-letter word which appears vertically in the first column, as a mnemonic feature for the inscription of the words in the rows.

c. The frequential-type system represented in Fig. 41*a* (enciphering matrix) and 41*b* (deciphering matrix) was described by Sacco,[3] who proposed that the dinomes inscribed in the enciphering matrix be thoroughly disarranged by applying a double transposition to the dinomes 00–99 as a means of suppressing any patent relationships among the variant values for the various plaintext letters; furthermore, the nulls incorporated in the matrix were to be used occasionally during the encryption of a message, in order to throw a cryptanalyst off the track. In this example the number of variant values for each plaintext letter has been established, of course, from the standpoint of Italian letter frequencies.

[3] Sacco, Generale Luigi, *Manuale di Crittografia,* 3d Ed., Rome, 1947, p. 22.

Enciphering Table

Nulls	A	E	I	M	Q	V	one	seven
48–56 21–09 76–54 42–12 64–74 55–14 83–90 63–06 47–45	03–25 52–62 79–69	18–35 37–65 71–78	10–23 53–75 82–87	39 68	20 77	02–86	44 66	46
	B 40 93	F 24 57	J 81	N 13–73	R 26–94	W 95	two 84	eight 29
	C 28 70	G 38 97	K 96	O 07–30 51–67 72–89	S 11–58	X 85	three 50	nine 31
					T 33–88	Y 22	four 27	zero 19 92
	D 08 80	H 17 43	L 05 49	P 41 98	U 00–15 36–99 01	Z 34 59	five 60–91	period 16–91
							six 04	comma 32

FIGURE 41*a*.

Deciphering Table

	1	2	3	4	5	6	7	8	9	Ø
1	S	–	N	–	U	period	H	E	zero	I
2	–	Y	I	F	A	R	four	C	eight	Q
3	nine	comma	T	Z	E	U	E	G	M	O
4	P	–	H	one	–	seven	–	–	L	B
5	O	A	I	–	–	–	F	S	Z	three
6	period	A	–	–	E	one	O	M	A	five
7	E	O	N	–	I	–	Q	E	A	C
8	J	I	–	two	X	V	I	T	O	D
9	five	zero	B	R	W	K	G	P	U	–
Ø	U	V	A	six	L	–	O	D	–	U

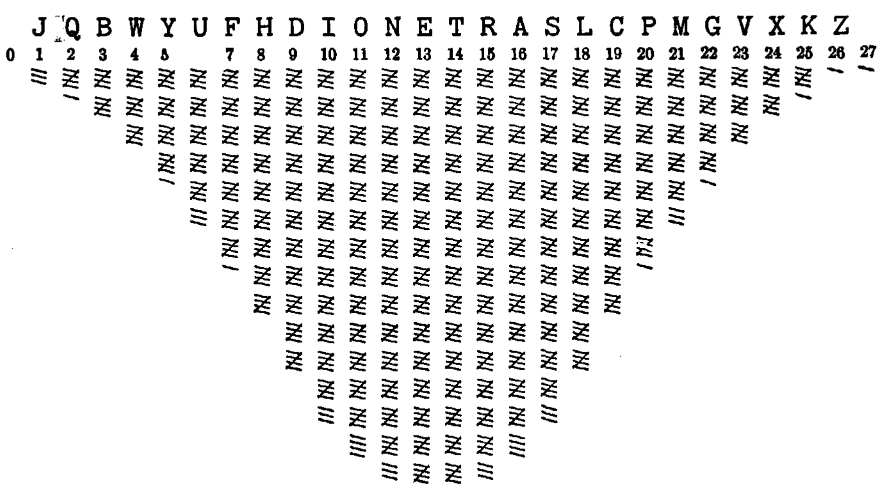
FIGURE 41*b*.

d. The Baconian cipher described in subpar. 53*a* may be used as a basis for superimposing additional complexities. For instance, the "a" elements may be represented by any one of the 20 consonants as variants, while the "b" elements may be represented by any one of the six vowels; or the letters A–M may be used to represent the "a" elements and the letters N–Z for the "b" elements; digits may be used for the "a" and "b" elements, either on the basis of the first five and last five digits, or on the basis of the odd and even digits; or the first 10 consonants (B–M) and the last 10 consonants (N–Z) may be used for the "a" and "b" elements, with the vowels used occasionally as nulls—thus the resultant cryptograms will resemble those of a fairly complex cryptosystem. However, once the cryptanalyst assumes the possibility of such a system, its complexity is more apparent than real. Similarly, variations of this genre may be superimposed on triliteral systems such as the Trithemian cipher illustrated in subpar. 53*b*; variants for the "1", "2", and "3" elements may be chosen in such a way as to provide a large number of equivalents for each basic triliteral combination.

e. Another scheme for a complex variant system is a *summing-trinome system.* In this cryptosystem, each plaintext letter is assigned a unique value of 1 to 26; this value is then expressed as a trinome, *the digits of which sum to the designated value of the letter.* For example, if a letter has been assigned the value "4", it may be represented by any one of the following permutations and combinations:[4]

004	031	112	202	301
013	040	121	211	310
022	103	130	220	400

Since the values toward the middle of the range 1–26 may be represented by a very considerable number of summing-trinomes (e. g., for the values 13 and 14 there are 75 variants each), such a system would offer a cryptographer wide latitude in the choice of cipher equivalents in enciphering, especially if the basic values of the plaintext letters were chosen to correspond with the scale of their relative frequencies, such as the following:



[4] The representations of an integer (i. e., a whole number) as the sum of integers in all possible ways are termed the *partitions* of that number. The members of the partitions in this subparagraph are one-digit numbers, including the digit Ø in order to form trinome equivalents out of all the possible permutations.

The tallies beneath each value represent the number of variants possible for the particular value. The unused values for Ø and 27 (uniquely represented by 000 and 999, respectively) may be used for punctuation marks, nulls, or other special-purpose symbols. Since such a system, once suspected, would offer little difficulty [5] to a cryptanalyst, certain modifications would be necessary in order to pose any real obstacles in the way of solution. For instance, if the numerical value of a letter is expressed by permutations of 3 letters (instead of digits) out of a set of the 10 letters A–J wherein the sequence of the letters A–J represents a *disarranged* sequence of the digits Ø–9, such a system may be among the most complex types of ciphers in the realm of monoalphabetic substitution, requiring the solution of many simultaneous equations. A further refinement would involve the use of all 26 letters as variants, in predetermined groups, to represent the digits Ø–9. Fortunately for the cryptanalyst, such systems are impracticable for field military use; but if they were encountered, a sufficiently large volume of text, coupled with Hitt's four essentials quoted in Chapter I, would eventually make a solution possible. The actual cryptanalytic complexity of certain apparently exceedingly complex cryptosystems is dependent on their being correctly used at all times, which is not always the case with military ciphers.

60. Analysis of simple examples.—*a*. The following cryptogram is available for study:

Q M D C V	P L F N F	D H N W J	W L K D K	N H B P V	R L T V M
B K L W D	W V H V K	S H B C L	P Q K J R	V W S M L	K G C N R
L R N K V	M G F X W	J R G M V	W G T J H	Q K X F N	Z V F D M
L T B P L	P V F L M	D C N W N	H B C V Z	N M L W Q	F D H D W
V Z B R V	K L C V C	V R D H L	R V T L F	N C D K G	M X W X M
D T S C B	C L Z L R	L M V T S	Z N K B W	V P B R N	C L R X R
D C N K V	P B T N T	G H J Z L	F Q F V K	B W D Z X	P N H S P
G H L K L	F V Z L T	V M L K D	P Q R N Z	L Z D T B	M N T G M
N Z V F X	K S F D C	L Z V T V	F D F V R	G C L P Q	P N C D W
V R J T N	H L Z L M	V W N P V	P D Z D W	J P N W L	R J K V M
X M D T S	M G F D R	D K L W J	F L P J M	S F Q W B	F N C B Z
D K V W G	Z S H B H	D H J C X			

The first thing that strikes the eye is the total absence of A, E, I, O, U, and Y, remarkable not only because six letters are missing (cf. the Λ test) in a text of this size, but also because all six of these letters fall into an identical limited category, namely, they are all vowels—a significant nonrandom phenomenon. Since a uniliteral substitution *alphabet* with six letters missing is highly improbable, the conclusion of multiliteral substitution is obvious. Upon closer inspection it is found that, if the cipher text is divided into pairs of letters, only ten consonants (B D G J L N Q S V X) are used as initial letters, and the remaining ten consonants

[5] The solution would involve simply dividing the cipher text into groups of 3 digits, summing the trinomes thus produced to yield 28 possible basic values, and solving these basic values as in any simple monoalphabetic substitution cipher.

(C F H K M P R T W Z) are used as final letters—thus the biliteral (and *bipartite*) characteristics of the cipher text are disclosed. A digraphic distribution is therefore constructed:[6]

	C	F	H	K	M	P	R	T	W	Z
B	3	1	1	1	1	2	2	1	2	1
D	5	1	5	5	1	1	1	4	5	2
G	2	2	2		3			1		1
J	1	1	1	1	1	1	2	1	1	1
L	1	5		5	4	5	[illegible]	3	4	5
N	5	1	5	4	1	1	1	3	4	4
Q		2		2	1	1	1		1	
S	1	2	2		2	1				1
V	1	5	1	4	5	5	5	4	5	4
X		1		1	2	1	1		2	

b. It is possible that the cryptogram under study may involve the use of a small enciphering matrix with variants for the rows and columns. Since there is available an easily-applied special solution which permits the determination of the row indicators which are equivalent (i. e., interchangeable variants) and the column indicators which are equivalent, merely from a study of the digraphic distribution, this possibility is examined. The special solution is based on the following considerations: in a message of moderate length for such a cryptosystem, it may be assumed that the various possible cipher digraphs for a given plaintext letter will be used with approximately equal frequency; for this reason, the column indicators which pair with one of the letters used to indicate any particular row of the enciphering matrix may be expected to pair equally often with any other cipher letter which has been used to indicate the same row. Thus, in the digraphic distribution of such a cryptogram, sets of rows appear which have similar "profiles" and, likewise, sets of similar columns.[7] First a study will be made of the rows of the distribution just compiled, in an attempt to locate and isolate those which *match* with each other; then, the same will be done with the columns of the distribution.

c. It is noted that the "L" and "V" distributions have pronounced similarities (Fig. 42*a*)—these rows came under consideration first because of the unique "heaviness" of their frequency characteristics. Likewise, the "D" and "N" rows have homologous attributes in their appearance (Fig. 42*b*). However, the further grouping of the rows by ocular inspection may present difficulties to the student, since he may not yet trust his eye in matching distributions; and he may feel the need for some kind of statistical assurance. In the following subparagraphs there is given the technique of a more precise method for matching, mathematical in nature.

---

[6] If it had not been noticed that the cryptogram should be divided into pairs for analysis, a *biliteral* distribution (see subpar. 23*d*) might have been made, in order to reveal contact affinities of the cipher letters.

[7] These similarities are especially pronounced when the encipherer uses a "check-off" procedure for choosing his variants for each letter, that is, when he systematically checks off the variants used during encryption to insure that all possible variants are used in approximately equal proportions.

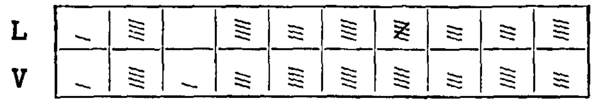


FIGURE 42*a*.

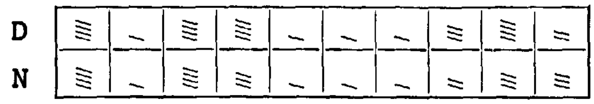


FIGURE 42*b*.

d. This method of matching in an attempt to "equate" interchangeable variants involves computing a separate value for each trial matching of a particular row (or column) against each of a series of other rows (or columns, as appropriate)—such a value is taken as an indication of the "goodness of match" exhibited by the particular trial, the theory being that the correct match will produce the highest value.[8] The value for a particular trial match is computed by multiplying the number of tallies in each cell of one row (or column) by the number of tallies in each corresponding cell in the other row (or column) and then totaling the products thus obtained. Because of the way in which it is produced, such a value is termed a "cross-products sum".

e. In subpar. *c* above, it was determined that the "L" and "V" rows were equivalent, and that the "D" and "N" rows also formed an equivalent pair. The next "heavy" row is the "G" row; this is to be tested for match with the five remaining unmatched rows. Let the "G" row be tested first against the "B" row. These two rows are given below, with their cross-products sum. For convenience, the cross-products sum is symbolized by $\chi(\theta^1,\theta^2)$, where θ^1 and θ^2 represent the designators of the distributions to be matched.[9]

```
    "G": 2 2 2 - 3 - - 1 - 1
    "B": 3 1 1 1 1 2 2 1 2 1
χ(G,B): 6 2 2 - 3 - - 1 - 1 = 15
```

The complete table of the comparisons of the "G" row with the five available rows is as follows:

```
χ(G,B): 6 2 2 - 3 - - 1 - 1 = 15
χ(G,J): 2 2 2 - 3 - - 1 - 1 = 11
χ(G,Q): - 4 - - 3 - - - - - =  7
χ(G,S): 2 4 4 - 6 - - - - 1 = 17
χ(G,X): - 2 - - 6 - - - - - =  8
```

The results indicate that the most probable match with the "G" row is the "S" row.

f. Since the next "heaviest" row to be tested is the "B" row, its matchings with the three remaining rows are made, and are given below:

```
χ(G,J): 3 1 1 1 1 2 4 1 2 1 = 17
χ(B,Q): - 2 - 2 1 2 2 - 2 1 = 12
χ(B,X): - 1 - 1 2 2 2 - 4 - = 12
```

[8] In this connection, note the considerations treated in subpar. 60*j*.

[9] The Greek letter χ (chi) is often used in cryptology to symbolize matching operations.

The correct matching of the "B" and "J" rows is indicated by the results. This leaves only the "Q" and "X" rows, which are presumed to go together, since not only is their cross-products sum satisfactory (when compared to the χ values for some of the other rows which have been matched), but, equally important, their patterns of *crests* and *troughs* are similar. Since we have not found more than *two* rows for any one set of interchangeable values, it appears that the original matrix had only five rows, with two variants for each row. The rows of the distribution diagram are therefore combined in the following diagram:

	C	F	H	K	M	P	R	T	W	Z
B J	4	2	2	2	2	3	4	2	3	2
D N	8	2	8	7	2	2	2	5	7	5
G S	3	4	4	–	5	1	–	1	–	2
L V	2	8	1	7	7	8	9	6	7	7
Q X	–	3	–	3	3	2	2	–	3	–

FIGURE 43.

g. Ocular inspection of the distributions of the columns of Fig. 43 quickly reveals that columns "C" and "H" may be matched as a pair, and likewise columns "F" and "M", and columns "P" and "R". In order to decide the groupings of the remaining columns, the six possible χ values are derived:

χ(K,T): 4 35 – 42 – = 81
χ(K,W): 4 49 – 49 9 = 113
χ(K,Z): 4 35 – 49 – = 88
χ(T,W): 6 35 – 42 – = 83
χ(T,Z): 4 25 2 42 – = 73
χ(W,Z): 6 35 – 49 – = 90

Combinations:
KT, WZ: 81 + 90 = 171
KW, TZ: 113 + 73 = 186
KZ, TW: 88 + 83 = 171

It appears that the proper pairings of the columns are "K" and "W", "T" and "Z".

h. The groupings of the columns having been determined, the frequency diagram is reduced to its basic 5 x 5 square, and the ϕ test is taken as further statistical assurance of the matchings

	C H	F M	K W	P R	T Z
B J	6	4	5	7	4
D N	16	4	14	4	10
G S	7	9	–	1	3
L V	3	15	14	17	13
Q X	–	6	6	4	–

$\phi_p = 1962$
$\phi_r = 1132$
$\phi_o = 1670$

Although ϕ_o in this case does not come up to the best expectations, we feel nevertheless that the matching has been carefully and correctly accomplished, and so the next step is continued with a conversion of the multiliteral text into uniliteral equivalents, using the following reduction square containing an arbitrary sequence:

		C	F	K	P	T
		H	M	W	R	Z
B	J	A	B	C	D	E
D	N	F	G	H	I	K
G	S	L	M	N	O	P
L	V	Q	R	S	T	U
Q	X	V	W	X	Y	Z

The converted cryptogram is now easily solved, using the principles set forth in Chapter VI. The first fifteen letters of the plaintext message are found to read "WEATHER FORECAST.....", and the original enciphering matrix is recovered, based on the key word ATMOSPHERIC, as follows:

		P	F	C	K	T
		R	M	H	W	Z
L	V	A	T	M	O	S
D	N	P	H	E	R	I
B	J	C	B	D	F	G
G	S	K	L	N	Q	U
Q	X	V	W	X	Y	Z

i. The method of matching rows and columns just described in the preceding subparagraphs applies equally well to all the matrices in Figs. 26–35, and similar variations. If in the process of equating indicators the cryptanalyst sees that the row indicators are falling into the same groupings as the column indicators, he might be able to accelerate the equating process by taking advantage of this feature alone, as would be the case if he had encountered a cryptogram involving a matrix with indicators arranged in a manner similar to that shown in Figs. 29 and 30. Furthermore, a cryptogram enciphered in a commutative system, wherein the equivalents have been taken in row-column and column-row order indiscriminately, may be recognized as such through a study of the digraphic distribution of the cryptogram since the "α" row of the distribution will have an appearance similar to the "α" column, the "β" row will be similar to the "β" column, etc;[10] this matter is discussed further in subpar. 61*d*.

[10] It is often convenient to use arbitrary symbols in cryptanalytic work, to prevent confusion with designations of actual elements of plain text, cipher text, or key (see footnote 1 on page 47). For this purpose Greek letters are often used; for reference, the 24 letters of the Greek alphabet and their names are appended in the chart below:

Α α alpha	Ε ε epsilon	Ι ι iota	Ν ν nu	Ρ ρ rho	Φ φ phi
Β β beta	Ζ ζ zeta	Κ κ kappa	Ξ ξ xi	Σ σ sigma	Χ χ chi
Γ γ gamma	Η η eta	Λ λ lambda	Ο ο omicron	Τ τ tau	Ψ ψ psi
Δ δ delta	Θ θ theta	Μ μ mu	Π π pi	Υ υ upsilon	Ω ω omega

j. It is important to point out that in matching, the cryptanalyst should begin with the "best" rows or columns—best not only from the standpoint of "heaviness" of the distribution, but also best from the point of view of a distinctive pattern of crests and troughs. If insufficient text is available to allow equating all the interchangeable coordinates of a particular enciphering matrix, it may still be possible that a conversion of the cipher test by means of a partially-reduced reconstruction matrix may yield enough idiomorphic patterns and other data to make possible an entry into the text. If the cryptographer has not used a "check-off" process in enciphering, but instead has favored certain equivalents for the various plaintext letters, matching may not be possible; nevertheless, an entry into the text may be facilitated in this case, because some of the resultant peaks in the cipher text may be correctly identified. Furthermore, since *no* variant system can possibly disguise the *letters of low frequency* in plain text, their low-frequency equivalents in the cipher text may provide possible approaches to solution. (See also subpar. 61*e*).

k. In addition to the method of solution by matching and combining rows and columns of a digraphic distribution of a multiliteral cipher, there is also the *general* approach applicable without exception to *any* variant system. This method, involving the correlation of cipher elements suspected to be the equivalents of specific but unknown plaintext letters, is treated in detail in pars. 61 and 62.

l. Systems such as the 4-level dinome cipher illustrated in Fig. 36 are susceptible to a very easy solution, if the dinomes have been inscribed in numerical order as indicated. Assuming such a case in a specific cryptogram, the first six groups of which are

68321 09022 48057 65111 88648 42036 . . .

a four-part frequency distribution of the entire message is taken, as illustrated in Fig. 44 below:

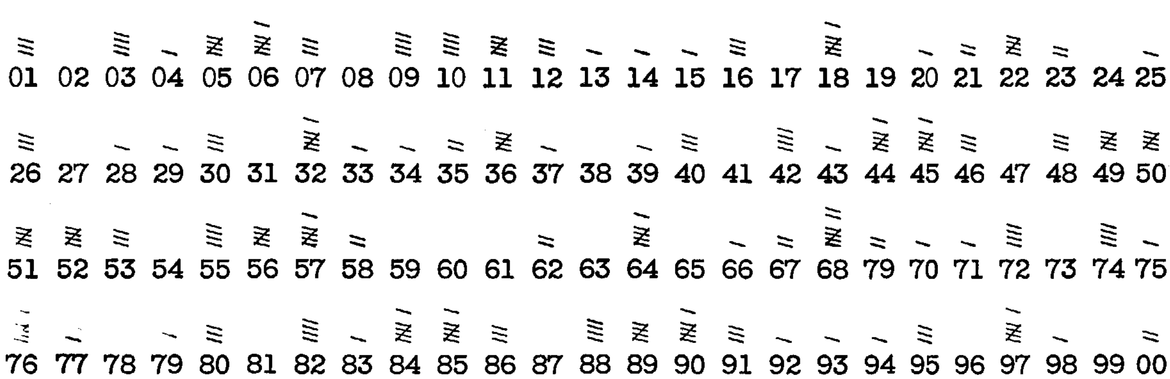


FIGURE 44.

If the student will bring to bear upon this problem the principles he learned in Chapter V of this text, he will soon realize that what he now has before him are four simple, monoalphabetic frequency distributions similar to those involved in a monoalphabetic substitution cipher using standard alphabets. The realization of this fact immediately provides the clue to the next step: "fitting each of the distributions to the normal". (See par. 31.) This can be done without difficulty in this case (remembering that a 25-letter alphabet is involved and assuming that I and J are combined) and the following alphabets result:

01—I-J	26—U	51—N	76—E
02—K	27—V	52—O	77—F
03—L	28—W	53—P	78—G
04—M	29—X	54—Q	79—H
05—N	30—Y	55—R	80—I-J
06—O	31—Z	56—S	81—K
07—P	32—A	57—T	82—L
08—Q	33—B	58—U	83—M
09—R	34—C	59—V	84—N
10—S	35—D	60—W	85—O
11—T	36—E	61—X	86—P
12—U	37—F	62—Y	87—Q
13—V	38—G	63—Z	88—R
14—W	39—H	64—A	89—S
15—X	40—I-J	65—B	90—T
16—Y	41—K	66—C	91—U
17—Z	42—L	67—D	92—V
18—A	43—M	68—E	93—W
19—B	44—N	69—F	94—X
20—C	45—O	70—G	95—Y
21—D	46—P	71—H	96—Z
22—E	47—Q	72—I-J	97—A
23—F	48—R	73—K	98—B
24—G	49—S	74—L	99—C
25—H	50—T	75—M	00—D

The key word is seen to be JUNE and the beginning of the cryptogram is deciphered as "EASTERN ENTRANCE. . . . ."

m. If instead of 25-element alphabets, a system such as that in Fig. 37 has been used, only a slight modification of the procedure in subpar. *l* would have been necessary, i. e., the distributions would have had to be considered on a basis of 26, and the process of fitting the distributions to the normal would have gone on as in the previous example.

n. One further application of principles learned in Chapter V deserves to be mentioned here, in connection with the solution of systems such as those of Fig. 36. Let the following short message be considered:

```
48226   88423   52099   93604   76059   05651
36683   52267   97114   54466   76
```

If it is known that the correspondents have been using a variant system such as that in Fig. 36, a special solution may be employed in those cases wherein there is insufficient cipher text to permit analysis by the method of fitting the frequency distributions to the normal. Thus, a short cryptogram may be solved by a variation of the plain-component completion method described in par. 34.[11] First, let the cryptogram be copied in dinomes, with an indication of the *level* (i. e., the "alphabet") the dinome would occupy in the 4-level matrix; thus:

48	22	68	84	23	52	09	99	36	04	76	05	90	56	51	36	68	35	22	67	97	11	45	44	66	76
2	1	3	4	1	3	1	4	2	1	4	1	4	3	3	2	3	2	1	3	4	1	2	2	3	4

The dinomes belonging to the four levels are as follows:

(1) 22 23 09 04 05 22 11
(2) 48 36 36 35 45 44
(3) 68 52 56 51 68 67 66
(4) 84 99 76 90 97 76

These dinomes are converted into terms of the plain component by setting each of the cipher sequences against the plain component at an arbitrary point of coincidence, such as in the following example:

A	B	C	D	E	F	G	H	IJ	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00

(1) 22=W; 23=X; 09=I; 04=D; 05=E; 22=W; 11=L
(2) 48=X; 36=L; 36=L; 35=K; 45=U; 44=T
(3) 68=S; 52=B; 56=F; 51=A; 68=S; 67=R; 66=Q
(4) 84=I; 99=Y; 76=A; 90=P; 97=W; 76=A

[11] It should be clear to the student that the reason this method can be applied in this instance is that both the plain component (ABC.....Z) and the cipher component (01, 02, 03 ..... 25; 26–50, 51–75, 76–00) are *known* sequences (or thus assumed).

o. The plain-component sequence is now completed on the letters of the four levels, as follows:

1st level	*2d level*	*3d level*	*4th level*
W X I D E W L	X L L K U T	S B F A S R Q	I Y A P W A
X Y K E F X M	Y M M L V U	T C G B T S R	K Z B Q X B
Y Z L F G Y N	Z N N M W V	U D H C U T S	L A C R Y C
Z A M G H Z O	A O O N X W	V E I D V U T	M B D S Z D
A B N H I A P	B P P O Y X	W F K E W V U	N C E T A E
B C O I K B Q	C Q Q P Z Y	X G L F X W V	O D F U B F
C D P K L C R	D R R Q A Z	Y H M G Y X W	P E G V C G
D E Q L M D S	E S S R B A	Z I N H Z Y X	Q F H W D H
E F R M N E T	F T T S C B	A K O I A Z Y	R G I X E I
F G S N O F U	G U U T D C	B L P K B A Z	S H K Y F K
G H T O P G V	H V V U E D	C M Q L C B A	T I L Z G L
H I U P Q H W	I W W V F E	D N R M D C B	U K M A H M
I K V Q R I X	K X X W G F	E O S N E D C	V L N B I N
K L W R S K Y	L Y Y X H G	F P T O F E D	W M O C K O
L M X S T L Z	M Z Z Y I H	G Q U P G F E	X N P D L P
M N Y T U M A	N A A Z K I	H R V Q H G F	Y O Q E M Q
N O Z U V N B	O B B A L K	I S W R I H G	Z P R F N R
O P A V W O C	P C C B M L	K T X S K I H	A Q S G O S
P Q B W X P D	Q D D C N M	L U Y T L K I	B R T H P T
Q R C X Y Q E	R E E D O N	M V Z U M L K	C S U I Q U
R S D Y Z R F	S F F E P O	N W A V N M L	D T V K R V
S T E Z A S G	T G G F Q P	O X B W O N M	E U W L S W
T U F A B T H	U H H G R Q	P Y C X P O N	F V X M T X
U V G B C U I	V I I H S R	Q Z D Y Q P O	G W Y N U Y
V W H C D V K	W K K I T S	R A E Z R Q P	H X Z O V Z

It is seen that the generatrices with the best assortment [12] of high-frequency letters for the four levels are:

1st level	*2d level*	*3d level*	*4th level*
E F R M N E T	R E E D O N	E O S N E D C	N C E T A E

If the letters of these generatrices are arranged in the order of appearance of their dinome equivalents, according to the way they fall into the various levels,

48	22	68	84	23	52	09	99	36	04	76	05	90	56	51	36	68	35	22	67	97	11	45	44	66	76
	E			F		R			M		N							E			T				
R								E							E		D					O	N		
		E			O								S	N		E			D					C	
			N				C			E		T								A					E

the plain text "REENFORCEMENTS NEEDED AT ONCE" is clearly seen. Or, more simply, if we examine the equivalents of 01, 26, 51, and 76 after the generatrix determination has been made,

[12] In evaluating generatrices, the sum of the arithmetical frequencies of the letters in each row may be used as an indication of their relative "goodness". A statistically much more accurate method of evaluating generatrices involves the use of logarithms of the probabilities of the plaintext letters forming the generatrices. This method is treated in detail in *Military Cryptanalytics, Part II.* (See also footnote 8 on p. 73.)

the key word JUNE is revealed. If an error had been made in the selection of a generatrix, the error could be resolved by hypothesizing the probable key word, or by deciphering the text on the basis of the assumed diagram and then noting and degarbling the systematic errors (which, it would be noticed, all come from one level).

p. The student should note that no one generatrix will yield plain text all the way across as in the example in par. 34. Instead, the generatrices must be considered separately for the four levels, since it is within each of the four levels that there is a homogeneous relationship of dinomes. Obviously if dinomes from more than one level were used to complete the plain component sequence, the generatrices would not consist of a homogeneous group of letters but instead would represent an assortment of letters from two or more "alphabets".

61. Analysis of more complicated examples.—*a.* As soon as a beginner in cryptography realizes the consequences of the fact that letters are used with greatly varying frequencies in normal plain text, a brilliant idea very speedily comes to him. Why not disguise the natural frequencies of letters by a system of substitution using many equivalents, and let the numbers of equivalents assigned to the various letters be more or less in direct proportion to the normal frequencies of the letters? Let E, for example, have 13 equivalents; T, 9; N, 8; etc., and thus (he thinks) the enemy cryptanalyst can have nothing in the way of telltale or characteristic frequencies to use as an entering wedge.

b. If the text available for study is small in amount and if the variant values are wholly independent of one another, the problem can become exceedingly difficult. But in practical military communications such methods are rarely encountered, *because the volume of text is usually great enough to permit of the establishment of equivalent values.* To illustrate what is meant, suppose a number of cryptograms produced by a monoalphabetic-variant method of the type mentioned above show the following two sets of groupings[13] of cipher elements in the text, Set "A" being assumed to be different representations of one particular underlying plaintext word or phrase and Set "B" assumed to be representations of another underlying plaintext word or phrase:

Set "A"	Set "B"
(12-37-02-79-68-13-03-37-77)	(71-12-02-51-23-05-77)
(82-69-02-79-13-68-23-37-35)	(11-82-51-02-03-05-35)
(82-69-51-16-13-13-78-05-35)	(11-91-02-02-23-37-35)
(91-05-02-01-68-42-78-37-77)	(97-12-51-02-78-69-77)

An examination of these groupings would lead to the following tentative conclusions with regard to probable equivalents:

(12,82,91)	(02,51)	(13,42,68)	(35,77)
(05,37,69)	(01,16,79)	(03,23,78)	(11,71,97)

The establishment of these equivalencies would sooner or later lead to the finding of additional sets of equal values. The completeness with which this can be accomplished will determine the ease or difficulty of solution. Of course, if many equivalencies can be established the problem can then be reduced practically to monoalphabetic terms and a speedy solution can be attained.

c. Theoretically, the determination of equivalencies may seem to be quite an easy matter, but practically it may be very difficult, because the cryptanalyst can never be *certain* that a

[13] The alert student might be able to determine the underlying plain text of the two sets of ciphertext groupings.

combination showing what may appear to be a variant value is really such and does not represent a part of a *different* plaintext sequence. For example, take the groups—

17–82–31–82–14–63, and
27–82–40–82–14–63

Here one might suspect that 17 and 27 represent the same letter, 31 and 40 another letter. But it happens that one group represents the word MANAGE, the other DAMAGE. There are hundreds of such cases in English and in other languages.

d. When reversible combinations are used as variants, the problem is perhaps a bit more

	K,Z	Q,V	B,H	M,R	D,L
W,S	N	H	A	O	E
F,X	D	T	M	F	P
G,J	Q	B	U	I	V
C,N	G	X	R	C	S
P,T	Z	L	Y	W	K

FIGURE 45.

simple. For example, using the accompanying Fig. 45 for encipherment, two messages with the same initial words, REFERENCE YOUR, may be enciphered as follows:

	R E F E R E N C E Y O U R					
(1)	N H W D R	X L S H C	D W W Z N	R S L H P	S R B J C	H
(2)	C H D W R	X S L H N	D W Z W N	R L S H P	R W J B N	H

The experienced cryptanalyst, noting the appearance of the very first few cipher groups, assumes that not only have the messages identical beginnings in their plain texts, but also that he is here confronted with a variant system involving biliteral reversible equivalents. One of the manifestations of such a cryptosystem is that in the digraphic distribution of the cipher text the "B" row will have an appearance similar to the "B" column, the "C" row will resemble the "C" column, etc.; thus the cryptanalyst will almost immediately realize that he has encountered a commutative system involving a matrix smaller than that indicated by the size of matrix necessary for making the digraphic distribution.

e. The probable-word method of solution may be used, but with a slight variation introduced because of the fact that, regardless of the system, *letters of low frequency in plain text remain infrequent in the cryptogram.* Hence, suppose a word containing low-frequency letters, but in itself a rather common word striking idiomorphic in character is sought as a "probable word"; for example, a word such as CAVALRY, ATTACK, or PREPARE. Such a word may be written on a slip of paper and slid one interval at a time under the text, which has been marked so that the high- and low-frequency characters are indicated. Each coincidence of a low-frequency letter of the text with a low-frequency letter of the assumed word is examined carefully to see whether the adjacent text letters correspond in frequency with the other letters of the assumed word, and whether there are correspondences between repetitions in the cipher text and those in the word. Many trials are necessary but this method will produce results when the difficulties are otherwise too much for the cryptanalyst to overcome.

62. Analysis involving the use of isologs.—*a.* In military communications it is not unusual that cryptograms are produced containing identical plain text but which have been subjected to different cryptographic treatment, thus yielding different cipher texts. This difference in cryptographic treatment may be caused by the use of an entirely different general system, or by the use of a different specific key, or merely by the choice of equivalents in a variant system. Messages which present different encrypted texts but which contain identical plain text are called *isologs* (from the Greek *isos*="equal" and *logos*="word"). One of the easily-noted indications of the possible presence of isologs is equality or near-equality in the lengths of two (or more) cryptograms. Isologs, no matter how the cryptographic treatment varies, are among the most powerful media available to the cryptanalyst for the successful solution of a difficult cryptosystem—and, in some cases, may provide the only possible entries into a complex cryptosystem. An inkling of the help afforded by isologs was revealed by the example contained in subpar. 61*d* above; however, a much more striking illustration is given in the next few subparagraphs.

b. The following two cryptograms, suspected to be isologs, are available for study:

Message "A"

8 2 2 6 5	6 3 1 0 3	7 4 8 3 9	6 9 8 4 2	3 2 5 2 9	7 0 1 1 5
8 0 2 7 7	8 9 1 0 6	9 4 0 0 0	1 3 8 2 8	5 4 0 8 2	4 0 0 6 5
6 3 6 2 9	3 3 9 1 8	4 3 1 5 8	8 1 0 4 8	2 6 4 5 8	4 5 0 3 9
8 1 7 1 3	5 2 5 3 8	7 3 3 0 9	2 0 7 4 9	6 1 7 5 2	1 6 4 7 6
3 8 7 2 8	9 1 1 4 7	9 9 9 2 6	4 1 4 6 8	1 3 3 6 5	3 3 8 8 1
8 9 6 9 7	9 3 8 1 6	5 1 7 5 0	5 7 0 7 4	1 1 8 0 4	4 3 2 5 5
2 8 1 2 0	2 7 7 3 0	3 1 1 9 9	7 9 9 6 2	2 7 8 6 5	6 0 6 5 3
9 0 8 7 0	4 0 8 6 7	4 6 5 9 4	1 9 8 5 5	1 0 8 2 2	2 2 9 8 7
4 6 7 2 9	3 6 2 4 5				

Message "B"

3 0 1 5 0	8 7 4 9 7	1 4 5 1 1	9 7 3 6 0	4 9 6 7 6	5 0 1 0 6
4 5 6 4 7	9 9 1 8 1	6 9 6 7 2	5 3 8 8 9	4 1 5 6 3	2 5 2 0 3
9 0 6 2 8	7 7 5 3 6	2 0 3 5 1	1 0 5 7 0	8 9 2 7 7	7 5 0 1 1
3 5 1 9 9	9 0 1 3 8	9 9 9 7 4	5 0 2 3 2	0 4 1 1 5	8 9 2 1 6
3 8 4 6 3	1 7 5 4 7	1 4 6 4 8	0 0 6 4 6	8 5 8 6 4	5 3 8 9 8
2 6 1 2 1	8 3 8 7 8	9 4 8 8 9	3 3 7 2 8	1 1 2 7 2	2 0 5 0 4
0 6 4 8 4	3 2 1 0 3	9 8 7 1 5	4 2 6 6 2	8 0 7 6 0	8 9 8 8 0
4 4 1 0 5	5 2 9 0 0	5 9 7 2 8	2 2 8 5 5	8 7 3 0 0	7 0 8 9 3
5 9 6 8 2	4 6 2 5 3				

CONFIDENTIAL

On the possibility that some dinome system (or systems) is involved, the messages are written under each other in dinomes to facilitate the examination of the *similarities and differences* of such a grouping of the cipher texts, as shown below:

					5					10					15
A	82	26	56	31	03	74	83	96	98	42	32	52	97	01	15
A'	30	15	08	74	97	14	51	19	73	60	49	67	65	01	06
B	80	27	78	91	06	94	00	01	38	28	54	08	24	00	65
B'	45	64	79	91	81	69	67	25	38	89	41	56	32	52	03
C	63	62	93	39	18	43	15	88	10	48	26	45	84	50	39
C'	90	62	87	75	36	20	35	11	05	70	89	27	77	50	11
D	81	71	35	25	38	73	30	92	07	49	61	75	21	64	76
D'	35	19	99	01	38	99	97	45	02	32	04	11	58	92	16
E	38	72	89	11	47	99	92	64	14	68	13	36	53	38	81
E'	38	46	31	75	47	14	64	80	06	46	85	86	45	38	98
F	89	69	79	38	16	51	75	05	70	74	11	80	44	32	55
F'	26	12	18	38	78	94	88	93	37	28	11	27	22	05	04
G	28	12	02	77	30	31	19	97	99	62	27	86	56	06	53
G'	06	48	43	21	03	98	71	54	26	62	80	76	08	98	80
H	90	87	04	08	67	46	59	41	98	55	10	82	22	29	87
H'	44	10	55	29	00	59	72	82	28	55	87	30	07	08	93
J	46	72	93	62	45										
J'	59	68	24	62	53										

The dinome distributions for the two messages are as follows:

	1	2	3	4	5	6	7	8	9	Ø
1	2	1	1	1	2	1	–	1	1	2
2	1	1	–	1	1	2	2	2	1	–
3	2	2	–	–	1	1	–	5	2	2
4	1	1	1	1	2	2	1	1	1	–
5	1	1	2	1	2	2	–	–	1	1
6	1	3	1	2	1	–	1	1	1	–
7	1	2	1	2	2	1	1	1	1	1
8	2	2	1	1	–	1	2	1	2	2
9	1	2	2	1	–	1	2	2	2	1
Ø	2	1	1	1	1	2	1	2	–	2

Distribution for Message "A"

	1	2	3	4	5	6	7	8	9	Ø
1	4	1	–	2	1	1	–	1	2	1
2	1	1	–	1	1	2	2	2	1	1
3	1	2	–	–	2	1	1	5	–	2
4	1	–	1	1	3	2	1	1	1	–
5	1	1	1	1	2	1	–	1	2	1
6	–	3	–	2	1	–	2	1	1	1
7	1	1	1	1	2	1	1	1	1	1
8	1	1	–	–	1	1	2	1	2	3
9	1	1	2	1	–	–	2	3	2	1
Ø	2	1	2	2	2	3	1	3	–	1

Distribution for Message "B"

c. Since a general absence of marked crests and troughs is noted in both distributions, if the division of these cryptograms into dinomes is correct, *and* if they are both monoalphabetic, it is quite probable that some type of variant system (or systems) has been used. With this in mind, the encrypted texts and their distributions are scrutinized further for some indication of the kind of relationship which exists between the methods of encipherment of the two messages. The distributions are seen to be strikingly similar, not only with respect to the location of the one predominant peak in each, but also in the close correlation of the locations of the blanks in each.[14] Furthermore, upon examination of the superimposed messages themselves, it is

[14] For the benefit of the student with a statistical background, it might be interesting to point out certain applications of cryptomathematics in connection with these two distributions. First of all, each of the two distributions is *much flatter* than that which would be expected for a sample of 125 dinomes of *random* text; i. e., a drawing (with replacement) and recording from an urn containing equal numbers of counters in each of 100 categories labeled 00–99 consecutively. That is, the samples at hand exhibit phenomena even *flatter* (or "worse") than that expected for random, approaching the theoretical (and fantastically nonrandom) "equilibrium" of exactly the same number of tallies in each cell of a distribution. The following table gives the observed number of x-fold repetitions in the two distributions, together with the expected numbers of x-fold repetitions in a sample of like size of random text, which expected numbers have been computed from tables of the Poisson exponential distribution (see *Military Cryptanalytics, Part III*):

x	Observed Msg. "A"	Observed Msg. "B"	Expected
Ø	14	17	29
1	51	52	36
2	33	23	22
3	1	6	9
4	–	1	3
5	1	1	1

It is to be noted that in the distribution for Message "A", the observed number of blanks (14) when compared with the expected number of blanks in random text (29) may be evaluated and found to represent a very small probability indeed. Likewise, the other entries besides Ø (in particular, the x-values of 1 and 2, and the *cumulative* values of 3-and-better) may be evaluated, and the conclusion would be reached that the two distributions have a most remote chance of being as flat as they are through mere chance. Moreover, the observed frequency distribution for Message "A" may be fitted against the expected distribution by means of the chi-square test, again getting an extremely small probability. In addition, by means of the chi-square test, the I. C. of Message "A" (found to be 0.59 as against the I. C. of random of 1.0) has an extremely small probability of occurring at random in a sample of this size. Similarly, the distribution for Message "B" could be studied, and it would be found that this too has characteristics that have a very small probability of occurrence by pure chance. Since the distributions of the two messages are much worse than would even be expected for random chance, the conclusion is drawn that the *dinome grouping* is highly significant and therefore *must* be correct, and further that the cryptosystem involves variants in sufficient numbers for the plaintext letters to permit the encipherer to select the cipher equivalents with a view to suppressing as much of the phenomena of repetition as possible. Furthermore, the χ test of the two distributions gives a χ value of 206, as against the expected χ value of 156 for a random matching of these two samples; the sigmage of this event could be computed and its signifiance estimated and the conclusion drawn that the ratio $\frac{206}{156}$ is extremely unlikely of happening by pure chance, i. e., if the cryptograms were not in the same general system and specific keys. Therefore, it is a foregone conclusion *statistically* that not only do the cryptosystems involve dinomes as the ciphertext grouping, but that the identical cryptosystem is involved in the two messages; and that because of the close correlation of the patterns of the two distributions, there is a good probability that the cryptograms contain identical plain text and therefore are isologs. This specific illustration of the potentialities of cryptomathematics indicates the important role that this branch of science may play in the art cryptanalysis.

observed that there are several instances wherein a value in Message "A" coincides with the same value in Message "B" (e. g., see positions A/A' 14, B/B' 9). This observation, taken in conjunction with the marked similarity of the distributions, strongly indicates that not only has the same *general* cryptosystem been used for the encryption of both messages, but that the *same enciphering matrix* has been used for both. Also, in the case of the value 38 and 62, it is noted that wherever either occurs in one message the same value occurs in the other message, a phenomenon explainable on the assumption that the plaintext equivalents of these values are of such low frequency that no variant values have been provided for these plaintext letters in the cryptosystem.

d. With the foregoing details determined, it is now realized that it should be possible to form, between the two messages, "chains" of those cipher values which represent identical plaintext letters, as exemplified below. Beginning with the first value in each message, 82 and 30, a partial chain of equivalent variants is started; now locating some other occurrence of either value elsewhere (e. g., 82 at position H'8), and noting the cipher value coinciding with it (in this case, 41), the partial chain may be extended (including now 82, 30, and 41). After this particular chain is extended to include as many values as possible, another chain is formed by starting with any value which has not already been included in the preceding chain, this procedure being repeated until all possible chains are completed. It is found that the following chains, arbitrarily arranged here according to length, may be derived from the two messages:

(06 14 15 26 28 31 35 73 74 81 89 98 99)
(02 07 20 22 43 44 63 90)
(12 37 48 51 69 70 83 94)
(03 30 41 54 65 82 97)
(05 10 24 32 49 87 93)
(16 18 36 76 78 79 86)
(27 45 53 64 80 92)
(11 39 75 88)
(21 58 77 84)
(46 59 68 72)
(00 52 67)
(04 55 61)
(08 29 56)
(19 71 96)
(01 25)
(13 85)
(42 60)

Single dinomes:

(38) (47) (50) (62) (91)

If we now make an arbitrary assignment of a different letter to represent each chain (and one for each single dinome) and convert either of the messages to uniliteral terms by means of these arbitrarily-assigned values, we note the pattern of the opening stereotype "REFERENCE YOUR MESSAGE.....", and quickly recover the plain text.

e. The plaintext values when inserted into a 10 x 10 matrix having arbitrarily-arranged coordinates yield the following:

	1	2	3	4	5	6	7	8	9	Ø
1	D	N	H	E	E	A	–	A	C	O
2	I	T	–	O	M	E	S	E	F	T
3	E	O	–	–	E	A	N	B	D	R
4	R	Y	T	T	S	L	V	N	O	–
5	N	U	S	R	P	F	–	I	L	X
6	P	W	T	S	R	–	U	L	N	Y
7	C	L	E	E	D	A	I	A	A	N
8	E	R	N	I	H	A	O	D	E	S
9	G	S	O	N	–	C	R	E	E	T
Ø	M	T	R	P	O	E	T	F	–	U

Manipulating the rows and columns with a view to uncovering some symmetry or systematic phenomena, the latent diagonal pattern of the equivalents for certain of the letters (such as E_p, N_p, O_p, R_p, and S_p) is revealed, and the rows and columns of the reconstruction diagram are permuted to yield the following original enciphering matrix:

	6	8	9	1	5	4	3	7	2	Ø
7	A	A	A	C	D	E	E	I	L	N
1	A	A	C	D	E	E	H	(K)	N	O
3	A	B	D	E	E	(H)	(J)	N	O	R
8	A	D	E	E	H	I	N	O	R	S
9	C	E	E	G	(I)	N	O	R	S	T
2	E	E	F	I	M	O	(Q)	S	T	T
Ø	E	F	I	M	O	P	R	T	T	U
5	F	I	L	N	P	R	S	(T)	U	X
6	(I)	L	N	P	R	S	T	U	W	Y
4	L	N	O	R	S	T	T	V	Y	Z

There are no observable relationships in or between the sequences of digits in the row and column coordinates; therefore for want of any visible phenomena or further information on the derivation (if any) of these digits, it is assumed that they must have been assigned at random. The student will note that the final matrix is identical to that of Fig. 39 in par. 59.

f. It should be emphasized that in the example of the preceding subparagraphs it was only possible to form chains of values from both messages *reciprocally* because the same enciphering matrix had been used for both. A nonreciprocal chaining procedure would have been required if only the general system had been the same for both but the enciphering matrices had differed in some respect, or if two completely different variant systems had been used (e. g., one using a frequential matrix and the other involving a less complex type of variant matrix, such as Fig. 29). Specifically, it would have been necessary to maintain two separate groups of chains, one group for each message; otherwise heterogeneous values would have become intermingled. For instance, if the two messages on p. 121 had been enciphered with two different matrices, then we would build up chains of equivalencies in Message "B" against *one* value of Message "A", and, likewise, chains of equivalencies in Message "A" against *one* value of Message "B". Thus, we note at position A2 we have $\genfrac{}{}{0pt}{}{26}{15}$, and at position C11 we have $\genfrac{}{}{0pt}{}{26}{89}$; this means that 15 and 89 are in one chain *in Message "B"*. Likewise, the $\genfrac{}{}{0pt}{}{28}{89}$ at position B1Ø and the $\genfrac{}{}{0pt}{}{26}{89}$ at position C11 demonstrates that 26 and 28 are in one chain *in Message "A"*. This process would of course be continued so as to expand chains wherever possible.

g. Although an analysis of but one isolated example by means of isologs was presented, the student should be able to appreciate the significance and potentially enormous value of isologs to a cryptanalyst. This value goes far beyond the simple variant encryption in a monoalphabetic substitution system; isologs produced by the use of two different code books, or two different enciphered code versions of the same underlying plain text, or two encryptions of identical plain text by two different "settings" of a cipher machine, may all prove of inestimable value in the attack on a difficult cryptosystem.

63. Further remarks on variant systems.—*a.* A few words should be added with regard to certain subterfuges which are sometimes encountered in monoalphabetic substitution with variants, and which, if not recognized in time, cause considerable delays. The considerations treated before in subpars. 52*i* and *j* on the disguise of the length of the basic multiliteral group apply equally here to multiliteral substitution with variants; thus, in dinome systems, a sum-checking digit or a null might be added in specified positions of the group to form a trinome. In complex variant systems, the presence of a *null* as one of the digits of a trinome would add greatly to the complexities of cryptanalysis of that system. The most important of the subterfuges have to deal with the use of nulls which are of a *different size than the real cryptographic units*, inserted occasionally to prevent the cryptanalyst from breaking up the text into its proper units. The student should take careful note of the last phrase; the mere insertion of symbols having the same characteristics as the symbols of the cryptographic text, except that they have no meaning, is not what is meant. *This* class of nulls rarely achieves the purpose intended. What is really meant can best be explained by an example. Suppose that a 5 x 5 variant matrix with the row and column indicators shown in Fig. 46 is adopted for encipherment. Normally, the cipher units would consist of 2-letter combinations of the indicators, invariably giving the row indicator first (by agreement).

	V A T F	G H O U	I P E R	W S B L	D M N C
V A T F	A	B	C	D	E
G H O U	F	G	H	IJ	K
I P E R	L	M	N	O	P
W S B L	Q	R	S	T	U
D M N C	V	W	X	Y	Z

FIGURE 46.

The phrase "COMMANDER OF SPECIAL TROOPS" might be enciphered thus:

C	O	M	M	A	N	D	E	R	O	F	. . .
VI	EB	PH	IU	FT	IE	AB	TM	WO	PW	GT	. . .

These would normally then be arranged in 5-letter groups, thus:

V I E B P   H I U F T   I E A B T   M W O P W   G T . . .

b. It will be noted, however, that only 20 of the 26 letters of the alphabet have been employed as row and column indicators, leaving J, K, Q, X, Y, and Z unused. Now, suppose these six letters are used as nulls, *not in pairs, but as individual letters inserted at random* just before the real text is arranged in 5-letter groups. Occasionally, a *pair* of letters might be inserted, in order to mask the characteristics of "avoidance" of these letters for each other. Thus, for example:

V I E X B   P H K I U   F J X T I   E A J B T   M W O Q P   W G K T Y

The cryptanalyst, after some study suspecting a biliteral cipher, proceeds to break up the text into pairs:

VI EX BP HK IU FJ XT IE AJ BT MW OQ PW GK TY

Compare this set of 2-letter combinations with the correct set. Only 4 of the 15 pairs are "proper" units. It is easy to see that without a knowledge of the *existence* of the nulls—and even with a knowledge, if he does not know *which* letters are nulls—the cryptanalyst would be confronted with a problem for the solution of which a fairly large amount of text might be necessary. The careful employment of the variants also very materially adds to the security of the method because repetitions can be rather effectively suppressed.

c. Similarly in the examples under par. 58, the letter J in Figs. 27 and 29 may be used as a null; the letter Y in Fig. 28; and the digit Ø in Figs. 33 and 34. In Fig. 30, any letters in the range of P–Z might be used as nulls, but this usage would be weak because of the extremely low frequency of these letters as compared with the letters A–O; this is an important point to consider in the examination of encrypted text for possible poor usages of nulls.

d. From the cryptographic standpoint, usage of nulls in the manner outlined above results in cryptographic text even more than twice as long as the plain text, thus constituting a serious disadvantage. From the cryptanalytic standpoint, the masking of the cipher units in the system described in subpar. *b* above constitutes the most important obstacle to solution; this, coupled with the use of variants, makes this system considerably more difficult to solve, despite its monoalphabeticity.

CHAPTER IX

POLYGRAPHIC SUBSTITUTION SYSTEMS

64. General remarks on polygraphic substitution.—*a.* The substitution systems dealt with thus far have involved plaintext units consisting of single elements (usually single letters). The major distinction between them has been made simply on the basis of the number of elements constituting the *ciphertext units* of each; i. e., those involving single-element ciphertext units were termed *uniliteral,* and those involving ciphertext units composed of two or more elements were termed *multiliteral.*[1] That is to say, when the terms "uniliteral", "biliteral", "triliteral", etc., were used, it was to have been inferred automatically that the plaintext units were composed of single elements.

b. This chapter of the text will deal with substitution systems involving *plaintext units* composed of more than one element; such systems are termed *polygraphic.*[2] (By comparing this new term with the terms "uniliteral " and "multiliteral" it may then be deduced—and correctly so—that a term involving the suffix "-literal" is descriptive of the composition of the ciphertext units of a cryptosystem, and that a term containing the suffix "-graphic" describes the composition of the plaintext units.[3]) Polygraphic systems in which the plaintext units are composed of two elements are called *digraphic,* those in which the plaintext units are composed of three elements are *trigraphic,* etc. The *ciphertext* units of polygraphic systems usually consist of the same number of elements as the plaintext units.[4] Thus, if a system is called "digraphic", it may be assumed that the ciphertext units of the system consist of two elements, as do the plaintext units; if this were not the case, the term "digraphic" by itself would not be adequate to describe the

[1] See also subpar. 52*a*.

[2] Systems involving plaintext units composed of single elements may, on this basis, be termed monographic; however, as has been stated in connection with the terms "uniliteral" and "multiliteral", the plaintext units of a system are understood (without restatement) to be monographic unless otherwise specified.

[3] In this connection, it is further pointed out that since the root "literal" derives from the Latin "litera", it is conventionally prefixed by modifiers of Latin origin, such as "uni-", "bi-", and "multi-"; similarly, "graphic", deriving from the Greek "graphikos", is prefixed by modifiers of Greek origin, such as "mono-", "di-", and "poly-".

[4] The qualifying adverb "usually" is employed because this correspondence is not essential. For example, if one should draw up a set of 676 arbitrary single signs, it would be possible to represent the 2-letter pairs from AA to ZZ by single symbols. This would still be a digraphic system.

system completely, and an additional modifying word or phrase would have to be used to indicate this fact.[5]

c. In polygraphic substitution, the *combinations* of elements which constitute the plaintext units are considered as indivisible compounds. The units are composite in character and the individual elements composing the units affect the equivalent cipher units *jointly*, rather than separately. The basic important factor in true polygraphic substitution is that *all* the letters of each plaintext unit participate in the determination of its cipher equivalent; the identity of *each* element of the plaintext unit affects the composition of the *whole* cipher unit.[6] Thus, in a certain digraphic system, $\overline{AB}_p$ may be enciphered as $\overline{XP}_c$; and $\overline{AC}_p$, on the other hand, may be enciphered as $\overline{NK}_c$; a difference in the identity of but one of the letters of the plaintext pair here produces a difference in the identity of *both* letters of the cipher pair.[7]

d. The fundamental purpose of polygraphic substitution is again the suppression or the elimination of the frequency characteristics of single letters of plain text, just as is the case in monoalphabetic substitution with variants; but here this is accomplished by a different method, the latter arising from a somewhat different approach to the problem involved in producing cryptographic security. When the substitution involves replacement of *single* letters in a monoalphabetic system, even a single cryptogram can be solved rather readily; basically the reason for this is that the principles of frequency and the laws of probability, applied to individual units (single letters) of the plain text, have a very good opportunity to manifest themselves. However, when the substitution involves replacement of plaintext units composed of two or more letters—that is, when the substitution is polygraphic in nature—the principles of frequency and laws of probability have a much lesser opportunity to manifest themselves. If the substitution is digraphic, then the units are pairs of letters and the normal frequencies of plaintext *digraphs* become of first consideration; if the substitution is trigraphic, the units are sets of three letters and the normal frequencies of plaintext trigraphs are involved. In these cases the data that can be employed in the solution are meager; that is why, generally speaking, the solution of polygraphic substitution ciphers is often extremely difficult.

e. By way of example, a given plaintext message of say N letters, enciphered by means of a uniliteral substitution system, affords N cipher characters, and the same number of cipher units. The same message, enciphered digraphically, still affords N cipher characters but only $\frac{N}{2}$ cipher units. Statistically speaking, the sample to which the laws of probability now are to be applied has been cut in half. Furthermore, from the point of view of frequency, the very noticeable diversity in the frequencies of individual letters, leading to the marked crests and troughs of the uniliteral frequency distribution, is no longer so strikingly in evidence in the frequencies of digraphs. Therefore, although digraphic encipherment, for example, simply cuts the cryptographic textual units in half, the number of cipher units which must be identified has been *squared*; and the difficulty of solution is not merely doubled but, if a matter of judgment arising

[5] See subpars. 65*e* and 66*f* for examples of two such systems and their names.

[6] An analogy is found in chemistry, when two elements combine to form a molecule, the latter usually having properties quite different from those of either of the constituent elements. For example: sodium, a metal, and chlorine, a gas, combine to form sodium chloride, common table salt. However, sodium and fluorine, also a gas similar in many respects to chlorine, combine to form sodium fluoride, which is much different from table salt.

[7] For this reason the two letters are marked by a ligature; that is, by a bar across their tops. In cryptologic notation, the symbol $\overline{\theta\theta}_p$ means "any plaintext digraph", the symbol $\overline{\theta\theta}_c$, "any ciphertext digraph". To refer specifically to the 1st, 2d, 3d, . . . member of a ligature, the exponent 1, 2, 3, . . . will be used. Thus θ_p^2 of $\overline{REM}_p$ is the letter E; θ_c^3 of $\overline{XRZ}_c$ is Z. See also footnote 1 on p. 47.

from practical experience can be expressed or approximated mathematically, squared or cubed.

f. The following two paragraphs will treat various polygraphic substitution methods. The most practical of these methods are digraphic in character and for this reason their treatment herein will be more detailed than that of trigraphic methods.

65. Polygraphic substitution methods employing large tables.—*a*. The simplest method of effecting polygraphic substitution involves the use of tables similar to that shown in Fig. 47*a*. This table merely provides equivalents for digraphs, by means of the coordinate system. Specifically, in obtaining the cipher equivalent of any plaintext digraph, the initial letter of the plaintext digraph is used to indicate the row in which the equivalent is found, and the final letter of the plaintext digraph indicates the column; the cipher digraph is then found at the intersection of the row and column thus indicated. For example, $\overline{KG}_p=\overline{FC}_c$; $\overline{WM}_p=\overline{OY}_c$; etc.

θ^2_p

θ^1_p	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	WG	EE	SN	TR	IA	NL	GC	HT	OI	UO	AM	RP	BY	KB	CD	DF	FH	JJ	LK	MQ	PS	QU	VV	XW	YX	ZZ
B	EG	SE	TN	IR	NA	GL	HC	OT	UI	AO	RM	BP	KY	CB	DD	FF	JH	LJ	MK	PQ	QS	VU	XV	YW	ZX	WZ
C	SG	TE	IN	NR	GA	HL	OC	UT	AI	RO	BM	KP	CY	DB	FD	JF	LH	MJ	PK	QQ	VS	XU	YV	ZW	WX	EZ
D	TG	IE	NN	GR	HA	OL	UC	AT	RI	BO	KM	CP	DY	FB	JD	LF	MH	PJ	QK	VQ	XS	YU	ZV	WW	EX	SZ
E	IG	NE	GN	HR	OA	UL	AC	RT	BI	KO	CM	DP	FY	JB	LD	MF	PH	QJ	VK	XQ	YS	ZU	WV	EW	SX	TZ
F	NG	GE	HN	OR	UA	AL	RC	BT	KI	CO	DM	FP	JY	LB	MD	PF	QH	VJ	XK	YQ	ZS	WU	EV	SW	TX	IZ
G	GG	HE	ON	UR	AA	RL	BC	KT	CI	DO	FM	JP	LY	MB	PD	QF	VH	XJ	YK	ZQ	WS	EU	SV	TW	IX	NZ
H	HG	OE	UN	AR	RA	BL	KC	CT	DI	FO	JM	LP	MY	PB	QD	VF	XH	YJ	ZK	WQ	ES	SU	TV	IW	NX	GZ
I	OG	UE	AN	RR	BA	KL	CC	DT	FI	JO	LM	MP	PY	QB	VD	XF	YH	ZJ	WK	EQ	SS	TU	IV	NW	GX	HZ
J	UG	AE	RN	BR	KA	CL	DC	FT	JI	LO	MM	PP	QY	VB	XD	YF	ZH	WJ	EK	SQ	TS	IU	NV	GW	HX	OZ
K	AC	RE	BN	KR	CA	DL	FC	JT	LI	MO	PM	QP	VY	XB	YD	ZF	WH	EJ	SK	TQ	IS	NU	GV	HW	OX	UZ
L	RG	BE	KN	CR	DA	FL	JC	LT	MI	PO	QM	VP	XY	YB	ZD	WF	EH	SJ	TK	IQ	NS	GU	HV	OW	UX	AZ
M	BG	KE	CN	DR	FA	JL	LC	MT	PI	QO	VM	XP	YY	ZB	WD	EF	SH	TJ	IK	NQ	GS	HU	OV	UW	AX	RZ
N	KG	CE	DN	FR	JA	LL	MC	PT	QI	VO	XM	YP	ZY	WB	ED	SF	TH	IJ	NK	GQ	HS	OU	UV	AW	RX	BZ
O	CG	DE	FN	JR	LA	ML	PC	QT	VI	XO	YM	ZP	WY	EB	SD	TF	IH	NJ	GK	HQ	OS	UU	AV	RW	BX	KZ
P	DG	FE	JN	LR	MA	PL	QC	VT	XI	YO	ZM	WP	EY	SB	TD	IF	NH	GJ	HK	OQ	US	AU	RV	BW	KX	CZ
Q	FG	JE	LN	MR	PA	QL	VC	XT	YI	ZO	WM	EP	SY	TB	ID	NF	GH	HJ	OK	UQ	AS	RU	BV	KW	CX	DZ
R	JG	LE	MN	PR	QA	VL	XC	YT	ZI	WO	EM	SP	TY	IB	ND	GF	HH	OJ	UK	AQ	RS	BU	KV	CW	DX	FZ
S	LG	ME	PN	QR	VA	XL	YC	ZT	WI	EO	SM	TP	IY	NB	GD	HF	OH	UJ	AK	RQ	BS	KU	CV	DW	FX	JZ
T	MG	PE	QN	VR	XA	YL	ZC	WT	EI	SO	TM	IP	NY	GB	HD	OF	UH	AJ	RK	BQ	KS	CU	DV	FW	JX	LZ
U	PG	QE	VN	XR	YA	ZL	WC	ET	SI	TO	IM	NP	GY	HB	OD	UF	AH	RJ	BK	KQ	CS	DU	FV	JW	LX	MZ
V	QG	VE	XN	YR	ZA	WL	EC	ST	TI	IO	NM	GP	HY	OB	UD	AF	RH	BJ	KK	CQ	DS	FU	JV	LW	MX	PZ
W	VG	XE	YN	ZR	WA	EL	SC	TT	II	NO	GM	HP	OY	UB	AD	RF	BH	KJ	CK	DQ	FS	JU	LV	MW	PX	QZ
X	XG	YE	ZN	WR	EA	SL	TC	IT	NI	GO	HM	OP	UY	AB	RD	BF	KH	CJ	DK	FQ	JS	LU	MV	PW	QX	VZ
Y	YG	ZE	WN	ER	SA	TL	IC	NT	GI	HO	OM	UP	AY	RB	BD	KF	CH	DJ	FK	JQ	LS	MU	PV	QW	VX	XZ
Z	ZG	WE	EN	SR	TA	IL	NC	GT	HI	OO	UM	AP	RY	BB	KD	CF	DH	FJ	JK	LQ	MS	PU	QV	VW	XX	YZ

$\overline{\theta\theta}_c$

FIGURE 47*a*.

b. In the preceding table two mixed sequences were employed to form the cipher equivalents, one sequence being based on the key phrase WESTINGHOUSE AIR BRAKE and the other on GENERAL ELECTRIC COMPANY. The table in Fig. 47*a* could have been drawn up in a slightly different manner, as shown in Fig. 47*b*, and still yield the same cipher equivalents as before. Using this latter table, θ^1_c for any plaintext digraph is found at the intersection of the row and column identified by θ^1_p and θ^2_p, respectively; θ^2_c, is found in the sequence below the table and is taken from the position directly under the column identified by θ^2_p. A few trial encipherments will illustrate that this table is cryptographically equivalent to that of Fig. 47*a*.

θ_p^2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	
B	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	
C	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	
D	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	
E	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	
F	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	
G	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	
H	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	
I	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	
J	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	
K	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	
L	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	
θ_p^1 M	B	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	θ_c^1
N	K	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	
O	C	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	
P	D	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	
Q	F	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	
R	J	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	
S	L	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	
T	M	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	
U	P	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	
V	Q	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	
W	V	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	
X	X	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	
Y	Y	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	
Z	Z	W	E	S	T	I	N	G	H	O	U	A	R	B	K	C	D	F	J	L	M	P	Q	V	X	Y	
θ_c^2	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	

FIGURE 47*b*.

c. Figs. 48 and 49, below, contain other possible types of tables for digraphic substitution. In Fig. 48, it will be seen that there are *two* vertical sequences to the left of this table and no horizontal sequence below it. θ_p^1 is located in the leftmost sequence, θ_c^1 being found directly to its side in the right-hand sequence; θ_c^2 is then found at the intersection of the row and column identified by θ_p^1 and θ_p^2, respectively. The table in Fig. 49 provides digraphic equivalents by means of the coordinate system (e. g., $\overline{RE}_p=\overline{JZ}_c$), in the same manner as in Fig. 47*a*, and a cursory examination of the inside of the table might disclose nothing new about this table at all. But, if one were to scan closely the diagonals formed by each θ_c^1 from upper right to lower left, he would see that each such diagonal changes below the "M_p row"; similarly, if the diagonals formed by θ_c^2 are scanned from upper left to lower right, it will be seen that each of them also changes after the "M_p row". In effect, the inside of the table is divided into two separate portions by an imaginary line extending horizontally between the M and N rows; but within each portion a straightforward type of symmetry is exhibited and the same two mixed sequences have been employed in each. Actually, in a 26 x 26 table, it is *not* possible to maintain the diagonals formed thus by θ_c^1 and θ_c^2 in a completely "unbroken" sequence without producing repeated digraphs within the table and without consequent cryptographic ambiguity; thus, Fig. 49 illustrates one type of limited diagonal symmetry which must be resorted to in the systematic construction of such a table.

θ_p^2

θ_p^1	θ_c^1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	W	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z
B	E	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G
C	S	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E
D	T	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N
E	I	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R
F	N	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A
G	G	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L
H	H	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C
I	O	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T
J	U	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I
K	A	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O
L	R	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M
M	B	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P
N	K	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y
O	C	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B
P	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D
Q	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F
R	J	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H
S	L	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J
T	M	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K
U	P	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q
V	Q	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S
W	V	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U
X	X	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V
Y	Y	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W
Z	Z	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X

θ_c^2

FIGURE 48.

d. All of the foregoing tables have exhibited a symmetry in the arrangement of their contents, which is undesirable from the standpoint of cryptographic security. This systematic internal arrangement could be detected by a cryptanalyst early in his attack on cryptograms produced through their use, permitting rapid reconstruction of the particular table involved; this subject will be given a more detailed treatment in par. 72. The table in Fig. 50 is an example of one type of table which would provide more security than the foregoing. This table is constructed by random assignment of values and shows no symmetry whatsoever in its arrangement of contents. It will be noted that this table is reciprocal in nature; that is $\overline{AF}_p=\overline{YG}_c$ and $\overline{YG}_p=\overline{AF}_c$. Thus, this single table serves for deciphering as well as for enciphering. Reciprocity is, however, not an essential factor; in fact, greater security is provided by nonreciprocal tables. But, in the case of such nonreciprocal, randomly constructed tables, each enciphering table must have its complementary deciphering table.

~~CONFIDENTIAL~~

θ_D^2

θ_D^1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	W Z	E X	S W	T V	I U	N S	G Q	H K	O J	U H	A F	R D	B B	K Y	C P	D M	F O	J I	L T	M C	P L	Q A	V R	X N	Y E	Z G
B	E G	S Z	T X	I W	N V	G U	H S	O Q	U K	A J	R H	B F	K D	C B	D Y	F P	J M	L O	M I	P T	Q C	V L	X A	Y R	Z N	W E
C	S E	T G	I Z	N X	G W	H V	O U	U S	A Q	R K	B J	K H	C F	D D	F B	J Y	L P	M M	P O	Q I	V T	X C	Y L	Z A	W R	E N
D	T N	I E	N G	G Z	H X	O W	U V	A U	R S	B Q	K K	C J	D H	F F	J D	L B	M Y	P P	Q M	V O	X I	Y T	Z C	W L	E A	S R
E	I R	N N	G E	H G	O Z	U X	A W	R V	B U	K S	C Q	D K	F J	J H	L F	M D	P B	Q Y	V P	X M	Y O	Z I	W T	E C	S L	T A
F	N A	G R	H N	O E	U G	A Z	R X	B W	K V	C U	D S	F Q	J K	L J	M H	P F	Q D	V B	X Y	Y P	Z M	W O	E I	S T	T C	I L
G	G L	H A	O R	U N	A E	R G	B Z	K X	C W	D V	F U	J S	L Q	M K	P J	Q H	V F	X D	Y B	Z Y	W P	E M	S O	T I	I T	N C
H	H C	O L	U A	A R	R N	B E	K G	C Z	D X	F W	J V	L U	M S	P Q	Q K	V J	X H	Y F	Z D	W B	E Y	S P	T M	I O	N I	G T
I	O T	U C	A L	R A	B R	K N	C E	D G	F Z	J X	L W	M V	P U	Q S	V Q	X K	Y J	Z H	W F	E D	S B	T Y	I P	N M	G O	H I
J	U I	A T	R C	B L	K A	C R	D N	F E	J G	L Z	M X	P W	Q V	V U	X S	Y Q	Z K	W J	E H	S F	T D	I B	N Y	G P	H M	O O
K	A O	R I	B T	K C	C L	D A	F R	J N	L E	M G	P Z	Q X	V W	X V	Y U	Z S	W Q	E K	S J	T H	I F	N D	G B	H Y	O P	U M
L	R M	B O	K I	C T	D C	F L	J A	L R	M N	P E	Q G	V Z	X X	Y W	Z V	W U	E S	S Q	T K	I J	N H	G F	H D	O B	U Y	A P
M	B P	K M	C O	D I	F T	J C	L L	M A	P R	Q N	V E	X G	Y Z	Z X	W W	E V	S U	T S	I Q	N K	G J	H H	O F	U D	A B	R Y
N	Z Z	Y G	X E	V N	Q R	P A	M L	L C	J T	F I	D O	C M	K P	B Y	R B	A D	U F	O H	H J	G K	N Q	I S	T U	S V	E W	W X
O	Y X	X Z	V G	Q E	P N	M R	L A	J L	F C	D T	C I	K O	B M	R P	A Y	U B	O D	H F	G H	N J	I K	T Q	S S	E U	W V	Z W
P	X W	V X	Q Z	P G	M E	L N	J R	F A	D L	C C	K T	B I	R O	A M	U P	O Y	H B	G D	N F	I H	T J	S K	E Q	W S	Z U	Y V
Q	V V	Q W	P X	M Z	L G	J E	F N	D R	C A	K L	B C	R T	A I	U O	O M	H P	G Y	N B	I D	T F	S H	E J	W K	Z Q	Y S	X U
R	Q U	P V	M W	L X	J Z	F G	D E	C N	K R	B A	R L	A C	U T	O I	H O	G M	N P	I Y	T B	S D	E F	W H	Z J	Y K	X Q	V S
S	P S	M U	L V	J W	F X	D Z	C G	K E	B N	R R	A A	U L	O C	H T	G I	N O	I M	T P	S Y	E B	W D	Z F	Y H	X J	V K	Q Q
T	M Q	L S	J U	F V	D W	C X	K Z	B G	R E	A N	U R	O A	H L	G C	N T	I I	T O	S M	E P	W Y	Z B	Y D	X F	V H	Q J	P K
U	L K	J Q	F S	D U	C V	K W	B X	R Z	A G	U E	O N	H R	G A	N L	I C	T T	S I	E O	W M	Z P	Y Y	X B	V D	Q F	P H	M J
V	J J	F K	D Q	C S	K U	B V	R W	A X	U Z	O G	H E	G N	N R	I A	T L	S C	E T	W I	Z O	Y M	X P	V Y	Q B	P D	M F	L H
W	F H	D J	C K	K Q	B S	R U	A V	U W	O X	H Z	G G	N E	I N	T R	S A	E L	W C	Z T	Y I	X O	V M	Q P	P Y	M B	L D	J F
X	D F	C H	K J	B K	R Q	A S	U U	O V	H W	G X	N Z	I G	T E	S N	E R	W A	Z L	Y C	X T	V I	Q O	P M	M P	L Y	J B	F D
Y	C D	K F	B H	R J	A K	U Q	O S	H U	G V	N W	I X	T Z	S G	E E	W N	Z R	Y A	X L	V C	Q T	P I	M O	L M	J P	F Y	D B
Z	K B	B D	R F	A H	U J	O K	H Q	G S	N U	I V	T W	S X	E Z	W G	Z E	Y N	X R	V A	Q L	P C	M T	L I	J O	F M	D P	C Y

$\theta\theta_c$

FIGURE 49.

e. Digraphic tables employing numerical equivalents instead of letter equivalents may be encountered. However, since 676 equivalents are required (there being 676, or 26 x 26 different pairs of letters), this means that combinations of three figures must be used; such systems are termed *trinome-digraphic* systems, indicating clearly the number of elements which comprise

(Showing only a partially filled table)

θ_p^2

θ_p^1	A	B	C	D	E	F	G	H	I	J	K	. . .	X	Y	Z	
A	FX	CH	XE	YY	ZA	YG	FB	CD	EF	XJ	ZX	. . .	EA	DJ	FH	A
B	NY	DC	NB	ZI	XX	DX						. . .				B
C	—			AH				AB				. . .		ND		C
D			BB		YA					AY		. . .	BF			D
E	AX					AI						. . .				E
F		AG			NZ			AZ				. . .	AA			F
:	:	:	:	:	:	:	:	:	:	:	:	. . .	:	:	:	
N		BC		CY								. . .		BA	FE	N
:	:	:	:	:	:	:	:	:	:	:	:	. . .	:	:	:	
X					AC					AJ		. . .	BE			X
Y	DE						AF					. . .		AD		Y
Z	AE								BD			. . .	AK			Z
	A	B	C	D	E	F	G	H	I	J	K		X	Y	Z	

FIGURE 50.

the cipher units. By way of an example, the following figure contains a fragment of a table [8] which provides trinome equivalents for the plaintext digraphs:

	J	U	P	I	T	E	R	A	B		X	Y	Z
V	001	002	003	004	005	006	007	008	009	... ...	024	025	026
E	027	028	029	030	031	032	033	034	035		050	051	052
N	053	054	055	056	057								
U	079	080	081	082									
S	105	106	107										
A	131	132											
B	157												
	...												
	...												
	...												
X	599	600									622	623	624
Y	625	626									648	649	650
Z	651	652								... ...	674	675	676

FIGURE 51.

[8] It is interesting to note that this comparatively bulky and unwieldy table can be reduced to the following two alphabets with numerical equivalents for the letters:

(1)	V	E	N	U	S	A	B	. . . . . .	X	Y	Z
	000	026	052	078	104	130	156	... ...	598	624	650
(2)	J	U	P	I	T	E	R	. . . . . .	X	Y	Z
	1	2	3	4	5	6	7	... ...	24	25	26

In enciphering, the first letter of the plaintext digraph is converted into its numerical value from alphabet (1), and the second plaintext letter is converted by means of alphabet (2); the two numerical values thus derived are added together, and their *sum* is taken as the cipher equivalent of the particular plaintext digraph. Of course, this simple reduction would not be possible if the trinomes, in ascending order, had been arranged in the table in, say, a diagonal manner.

f. All of the foregoing tables have been digraphic in nature, but a kind of false trigraphic substitution may also be accomplished by means of similar tables, as illustrated in Fig. 52, wherein the table is the same as that in Fig. 48 with the addition of one more sequence at the top of the table. In using this table, θ_p^1 is located in sequence I, and its equivalent, θ_c^1, taken from sequence II; θ_p^2 is located in sequence III, and its equivalent, θ_c^2, taken from sequence IV; θ_c^3 is the letter lying at the intersection of the row indicated by θ_p^3 in sequence I and the column determined by θ_p^2. Thus, FIRE LINES would be enciphered $\overline{\text{NNZ}}$ $\overline{\text{IEQ}}$ $\overline{\text{KOV}}$. Various other agreements may be made with respect to the alphabets in which each plaintext letter will be sought in such a table, but the basic cryptographic principles are the same as in the case described.

	III.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	IV.	R	A	D	I	O	C	P	T	N	F	M	E	B	G	H	J	K	L	Q	S	U	V	W	X	Y	Z
I.	II.																										
A	W	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z
B	E	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G
C	S	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E
D	T	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N
E	I	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R
F	N	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A
G	G	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L
H	H	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C
I	O	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T
J	U	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I
K	A	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O
L	R	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M
M	B	Y	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P
N	K	B	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y
O	C	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B
P	D	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D
Q	F	H	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F
R	J	J	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H
S	L	K	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J
T	M	Q	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K
U	P	S	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q
V	Q	U	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S
W	V	V	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U
X	X	W	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V
Y	Y	X	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W
Z	Z	Z	G	E	N	R	A	L	C	T	I	O	M	P	Y	B	D	F	H	J	K	Q	S	U	V	W	X

FIGURE 52.

g. Tables such as those illustrated in Figs. 47–52, above, have been encountered in operational systems, but their use has not been very widespread because of their relatively large size and the inconvenience in their production and handling. In lieu of these large tables it is possible to employ much smaller matrices or geometrical designs to accomplish digraphic substitution; methods involving their use will be discussed in the following paragraph.

66. Polygraphic substitution methods employing small matrices.[9]—*a.* A simple method for accomplishing digraphic substitution involves the use of the *four-square matrix,* a matrix consisting of four 5 x 5 squares in which the letters of a 25-element alphabet (combining I and J) are inserted in any prearranged order. In a four-square matrix, θ_p^1 of $\overline{\theta\theta}_p$ is sought in Section 1; θ_p^2, in Section 2. Thus, θ_p^1 and θ_p^2 will always form the northwest-southeast corners of an imaginary rectangle delimited by these two letters as located in these two sections of the square. Then θ_c^1 and θ_c^2 are, respectively, the letters at the northeast-southwest corners of this same rectangle. Thus, $\overline{TG}_p=\overline{XS}_c$; $\overline{WD}_p=\overline{CH}_c$; $\overline{OR}_p=\overline{YV}_c$; $\overline{UR}_p=\overline{XB}_c$; etc. In decrypting, θ_c^1 and θ_c^2 are sought in Sections 3 and 4, respectively, and their equivalents, θ_p^1 and θ_p^2, noted in Sections 1 and 2, respectively.

	A	B	C	D	E	F	O	U	R	T	
	F	G	H	I	K	L	M	P	Q	E	
Sec. 1 (θ_p^1)	L	M	N	O	P	K	Y	Z	S	N	Sec. 3 (θ_c^1)
	Q	R	S	T	U	I	X	W	V	A	
	V	W	X	Y	Z	H	G	D	C	B	
	T	H	I	R	E	A	B	C	D	E	
	O	P	Q	S	N	F	G	H	I	K	
Sec. 4 (θ_c^2)	M	Y	Z	U	A	L	M	N	O	P	Sec. 2 (θ_p^2)
	L	X	W	V	B	Q	R	S	T	U	
	K	G	F	D	C	V	W	X	Y	Z	

FIGURE 53.

b. It is possible to effect digraphic substitution with a matrix consisting of but two sections by a modification in the method of finding equivalents. In a *horizontal two-square matrix,* such as that shown in Fig. 54, θ_p^1 of $\overline{\theta\theta}_p$ is located in the square at the left; θ_p^2, in the square at the right.

	M	A	N	U	F	A	U	T	O	M	
	C	T	R	I	G	B	I	L	E	S	
$\theta_p^1\theta_c^2$	B	D	E	H	K	C	D	F	G	H	$\theta_p^2\theta_c^1$
	L	O	P	Q	S	K	N	P	Q	R	
	V	W	X	Y	Z	V	W	X	Y	Z	

FIGURE 54.

When θ_p^1 and θ_p^2 are at the opposite ends of the diagonal of an imaginary rectangle defined by these letters, the ciphertext equivalent comprises the two letters appearing at the opposite ends of the other diagonal of the same rectangle; θ_c^1 is the particular one which is in the same row as θ_p^1, and θ_c^2 is the one in the same row as θ_p^2. For example, $\overline{AL}_p=\overline{TT}_c$; $\overline{DO}_p=\overline{GA}_c$. When θ_p^1 and θ_p^2 happen to be in the same row, the ciphertext equivalent is merely the reverse of the plaintext digraph; for example, $\overline{AT}_p=\overline{TA}_c$ and $\overline{EH}_p=\overline{HE}_c$.

[9] The word *matrix* as employed in this paragraph refers to checkerboard-type diagrams smaller than the tables illustrated in the preceding paragraph. These matrices are usually composed of sections containing 25 cells each.

c. Digraphic substitution may also be effected by means of *vertical two-square matrices*, in which one section is directly above the other, as in Fig. 55; it will be noted that matrices of this type have a feature of reciprocity when employed according to the usual rules, which follow.

	M	A	N	U	F
	C	T	R	I	G
$\theta_p^1\theta_c^1$	B	D	E	H	K
	L	O	P	Q	S
	V	W	X	Y	Z
	A	U	T	O	M
	B	I	L	E	S
$\theta_p^2\theta_c^2$	C	D	F	G	H
	K	N	P	Q	R
	V	W	X	Y	Z

FIGURE 55.

When θ_p^1 and θ_p^2 are at the opposite ends of a diagonal, the rule for encipherment is the same as that for horizontal two-square encipherment (e. g., $\overline{MO}_p=\overline{UA}_c$ and $\overline{UA}_p=\overline{MO}_c$); when both θ_p^1 and θ_p^2 happen to be in the same column, the plaintext digraphs are self-enciphered (e. g., $\overline{MA}_p=\overline{MA}_c$ and $\overline{EL}_p=\overline{EL}_c$), a fact which constitutes an important weakness of this method.[10] This disadvantage is only slightly less obvious in the preceding case of horizontal two-square methods wherein the cipher equivalent of $\overline{\theta\theta}_p$ consists merely of the plaintext letters in reversed order.

d. *One*-square digraphic methods, with a necessary modification of the method for finding equivalents, are also possible. The first of this type to appear as a practical military system was that known as the *Playfair cipher*.[11] It was used for a number of years as a field cipher by the British Army, before and during World War I, and for a short time, also during that war, by certain units of the American Expeditionary Forces. Fig. 56 shows a typical Playfair square.

M	A	N	U	F
C	T	R	I	G
B	D	E	H	K
L	O	P	Q	S
V	W	X	Y	Z

FIGURE 56.

The modification in the method of finding cipher equivalents has been found useful in imparting a greater degree of security than that afforded in the preceding small matrix methods. The usual method of encipherment can be best explained by examples given under four categories:

(1) Members of the plaintext pair, θ_p^1 and θ_p^2, are at opposite ends of the diagonal of an imaginary rectangle defined by the two letters; the members of the ciphertext pair, θ_c^1 and θ_c^2,

[10] See subpar. 73*b* on other enciphering conventions which remove this weakness.

[11] This cipher was really invented by Sir Charles Wheatstone but receives its name from Lord Playfair, who apparently was its sponsor before the British Foreign Office. See Wemyss Reid, *Memoirs of Lyon Playfair*, London, 1899. It is of interest to note that, to students of electrical engineering, Wheatstone is generally not known for his contributions to cryptography but is famed for something he did not invent—the so-called "Wheatstone bridge", really invented by Samuel H. Christie.

are at the opposite ends of the other diagonal of this imaginary rectangle. Examples: $\overline{MO}_p=\overline{AL}_c$; $\overline{MI}_p=\overline{UC}_c$; $\overline{LU}_p=\overline{QM}_c$; $\overline{VI}_p=\overline{YC}_c$.

(2) θ_p^1 and θ_p^2 are in the same row; the letter immediately to the *right* of θ_p^1 forms θ_c^1; the letter immediately to the right of θ_p^2 forms θ_c^2. When either θ_p^1 or θ_p^2 is at the extreme right of the row, the first letter in the row becomes its cipher equivalent. Examples: $\overline{MA}_p=\overline{AN}_c$; $\overline{MU}_p=\overline{AF}_c$; $\overline{AF}_p=\overline{NM}_c$; $\overline{FA}_p=\overline{MN}_c$.

(3) θ_p^1 and θ_p^2 are in the same column; the letter immediately *below* θ_p^1 forms θ_c^1, the letter immediately below θ_p^2 forms θ_c^2. When either θ_p^1 or θ_p^2 is at the bottom of the column, the top letter in that column becomes its cipher equivalent. Examples: $\overline{MC}_p=\overline{CB}_c$; $\overline{AW}_p=\overline{TA}_c$; $\overline{WA}_p=\overline{AT}_c$; $\overline{QU}_p=\overline{YI}_c$.

(4) θ_p^1 and θ_p^2 are identical; they are to be separated by inserting a null, usually the letter X or Q, and subsequently enciphered by the pertinent rule from above. For example, the word BATTLES would be enciphered thus:

BA TX TL ES
DM RW CO KP

The Playfair square is automatically reciprocal so far as encipherments of type (1) above are concerned; but this is not true of encipherments of type (2) and (3).

e. It is not essential that the small matrices used for digraphic substitution be in the shape of perfect squares; rectangular designs will serve equally well, with little or no modification in procedure.[12] For example, each section of, say, a four-square matrix could be constructed with four rows containing six letters each by having U_p serve for V_p, as well as I_p for J_p. Furthermore, it is possible to expand the sections of a digraphic matrix to 28, 30, or more characters by the following subterfuge, without introducing digits or symbols into the cipher text.[13] One of the letters of the alphabet may be omitted from the set of 26 letters, and this letter may then be replaced by 2, 3, or more *pairs* of letters, each pair having as one of its members the omitted single letter. The 5 x 6 Playfair square of Fig. 57*a* has been derived thus; the letter K has

W	A	S	H	I	N
G	T	O	B	C	D
E	F	J	KA	KE	KI
KO	KU	L	M	P	Q
R	U	V	X	Y	Z

FIGURE 57*a*.

been omitted as a single letter, and the number of characters in the rectangle has been made a total of 30 by the addition of five combinations of K with other letters. An interesting consequence of this modification is that certain irregularities are introduced in any cryptogram produced through its use; for example, (1) occasionally a plaintext digraph is replaced by a ciphertext trigraph or tetragraph, such as $\overline{AM}_p=\overline{HKU}_c$ and $\overline{EP}_p=\overline{KEKO}_c$; and (2) variant values may

[12] However, because the terms "four-*square* matrix", "two-*square* matrix", and "Playfair *square*" have become firmly fixed in cryptologic literature and practice, they continue to be applied to all such matrices, even when the "squares" of such matrices do not contain an equal number of rows and columns (that is, even when they are not square).

[13] The addition of any symbols such as the digits 1, 2, 3, . . . into a matrix solely to augment the number of elements to 27, 28, 30, 32, or 36 characters would not be considered practicable, since such a procedure would result in producing cryptograms containing intermixtures of letters and figures.

appear—$\overline{BKE}_c$, $\overline{DKE}_c$, $\overline{KEP}_c$, $\overline{GP}_c$, and $\overline{TP}_c$ all may be used to represent $\overline{CK}_p$. As far as the deciphering is concerned, there is no difficulty because any K occurring in the cipher text is considered as invariably forming a ligature with the succeeding letter, taking the pair of letters as a unit; and, when a plaintext unit is obtained containing one of the K-pairs, the letter after

	B	2	E	5	R	L	A	B	C	D	E	F	
	I	9	N	A	1	C	G	H	I	J	KA	KE	
θ_p^1	3	D	4	F	6	G	KI	KO	KU	KY	L	M	θ_c^1
	7	H	8	J	Ø	K	N	O	P	QA	QE	QI	
	M	O	P	Q	S	T	QO	QU	QY	R	S	T	
	U	V	W	X	Y	Z	U	V	W	X	Y	Z	
	A	B	C	D	E	F	M	U	N	I	9	C	
	G	H	I	J	KA	KE	3	H	8	A	1	B	
θ_c^2	KI	KO	KU	KY	L	M	2	D	4	E	5	F	θ_p^2
	N	O	P	QA	QE	QI	6	G	7	J	Ø	K	
	QO	QU	QY	R	S	T	L	O	P	Q	R	S	
	U	V	W	X	Y	Z	T	V	W	X	Y	Z	

FIGURE 57*b*.

the K is disregarded; for example, $\overline{CKO}_p$ is read as CK. The four-square matrix in Fig. 57*b* has also been constructed using the foregoing subterfuge. With this latter matrix, numbers in the plain text may be enciphered, still without producing *cipher* text containing numbers; for example, the plain text "HILL 3406" would be represented by the cipher QAB AT KUKI NQE which would be regrouped into groups of five letters and sent as QABAT KUKIN QE. . .

f. Fig. 58 shows a numerical four-square matrix which presents a rather interesting feature in that it makes possible the substitution of 3-figure combinations for digraphs in a unique manner. To encipher a message one proceeds as usual to find the numerical equivalents of a pair, and then these numbers are added together. Thus:

Plain text:	PR	OC	EE	DI	NG
	275	350	100	075	325
	9	13	24	18	7
Cipher text:	284	363	124	093	332

	A	B	C	D	E	000	025	050	075	100	
	F	G	H	I	K	125	150	175	200	225	
Sec. 1 (θ_p^1)	L	M	N	O	P	250	275	300	325	350	Sec. 3 (θ_c^1)
	Q	R	S	T	U	375	400	425	450	475	
	V	W	X	Y	Z	500	525	550	575	600	
	Ø	1	2	3	4	V	Q	L	F	A	
	5	6	7	8	9	W	R	M	G	B	
Sec. 4 (θ_c^2)	10	11	12	13	14	X	S	N	H	C	Sec. 2 (θ_c^2)
	15	16	17	18	19	Y	T	O	I	D	
	20	21	22	23	24	Z	U	P	K	E	

FIGURE 58.

In deciphering, the greatest multiple of 25 contained in the group of three digits is determined; then this multiple and its remainder are used to form the elements for determining the plaintext pair in the usual manner. Thus, 284=275+9=PR.

g. Thus far all the small-matrix methods have involved only digraphic substitution. The two matrices together illustrated in Figs. 59*a* and *b* may be used to provide a system for encipherment which is partly trigraphic; the adverb "partly" has been used because this particular system will yield trigraphic encipherment approximately 88.5% of the time in ordinary text and digraphic encipherment approximately 11.5% of the time.[14] In this case the cipher equivalents of the trigraphs (or digraphs, as the case may be) are tetranomes. Encipherment is best illustrated by an example; this is given in the next subparagraph.

	H_1	H_2	H_3	H_4	Y_1	Y_2	Y_3	Y_4	D_1	D_2	00	01	02	03	04	05	06	07	08	09	
	D_3	D_4	R_1	R_2	R_3	R_4	A_1	A_2	A_3	A_4	10	11	12	13	14	15	16	17	18	19	
	U_1	U_2	U_3	U_4	L_1	L_2	L_3	L_4	I_1	I_2	20	21	22	23	24	25	26	27	28	29	
	I_3	I_4	C_1	C_2	C_3	C_4	B_1	B_2	B_3	B_4	30	31	32	33	34	35	36	37	38	39	
	E_1	E_2	E_3	E_4	F_1	F_2	F_3	F_4	G_1	G_2	40	41	42	43	44	45	46	47	48	49	
Sec. 1	G_3	G_4	K_1	K_2	K_3	K_4	M_1	M_2	M_3	M_4	50	51	52	53	54	55	56	57	58	59	Sec. 3
	N_1	N_2	N_3	N_4	O_1	O_2	O_3	O_4	P_1	P_2	60	61	62	63	64	65	66	67	68	69	
	P_3	P_4	Q_1	Q_2	Q_3	Q_4	S_1	S_2	S_3	S_4	70	71	72	73	74	75	76	77	78	79	
	T_1	T_2	T_3	T_4	V_1	V_2	V_3	V_4	W_1	W_2	80	81	82	83	84	85	86	87	88	89	
	W_3	W_4	X_1	X_2	X_3	X_4	Z_1	Z_2	Z_3	Z_4	90	91	92	93	94	95	96	97	98	99	
	00	01	02	03	04	05	06	07	08	09	Q_1	Q_2	Q_3	Q_4	U_1	U_2	U_3	U_4	E_1	E_2	
	10	11	12	13	14	15	16	17	18	19	E_3	E_4	S_1	S_2	S_3	S_4	T_1	T_2	T_3	T_4	
	20	21	22	23	24	25	26	27	28	29	I_1	I_2	I_3	I_4	O_1	O_2	O_3	O_4	N_1	N_2	
	30	31	32	33	34	35	36	37	38	39	N_3	N_4	A_1	A_2	A_3	A_4	B_1	B_2	B_3	B_4	
Sec. 4	40	41	42	43	44	45	46	47	48	49	L_1	L_2	L_3	L_4	Y_1	Y_2	Y_3	Y_4	C_1	C_2	Sec. 2
	50	51	52	53	54	55	56	57	58	59	C_3	C_4	D_1	D_2	D_3	D_4	F_1	F_2	F_3	F_4	
	60	61	62	63	64	65	66	67	68	69	G_1	G_2	G_3	G_4	H_1	H_2	H_3	H_4	K_1	K_2	
	70	71	72	73	74	75	76	77	78	79	K_3	K_4	M_1	M_2	M_3	M_4	P_1	P_2	P_3	P_4	
	80	81	82	83	84	85	86	87	88	89	R_1	R_2	R_3	R_4	V_1	V_2	V_3	V_4	W_1	W_2	
	90	91	92	93	94	95	96	97	98	99	W_3	W_4	X_1	X_2	X_3	X_4	Z_1	Z_2	Z_3	Z_4	

FIGURE 59*a*.

	1	2	3	4
1	–	E	T	N
2	R	O	A	I
3	S	D	L	H
4	C	F	P	U

FIGURE 59*b*.

h. Let the text to be enciphered be a message beginning with the words "REFERRING TO YOUR MESSAGE NUMBER FIVE STOP . . ." This is rewritten into trigraphs, with the proviso that the third letter of the trigraph be one of the letters contained in the small square in Fig. 59*b*; if the third letter is not one of these 15 letters, the plaintext grouping is left as a digraph; then the grouping into trigraphs (or digraphs) continues. Thus, the foregoing plain text would be written as follows:

REF ERR IN– GTO YOU RME SSA GEN UM– BER FI– VES TOP ...

In encipherment, it is to be noticed that R_p occurs four times in Section 1 (as do all the letters) and E_p occurs four times in Section 2; the proper combination of the 16 possibilities is determined by the *coordinates of the third letter of the trigraph* as indicated in the small square, Fig. 59*b*.

[14] These figures are based on the number of trigraphs ending in one of the 15 highest-frequency letters (ETNROAISDLHCFPU), and on the number of trigraphs ending with other letters.

Since the coordinates of F_p in this square are 42, then it is the *4th* occurrence of R_p in Section 1 and the *2d* occurrence of E_p in Section 2 which are used to obtain the equivalent for the trigraph $\overline{REF}_p$; this equivalent is 1905. When the plaintext unit as obtained above is only a digraph, it is the *1st* occurrence of θ_p^1 which is used in Section 1 and the *1st* occurrence of θ_p^2 which is used in Section 2; thus, "IN–" from the sample message beginning, above, would be enciphered 2828. The encipherment of the plaintext example above is then

REF	ERR	IN–	GTO	YOU	RME	SSA	GEN	UM–	BER	FI–	VES	TOP
1905	4081	2828	4719	0727	1372	7417	4118	2270	3807	4024	8806	8623

The cipher text could then be transmitted in groups of four digits, or, as a subterfuge to conceal the basic group length, the transmission could be in five-digit groups. In decipherment, the ciphertext tetranome is deciphered in the manner of the usual four-square matrix, and the location of the particular values for θ_p^1 and θ_p^2 will indicate the identity of the third plaintext letter, if any.

i. Now that the student has become familiar with the details of typical polygraphic substitution systems, he is ready to continue his cryptanalytic study with the treatment of methods for recognizing polygraphic substitution; these methods are described in the next paragraph.

67. Methods for recognizing polygraphic substitution.—*a*. The methods used to determine whether or not a given cryptogram is digraphic in character are usually rather simple. If there are many repetitions in a cryptogram or a set of cryptograms and yet the uniliteral frequency distribution gives no clear-cut indications of monoalphabeticity; if most of the repetitions contain an even number of letters and these *repetitions for the most part begin on the odd letters and end on the even letters of the message*, yet the cipher text does not yield to solution as a biliteral cipher when the procedures outlined in Chapters VII and VIII are applied to it; if the cryptograms usually contain an even number of letters (exclusive of nulls); and if the cipher text is in letters and all 26 letters are not present and J or U are among the absent letters (or if the cipher is in digits and there is a limitation in the range of the text when divided into trinomes, this range usually being not greater than 001–676); then the encipherment may be assumed to be digraphic in nature.

b. Although the foregoing general remarks are true as far as they go, occasionally they may be difficult to apply with any clear-cut results unless a large volume of cipher text is available for study. To supplement them there are statistical tests which may be applied for the recognition of digraphic substitution. Just as the ϕ test and the Λ test may be applied to the uniliteral distribution of a cryptogram to help determine whether it is monoalphabetic with respect to single-letter plaintext units, so may these same tests be applied to the *digraphic* distribution of a cryptogram for the purpose of determining whether the cryptogram in question is monoalphabetic when considered as a digraphic cipher.

c. The basic *form* of the ϕ test is the same when applied to digraphic distributions as when applied to monographic—that is, uniliteral—distributions (see par. 27). It is only the plain and random constants that change, and "N" in the formulas now pertains to the number of *digraphs* under consideration, instead of the number of single letters. To illustrate this, the formulas for computing the "digraphic phi plain" ($_2\phi_p$) and "the digraphic phi random" ($_2\phi_r$) are shown below:[15]

$$_2\phi_p = .0069\ N(N-1)$$

$$_2\phi_r = .0015\ N(N-1)$$

The "digraphic phi observed" ($_2\phi_o$) is calculated in the usual manner, that is, by multiplying each *f* (which in this case is found in each one of the cells of a digraphic distribution) by *f–1*, and then totalling all the values thus derived.

d. The digraphic Λ test (or the "digraphic blank-expectation test") may be applied to a digraphic distribution just as easily as its monographic counterpart is applied to a uniliteral frequency distribution. For this purpose, Chart 8 is given below, showing the average number of blanks theoretically expected in digraphic distributions for plain text and for random text containing various numbers of digraphs (up to 200 digraphs). As can be seen, the chart contains two curves. The one labeled P applies to the average number of blanks theoretically expected in digraphic distributions based upon normal *plaintext* messages containing the indicated number of digraphs. The other curve, labeled R, applies to the average number of blanks theoretically expected in digraphic distributions based upon perfectly random assortments of digraphs. In using this chart one finds the point of intersection of the vertical line corresponding to the number of digraphs in the message, with the horizontal line corresponding to the observed number of blanks in the digraphic distribution for the message. If this point of intersection falls closer to curve P than it does to curve R, this is evidence that the cryptogram is digraphic in nature [16]; if it falls closer to curve R than to curve P, this is evidence that the cryptogram is not digraphic in character.

e. Although it may not be necessary to resort to the use of the digraphic ϕ and Λ tests to determine whether or not a particular cryptogram has been digraphically enciphered, it is well to know the application of these tests, since use has been made of them in difficult cases in operational practice. They may be helpfully employed in cases where the cryptanalyst is uncertain as to whether or not a single null has been added at the *beginning* of a cryptogram suspected to

---

[15] The digraphic plain constant, .0069, was obtained by summing the squares of the probabilities of digraphs in English plain text; the digraphic random constant, .0015 (or .00148 to three significant figures), is merely the decimal equivalent of 1/676. The digraphic I. C. for English plain text is 4.66, i. e., $\frac{.0069}{.00148}$, as compared with the digraphic I. C. for random text of 1.0, i. e., $\frac{.00148}{.00148}$. Further elaboration on the use of these constants, among others, will be given in *Military Cryptanalytics, Part III.*

[16] Unfortunately, such would also be the case if the cryptogram under consideration were a polyalphabetic cipher involving two alphabets. However, to distinguish between a digraphic cipher and a polyalphabetic cipher with two alphabets, a digraphic distribution could be made "off the cut", that is, made of those ciphertext digraphs which are formed by omitting the first letter of text and then dividing the remaining text into groups of two letters. If the system were digraphic, such a distribution would exhibit a poor $_2\phi_o$; if the system were a two-alphabet substitution system, the $_2\phi_o$ would be as satisfactory as that of the regular distribution, taken "on the cut".

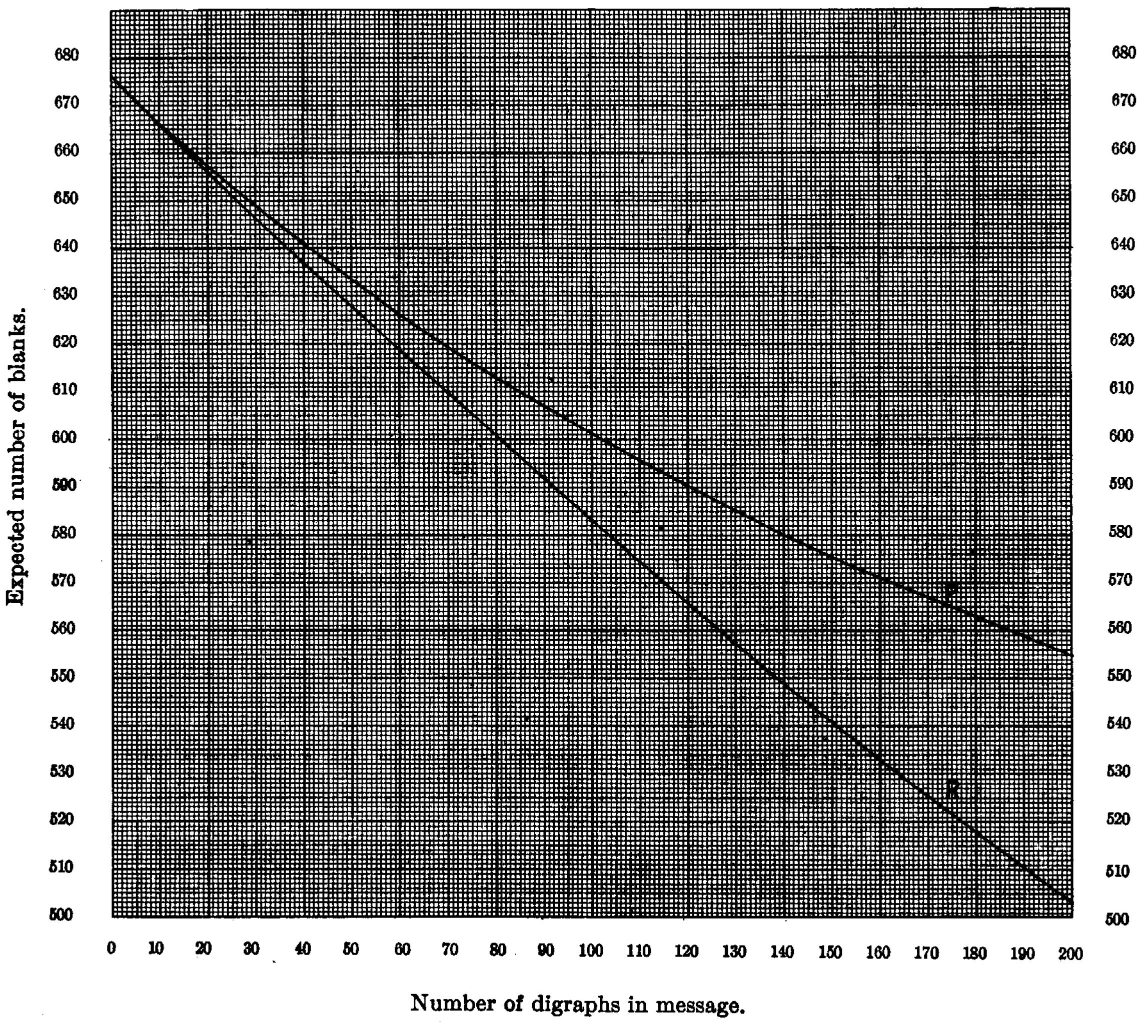


CHART 8. Curves showing the average number of blanks theoretically expected in digraphic distributions for plain text (P) and for random text (R) for messages comprising various numbers of digraphs. (See subpar. 67*d*.)

be a digraphic cipher; and these tests may also be found useful in the analysis of complex cases where the digraphic encipherment has been applied, not to adjacent letters of the plaintext message, but to digraphs composed of more-or-less *separated* letters in the message. Elaborations of these ideas will be treated in *Military Cryptanalytics, Part II.*

f. As for the recognition of *trigraphic* substitution ciphers—if most of the repetitions are a multiple of three letters in length, if these repetitions for the most part begin (when the cipher text is divided into trigraphs) with the first letters and end with the third letters of the trigraphs, and if the length of the cryptograms is for the most part a multiple of three letters, yet the cipher text does not yield to solution as a triliteral cipher, then the encipherment may be assumed to be trigraphic in nature.

g. Just as the ϕ test may be used as an aid in the recognition of digraphicity, it may theoretically be used for recognizing the trigraphic, tetragraphic, etc., nature of cryptograms, but its use for these latter purposes is much more limited because of the large amount of text which would be required to permit a valid application of the pertinent polygraphic ϕ test.

68. General procedure in the identification and analysis of polygraphic substitution ciphers.—*a.* Certain systems which at first glance seem to be polygraphic, in that groupings of plaintext letters are treated as units, are on closer inspection seen to be only partly polygraphic in character. Such is true of systems involving large tables of the type illustrated in Figs. 47*a* and *b*, and 48 (in par. 65, above), wherein encipherment is by pairs but one of the letters in each pair is enciphered monoalphabetically, making these systems only *pseudo*-polygraphic. For example, using the table in Fig. 48, any plaintext digraph beginning with "A" must be enciphered by a ciphertext digraph beginning with "W"; any plaintext digraph beginning with "B" must be enciphered by a ciphertext digraph beginning with "E"; etc. A cryptogram involving the use of this table may then be identified as such merely from a study of the uniliteral frequency distribution made on the initial letters of the cipher digraphs, since such a distribution would perforce be monoalphabetic.[17]

b. In certain other systems—namely, the four-square, two-square, and Playfair square systems of par. 66, above—the method of encipherment is by pairs, but the encipherments of the left-hand and right-hand members of the pairs show group relationships; this is not pseudo-polygraphic but, rather, *partially*-polygraphic. Cryptograms enciphered by means of systems of this latter type may not be readily identified as such merely through an examination of their cipher text, but their solution may be effected rather rapidly as soon as a few correct plaintext assumptions have been made therein. A more detailed treatment of this matter will be given in succeeding paragraphs of this chapter.

c. The analysis of cryptograms which have been produced by digraphic substitution is accomplished largely by the application of the simple principles of frequency of digraphs,[18] with the additional aid of digraphic idiomorphs and such special circumstances as may be known to or suspected by the cryptanalyst. The latter refer to peculiarities which may be the result of the particular method employed in obtaining the equivalents of the plaintext digraphs in the encrypting process, such as those mentioned in subpars. *a* and *b*, above. In general, if there is sufficient text to disclose the normal phenomena of repetition and idiomorphism, or if cribs are available to be used as an entering wedge, solution will be feasible. The foregoing general statements will be expanded upon in the following two subparagraphs, *d* and *e*.

d. When a digraphic system is employed in regular service, there is little doubt that traffic will rapidly accumulate to an amount more than sufficient to permit of solution by simple principles of frequency. Sometimes only two or three long messages, or a half-dozen of average length, are sufficient. For with the identification of only a few cipher digraphs, larger portions of messages may be read because the skeletons of words formed from the few high-frequency

---

[17] For this purpose, the simplest and most economical way to obtain the uniliteral distributions for the initial and final letters of digraphs is to make a digraphic distribution and then add the tallies in each row to yield the distribution for the initial letters, and add the tallies in each column to obtain the distribution for the final letters.

[18] In this connection, it would be well for the student to familiarize himself with that portion of Appendix 2 which contains digraphic frequency data, if he has not already done so.

digraphs very definitely limit the values that can be inserted for the intervening unidentified digraphs. For example, suppose that the plaintext digraphs RE, IN, ON, ND, NO, SI, NT, and TO are among those that have been identified by frequency considerations, corroborated by a tentatively identified long repetition; and suppose also that the enemy is known to be using a large table of 676 cells containing digraphs showing reciprocal equivalence between plaintext and ciphertext digraphs. Suppose the message begins as follows (in which the assumed values have been inserted):

XQ	VO	ZI	LK	AP	OL	ZX	PV	CK	IK	OL	UK	AT	HN	LK
		ND	IN		NT		RE			NT	NO			IN

VL	BN	OZ	BZ	DY	TY	LE	GI
	SI		ON	TO			

The initial words SECOND INFANTRY REGIMENT are readily recognized. Furthermore, if $\overline{CK}_c=\overline{GI}_p$ then $\overline{GI}_c=\overline{CK}_p$, which suggests ATTACK as the last word in the message beginning. This fragment of the message may now be completely recovered: SECOND INFANTRY REGI- MENT NOT YET IN POSITION TO ATTACK......

e. Just as the choice of probable words in the solution of uniliteral systems is aided or limited by the positions of repeated letters (see subpar. 49*d*), so, in digraphic ciphers, is the placing of cribs aided or limited by the positions of repeated *digraphs.* In this connection, several frequent words and phrases containing repeated digraphs have been tabulated for the student's aid, and this list of *digraphic idiomorphs* is presented as Section D in Appendix 3 (q. v.). Thus, if one is confronted by a ciphertext message containing the following repeated sequence (therefore likely to represent an entire word)

VI FW HM AZ FF FW RO

he may refer to the appropriate section of Appendix 3 which will disclose, on the basis of the idiomorphic pattern "AB -- -- -- AB" starting with the second cipher digraph, that the underlying plaintext word may be RE EN FO RC EM EN T, among others. Once a good start has been made and a few words have been solved, subsequent work is quite simple and straightforward. A knowledge of enemy correspondence, including data regarding its most common words and phrases, is of as much assistance in breaking down digraphic systems as it is in the solution of any other cryptosystems.

f. In the case of trigraphic substitution, analysis is made considerably more complex by the large amount of traffic required, not only for the initial entries, but also for further exploitation of the entering wedges. In effect, the solution of a trigraphic system closely parallels the solution of the syllabary portion of a large two-part code; these techniques will be discussed in *Military Cryptanalytics, Part V.*

69. Analysis of four-square matrix systems.—*a.* In all the small-matrix methods illustrated in par. 66, the encipherment is only partially digraphic because there are certain relationships between those plaintext digraphs which have common elements and their corresponding ciphertext digraphs, which will also have common elements. For example, in the four-square matrix given in Fig. 53, it will be noted that $\overline{AA}_p=\overline{FT}_c$, $\overline{AF}_p=\overline{FO}_c$, $\overline{AL}_p=\overline{FM}_c$, $\overline{AQ}_p=\overline{FL}_c$, and $\overline{AV}_p=\overline{FK}_c$. In each of these cases when A_p is the initial letter of the plaintext pair, the initial letter of the ciphertext equivalent is F_c. This, of course, is the direct result of the method; it means that the encipherment is monoalphabetic for the first half of each of *these five* plaintext pairs. This relationship holds true for *four* other groups of five pairs beginning with A_p; in effect, there are five cipher alphabets employed, not 25. Thus, this case differs from the case discussed under subpar. 68*a* only in that the monoalphabeticity is complete, not for half of all the pairs but only among the members of certain groups of pairs. In a *true* digraphic system, such as a system making use of a 676-cell randomized table, relationships of the foregoing type are entirely absent, and for this reason such a system is cryptographically more secure than small-matrix systems.

b. From the foregoing it is clear that when solution has progressed sufficiently to disclose a few values, the insertion of letters within the cells of the matrix to give the plaintext-ciphertext relationships indicated by the solved values immediately leads to the disclosure of additional values. Thus, the solution of only a few values soon leads to the breakdown of the entire matrix.

c. The following example will serve to illustrate the procedure.

(1) Let the message be as follows:

```
A. H F C A P  G O Q I L  B S P K M  N D U K E  O H Q N F  B O R U N
B. Q C L C H  Q B Q B F  H M A F X  S I O K O  Q Y F N S  X M C G Y
C. X I F B E  X A F D X  L P M X H  H R G K G  Q K Q M L  F E Q Q I
D. G O I H M  U E O R D  C L T U F  E Q Q C G  Q N H F X  I F B E X
E. F L B U Q  F C H Q O  Q M A F T  X S Y C B  E P F N B  S P K N U
F. Q I T X E  U Q M L F  E Q Q I G  O I E U E  H P I A N  Y T F L B
G. F E E P I  D H P C G  N Q I H B  F H M H F  X C K U P  D G Q P N
H. C B C Q L  Q P N F N  P N I T O  R T E N C  C B C N T  F H H A Y
J. Z L Q C I  A A I Q U  C H T P C  B I F G W  K F C Q S  L Q M C B
K. O Y C R Q  Q D P R X  F N Q M L  F I D G C  C G I O G  O I H H F
L. I R C G G  G N D L N  O Z T F G  E E R R P  I F H O T  F H H A Y
M. Z L Q C I  A A I Q U  C H T P
```

(2) The cipher having been tested for standard alphabets (by the method of completing the plain-component sequence) and found to give negative results, a uniliteral frequency distribution is made. It is as follows:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
11	15	26	8	16	30	17	23	23	0	8	14	11	18	15	16	33	9	6	11	11	0	1	12	7	3

(3) At first glance this may appear to the untrained eye to be a monoalphabetic frequency distribution, but upon closer inspection it is noted that, aside from the frequencies of four or five letters, the frequencies for the remaining letters are not very dissimilar. There are, in

reality, no very marked crests and troughs—certainly not as many as would be expected in a a monoalphabetic substitution cipher of equal length. The ϕ test, if taken (this test, as a rule, is not necessary with samples of text of sizes such as this), would show unsatisfactory results (ϕ_o=6082, as against ϕ_p=7870 and ϕ_r=4543).

(4) The message is carefully examined for repetitions of 4 or more letters, and all of them are listed:

	Frequency	*Located in lines*
TFHHAYZLQCIAAIQUCHTP (20 letters)	2	H and L.
QMLFEQQIGOI (11 letters)	2	C and F.
XIFBEX (6 letters)	2	C and D.
FEQQ	3	C, D, F.
QMLF	3	C, F, K.
BFHM	2	B and G.
BSPK	2	A and E.
GOIH	2	D and K.

Since there are quite a few repetitions, two of considerable length, since all but one of them contain an even number of letters, since these repetitions with but two exceptions begin on odd letters and end on even letters, and since the message also contains an even number of letters (344), the cryptogram is retranscribed into 2-letter groups for further study. It is as follows:

					5					10					15
A	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM	LF	EQ	QI
D	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E	FL	BU	QF	CH	QO	QM	AF	TX	SY	CB	EP	FN	BS	PK	NU
F	QI	TX	EU	QM	LF	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN
H	CB	CQ	LQ	PN	FN	PN	IT	OR	TE	NC	CB	CN	TF	HH	AY
J	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM	CB
K	OY	CR	QQ	DP	RX	FN	QM	LF	ID	GC	CG	IO	GO	IH	HF
L	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M	ZL	QC	IA	AI	QU	CH	TP								

It is noted that all the repetitions listed above break up properly into digraphs except in one case, *viz.*, FEQQ in lines C, D, and F. This latter seems rather strange, and at first thought one might suppose that a letter was dropped out or was added in the vicinity of the FEQQ in line D. But it may be assumed that the FE QQ in line D has no relation at all to the .F EQ Q. in lines C and F and is merely an accidental repetition.

(5) A digraphic distribution is made as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A						///			//																//		7
B						//											/		//		/						6
C	/	~~////~~					////	///				/		/			//	/									18
D																/					/			/			3
E															/	//	//	/			/			//			9
F		///			//							/		////													10
G			/		/		/								///		/						/		/		9
H						////		///					//		/	//	/										13
I	///			//	/	//		///				/			//			/		/							16
J																											–
K					/	/	/								/						/						5
L		/	/			///								/		/	/										8
M			/											/							/			/			4
N			/	/													/				/				/		5
O								/									/	//							/	/	6
P				/			/				//			///													7
Q			///			/			///			/	~~////~~ /	//	/		//				//				/		22
R				/			/									/								/			4
S												/												/	/		3
T					/	////										//					/			//			10
U					/									/													2
V																											–
W																											–
X			/						//										/								4
Y																											–
Z												//															2
	4	9	8	5	7	20	8	10	7	–	3	6	8	13	9	9	12	5	3	1	9	–	1	8	7	1	

FIGURE 60.

(6) The appearance of the foregoing distribution for this message is quite characteristic of that for a digraphic substitution cipher Although there are 676 possible digraphs, only 107 are present in the distribution; this parallels what is expected of normal plain text, since out of the 676 possible two-letter combinations (including "impossible plaintext digraphs" such as QQ, JK, etc., which *might* have been used for special indicators, punctuation marks, etc.) only about 300 are usually used in the construction of plain text.[19] The number of blank cells, 569, closely approximates the 566 which would be expected in a distribution made on a sample of plain text of this size, as shown by Chart 8. Furthermore, although there are many cases in

[19] The 300 most frequent digraphs comprise 95% of normal English plain text (Appendix 2, Table 7–A).

which a digraph appears only once, there are quite a few in which a digraph appears two or three times, four cases in which a digraph appears four times, one case in which a digraph appears five times, and one in which a digraph appears six times. All of the foregoing observations concerning the distribution are reflected by the ϕ test: the observed digraphic phi value, 210, compares very favorably with the expected plain value ($=.0069\times172\times171 = 203$) as against the expected random value ($=.0015\times172\times171=44$). Thus all indications point to a *digraphic* substitution system.

(7) Since neither the ϕ_o (1780) and Λ_o (4) for the initial letters of the cipher digraphs nor the ϕ_o (1496) and Λ_o (2) for the final letters are too satisfactory in their approximation to the values expected for monoalphabetic distributions ($\phi_p=1962$ and $\phi_r=1133$; $\Lambda_p=5$ and $\Lambda_r=0$), the possibility of a *pseudo*-digraphic system is ruled out for the time being. There remain the possibilities of a *partially*-digraphic system employing a small matrix, or a *true* digraphic system employing a large, randomized table. In one common type of small-matrix system, the Playfair cipher, one of the telltale indications besides the absence of (usually) the letter J is the absence of cipher doublets, that is, two successive identical cipher letters. The occurrence of the double letters GG, HH, and QQ in the message under investigation eliminates the possibility of its being a normal Playfair cipher. For want of more accurate diagnostic criteria[20] *at this stage*,[21] the simplest thing to assume, from among the various hypotheses that remain to be considered, is that a four-square matrix is involved. One with normal alphabets (as being the simplest case) in Sections 1 and 2 is therefore set down (Fig. 61*a*).

	A	B	C	D	E						
	F	G	H	I	K						
1	L	M	N	O	P						3
	Q	R	S	T	U						
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I	K	
4						L	M	N	O	P	2
						Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 61*a*.

[20] Even a medical practitioner often cannot successfully diagnose a condition on the first visit. Cryptanalytically speaking, we are still on our "first visit". Subsequent probing will, we hope, reject or substantiate this or that hypothesis or assumption, until the patient (the cipher text) is recovered (i. e., brought back to plain text).

[21] However, see the treatment on the diagnosis of various types of digraphic systems in subpar. 73*j*.

(8) The recurrence of the group QMLF, three times, and at intervals suggesting that it might be a sentence separator, leads to the assumption that it represents the word STOP. The letters Q, M, L, and F are therefore inserted in the appropriate cells in Sections 3 and 4 of the diagram. Thus (Fig. 61*b*):

	A	B	C	D	E						
	F	G	H	I	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U				Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

Figure 61*b*.

These placements seem rather good from the standpoint that keyword-mixed sequences may have been used in these two sections. Moreover, in Section 3 the number of cells between L and Q is just one less than enough to contain all the letters M to P, inclusive; this suggests that one of these letters, probably N or O, is in the keyword portion of the sequence; that is, near the top of Section 3. Without making a commitment in the matter, let us suppose that M follows L and that P precedes Q; then let both N and O, for the present, be inserted in the cell between M and P. Thus (Fig. 61*c*):

	A	B	C	D	E						
	F	G	H	I	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N/O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

Figure 61c.

(9) Now, *if* the placement of P in Section 3 is correct, the cipher equivalent of $\overline{TH}_p$ will be $\overline{P\theta}_c$, and there should be a group of adequate frequency to correspond. Noting that $\overline{PN}_c$ occurs three times, it is assumed to represent $\overline{TH}_p$ and the letter N is inserted in the appropriate cell in Section 4. Thus (Fig. 61*d*):

	A	B	C	D	E						
	F	G	H	I	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 61*d*.

(10) It is about time to try out these assumed values in the message. The proper insertions are made, with the following results:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM ST	LF OP	EQ	QI
D	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E	FL	BU	QF	CH	QO	QM ST	AF	TX	SY	CB	EP	FN	BS	PK	NU
F	QI	TX	EU	QM ST	LF OP	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN TH
H	CB	CQ	LQ	PN TH	FN	PN TH	IT	OR	TE	NC	CB	CN	TF	HH	AY
J	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM ST	CB
K	OY	CR	QQ	DP	RX	FN	QM ST	LF OP	ID	GC	CG	IO	GO	IH	HF
L	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M	ZL	QC	IA	AI	QU	CH	TP								

(11) So far no impossible combinations are in evidence. Beginning with group H4 in the message is seen the following sequence:

```
PN  FN  PN
TH  ..  TH
```

Assume it to be THAT THE. Then $\overline{AT}_p=\overline{FN}_c$, and the letter N is to be inserted in row 4 column 1 of Section 4. But this is inconsistent with previous assumptions, since N in Section 4 has already been tentatively placed in row 2 column 4. Other assumptions for $\overline{FN}_c$ are made: that it is $\overline{IS}_p$ (THIS TH...); that it is $\overline{EN}_p$ (THEN TH...); but the same inconsistency is apparent. In fact the student will see that $\overline{FN}_c$ must represent a digraph ending in F, G, H, I–J, or K, since N_c is tentatively located on the same line as these letters in Section 2. Now $\overline{FN}_c$ occurs 4 times in the message. The digraph it represents *must* be one of the following:

```
DF, DG, DH, DI, DJ, DK        OF, OG, OH, OI, OJ,
IF, IG, IH, II, IJ, IK        TK,
JF, JG, JH, JI, JJ, JK        YF, YG, YH, YI, YJ, YK
```

Of these the only one likely to be repeated 4 times is OF, yielding

```
PN FN PN
TH OF TH
```
which may be a part of

```
CQ LQ PN FN PN IT        CQ LQ PN FN PN IT
.N OR TH OF TH E.   or   .S OU TH OF TH E.
```

In either case, the position of the F in Section 3 is excellent: F . . . L in row 3. There are 3 cells intervening between F and L, into which G, H, I–J, and K may be inserted. It is not nearly so likely that G, H, and K are in the key word as that I should be in it. Let it be assumed that this is the case, and let the letters G, H, and K be placed in the appropriate cells in Section 3. Thus (Fig. 61*e*):

	A	B	C	D	E						
	F	G	H	I	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N/O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
					N	F	G	H	I	K	
4					F	L	M	N	O	P	2
				M	Q	Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 61*e*.

Let the resultant derived values be checked against the frequency distribution. If the position of H in Section 3 is correct, then the digraph $\overline{ON}_p$, normally of high frequency, should be represented several times by $\overline{HF}_c$. Reference to Fig. 60 shows $\overline{HF}_c$ to have a frequency of 4. And $\overline{HM}_c$, with 2 occurrences, represents $\overline{NS}_p$. There is no need to go through all the possible corroborations.

(12) Going back to the assumption that
PN FN PN
TH .. TH is part of the expression

```
CQ LQ PN FN PN IT        CQ LQ PN FN PN IT
.N OR TH OF TH E.   or   .S OU TH OF TH E.,
```

it is seen at once from Fig. 61*e* that the latter is apparently correct and not the former, because $\overline{LQ}_c$ equals $\overline{OU}_p$ and not $\overline{OR}_p$. If $\overline{OS}_p=\overline{CQ}_c$, this means that the letter C of the digraph $\overline{CQ}_c$, must be placed in row 1 column 3 or row 2 column 3 of Section 3. Now the digraph $\overline{CB}_c$ occurs 5 times; $\overline{CG}_c$, 4 times; $\overline{CH}_c$, 3 times; $\overline{CQ}_c$, 2 times. Let an attempt be made to deduce the exact position of C in Section 3 and the positions of B, G, and H in Section 4. Since F is already placed in Section 4, assume G and H directly follow it, and that B comes before it. How much before? Suppose a trial be made. Thus (Fig. 61*f*):

	A	B	C	D	E			C?			
	F	G	H	I	K			C?			
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N/O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I	K	
4	B?	B?	B?	F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 61*f*.

By referring now to the frequency distribution, Fig. 60, after a very few minutes of experimentation it becomes apparent that the following is correct:

	A	B	C	D	E			C			
	F	G	H	I	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N/O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I	K	
4	B			F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 61*g*.

(13) The identifications given by these placements are inserted in the text, and solution is very rapidly completed. The final matrix and deciphered text are given below.

	A	B	C	D	E	S	O	C	I	E	
	F	G	H	I	K	T	Y	A	B	D	
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N	P	Q	R	
	V	W	X	Y	Z	U	V	W	X	Z	
	E	X	P	U	L	A	B	C	D	E	
	S	I	O	N	A	F	G	H	I	K	
4	B	C	D	F	G	L	M	N	O	P	2
	H	K	M	Q	R	Q	R	S	T	U	
	T	V	W	Y	Z	V	W	X	Y	Z	

FIGURE 61*h*.

					5					10					15
A	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
	ON	EH	UN	DR	ED	FI	RS	TF	IE	LD	AR	TI	LL	ER	YF
B	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
	RO	MP	OS	IT	IO	NS	IN	VI	CI	NI	TY	OF	BA	RL	OW
C	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM	LF	EQ	QI
	WI	LL	BE	IN	GE	NE	RA	LS	UP	PO	RT	ST	OP	DU	RI
D	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
	NG	AT	TA	CK	SP	EC	IA	LA	TT	EN	TI	ON	WI	LL	BE
E	FL	BU	QF	CH	QO	QM	AF	TX	SY	CB	EP	FN	BS	PK	NU
	PA	ID	TO	AS	SI	ST	IN	GA	DV	AN	CE	OF	FI	RS	TB
F	QI	TX	EU	QM	LF	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
	RI	GA	DE	ST	OP	DU	RI	NG	AD	VA	NC	EI	TW	IL	LP
G	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN
	LA	CE	CO	NC	EN	TR	AT	IO	NS	ON	WO	OD	SN	OR	TH
H	CB	CQ	LQ	PN	FN	PN	IT	OR	TE	NC	CB	CN	TF	HH	AY
	AN	DS	OU	TH	OF	TH	AY	ER	FA	RM	AN	DH	IL	LS	IX
J	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM	CB
	ZE	RO	EI	GH	TD	AS	HA	AN	DO	NW	OO	DS	EA	ST	AN
K	OY	CR	QQ	DP	RX	FN	QM	LF	ID	GC	CG	IO	GO	IH	HF
	DW	ES	TT	HE	RE	OF	ST	OP	CO	MM	EN	CI	NG	AT	ON
L	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
	ET	EN	PM	SM	OK	EW	IL	LB	EU	SE	DO	NH	IL	LS	IX
M	ZL	QC	IA	AI	QU	CH	TP								
	ZE	RO	EI	GH	TD	AS	HA								

d. In the solution of four-square cryptograms, advantage may be taken not only of the general type of digraphic idiomorphs mentioned in subpar. 68*e*, above, but also of a special type of *partial* idiomorphism present in any four-square cryptograms involving the use of a matrix in which the *plain components consist of normal alphabets* normally inscribed.[22] As an illustration, let the digraphs $\overline{SO}$ $\overline{UT}$ (H.) be enciphered by means of any four-square having normal alphabets in Sections 1 and 2, and it will be found that *in the encipherment* the initial letter of the cipher digraph representing $\overline{SO}_p$ will be identical to the initial letter of the cipher digraph representing $\overline{UT}_p$, regardless of how the cipher components are constructed. On this basis, a brief list of *specialized single-letter patterns* have been compiled for use in the solution of such a digraphic system; this list of "four-square digraphic idiomorphs" constitutes Section F of Appendix 3.

e. It is interesting to note how much simpler the technique of analysis is in the case of so-called *inverse four-square* ciphers, which involve the use of a matrix wherein the *ciphertext* sections contain normal alphabets, the plain components being mixed. For example, referring to Fig. 53, suppose that Sections 3 and 4 are used as the source of the plaintext pairs, and Sections 1 and 2 as the source of the ciphertext pairs; then $ON_p = ET_c$, $EH_p = GE_c$, etc. The simplicity of the analytic procedure will be made clear by the following exposition.

(1) To solve a message enciphered with an inverse four-square matrix, it is necessary to perform two steps. First, convert the ciphertext pairs into their plain-component equivalents by "deciphering" the message with a matrix in which all four sections contain normal alphabets; this operation yields two uniliteral substitution "ciphers", one composed of the odd letters, the other of the even letters. The second step is to solve these two monoalphabetic portions.

(2) As an example, let us consider the following cipher text, known (or assumed) to have been encrypted with a trinome-digraphic [23] system incorporating a four-square matrix similar to that illustrated in Fig. 58, except that the plain-component sections have been changed:

20323	85081	83450	27934	11503	09168
27835	41804	50413	27416	33091	01092
20805	74135	35473	32626	91160	03218
46818	33930	91393	41104	41331	17296
24302	83832	28359	38022	61043	69130
15313	61041	00144	10101	82403	36168
46536	62663	44007	18345	01402	88152
47821	73933	81193	47924	04032	41306
08703	70914	19391	11607	71371	53595
00741	33381	33593	39340	63531	88133

[22] If any other *known* plain components were involved, the procedure of deriving a list of idiomorphic patterns would be modified to fit the particular case.

[23] If the cipher text were being examined "from cryptanalytic scratch", the limitations (003–595) of the cipher text when the latter is divided into trinomes for examination would have at once indicated that this grouping is the one which merits detailed analysis. The digraphic ϕ test would then give an indication of the digraphic nature of the cryptographic treatment.

(3) The first thing to be done is to construct a four-square matrix with the known cipher-text sections, and inscribe arbitrary alphabets in the plaintext sections, as follows:

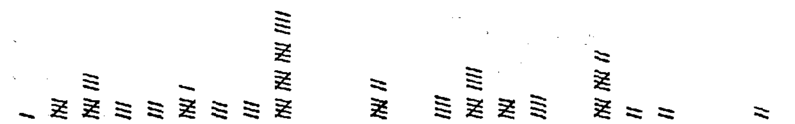
A	B	C	D	E	000	025	050	075	100
F	G	H	I	K	125	150	175	200	225
L	M	N	O	P	250	275	300	325	350
Q	R	S	T	U	375	400	425	450	475
V	W	X	Y	Z	500	525	550	575	600
Ø	1	2	3	4	A	B	C	D	E
5	6	7	8	9	F	G	H	I	K
10	11	12	13	14	L	M	N	O	P
15	16	17	18	19	Q	R	S	T	U
20	21	22	23	24	V	W	X	Y	Z

(4) The cipher text is then written in trinomes, and these trinomes are "deciphered" by means of the foregoing matrix, yielding the converted cipher text as follows:

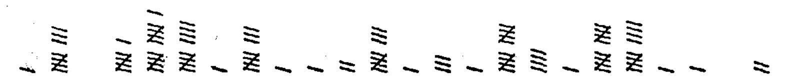
					5					10					15
A	203	238	508	183	450	279	341	150	309	168	278	354	180	450	413
	ID	IP	YF	IH	QD	PB	MT	FB	PH	IR	OB	PE	FH	QD	TM
B	274	163	309	101	092	208	057	413	535	473	326	269	116	003	218
	PV	IM	PH	BE	CT	II	CH	TM	VM	TY	MD	PQ	BU	DA	IT
C	468	183	393	091	393	411	044	133	117	296	243	028	383	228	359
	TT	IH	TQ	BT	TQ	RM	ER	IF	CU	MW	IU	DB	TF	IE	PK
D	380	226	104	369	130	153	136	104	100	144	101	018	240	336	168
	QF	GE	EE	PU	FF	IB	GL	EE	AE	KQ	BE	DQ	FU	MO	IR
E	465	366	266	344	007	183	450	140	288	152	478	217	393	381	193
	QT	MU	MQ	PT	CF	IH	QD	FQ	OM	HB	TE	HT	TQ	RF	IS
F	479	240	403	241	306	087	037	091	419	391	116	077	137	153	595
	UE	FU	TB	GU	MH	CO	CM	BT	UR	RQ	BU	CD	HL	IB	VY
G	007	413	338	133	593	393	406	353	188	133					
	CF	TM	OO	IF	YT	TQ	RG	OE	IN	IF					

CONFIDENTIAL

The distributions of the letters constituting the initial letters and final letters of the converted digraphs are as follows:



(Initial Letters) A B C D E F G H I K L M N O P Q R S T U V W X Y Z



(Final Letters) A B C D E F G H I K L M N O P Q R S T U V W X Y Z

(5) Using straightforward principles of frequency and partial idiomorphs,[24] the plain text (beginning with the opening words ENEMY RECONNAISSANCE. . .) is recovered, and the following equivalents are obtained for the converted cipher letters of the two alphabets:

(Initial Letters)		A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	C:	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	P:	B	R	A	H	M	S	C	D	E	F		I		L	N	O	P		T	U	V			Y	

(Final Letters)		A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	C:	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	P:	W	A		N	E	R	B	C	D	F	H	I	K	L	M	O	P	Q	S	T	U	V		Y	

Keyword-mixed sequences directly manifest themselves because the original enciphering matrix contained such sequences in Sections 1 and 2, inscribed in the same manner as were the arbitrary A–Z sequences which were used for the conversion. In fact, the key words of the two distributions might have been recovered from an analysis of the "profiles" of the distributions above, as described in subpar. *54e.*

(6) The original enciphering matrix is then reconstructed, thus:

B	R	A	H	M	000	025	050	075	100
S	C	D	E	F	125	150	175	200	225
G	I	K	L	N	250	275	300	325	350
O	P	Q	T	U	375	400	425	450	475
V	W	X	Y	Z	500	525	550	575	600
Ø	1	2	3	4	W	A	G	N	E
5	6	7	8	9	R	B	C	D	F
10	11	12	13	14	H	I	K	L	M
15	16	17	18	19	O	P	Q	S	T
20	21	22	23	24	U	V	X	Y	Z

[24] Note the ABA pattern of the first word in the message (ENEMY), made patent by the two-alphabet conversion process. Also note the 3-fold repetition (representing the plaintext word STOP) which, although hidden in the original cipher text, now comes to light.

CONFIDENTIAL

(7) Although the example illustrated was that of a numerical digraphic system, it is obvious that this technique of solution also applies to *literal* four-square systems in which the cipher components are known sequences. It should be clear to the student the tremendous difference it makes when it is possible to convert a *digraphic* system into a *two-alphabet* system; in a digraphic system, we are plagued by a potential 676 different elements in the cipher, whereas in a two-alphabet system we still have only 26 elements (in each of two sets, it is true) in the cipher text to be solved. This principle of conversion of cipher text into a secondary cipher text has application in some of the most complex types of cryptosystems; the student would do well to keep this in mind.

(8) As a further observation on inverse four-square systems, it is pointed out that where the *same* mixed alphabet is present in the two plaintext sections, the problem is still easier, since the letters resulting from the conversion into plain-component equivalents all belong to the same, *single* mixed alphabet; thus such a digraphic system is reduced to an ordinary simple substitution cipher.

f. The solution of cryptograms enciphered by other types of small matrices is accomplished along lines very similar to those set forth in subpar. *c* on the solution of a four-square cipher; this will be illustrated in subsequent paragraphs. There are, unfortunately, few means or tests which can be applied to determine in the early stages of the analysis *exactly* what type of digraphic system is involved in the first case under study. The author freely admits that the solution outlined in subpar. *c* is quite artificial in that nothing is demonstrated in step (7) that obviously leads to or warrants the assumption that a four-square matrix is involved. The point was passed over with the quite bald statement that this was "from among the various hypotheses that remain to be considered"—and then the solution proceeded exactly as though this mere hypothesis had been definitely established. For example, the very first results obtained were based upon our assuming that a certain 4-letter repetition represented the word STOP and *immediately inserting certain letters in appropriate cells in a four-square matrix with normal sequences in Sections 1 and 2*. Several more assumptions were built on top of that, and very rapid strides were made. What if it had not been a four-square matrix at all? What if it had been some other type of not readily identifiable digraphic system? The only defense that can be made of what may seem to the student to be purely arbitrary procedure based upon the author's advance information or knowledge is the following: In the first place, in order to avoid making the explanation a too-long-drawn-out affair, it is necessary (and pedagogical experience warrants) that certain alternative hypotheses be passed over in silence. In the second place it may now be added, *after* the principles and procedure have been elucidated (which at this stage is the primary object of this text), that if good results do not follow from a first hypothesis, the only thing the cryptanalyst can do is to reject that hypothesis and formulate a second hypothesis. In actual practice he may have to reject a second, third, fourth, . . . *n*th hypothesis. In the end he may strike the right one—or he may not. There is no assurance of success in the matter. In the third place, one of the objects of this text is to show how certain cryptosystems, if employed for military purposes, can readily be broken down. Assuming that some type of digraphic system is in use, and that daily changes in key words are made, it is possible that the traffic of the first day might give considerable difficulty in solution if the *specific* type of digraphic system were not known to the cryptanalyst. But by the time two or three days' traffic had accumulated it would be easy to solve, because probably by that time the cryptanalytic personnel would have successfully analyzed the cryptosystem and thus learned what type of matrix or table the enemy is using.

70. Analysis of two-square matrix systems.—*a.* Cryptosystems involving either vertical two-square or horizontal two-square matrices may be identified as such and solved by capitalizing on the cryptographic peculiarities and idiosyncracies of these systems. It will be noted that, considering the mechanics of the cryptosystems, in vertical two-square matrices employing the normal enciphering conventions,[25] exactly 20% of the 625 "possible" plaintext digraphs will be "transparent" (i. e., self-enciphered) in cipher text; in horizontal two-square systems, exactly 20% of the 625 digraphs will be characterized by an "inverse transparency" (i. e., enciphered by the same digraphs reversed).[26] Therefore, if an examination of a cryptogram or a set of cryptograms discloses a goodly portion of what appear to be *direct* transparencies (cipher digraphs which could well be plaintext digraphs), it may then be assumed that a *vertical* two-square matrix has been used for the encryption. On the other hand, if a large number of cipher digraphs could be "good" plaintext digraphs when the positions of the letters were *reversed*, then it may be assumed that the cryptosystem involved a *horizontal* two-square matrix. Sometimes skeletons of words or even of whole phrases are self-evident in such cipher text, thus affording an easy entering wedge into the cryptosystem.

b. An example will best serve to illustrate the techniques of identification and subsequent solution of a two-square matrix cipher. The following naval message is to be studied:

```
U O D L C   E N O A N   S I G L B   B E I R I   R C R G L   N M M L C
P T E R G   R B B O E   G P A B Q   W N N K S   I P C R M   M O R A P
D E A M H   A N X R A   I E D A I   R M A G B   E K H S L   C D D L C
T Q O R E   N D T M D   T I A Q F   I E Q T A   N N B F N   O U O O S
S N N N R   K T A S E   S N H L P   O N N K S   I P C R C   E N O I S
H L I R K   P L O N O   N Z U C T   A L T O I   I H O C N   O C E R A
O S D I N   O E E K R   L C U B R   A O S D I   I P D A R   C O G G R
O L N O C   W D I L P   O I L N Q   X D I G L   R B B Q Y   F S S R A
V Y O I G   R S L X X
```

Preliminary steps in analysis are made according to the procedures already described in this text, and the hypothesis of monographic, uniliteral encipherment (with either standard or

[25] That is, for vertical two-square systems, digraphs are self-enciphered if θ_p^1 and θ_p^2 fall in the same column in the matrix; and, for horizontal two-square systems, if θ_p^1 and θ_p^2 are in the same row, the ciphertext digraphs are the *reversed* plaintext digraphs.

[26] Although 625 "possible" plaintext digraphs are involved, the identity of digraphs actually used in plain text limit this figure considerably. Furthermore, the *frequencies* of the plaintext digraphs actually used come into consideration, in conjunction with the location of the letters of these digraphs in any particular two-square matrix. Thus, from the cryptanalyst's standpoint, there are "excellent" two-square matrices giving a high self-encipherment rate for high-frequency plaintext digraphs, and there are "poor" two-square matrices which have a potentially high self-encipherment rate only for those low-frequency plaintext digraphs which may not occur at all in a given cryptogram.

mixed cipher alphabets) is rejected. Multiliteral substitution, or digraphic substitution, comes next into consideration. The cipher text is written in digraphs, as follows:

					5					10					15
A	UO	DL	CE	NO	AN	SI	GL	BB	EI	RI	RC	RG	LN	MM	LC
B	PT	ER	GR	BB	OE	GP	AB	QW	NN	KS	IP	CR	MM	OR	AP
C	DE	AM	HA	NX	RA	IE	DA	IR	MA	GB	EK	HS	LC	DD	LC
D	TQ	OR	EN	DT	MD	TI	AQ	FI	EQ	TA	NN	BF	NO	UO	OS
E	SN	NN	RK	TA	SE	SN	HL	PO	NN	KS	IP	CR	CE	NO	IS
F	HL	IR	KP	LO	NO	NZ	UC	TA	LT	OI	IH	OC	NO	CE	RA
G	OS	DI	NO	EE	KR	LC	UB	RA	OS	DI	IP	DA	RC	OG	GR
H	OL	NO	CW	DI	LP	OI	LN	QX	DI	GL	RB	BQ	YF	SS	RA
J	VY	OI	GR	SL	XX										

FIGURE 62.

Noting the 8-letter repetition 90 letters apart, the 6-letter repetition 16 letters apart, and the 4-letter repetition at an interval of 220 letters, and that those repetitions begin on odd letters and end on even letters, credence is given to the grouping of the cipher text into pairs of letters. A digraphic distribution is then made, illustrated in Fig. 63.

c. The ${}_2\phi_o$, 152, is most satisfactory when compared with ${}_2\phi_p$ (107) and ${}_2\phi_r$ (23). Since the cryptogram has all the earmarks of a digraphic cipher, and no manifestations are found to support the hypothesis of a multiliteral system, the next problem is the specific determination of the particular kind of digraphic system involved. It may be noted that there are quite a few digraphs in the cipher text which resemble good plaintext digraphs, proportionally more so than, for instance, in the cryptogram in subpar. 69*c*; the cryptologic finger points to the possibility of a two-square system. However, since the words "good digraphs" are semantically elusive, let us attempt to determine statistically whether or not a two-square system might be involved and, if a two-square, whether it is more probably a vertical or a horizontal two-square.[27]

[27] The test to be described in the following subparagraphs is based on an evaluation of those instances wherein the observed frequency of any particular ciphertext digraph approximates the frequency with which the particular digraph, or its reversal, would be expected to occur if considered as a plaintext digraph. Any such correlation which occurs in a four-square or Playfair cipher, or in a cryptogram produced by a large randomized digraphic table, is purely accidental because it is not a result of the mechanics of the system. However, in two-square cryptograms such correlation *is* caused by the mechanics of the system in the encipherment of 20% of the possible plaintext digraphs, and these causal instances of correlation occur *in addition* to any accidental instances which may arise in the encipherment of the remaining 80%. Thus, if a digraphic cipher exhibits *merely the random* expectation of correlation both when the particular ciphertext digraphs are considered as they are *and* when their reversals are considered, the cryptogram may be assumed to involve a system other than two-square. If a digraphic cipher exhibits more than the random expectation of correlation, either when the particular digraphs are considered direct or when considered reversed, it may be assumed to involve two-square encipherment; and the particular consideration—that of the digraphs direct or that of the digraphs reversed—which gives rise to the greater degree of correlation indicates whether the cryptogram involves a vertical two-square or a horizontal two-square, respectively.

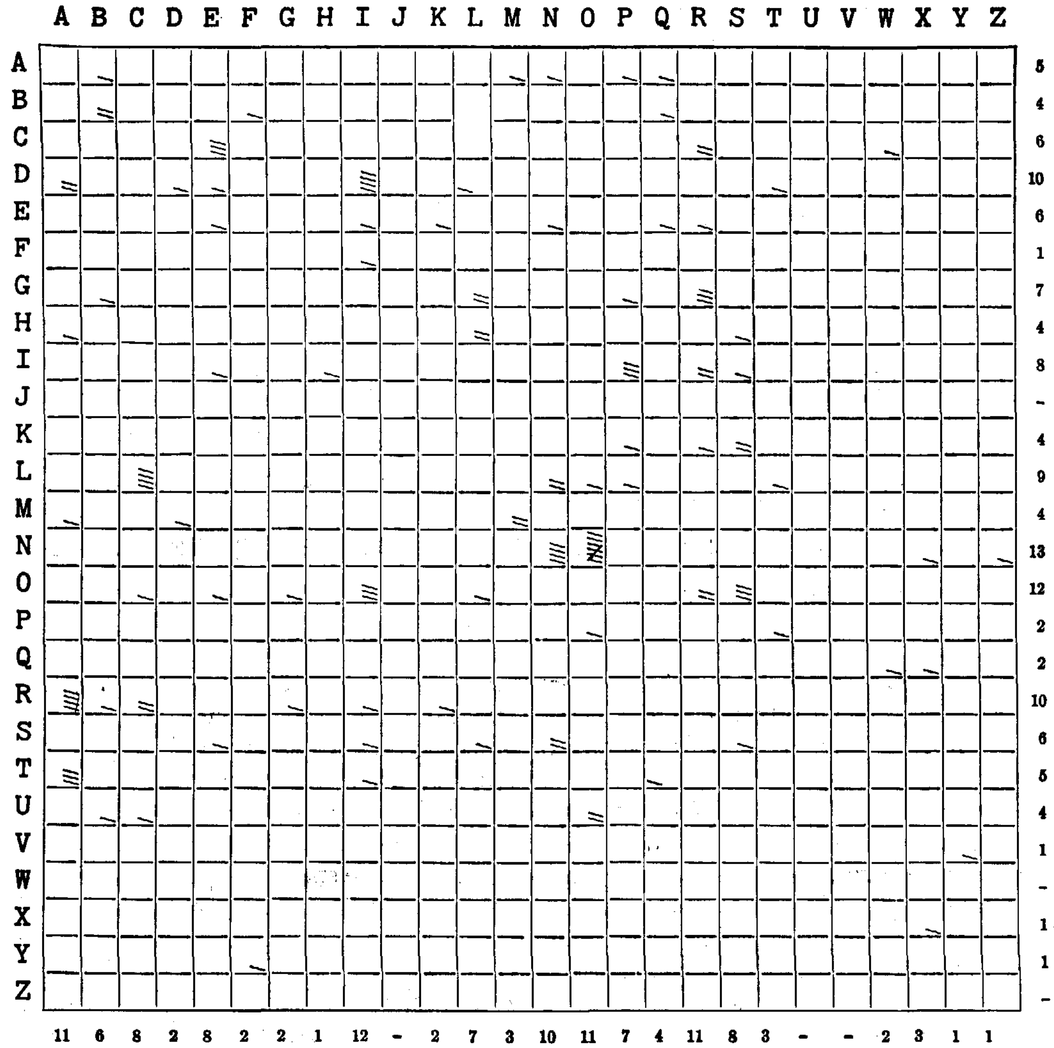


FIGURE 63.

d. First, for the purpose of determining whether "direct transparencies" or "inverse transparencies" predominate in this cryptogram, the digraphs of the distribution in Fig. 63 will be set down in tabular form, with an indication of their frequency in the cryptogram, and with data relative to the probability of these digraphs as *plaintext digraphs*, and as plaintext digraphs when *reversed*. In the table on p. 163, col. (1) is a listing of the ciphertext digraphs; col. (2) is the frequency of the ciphertext digraph as it occurs in the cryptogram; col. (3) is the logarithm of the theoretical *plaintext* frequency of the particular digraph (from Table 15, Appendix 2); col. (4) represents the products of the entries in cols. (2) and (3); col. (5) is the logarithm of the theoretical plaintext frequency of the *reversed* digraph (from Table 15, Appendix 2); and col. (6) represents the products of the entries in cols. (2) and (5). From this, the sum of the values in col. (4), 58.42, is taken to be the "direct transparency" value, and the sum of the values in col. (6), 62.76, is taken to be the "inverse transparency" value. Thus, since this particular cryptogram has an "inverse transparency" value which is higher than the "direct transparency" value, it may be assumed [28] to involve a *horizontal* two-square—if, indeed, two-square encipherment has been employed. It is now for us to establish whether or not this latter *is* the case, and this will be done by determining whether or not the foregoing observed value, 62.76, is representative of the degree of transparency which may be expected in a horizontal two-square cipher. (If the "*direct* transparency" value had been the higher of the two, then it would have been more probable that a *vertical* two-square were involved, and it would be necessary to determine whether or not *this* observed value was representative of the degree of transparency expected in a *vertical* two-square cipher.)

(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
AB	1	.45	0.45	.38	0.38	HA	1	.67	0.67	.25	0.25	OR	2	.89	1.78	.74	1.48
AM	1	.61	0.61	.78	0.78	HL	2	.13	0.26	.13	0.26	OS	3	.61	1.83	.62	1.86
AN	1	.89	0.89	.72	0.72	HS	1	.38	0.38	.72	0.72	PO	1	.64	0.64	.72	0.72
AP	1	.58	0.58	.61	0.61	IE	1	.59	0.59	.73	0.73	PT	1	.51	0.51	.25	0.25
AQ	1	.00	0.00	.00	0.00	IH	1	.00	0.00	.77	0.77	QW	1	.00	0.00	.00	0.00
BB	2	.00	0.00	.00	0.00	IP	3	.48	1.44	.45	1.35	QX	1	.00	0.00	.00	0.00
BF	1	.00	0.00	.00	0.00	IR	2	.73	1.46	.75	1.50	RA	4	.80	3.20	.82	3.28
BQ	1	.00	0.00	.00	0.00	IS	1	.78	0.78	.77	0.77	RB	1	.25	0.25	.25	0.25
CE	3	.76	2.28	.76	2.28	KP	1	.00	0.00	.00	0.00	RC	2	.53	1.06	.38	0.76
CR	2	.38	0.76	.53	1.06	KR	1	.00	0.00	.13	0.13	RG	1	.48	0.48	.42	0.42
CW	1	.13	0.13	.00	0.00	KS	2	.13	0.26	.13	0.26	RI	1	.75	0.75	.73	0.73
DA	2	.76	1.52	.73	1.46	LC	4	.33	1.32	.42	1.68	RK	1	.13	0.13	.00	0.00
DD	1	.51	0.51	.51	0.51	LN	2	.13	0.26	.42	0.84	SE	1	.84	0.84	.86	0.86
DE	1	.77	0.77	.88	0.88	LO	1	.59	0.59	.67	0.67	SI	1	.77	0.77	.78	0.78
DI	4	.73	2.92	.45	1.80	LP	1	.33	0.33	.59	0.59	SL	1	.25	0.25	.45	0.45
DL	1	.33	0.33	.53	0.53	LT	1	.51	0.51	.42	0.42	SN	2	.38	0.76	.71	1.42
DT	1	.62	0.62	.45	0.45	MA	1	.78	0.78	.61	0.61	SS	1	.67	0.67	.67	0.67
EE	1	.81	0.81	.81	0.81	MD	1	.13	0.13	.42	0.42	TA	3	.74	2.22	.83	2.49
EI	1	.73	0.73	.59	0.59	MM	2	.59	1.18	.59	1.18	TI	1	.82	0.82	.73	0.73
EK	1	.00	0.00	.45	0.45	NN	4	.51	2.04	.51	2.04	TQ	1	.13	0.13	.00	0.00
EN	1	.99	0.99	.87	0.87	NO	7	.66	4.62	.92	5.74	UB	1	.33	0.33	.25	0.25
EQ	1	.58	0.58	.00	0.00	NX	1	.00	0.00	.13	0.13	UC	1	.33	0.33	.38	0.38
ER	1	.94	0.94	.96	0.96	NZ	1	.00	0.00	.00	0.00	UO	2	.13	0.26	.79	1.58
FI	1	.80	0.80	.55	0.55	OC	1	.51	0.51	.80	0.80	VY	1	.00	0.00	.00	0.00
GB	1	.00	0.00	.00	0.00	OE	1	.33	0.33	.58	0.58	XX	1	.00	0.00	.00	0.00
GL	2	.25	0.50	.13	0.26	OG	1	.25	0.25	.45	0.45	YF	1	.56	0.56	.13	0.13
GP	1	.25	0.25	.00	0.00	OI	3	.42	1.26	.80	2.40		125		58.42		62.76
GR	3	.42	1.26	.48	1.44	OL	1	.67	0.67	.59	0.59						

(1) Identity of cipher digraph appearing in the cryptogram.
(2) Frequency of the particular digraph as it occurs in the cryptogram.
(3) Logarithm of theoretical *plaintext* frequency of the particular digraph (from Table 15, Appendix 2).
(4) Product of entries in columns (2) and (3).
(5) Logarithm of theoretical *plaintext* frequency of the digraph's *reversal* (from Table 15, Appendix 2).
(6) Product of entries in columns (2) and (5).

[28] The difference between the higher inverse transparency value and the direct value is indicative of the degree of probability of the horizontal hypothesis over the vertical hypothesis. In this case, the difference of 4.34 (i. e., 62.76−58.42) represents a difference of *log scores*; but since the cipher text is expected to contain 20% plaintext digraphs (or their reversals) "diluted" with 80% random digraphs, it can be proved mathematically that the correct allowance to compensate for this is to divide the log score by 5—that is, $\frac{4.34}{5}$ or 0.87. This adjusted value is then employed as an exponent of the log base (224); the number produced, 110 (i. e., $224^{0.87}$), is the factor in favor of the hypothesis of a horizontal two-square. Statistical interpretation of scoring techniques will be treated in detail in *Military Cryptanalytics, Part III.*

e. The observed "inverse transparency" value (selected in this case because it is the higher observed value) will be compared with the value *expected* from a horizontal two-square cryptogram of the same size, and if this observed value is *as great as or greater than* the transparency value expected for horizontal two-squares, the cryptogram may be considered to be a horizontal two-square cipher; if the observed value is lower than the expected two-square value, decision will have to be suspended.[29] The transparency value expected in a horizontal two-square cipher containing N digraphs is computed by multiplying N by .3388, which in this case yields 42.35 (=.3388×125).[30] The observed value for the cryptogram, 62.76, is much higher than the expected value, 42.35. Thus, it has been proven statistically that the cryptogram at hand involves two-square encipherment, particularly, *horizontal* two-square encipherment.

f. Having now proved that the cryptogram at hand is a horizontal two-square cipher, the next step is to assume some plain text in the message, guided by probable inverse transparencies (*inverse* because the system has been identified as a horizontal two-square) in the cipher text. Referring to the work sheet in Fig. 62, the repeated sequence at B9 and E9 is assumed to represent the plain text TA SK FO RC (E–), on the basis of $\overline{KS}_c=\overline{SK}_p$, and $\overline{CR}_c=\overline{RC}_p$. The plaintext-ciphertext

---

[29] For the benefit of the student with a background in statistics, it is pointed out that by abiding by the stipulation "as great or greater", some cryptograms which actually *are* the result of two-square encipherment may be rejected by this stipulation, but it will insure that only a relatively few non-two-square cryptograms will be accepted. A better approach of a statistical nature would involve, first, computing the expected value for non-two-squares as well as that for two-squares. Then, any observed value falling below the expected two-square value could be expressed in terms of the number of standard deviations (i. e., the sigmage) from this expected two-square value and from the expected *non*-two-square value. Finally, the particular expected value which would be considered as significant would be the one from which the observed value differed by the smaller number of standard deviations. The concept of standard deviation will be treated in *Military Cryptanalytics, Part III.*

[30] In the case of *vertical* two-squares, N would be multiplied by the constant .3610. The mathematical considerations underlying this test and their proofs (involving Bayes' theorem and Bayes' factors) are beyond the scope of this text; however, for the benefit of the mathematician, the derivation of the foregoing constants is explained below, along with the derivation of the constant used for computing the expected transparency value for *non*-two-squares. In the formulas, below,

$\sum_{AB}$ = the summation over all digraphs AA–ZZ

F_{AB} = the frequency of a given digraph AB as found in Table 6A, Appendix 2

α_{AB} = the logarithm (to the base 224) of the frequency of a given digraph AB as found in Table 15, Appendix 2

For vertical two-squares,

$$k=\sum_{AB}\alpha_{AB}\left[.80(.0015)+\frac{.20\,F_{AB}}{5000}\right]=.3610$$

For horizontal two-squares,

$$k=\sum_{AB}\alpha_{BA}\left[.80(.0015)+\frac{.20\,F_{AB}}{5000}\right]=.3388$$

For *non*-two square digraphic systems,

$$k=\frac{\alpha_{AB}}{676}=.2737$$

values are now recorded[31] in a skeleton reconstruction diagram as illustrated in Fig. 64*a*. At A3, the assumption of (–R) EC ON NA IS SA NC (E–) is tossed off without much ado, since four of the six diagraphs concerned are transparent. The plain-cipher relationships from this assumption are added to the reconstruction diagram, as shown in Fig. 64*b*. Continuing in this vein, the plain text (–A) IR CR AF (T–) is inserted at A1Ø, and the plain text (–B) AT TL

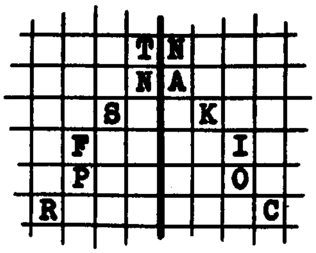


Figure 64*a*.

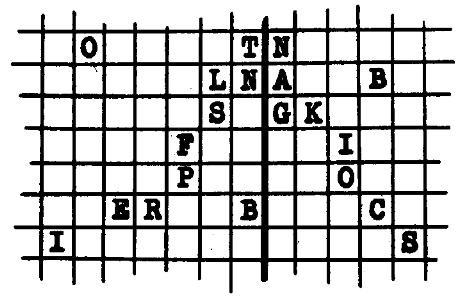


Figure 64*b*.

ES HI (P–) is inserted at F8; the successive cumulative reconstruction diagrams for these two assumptions are shown in Figs. 64*c* and *d* below. It is to be noted that at F12, $\overline{OC}_c = \overline{P\theta}_p$; but

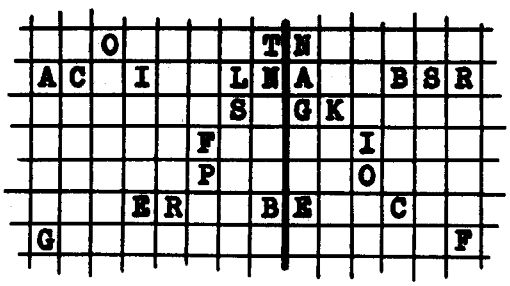


Figure 64*c*.

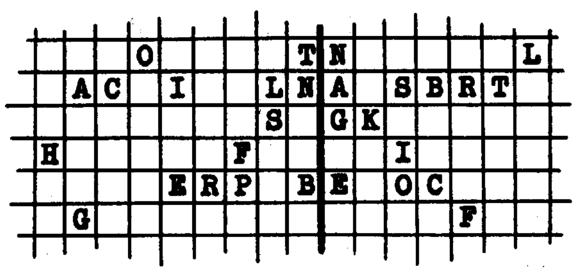


Figure 64*d*.

since in Fig. 64*d* it has already been determined that $\overline{OC}_c = \overline{\theta S}_p$, then $\overline{OC}_c$ must equal $\overline{PS}_p$, making the word BATTLESHIP<u>S</u> rather than BATTLESHIP.

---

[31] During the reconstruction of the squares of the matrix, the student should keep clear in his skeleton diagram which letters are in the same row, and which are in the same column. It will be found expeditious to draw a dividing line (either horizontal or vertical, depending on the type of two-square matrix involved) on the page to keep the elements of the two squares independent, recording the values which are in the same row or column and writing down the letters as they are assumed. In the early stages of this process the student must exercise care in recording the letters so that no false relationships are formed; in other words, the values should be written down so that they are not in the same row or column with any letters other than those with which they are known to be related. This will entail spreading the work rather widely over the page initially, then gradually telescoping and reducing the size of the reconstruction diagram as the work progresses, until in the end it will be reduced to a concise matrix of two 5 x 5 squares.

g. At this point the partially filled-in work sheet will look as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	UO	DL	CE	NO	AN	SI	GL	BB	EI	RI	RC	RG	LN	MM	LC
		-R	EC	ON	NA	IS	SA	NC	EA	IR	CR	AF	T-	E-	
B	PT	ER	GR	BB	OE	GP	AB	QW	NN	KS	IP	CR	MM	OR	AP
		RE	-E	NC	EO	-E	NE		TA	SK	FO	RC	E-	RO	-E
C	DE	AM	HA	NX	RA	IE	DA	IR	MA	GB	EK	HS	LC	DD	LC
					AR	-O		-O		-E					
D	TQ	OR	EN	DT	MD	TI	AQ	FI	EQ	TA	NN	BF	NO	UO	OS
		RO	BA			IT		-R		AT	TA		ON		
E	SN	NN	RK	TA	SE	SN	HL	PO	NN	KS	IP	CR	CE	NO	IS
	NS	TA		AT	IO	NS			TA	SK	FO	RC	EC	ON	
F	HL	IR	KP	LO	NO	NZ	UC	TA	LT	OI	IH	OC	NO	CE	RA →
		-O		OL	ON		-B	AT	TL	ES	HI	PS	ON	EC	AR
G	← OS	DI	NO	EE	KR	LC	UB	RA	OS	DI	IP	DA	RC	OG	GR
			ON	EE				AR			FO		CR		-E
H	OL	NO	CW	DI	LP	OI	LN	QX	DI	GL	RB	BQ	YF	SS	RA
	-S	ON				ES	T-			SA	N-			L-	AR
J	VY	OI	GR	SL	XX										
		ES	-E	LS											

Skeletons of additional plain text, such as the word OUR at A1, PRESENCE OF ENEMY at B1, PROBABLE at D1, ATTACK ON OUR INSTALLATIONS at D10, CARRIER at F14, and VESSELS at J1, may now clearly be seen. The complete recovery of the plain text follows, and the reconstruction diagram is completed and telescoped into the form shown in Fig. 64*e*. Since phenomena of keyword-mixed sequences are observed, the rows and columns of Fig. 64*e* are permuted to yield the original two-square matrix as shown in Fig. 64*f*.

Q	M	O	K	T	N	–	Q	L	P
A	I	C	L	N	A	B	S	R	T
G	D	F	S	H	G	K	I	F	H
U	E	P	R	B	E	C	O	D	M
Y	–	X	V	–	W	Z	Y	–	X

FIGURE 64*e*.

R	E	P	U	B	D	E	M	O	C
L	I	C	A	N	R	A	T	S	B
S	D	F	G	H	F	G	H	I	K
K	M	O	Q	T	L	N	P	Q	U
V	W	X	Y	Z	V	W	X	Y	Z

FIGURE 64*f*.

h. The solution of vertical two-square systems follows analogous lines, with the necessary modifications of the reconstruction diagram in consonance with the difference in mechanics between horizontal and vertical two-square systems.

i. A few additional remarks concerning the test applied in subpars. *d* and *e*, above, are in order. First, the exceptionally high transparency value observed in this cryptogram is a direct result of the very favorable manner in which the keyword-mixed sequences in the two squares interact; in the foregoing cryptogram, 47 of the 125 digraphs present (approx. 38%) were inverse transparencies. It is also pointed out that, although some actual two-square cryptograms may be rejected by that portion of the test which was described in subpar. *e*, the other phase of the test (described in subpar. *d*)—by which one may determine whether a cryptogram is more probably a vertical two-square encipherment or more probably a horizontal two-square encipherment—is sensitive and accurate to a high degree. The foregoing statistical method is not merely valuable *per se* as an application of cryptomathematics in the analysis of two-square matrix systems, but is included as being illustrative of the general principles of special techniques that may be developed in the attack on any particular cryptosystem, the mechanics of which are known to the cryptanalyst. The field of actual operational cryptanalysis is replete with special methods of attack of this nature.

71. Analysis of Playfair cipher systems.—*a*. Of all digraphic cryptosystems employing small matrices, the one which has been most frequently encountered is the Playfair cipher. Certain variations of this cipher have been incorporated in several complex manual ciphers used in actual operational practice; because of this it is important that the student gain familiarity with the methods of solution of the classic Playfair system.

b. The first published solutions [32] for this cipher are quite similar basically and vary only in minor details. The earliest, that by Lieut. Mauborgne (later to become Chief Signal Officer of the U. S. Army), used straightforward principles of frequency to establish the values of three or four of the most frequent digraphs. Then, on the assumption that in most cases in which a key word appears on the first and second rows the last five letters of the normal alphabet, VWXYZ, will rarely be disturbed in sequence and will occupy the last row of the square, he "juggles" the letters given by the values tentatively established from frequency considerations, placing them in various positions in the square, together with VWXYZ, to correspond to the plaintext-ciphertext relationships tentatively established. A later solution by Lieut. Frank Moorman, as described in Hitt's manual, assumes that in a Playfair cipher prepared by means of a square in which the key word occupies the first and second rows, if a digraphic frequency distribution is made, it will be found that the letters having the greatest combining power are very probably letters of the key. A still later solution, by Lieut. Commander Smith, is perhaps the most lucid and systematized of the three. He sets forth in definite language certain considerations which the other two writers certainly entertained but failed to indicate.

c. The following details have been summarized from Smith's solution:

(1) The Playfair cipher may be recognized by virtue of the fact that it always contains an even number of letters, and that when divided into groups of two letters each, no group contains a repetition of the same letter, as NN or EE. Repetitions of digraphs, trigraphs, and polygraphs will be evident in fairly long messages.

(2) Using the square [33] shown in Fig. 65, there are two general cases to be considered, as regards the results of encipherment:

---

[32] Mauborgne, Lieut. J. O., U. S. A. *An advanced problem in cryptography and its solution*, Leavenworth, 1914.
Hitt, Captain Parker, U. S. A. *Manual for the solution of military ciphers*, Leavenworth, 1918.
Smith, Lieut. Commander W. W., U. S. N. In *Cryptography* by André Langie, translated by J. C. H. Macbeth, New York, 1922.

[33] The Playfair square accompanying Smith's solution is based upon the key word BANKRUPTCY "to be distributed between the first and fourth lines of the square." This is a simple departure from the original Playfair scheme in which the letters of the key word are written from left to right and in consecutive lines from the top downward.

B	A	N	K	R
D	E	F	G	H
I	L	M	O	Q
U	P	T	C	Y
S	V	W	X	Z

FIGURE 65.

Case 1. Letters at opposite corners of a rectangle. The following illustrative relationships are found:

$$\overline{TH}_p=\overline{YF}_c$$
$$\overline{HT}_p=\overline{FY}_c$$
$$\overline{YF}_p=\overline{TH}_c$$
$$\overline{FY}_p=\overline{HT}_c$$

Reciprocity and reversibility.[34]

Case 2. Two letters in the same row or column. The following illustrative relationships are found:

$$\overline{AN}_p=\overline{NK}_c$$
$$\overline{NA}_p=\overline{KN}_c$$

But $\overline{NK}_p$ does not equal $\overline{AN}_c$ nor does $\overline{KN}_p=\overline{NA}_c$.

Reversibility only.

(3) The foregoing gives rise to the following:

Rule I. (*a*) Regardless of the position of the letters in the square, if

1.2=3.4, then
2.1=4.3

This rule is of particular aid in selecting probable words in the solution of Playfair ciphers, as will be shown shortly.[35]

(*b*) If 1 and 2 form opposite corners of a rectangle, the following equations obtain:

1.2=3.4
2.1=4.3
3.4=1.2
4.3=2.1

(4) A letter considered as occupying a position in a row can be combined with but four other letters in the same row; the same letter considered as occupying a position in a column can be combined with but four other letters in the same column. Thus, this letter can be combined with only 8 other letters all told, under Case 2, above. But the same letter considered as occupying a corner of a rectangle can be combined with 16 other letters, under Case 1, above. Smith derives from these facts the conclusion that "it would appear that Case 1 is twice as probable as Case 2". He continues thus (notation my own):

[34] By way of explaining what is meant by *reciprocity* and by *reversibility*, in the case of digraphic systems, the following examples are given: $\overline{TH}_p=\overline{YF}_c$ and $\overline{YF}_p=\overline{TH}_c$ constitute a *reciprocal* relationship; $\overline{TH}_p=\overline{YF}_c$ and $\overline{HT}_p=\overline{FY}_c$ constitute a *reversible* relationship.

[35] In this connection, a list of frequently-encountered words and phrases which contain reversed digraphs (so-called "ABBA patterns") has been compiled and is included as Section E, "Digraphic idiomorphs: Playfair", in Appendix 3.

"Now in the square, note that:

$\overline{AN}_p=\overline{NK}_c$		$\overline{EN}_p=\overline{FA}_c$
$\overline{GN}_p=\overline{FK}_c$		$\overline{EM}_p=\overline{FL}_c$
$\overline{ON}_p=\overline{MK}_c$	also	$\overline{ET}_p=\overline{FP}_c$
$\overline{CN}_p=\overline{TK}_c$		$\overline{EW}_p=\overline{FV}_c$
$\overline{XN}_p=\overline{WK}_c$		$\overline{EF}_p=\overline{FG}_c$

"From this it is seen that of the 24 equations that can be formed when each letter of the square is employed either as the initial or final letter of the group, five will indicate a repetition of a corresponding letter of plain text.

"Hence, *Rule II.* After it has been determined, in the equation 1.2=3.4, that, say, $\overline{EN}_p=\overline{FA}_c$, there is a probability of one in five that any other group beginning with F_c indicates $\overline{E\theta}_p$, and that any group ending in A_c indicates θN_p.[36]

"After such combinations as $\overline{ER}_p$, $\overline{OR}_p$, and $\overline{EN}_p$ have been assumed or determined, the above rule may be of use in discovering additional digraphs and partial words."

[36] *The probability of "one in five" is only an approximation. Take for example, the 24 equations having* F *as an initial letter:*

Case		Case		Case		Case	
1.	$FB_c=DN_p$	2.	FE=ED	2.	FT=NM	1.	FX=GW
2.	FD =EH	1.	FL=EM	2.	FW=NT	1.	FR=HN
1.	FI =DM	1.	FP=ET	1.	FK=GN	2.	FH=EG
1.	FU =DT	1.	FV=EW	2.	FG=EF	1.	FQ=HM
1.	FS =DW	2.	FN=NW	1.	FO=GM	1.	FY=HT
1.	FA =EN	2.	FM=NF	1.	FC=GT	1.	FZ=HW

Here, the initial letter F_c *represents the following initial letters of plaintext digraphs:*

$D\theta_p$, $E\theta_p$, $N\theta_p$, $G\theta_p$, and $H\theta_p$.

It is seen that F_c *represents* D_p, N_p, G_p, H_p *4 times each, and* E_p, *8 times. Consequently, supposing that it has been determined that* $FA_c=EN_p$, *the probability that* F_c *will represent* E_p *is not 1 in 5 but 8 in 24, or 1 in 3; but supposing that it has been determined that* $FW_c=NT_p$, *the probability that* F_c *will represent* N_p *is 4 in 24 or 1 in 6. The difference in these probabilities is occasioned by the fact that the first instance,* $FA_c=EN_p$ *corresponds to a Case 1 encipherment, the second instance,* $FW_c=NT_p$, *to a Case 2 encipherment. But there is no way of knowing initially, and without other data, whether one is dealing with a Case 1 or Case 2 encipherment. Only as an approximation, therefore, may one say that the probability of* F_c *representing a given* θ_p *is 1 in 5. A probability of 1 in 5 is of almost trivial importance in this situation, since it represents such a "long shot" for success. The following rule might be preferable: If the equation 1.2=3.4 has been established, where all the letters represented by 1, 2, 3, and 4 are different, then there is a probability of 4/5 that a Case 1 encipherment is involved. Consequently, if at the same time another equation, 3.6=5.2, has been established, where 2 and 3 represent the same letters as in the first equation, and 5 and 6 are different letters, also different from 2 and 3, there is a probability of 16/25 that the equation 1.6=5.4 is valid; or if at the same time that the equation 1.2=3.4 has been determined, the equation 1.6=5.4 has also been established, then there is a probability of 16/25 that the equation 3.6=5.2 is valid. (Check this by noting the following equations based upon Fig. 65:* $\overset{1\,2}{CE}=\overset{3\,4}{PG}$, $\overset{3\,6}{PH}=\overset{5\,2}{YE}$, $\overset{1\,6}{CH}=\overset{5\,4}{YG}$. *Note the positions occupied in Fig. 65 by the letters involved.) Likewise, if the equations 1.2=3.4 and 1.6=3.5 have been simultaneously established, then there is a probability that the equation 2.5=4.6 is valid; or if the equations 1.2=3.4 and 2.5=4.6 have been simultaneously established, then there is a probability that the equation 2.5=4.6 is valid. (Check this by noting the following equations:* $\overset{1\,2}{CE}=\overset{3\,4}{PG}$; $\overset{1\,6}{CA}=\overset{3\,5}{PK}$; $\overset{2\,5}{EK}=\overset{4\,6}{GA}$; *note the positions occupied in Fig. 65 by the letters involved.) However, it must be added that these probabilities are based upon assumptions which fail to take into account any considerations whatever as to frequency of letters or specificity of composition of the matrix. For instance, suppose the 5 high-frequency letters E, T, N, R, O all happen to fall in the same row or column in the matrix; the number of Case 2 encipherments would be much greater than expectancy and the probability that the equation 1.2=3.4 represents a Case 1 encipherment falls much below 4/5.*

Rule III. In the equation 1.2=3.4, 1 and 3 can never be identical, nor can 2 and 4 ever be identical. Thus, $\overline{\text{AN}}_p$ could not possibly be represented by $\overline{\text{AY}}_c$, nor could $\overline{\text{ER}}_p$ be represented by $\overline{\text{KR}}_c$. This rule is useful in elimination of certain possibilities when a specific message is being studied.

Rule IV. In the equation $1.2_p=3.4_c$, if 2 and 3 are identical, the letters are all in the same row or column, and in the relative order 1–2–4 from left to right or top to bottom, respectively. In the square shown, $\overline{\text{AN}}_p=\overline{\text{NK}}_c$ and the absolute order is ANK. The relative order 1–2–4 includes five absolute orders which are cyclic permutations of one another. Thus: ANK.., NK..A, K..AN, ..ANK, and .ANK..

Rule V. In the equation $1.2_p=3.4_c$, if 1 and 4 are identical, the letters are all in the same row or column, and in the relative order 2–4–3 from left to right or top to bottom. In the square shown, $\overline{\text{KN}}_p=\overline{\text{RK}}_c$ and the absolute order is NKR. The relative order 2–4–3 includes five absolute orders which are cyclic permutations of one another. Thus NKR.., KR..N, R..NK, ..NKR, and .NKR..

Rule VI. "Analyze the message for group recurrences. Select the groups of greatest recurrence and assume them to be high-frequency digraphs.[37] Substitute the assumed digraphs throughout the message, testing the assumptions in their relation to other groups of the cipher. The reconstruction of the square proceeds simultaneously with the solution of the message and aids in hastening the translation of the cipher."

d. (1) When solutions for the Playfair cipher system were first developed, based upon the fact that the letters were inserted in the cells in keyword-mixed order, cryptographers thought it desirable to place stumbling blocks in the path of such solution by departing from strict, keyword-mixed order. One of the simplest methods is illustrated in Fig. 65, wherein it will be noted that the last five letters of the key word proper are inserted in the fourth row of the square instead of the second, where they would naturally fall. Another method involves inserting the letters within the cells from left to right and top downward but using a sequence that is derived from a columnar transposition instead of a keyword-mixed sequence. Thus, using the keyword BANKRUPTCY:

2	1	5	4	7	9	6	8	3	10
B	A	N	K	R	U	P	T	C	Y
D	E	F	G	H	I	L	M	O	Q
S	V	W	X	Z					

Sequence: A E V B D S C O K G X N F W P L R H Z T M U I Y Q

The Playfair square is as follows:

A	E	V	B	D
S	C	O	K	G
X	N	F	W	P
L	R	H	Z	T
M	U	I	Y	Q

FIGURE 66*a*.

[37] A more accurate guide to the determination of the plaintext equivalents of high-frequency cipher digraphs would involve the consideration of the difference in frequency of a particular digraph and its reversal. Thus, an example of a high-frequency $\overline{\theta\theta}_p$ which is also high-frequency in its reversal, is $\overline{\text{RE}}_p$; an example of a high-frequency $\overline{\theta\theta}_p$ which is rarely found in its reversed form, is $\overline{\text{TH}}_p$.

(2) Note the following three squares:

Z	T	L	R	H
Y	Q	M	U	I
B	D	A	E	V
K	G	S	C	O
W	P	X	N	F

FIGURE 66*b*.

O	K	G	S	C
F	W	P	X	N
H	Z	T	L	R
I	Y	Q	M	U
V	B	D	A	E

FIGURE 66*c*.

N	F	W	P	X
R	H	Z	T	L
U	I	Y	Q	M
E	V	B	D	A
C	O	K	G	S

FIGURE 66*d*.

At first glance they all appear to be different, but closer examination shows them to be *cyclic permutations* of one another and of the square in Fig. 66*a*. They yield identical cryptographic equivalents in all cases. However, if an attempt be made to reconstruct the original key word, it would be much easier to do so from Fig. 66*a* than from any of the others, because in Fig. 66*a* the original mixed sequence has not been disturbed as much as in Figs. 66*b*, *c*, and *d*. In working with Playfair ciphers, the student should be on the lookout for such instances of cyclic permutation of the original Playfair square, for during the course of solution he will not know whether he is building up the original or an equivalent cyclic permutation of the original matrix; usually only after he has completely reconstructed the matrix will he be able to determine this point.

e. (1) The steps in the solution of a typical example of this cipher will now be illustrated. Let the message be as follows:

```
VTQEU  HIOFT  CHXSC  AKTVT  RAZEV  TAGAE
OXTYM  HCRLZ  ZTQTD  UMCYC  XCTGM  TYCZU
SNOPD  GXVXS  CAKTV  TPKPU  TZPTW  ZFNBG
PTRKX  IXBPR  ZOEPU  TOLZE  KTTCS  NHCQM
VTRKM  WCFZU  BHTVY  ABGIP  RZKPC  QFNLV
OXOTU  ZFACX  XCPZX  HCYNO  TYOLG  XXIIH
TMSMX  CPTOT  CXOTT  CYATE  XHFAC  XXCPZ
XHYCT  XWLZT  SGPZT  VYWCE  TWGCC  MBHMQ
YXZPW  GRTIV  UXPUM  QRKMW  CXTMR  SWGHB
XCPTO  TCXOT  MIPYD  NFGKI  TCOLX  UETPX
XFSRS  UZTDB  HOZIG  XRKIX  ZPPVZ  IDUHQ
OTKTK  CCHXX
```

(2) Without going through the preliminary tests in detail, with which it will be assumed that the student is now familiar,[38] the conclusion is reached that the cryptogram is digraphic in nature. The digraphic frequency distribution for the cryptogram is shown in Fig. 67.

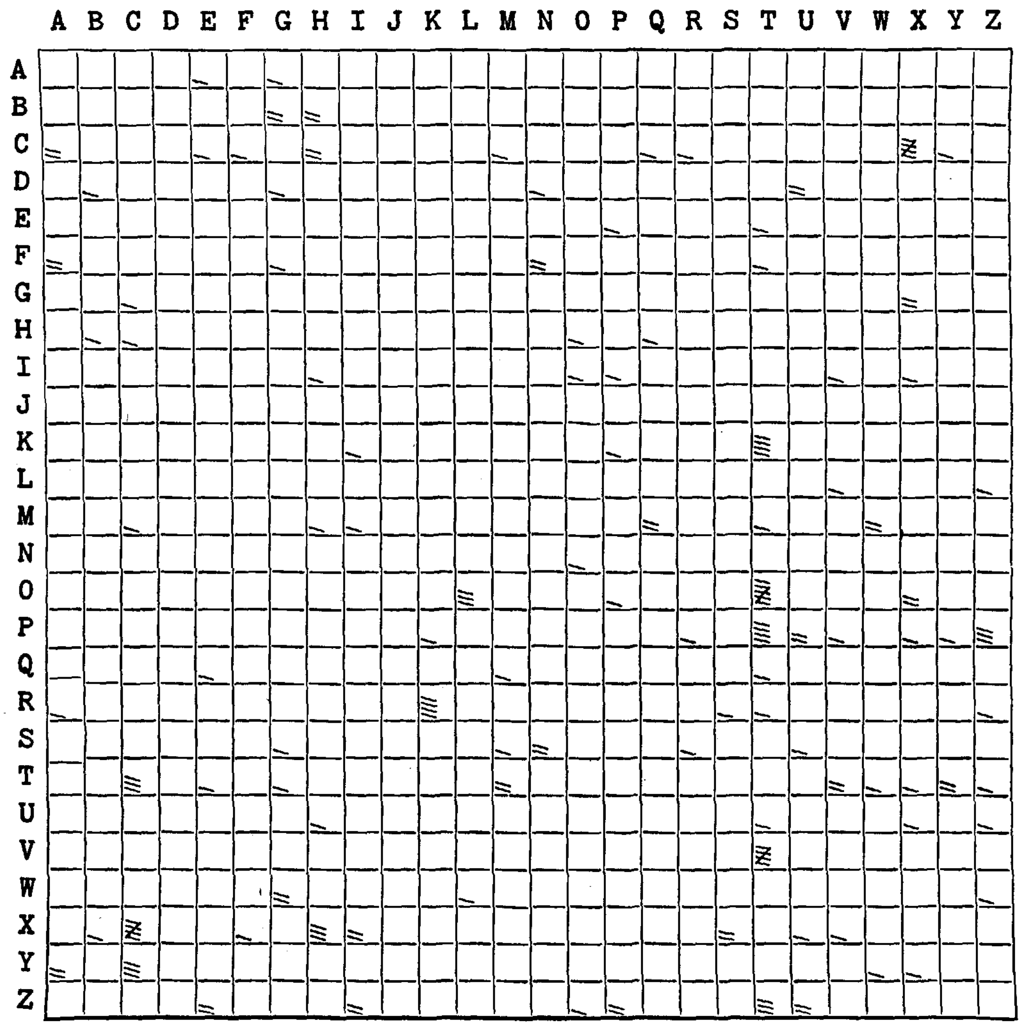


FIGURE 67.

Since there are no double-letter groups (termed "doublets"), the conclusion is reached that a Playfair cipher is involved. The message, having been rewritten in digraphs, is given below.

[38] See par. 69*c*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	VT	QE	UH	IO	FT	CH	XS	CA	KT	VT	RA	ZE	VT	AG	AE
B	OX	TY	MH	CR	LZ	ZT	QT	DU	MC	YC	XC	TG	MT	YC	ZU
C	SN	OP	DG	XV	XS	CA	KT	VT	PK	PU	TZ	PT	WZ	FN	BG
D	PT	RK	XI	XB	PR	ZO	EP	UT	OL	ZE	KT	TC	SN	HC	QM
E	VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY	OL	GX	XI	IH
G	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ
H	XH	YC	TX	WL	ZT	SG	PZ	TV	YW	CE	TW	GC	CM	BH	MQ
J	YX	ZP	WG	RT	IV	UX	PU	MQ	RK	MW	CX	TM	RS	WG	HB
K	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET	PX
L	XF	SR	SU	ZT	DB	HO	ZI	GX	RK	IX	ZP	PV	ZI	DU	HQ
M	OT	KT	KC	CH	XX										

(3) The following three fairly lengthy repetitions are noted:

Lines									
F:	OT	UZ	FA	CX	XC	PZ	XH	CY	NO
G:	TE	XH	FA	CX	XC	PZ	XH	YC	TX
A:	FT	CH	XS	CA	KT	VT	RA	ZE	VT
C:	DG	XV	XS	CA	KT	VT	PK	PU	TZ
G:	TM	SM	XC	PT	OT	CX	OT	TC	YA
K:	WG	HB	XC	PT	OT	CX	OT	MI	PY

The first long repetition, with the sequent reversed digraphs **CX** and **XC** immediately suggests the word **BATTALION** (see Section E, Appendix 3), split up into **–B AT TA LI ON** and the sequence containing this repetition in lines F and G becomes as follows:

Line F------------	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
				B	AT	TA	LI	ON			
Line G------------	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
				B	AT	TA	LI	ON			

(4) Because of the frequent use of numerals before the word BATTALION (as mentioned in Section B of Appendix 4) and because of the appearance of $\overline{ON}_p$ before this word in line G, the possibility suggests itself that the word before BATTALION in line G is either ONE or SECOND. The identical cipher digraph $\overline{FA}$ in both cases gives a hint that the word BATTALION in line F may also be preceded by a numeral; if ONE is correct in line G, then THREE is possible in line F. On the other hand, if SECOND is correct in line G, then THIRD is possible in line F. Thus:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
1st hypothesis	—	TH	RE	EB	AT	TA	LI	ON			
2d hypothesis	—	TH	IR	DB	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
1st hypothesis	—	—	ON	EB	AT	TA	LI	ON			
2d hypothesis	–S	EC	ON	DB	AT	TA	LI	ON			

First, note that if either hypothesis is true, then $\overline{OT}_c=\overline{TH}_p$. The frequency distribution shows that $\overline{OT}$ occurs 6 times and is in fact the most frequent digraph in the message. Moreover, by Rule I of subpar. *b*, if $\overline{OT}_c=\overline{TH}_p$ then $\overline{TO}_c=\overline{HT}_p$. Since $\overline{HT}_p$ is a very rare digraph in normal plain text, $\overline{TO}_c$ should either not occur at all in so short a message or else it should be very infrequent. The frequency distribution shows that it does not occur. Hence, there is nothing inconsistent with the supposition that the word in front of BATTALION in line F is THREE or THIRD and there is some evidence that it is actually one of these.

(5) But can evidence be found for the support of one hypothesis against the other? Let the frequency distribution be examined with a view to throwing light upon this point. If the first hypothesis is true, then $\overline{UZ}_c=\overline{RE}_p$, and, by Rule I, $\overline{ZU}_c=\overline{ER}_p$. The frequency distribution shows but one occurrence of $\overline{UZ}_c$ and but two occurrences of $\overline{ZU}_c$. These do not look very good for $\overline{RE}_p$ and $\overline{ER}_p$. On the other hand, if the second hypothesis is true, then $\overline{UZ}_c=\overline{IR}_p$, and, by Rule I, $\overline{ZU}_c=\overline{RI}_p$. The frequencies are much more favorable in this case. Is there anything inconsistent with the assumption, on the basis of the second hypothesis, that $\overline{TE}_c=\overline{EC}_p$? The frequency distribution shows no inconsistency, for $\overline{TE}_c$ occurs once and $\overline{ET}_c(=\overline{CE}_p$, by Rule I) occurs once. As regards whether $\overline{FA}_c=\overline{EB}_p$ or $\overline{DB}_p$, both hypotheses are tenable; possibly the second hypothesis is a shade better than the first, on the following reasoning: By Rule I, if $\overline{FA}_c=\overline{EB}_p$ then $\overline{AF}_c=\overline{BE}_p$, or if $\overline{FA}_c=\overline{DB}_p$ then $\overline{AF}_c=\overline{BD}_p$. The fact that no $\overline{AF}_c$ occurs, whereas at least one $\overline{BE}_p$ may be expected in this message, inclines one to the second hypothesis, since $\overline{BD}_p$ is very rare.

(6) Let the second hypothesis be assumed to be correct. The additional values are tentatively inserted in the text, and in lines G and K two interesting repetitions are noted:

Line G	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ	XH
			TA		TH	AT	TH		–S	EC	ON	DB	AT	TA	LI	ON
Line K	WG	HB	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET
			TA		TH	AT	TH									

This certainly looks like STATE THAT THE . . . , which would make $\overline{TE}_p=\overline{PT}_c$. Furthermore, in line G the sequence STATETHATTHE. . SECONDBATTALION can hardly be anything else than

STATE THAT THEIR SECOND BATTALION, which would make $\overline{TC}_c=\overline{EI}_p$ and $\overline{YA}_c=\overline{RS}_p$. Also $\overline{SM}_c=\overline{-S}_p$.

(7) It is perhaps high time that the whole list of tentative equivalent values be studied in relation to their consistency with the positions of letters in the Playfair square; moreover, by so doing, additional values may be obtained in the process. The complete list of values is as follows:

Assumed values	*Derived by Rule I*
$\overline{AT}_p=\overline{CX}_c$	$\overline{TA}_p=\overline{XC}_c$
$\overline{LI}_p=\overline{PZ}_c$	$\overline{IL}_p=\overline{ZP}_c$
$\overline{ON}_p=\overline{XH}_c$	$\overline{NO}_p=\overline{HX}_c$
$\overline{TH}_p=\overline{OT}_c$	$\overline{HT}_p=\overline{TO}_c$
$\overline{IR}_p=\overline{UZ}_c$	$\overline{RI}_p=\overline{ZU}_c$
$\overline{DB}_p=\overline{FA}_c$	$\overline{BD}_p=\overline{AF}_c$
$\overline{EC}_p=\overline{TE}_c$	$\overline{CE}_p=\overline{ET}_c$
$\overline{TE}_p=\overline{PT}_c$	$\overline{ET}_p=\overline{TP}_c$
$\overline{EI}_p=\overline{TC}_c$	$\overline{IE}_p=\overline{CT}_c$
$\overline{RS}_p=\overline{YA}_c$	$\overline{SR}_p=\overline{AY}_c$
$\overline{-S}_p=\overline{SM}_c$	$\overline{S-}_p=\overline{MS}_c$

(8) By Rule V, the equation $\overline{TH}_p=\overline{OT}_c$ means that H, O, and T are all in the same row or column and in the absolute order HTO; similarly, C, E, and T are in the same row or column and in the absolute order CET. Further, E, P, and T are in the same row and column, and their absolute order is ETP. That is, these sequences must occur some place in the square, in either rows or columns, taking into consideration of course the probability of cyclic displacements of these sequences within the square:

(a) HTO (b) CET (c) ETP

(9) Noting the common letters E and T in the second and third sequences, these two sequences may be combined into one sequence of four letters, *viz.*, CETP. Since only one position remains to be filled in this row (or column) of the square, and noting in the list of equivalents that $\overline{EI}_p=TC_c$, it is obvious that the letter I belongs to the CETP sequence; the complete sequence is therefore ICETP.

(10) Since the sequence HTO has a common letter (T) with the sequence ICETP, it follows that if the HTO sequence occupies a row, then the ICETP sequence must occupy a column; or, if the HTO sequence occupies a column, then the ICETP sequence must occupy a row; and they may be combined by means of their common letter, T, *viz.*:

```
              ┌ - - ┐
              ¦  I  ¦
              ¦  C  ¦
┌ - - - - ┐   ¦  E  ¦
¦      H  ¦   ¦H T O¦
¦I C E T P¦   ¦  P  ¦
¦      O  ¦   └ - - ┘
└ - - - - ┘
```

The proof of whether the ICETP sequence, for example, properly belongs in a row or a column of the Playfair square lies in the establishment of a *rectangular* relationship, instead of the *linear* relationships constructed thus far.

(11) We note that, from the assumptions in subpar. *d*(6), $\overline{AT}_p=\overline{CX}_c$ and $\overline{ON}_p=\overline{XH}_c$. The relationship $\overline{ON}_p=\overline{XH}_c$ might be either a rectangular one, such as

```
O   X
 \ /
  X
 / \
H   N
```

or it might be linear, *viz.*, HTOXN or

```
H
T
O
X
N
```

Since, however, $\overline{AT}_p=\overline{CX}_c$ must be a rectangular relationship, then only the configuration

	I			
	C		A	
	E			
H	T	O	X	N
	P			

will be valid, since the alternative form

			H	
I	C	E	T	P
			O	
	A		X	
			N	

will not satisfy the equation $\overline{AT}_p=\overline{CX}_c$.

(12) The fragmentary Playfair square [39] has been established, in one of its 25 possible cyclic permutations, as follows:

	I			
	C		A	
	E			
H	T	O	X	N
	P			

FIGURE 68*a*.

Scanning the list of plain-cipher equivalents given in subpar. *d*(7) in order to insert possible additional values, note is made of $\overline{IR}_p=\overline{UZ}_c$, which means that U must be in the same row as I; and since Z cannot be in the same column as I the square must be one of the two following possibilities:

	I	U	R	Z
	C		A	
	E			
H	T	O	X	N
	P			

FIGURE 68*b*.

Z	I	U		R
	C		A	
	E			
H	T	O	X	N
	P			

FIGURE 68*c*.

[39] In actual practice, it is more usual to start with a much larger diagram than a simple 5 x 5 square; as relationships develop, the diagram is gradually condensed, until finally a 5 x 5 square emerges. This procedure is quite similar to that employed in the reconstruction diagrams for two-square matrices.

(13) Now note that $\overline{RS}_p=\overline{YA}_c$; this eliminates one of the two squares above, thus the correct square is now

Y	I	U	R	Z
S	C		A	
	E			
H	T	O	X	N
	P			

FIGURE 68*d*.

Since $\overline{LI}_p=\overline{PZ}_c$, this places L in the square:

Y	I	U	R	Z
S	C		A	
	E			
H	T	O	X	N
	P			L

FIGURE 68*e*.

Finally, since $\overline{DB}_p=\overline{FA}_c$, the new letters can be placed in the square in the three following ways:

Y	I	U	R	Z
S	C	B	A	
	E	F	D	
H	T	O	X	N
	P			L

FIGURE 68*f*.

Y	I	U	R	Z
S	C		A	B
	E		D	F
H	T	O	X	N
	P			L

FIGURE 68*g*.

Y	I	U	R	Z
S	C	B	A	
	E			
H	T	O	X	N
	P	F	D	L

FIGURE 68*h*.

Checking back to the cipher text at A5, of the three possibilities for $\overline{FT}_c$ (=$\overline{EO}_p$, $\overline{EN}_p$, or $\overline{PO}_p$), the obvious choice is $\overline{PO}_p$ in the word –O UT PO ST, so this confirms Fig. 68*h* as the correct square of the three possibilities.

(14) It is now a simple matter to decipher the cryptogram and make the few assumptions in the text necessary to permit filling in the remaining six letters in the square, which will result in its completion as follows:

Y	I	U	R	Z
S	C	B	A	G
M	E	Q	K	V
H	T	O	X	N
W	P	F	D	L

FIGURE 68*i*.

~~CONFIDENTIAL~~

f. Reconstruction of the square in Playfair ciphers is normally carried on concurrently with the synthesis of the plain text, once a few correct assumptions have been made. Now, having just reconstructed the square as shown in Fig. 68*i*, the question to be answered is whether this square is identical with the original enciphering matrix or whether it is a cyclic permutation of the original square (which may have contained, say, a transposition-mixed sequence). Even though the cryptogram in subpar. 71*e* has been solved, this point is still of interest.

(1) The square that is derived may not necessarily be the original enciphering square; more than likely it will be one of the 24 possible cyclic permutations of the original square. If the Playfair square consisted of a keyword-mixed sequence, a permutation of the square will cause no difficulty in recovering the original matrix and hence the key word. For example, if the square derived in some other instance is

```
Q T L N O
X Z U V W
A M P Y R
B C I D S
H K E F G
```

then the square

```
P Y R A M
I D S B C
E F G H K
L N O Q T
U V W X Z
```

is easily recovered because of the telltale letters UVWXZ occurring in a row of the derivative square. But when the Playfair square consists of a transposition-mixed sequence, then a different procedure must be adopted.

(2) As an example, let us take the transposition matrix

```
5 8 6 1 4 3 2 7
P Y R A M I D S
B C E F G H K L
N O Q T U V W X
Z
```

from which

```
A F T D K
W I H V M
G U P B N
Z R E Q S
L X Y C O
```

is the original square. Using the methods illustrated in par. 51*g*, scanning successive rows of the square will disclose sequences of letters which could have appeared as columns in the transposition matrix. For example, discovery of the columns

```
I|D|S
H|K|L
V|W|X
```

will afford rapid recovery of the key word. But if instead of the original square we had one of its permutations such as

```
Q S Z R E
C O L X Y
D K A F T
V M W I H
B N G U P
```

then treatment of the "columns", e. g.,

```
F|V|O|L|Q
T|M|L|X|S
V|W|X|Y|Z
```

of the tentative transposition matrix (assuming that some or all of the letters V, W, X, Y, Z are in the last row of the transposition matrix) will be without significance; therefore the procedure above is inapplicable without a slight modification.

(3) Since it will be noted that a permutation of the rows will not affect the procedure of keyword recovery, then we construct a 9 x 5 rectangle

```
Q S Z R E Q S Z R
C O L X Y C O L X
D K A F T D K A F
V M W I H V M W I
B N G U P B N G U
```

which contains

~~CONFIDENTIAL~~

the five squares which result simply from successive permutations of the columns. A 5 x 5 cut-out square will be found convenient in testing each permutation in turn. Recovery of the key word will be possible when the correct permutation is reached, which in this case is the third square in the rectangle, namely,

```
Z R E Q S
L X Y C O
A F T D K
W I H V M
G U P B N
```

After recovery of the key word from this permuted square it is probable then that the original enciphering square must have been

```
A F T D K
W I H V M
G U P B N
Z R E Q S
L X Y C O
```

(4) In the case of the square recovered in Fig. 68*i*, it is found that, following the procedure outlined in subpars. (1), (2), and (3), above, the key word is based on COMPANY, recoverable from the following diagram:

```
2 5 3 6 1 4 7
C O M P A N Y
B D E F G H I
K L Q R S T U
V W X Z
```

The original square must have been this:

A	G	S	C	B
K	V	M	E	Q
X	N	H	T	O
D	L	W	P	F
R	Z	Y	I	U

FIGURE 68*j*.

g. Continued practice in the solution of Playfair ciphers will make the student quite expert in the matter and will enable him to solve shorter and shorter messages.[40] Also, with practice it will become a matter of indifference to him as to whether the letters are inserted in the square with any sort of regularity, such as simple keyword-mixed order, transposition-mixed order, or in a purely random order.

h. It may perhaps seem to the student that the foregoing steps are somewhat too artificial, a bit too "cut and dried" in their accuracy to portray the process of analysis as it is applied in practice. For example, the critical student may well object to some of the assumptions and the reasoning in subpar. *e* (5), above, in which the words THREE and ONE (1st hypothesis) were rejected in favor of the words THIRD and SECOND (2d hypothesis). This rested largely upon the rejection of $\overline{RE}_p$ and $\overline{ER}_p$ as the equivalents of $\overline{UZ}_c$ and $\overline{ZU}_c$, and the adoption of $\overline{IR}_p$ and $\overline{RI}_p$ as their equivalents. Indeed, if the student will examine the final plain text with a

[40] The author once had a student who "specialized" in Playfair ciphers and became so adept that he could solve messages containing as few as 50–60 letters within 30 minutes.

critical eye, he will find that while the bit of reasoning in step (5) is perfectly logical, the assumption upon which it is based is in fact wrong; for it happens that in this case $\overline{ER}_p$ occurs only once and $\overline{RE}_p$ does not occur at all. Consequently, although most of the reasoning which led to the rejection of the first hypothesis and the adoption of the second was logical, it was in fact based upon erroneous assumption. In other words, despite the fact that the assumption was incorrect, a correct deduction was made. *The student should take note that in cryptanalysis situations of this sort are not at all unusual.* Indeed they are to be expected, and a few words of explanation at this point may be useful.

i. Cryptanalytics is a science in which deduction, based upon observational data, plays a very large role. But it is also true that in this science most of the deductions usually rest upon assumptions. It is most often the case that the cryptanalyst is forced to make his assumptions based upon a quite limited amount of text. It cannot be expected that assumptions based upon statistical generalizations will always hold true when applied to data comparatively very much smaller in quantity than the total data used to derive the generalized rules. Consequently, as regards assumptions made in specific messages, *most of the time* they will be correct, but *occasionally* they will be incorrect.[41] In cryptanalysis it is often found that among the correct deductions there will be cases in which subsequently discovered facts do not bear out the assumptions on which the deduction was based. Indeed, it is sometimes true that if the *facts* had been known *before* the deduction was made, this knowledge would have prevented making the correct deduction. For example, suppose the cryptanalyst had somehow or other divined that the message under consideration contained no RE, only one ER, one IR, and two RI's (as is actually the case). He would certainly not have been able to choose between the words THREE and ONE (1st hypothesis) as against THIRD and SECOND (2d hypothesis). But because he assumes that there should be more $\overline{ER}_p$'s and $\overline{RE}_p$'s than $\overline{IR}_p$'s and $\overline{RI}_p$'s in the message, he deduces that $\overline{UZ}_c$ cannot be $\overline{RE}_p$, rejects the first hypothesis and takes the second. It later turns out, after the problem has been solved, that the deduction was correct, although the assumption on which it was based (expectation of more frequent appearance of $\overline{RE}_p$ and $\overline{ER}_p$) was, in fact, *not* true in this particular case. The cryptanalyst can only hope that the number of times when his deductions are correct, even though based upon assumptions which later turn out to be erroneous, will abundantly exceed the number of times when his deductions are wrong, even though based upon assumptions which later prove to be correct. If he is lucky, the making of an assumption which is really not true will make no difference in the end and will not delay solution; but if he is specially favored with luck, it may actually help him solve the message—as was the case in this particular example.

j. Another comment of a general nature may be made in connection with this specific example. The student may ask what would have been the procedure in this case if the message had not contained such a telltale repetition as the word BATTALION, which formed the point of departure for the solution, or, as it is often said, permitted an "entering wedge" to be driven into the message. The answer to his query is that if the word BATTALION had not been repeated, there would probably have been some other repetition which would have permitted the same sort of attack. If the student is looking for cut and dried, straightforward, unvarying methods of attack, he should remember that cryptanalytics, while considered a branch of mathematics by some, is not a science which has many "general solutions" such as are found and expected in

[41]See footnote 19 on p. 43.

mathematics proper. It is inherent in the very nature of cryptanalytics that, *as a rule,* only general principles can be established; their practical application must take advantage of peculiarities and particular situations which are noted in specific messages. This is especially true in a text on the subject. The illustration of a general principle requires a specific example, and the latter must of necessity manifest characteristics which make it different from any other example. The word BATTALION was not purposely repeated in this example in order to make the demonstration of solution easy; "it just happened that way". In another example, some other entering wedge would have been found. The student can be expected to learn only the *general principles* which will enable him to take advantage of the *specific characteristics* manifested in *specific cases.* Here it is desired to illustrate the general principles of solving Playfair ciphers and to point out the fact that entering wedges must and can be found. The specific nature of the entering wedge varies with specific examples.

72. Analysis of polygraphic systems involving large tables.—*a.* The analysis of systems incorporating large digraphic tables is accomplished by entering, within the appropriate cells of a 26 x 26 chart, data corresponding to the plain-cipher relationships of assumed cribs, and examining the charts for evidences of symmetry or systematic construction in their compilation. The initial plaintext entries may, in the absence of cribs, be made on the basis of digraphic frequency considerations, aided by idiomorphisms and repetitions.

b. In pseudo-digraphic systems, such as those incorporating tables similar to Figs. 47*a* and *b*, and 48, the identification of the monoalphabetically-enciphered component of cipher digraphs will greatly accelerate plaintext entries, since advantage may be taken of this monoalphabeticity. Tables with a feature of reciprocity, such as the example in Fig. 50, may be exploited on the basis of *this* weakness, even if the reciprocal pairs are assigned at random. Tables such as that in Fig. 49 and the one for trinome digraphic encipherment shown in Fig. 51 may also be exploited with facility, once enough plain text has been correctly assumed and inserted to disclose their systematic construction. A word of warning is inserted here against making incautious assumptions concerning the exact internal composition of tables such as that in Fig. 49, since their unusual construction could easily mislead the analyst who jumps to premature conclusions. In the case of a table such as Fig. 51 wherein the trinomes have been inscribed in straight horizontals, if the dimensions of the table have been correctly assumed the simplest solution involves a reduction to two alphabets, reflecting the sequences of letters for the side and top of the matrix; this solution closely parallels that of the numerical four-square system described in subpar. 69*e*.

c. Because the foregoing principles are rather straightforward, it is not considered necessary to illustrate their application with examples. Of course, when digraphic tables of random construction have been used, no refinements in solution are possible. However, the recording of as few as 225 different plaintext digraphs and their ciphertext equivalents will theoretically enable the automatic decryption of approximately 92% of the cipher digraphs of messages, and the recording of 335 plaintext-ciphertext values will enable the automatic decryption of 98% of the cipher digraphs; thus almost every message may be read in its entirety without recourse to further assumptions. Actually, it should be pointed out that having only 122 matched plaintext-ciphertext equivalencies will theoretically enable the decryption of 75% of the cipher digraphs, and enough skeletons of plain text may then be manifest to permit the decryption of the complete message texts.

d. It might be well to point out in connection with large digraphic tables that there exist literal types which give rise to monoalphabetic distributions for *both* the initial letters and final letters of pairs. Such a table is illustrated in Fig. 69 below:

θ_p^2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	HQ	YQ	DQ	RQ	AQ	UQ	LQ	IQ	CQ	BQ	EQ	FQ	GQ	JQ	KQ	MQ	NQ	OQ	PQ	QQ	SQ	TQ	VQ	WQ	XQ	ZQ
B	HU	YU	DU	RU	AU	UU	LU	IU	CU	BU	EU	FU	GU	JU	KU	MU	NU	OU	PU	QU	SU	TU	VU	WU	XU	ZU
C	HE	YE	DE	RE	AE	UE	LE	IE	CE	BE	EE	FE	GE	JE	KE	ME	NE	OE	PE	QE	SE	TE	VE	WE	XE	ZE
D	HS	YS	DS	RS	AS	US	LS	IS	CS	BS	ES	FS	GS	JS	KS	MS	NS	OS	PS	QS	SS	TS	VS	WS	XS	ZS
E	HT	YT	DT	RT	AT	UT	LT	IT	CT	BT	ET	FT	GT	JT	KT	MT	NT	OT	PT	QT	ST	TT	VT	WT	XT	ZT
F	HI	YI	DI	RI	AI	UI	LI	II	CI	BI	EI	FI	GI	JI	KI	MI	NI	OI	PI	QI	SI	TI	VI	WI	XI	ZI
G	HO	YO	DO	RO	AO	UO	LO	IO	CO	BO	EO	FO	GO	JO	KO	MO	NO	OO	PO	QO	SO	TO	VO	WO	XO	ZO
H	HN	YN	DN	RN	AN	UN	LN	IN	CN	BN	EN	FN	GN	JN	KN	MN	NN	ON	PN	QN	SN	TN	VN	WN	XN	ZN
I	HA	YA	DA	RA	AA	UA	LA	IA	CA	BA	EA	FA	GA	JA	KA	MA	NA	OA	PA	QA	SA	TA	VA	WA	XA	ZA
J	HB	YB	DB	RB	AB	UB	LB	IB	CB	BB	EB	FB	GB	JB	KB	MB	NB	OB	PB	QB	SB	TB	VB	WB	XB	ZB
K	HL	YL	DL	RL	AL	UL	LL	IL	CL	BL	EL	FL	GL	JL	KL	ML	NL	OL	PL	QL	SL	TL	VL	WL	XL	ZL
L	HY	YY	DY	RY	AY	UY	LY	IY	CY	BY	EY	FY	GY	JY	KY	MY	NY	OY	PY	QY	SY	TY	VY	WY	XY	ZY
M	HC	YC	DC	RC	AC	UC	LC	IC	CC	BC	EC	FC	GC	JC	KC	MC	NC	OC	PC	QC	SC	TC	VC	WC	XC	ZC
θ_p^1 N	HD	YD	DD	RD	AD	UD	LD	ID	CD	BD	ED	FD	GD	JD	KD	MD	ND	OD	PD	QD	SD	TD	VD	WD	XD	ZD
O	HF	YF	DF	RF	AF	UF	LF	IF	CF	BF	EF	FF	GF	JF	KF	MF	NF	OF	PF	QF	SF	TF	VF	WF	XF	ZF
P	HG	YG	DG	RG	AG	UG	LG	IG	CG	BG	EG	FG	GG	JG	KG	MG	NG	OG	PG	QG	SG	TG	VG	WG	XG	ZG
Q	HH	YH	DH	RH	AH	UH	LH	IH	CH	BH	EH	FH	GH	JH	KH	MH	NH	OH	PH	QH	SH	TH	VH	WH	XH	ZH
R	HJ	YJ	DJ	RJ	AJ	UJ	LJ	IJ	CJ	BJ	EJ	FJ	GJ	JJ	KJ	MJ	NJ	OJ	PJ	QJ	SJ	TJ	VJ	WJ	XJ	ZJ
S	HK	YK	DK	RK	AK	UK	LK	IK	CK	BK	EK	FK	GK	JK	KK	MK	NK	OK	PK	QK	SK	TK	VK	WK	XK	ZK
T	HM	YM	DM	RM	AM	UM	LM	IM	CM	BM	EM	FM	GM	JM	KM	MM	NM	OM	PM	QM	SM	TM	VM	WM	XM	ZM
U	HP	YP	DP	RP	AP	UP	LP	IP	CP	BP	EP	FP	GP	JP	KP	MP	NP	OP	PP	QP	SP	TP	VP	WP	XP	ZP
V	HR	YR	DR	RR	AR	UR	LR	IR	CR	BR	ER	FR	GR	JR	KR	MR	NR	OR	PR	QR	SR	TR	VR	WR	XR	ZR
W	HV	YV	DV	RV	AV	UV	LV	IV	CV	BV	EV	FV	GV	JV	KV	MV	NV	OV	PV	QV	SV	TV	VV	WV	XV	ZV
X	HW	YW	DW	RW	AW	UW	LW	IW	CW	BW	EW	FW	GW	JW	KW	MW	NW	OW	PW	QW	SW	TW	VW	WW	XW	ZW
Y	HX	YX	DX	RX	AX	UX	LX	IX	CX	BX	EX	FX	GX	JX	KX	MX	NX	OX	PX	QX	SX	TX	VX	WX	XX	ZX
Z	HZ	YZ	DZ	RZ	AZ	UZ	LZ	IZ	CZ	BZ	EZ	FZ	GZ	JZ	KZ	MZ	NZ	OZ	PZ	QZ	SZ	TZ	VZ	WZ	XZ	ZZ

FIGURE 69.

In effect, encipherment by means of such a system yields the equivalent of a two-alphabet cipher, with a transposition within each of the pairs of letters. The cipher text produced by such a system may be characterized by a large number of repetitions which begin with the initial letter of digraphs and end on the final letter of digraphs and which are preceded by digraphs having repeated *initial* letters or which are followed by digraphs having repeated *final* letters; for example, ciphertext passages of the following type might often arise: SF BD GB HK and SQ BD GB WK (wherein the repeated plain text is actually represented by SDBBGK, affected by the transposition). This system is included here as being illustrative of many simple systems which are capable of leading the student very much astray; in this instance, if one were unaware of the transposition feature involved and were to attempt what appears to be the simple task of fitting plain text into the two monoalphabetic portions on the basis of single-letter frequency considerations, he could spend a great deal of time without success—probably without any idea of what was causing his difficulties.

e. A pseudo-trigraphic cipher involving a table such as that in Fig. 52 may be readily recognized as such, since two letters of each trigraph enciphered by means of such a table are treated monoalphabetically. If three separate uniliteral frequency distributions are made—one for each of the three letters of the cipher trigraphs—two of the distributions should be monoalphabetic. Then, exploiting the monoalphabeticity (i. e., the *positional* monoalphabeticity) thus disclosed in the cipher text, plain text can be fitted to the cipher on the basis of single-letter frequency considerations; in addition, advantage may be taken of *partial* idiomorphisms, if these idiomorphisms involve the particular positions of the trigraphs which have been treated monoalphabetically.

f. Fortunately, it is unlikely that trigraphic systems other than the foregoing pseudo-trigraphic type will be encountered, because they are difficult to manipulate without extensive tables or complicated rules for encryption.[42] The subject can be passed over with the simple statement that their analysis requires much text to permit of solution by the frequency method—and blood, sweat, toil, and tears.[43]

73. Further remarks on polygraphic substitution systems.—*a*. In the treatment of the cryptography of the various digraphic systems in this chapter, the rules for encryption and decryption which have been illustrated are the "standard" rules (i. e., the rules extant in cryptologic literature, or the rules most commonly encountered in operational practice). Needless to say, however, there is no cryptologic counterpart of the Geneva Convention making these rules sacrosanct, nor forbidding the use of other rules for enciphering and deciphering.

b. In two-square systems and Playfair systems there are possible (and, in fact, there have been encountered in operational practice) modifications of the usual enciphering and deciphering rules which, if not suspected, may pose difficulties in the identification of such systems and in their cryptanalysis. For example, in a vertical two-square system, when two plaintext letters fall in the same column, their cipher equivalents might be taken as the letters immediately to the right of or immediately below these plaintext letters. Similarly, in a horizontal two-square system, if two plaintext letters are in the same row, their cipher equivalents might be taken as those immediately below or to the right of these letters. In Playfair cipher systems, two plaintext letters in the same row might be represented by the letters immediately below; two plaintext letters in the same column might be represented by the letters immediately to the right; a plaintext doublet might be represented by a ciphertext *doublet* formed by doubling the letter immediately to the right, or below, or diagonally to the right and below, thus removing one of the identifying ciphertext characteristics of the *normal* Playfair system. In one case encountered, instead of the normal Playfair linear relationship $\overline{AB}_p=\overline{BC}_c$, the rule was changed to $\overline{AB}_p=\overline{CB}_c$ (thus allowing a letter to "represent itself"—an "impossibility" in Playfair encipherment); even this simple modification caused difficulties in cryptanalysis because variant rules for encryption had not been considered.

c. The placing of cribs in small-matrix digraphic systems may be guided by the cryptographic peculiarities of these systems, when the general system is known to, or suspected by the cryptanalyst; conversely, the placing of a known crib may assist in the determination of the type of cryptosystem, or in the rejection of other types of systems. For example, cribs may be placed in Playfair ciphers on the basis of the "non-crashing" feature of the normal Playfair; that is, on the basis that in the equation 1.2=3.4 neither 1 and 3 nor 2 and 4 can be identical. Further-

[42] However, see in this connection subpar. 73*h*, which treats of a relatively simple mathematical method for enciphering polygraphs of *any* size.

[43] If a trigraphic system is encountered in operational cryptanalysis, special solutions would be made possible by the application of cribs, the aid furnished by isologs (not only in the same system, but also *between* systems), etc.

more, in the normal Playfair, $\alpha\beta_c$ cannot equal $\beta\alpha_p$. In horizontal two-square systems, if $\alpha\beta_c = -\alpha_p$, then $\alpha\beta_c$ must equal $\beta\alpha_p$; and if $\alpha\beta_c = \beta-_p$, then $\alpha\beta_c$ must equal $\beta\alpha_p$. If, by placing a known crib in a cryptogram, evidence of *non*-reciprocity is disclosed (e. g., if $\overline{AB}_p = \overline{CD}_c$, but $\overline{CD}_p = \overline{XY}_c$), the cryptogram may be assumed to be other than a vertical two-square cipher, since vertical two-square encipherment yields complete reciprocity. In either type of two-square system, if one of the two squares is known (for example, a vertical two-square might be employed in which the upper square is always a normal alphabet), the placement of cribs is materially facilitated.

d. The ϕ test performed separately on the initial letters and final letters of ciphertext pairs from cryptograms produced by small-matrix digraphic systems will give results neither close to that expected for plain text, nor close to that for random text. The reason for the comparative "roughness" or pronounced differences among the relative frequencies in these distributions, as contrasted with the "smoothness" expected of random, is that small-matrix digraphic systems are only *partially* digraphic in nature and that the encryption involves characteristics similar to those of monoalphabetic substitution with variants. This roughness of the uniliteral frequency distributions for the initial and final letters, and, for that matter, for the over-all cipher text, reflects the partially digraphic nature of the encipherment.

e. If the cipher letters **V**, **W**, **X**, **Y**, and **Z** are of very low frequency in the over-all uniliteral frequency distribution of a digraphic cryptogram or set of cryptograms, this may be taken as evidence that the cryptosystem is a small-matrix digraphic system employing keyword-mixed sequences in the matrix or matrices. Furthermore, in small-matrix systems involving keyword-mixed squares, if θ^1_c of $\overline{\theta\theta}_c$ is one of the letters **VWXYZ**, the θ^1_p of the corresponding $\overline{\theta\theta}_p$ is likely to be one of these same letters. Similarly, if θ^2_c is one of the letters **VWXYZ**, then θ^2_p of the corresponding $\overline{\theta\theta}_p$ is likely to be one of these letters.

f. In trinome-digraphic systems employing large tables, the trinomes may run from 001 to 676, as in Fig. 51, or any consecutive set of 676 trinomes in the scale of 1000 possible trinomes may be used. For that matter, the entire span of trinomes 000–999 might be used in such a table, with occasional gaps, to hide the limitations of this system. As another means of disguising the limitation of 676 trinomes in such a system, three of the initial digits of the trinomes might have one variant each—thus no limitation would exist in the first position of trinomes. The 001, or other starting point in the cyclic scale, need not be at the upper left-hand corner of the table. The 676 trinomes in such tables may be inscribed in straight horizontals (i. e., in the normal manner of writing) as in Fig. 51, or they might be inscribed according to some other route; they probably would *not* be inscribed in a random manner because clumsy "deciphering tables" would then be necessary. It is also possible that the trinomes in a trinome-digraphic system might be converted into tetranomes by the addition of a sum-check (to assist in error-correction).

g. The cryptanalysis of tetranome-trigraphic systems with matrices similar to that illustrated in Fig. 59 involves a modification of the technique used in solving inverse four-square systems. If the plain-component and cipher-component sections of the large square have been inscribed according to the normal manner of writing (or any other manner, *if known*), the first two elements of the trigraphs may be reduced to a pair of cipher alphabets, and these two monoalphabetic substitutions may be solved as indicated in subpar. 69*e*. The applicability of inverse four-square solution principles to this tetranome-trigraphic system of course rests on the fact that the ciphertext sections are known or assumed to contain the dinomes 00–99 in numerical order, inscribed in the normal manner of writing; the conversion of the first two elements of the trigraphs depends upon the knowledge of the manner of inscription of the letters of the plain component sections, in order that the *four occurrences of the initial letters* and the *four occurrences* *of the final letters* may be correctly combined into two monoalphabetic distributions. Of course, if the composition of the small square (for the third element of trigraphs) is known, the third letter of trigraphs may be automatically deciphered. If the composition of the small square is not known, a consideration of the frequencies of the converted dinomes for the small square (i. e., the coordinates of the square to indicate the third member of trigraphs) may be used to obtain an entering wedge into this *third* monoalphabetic substitution.

h. There are but a very limited number of known cipher mechanisms which employ the polygraphic encipherment principle in any form. U. S. Patent No. 1515680 issued to A. Henkels in 1924 and U. S. Patent No. 1845947 issued to Weisner and Hill in 1932 describe two such mechanisms which produce polygraphic substitution. The latter, that of Weisner and Hill, is of particular interest because it is based on a rather simple mathematical process which can yield true polygraphic encipherment for polygraphs of any size. The underlying mathematical process, invented by Prof. Lester S. Hill of Hunter College and described in the "American Mathematical Monthly" in 1929 (Vol. XXXVI, p. 306) and 1931 (Vol. XXXVIII, p. 135), is treated briefly, below.

(1) Since Professor Hill's system is mathematical in nature, the first step in its use involves the conversion of the plaintext letters into numbers by means of a conversion alphabet which shows a correspondence between the 26 letters of the alphabet and the 26 numbers from 0 to 25, such as the following:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	9	3	5	24	6	18	8	11	1	21	14	15	12	4	10	25	17	7	19	20	2	22	16	23	13

(2) The numbers obtained through the conversion of the plaintext letters are next treated arithmetically through the application of algebraic linear functions, this treatment being performed by means of mod 26 arithmetic.[43] The numerical results yielded by the algebraic treatment are then converted back into letters by means of the conversion alphabet, to yield the cipher equivalent of the original plain text.

(3) For example, suppose that the message, "NOTHING TO REPORT" is to be enciphered by trigraphs, and that, for this purpose, the enciphering keys[44] are (1, 2, 1); (5, 11, 3); (2, 4, 13). The message would be divided into trigraphs NOT–HIN–GTO–REP–ORT and the letters which result from the following operation would be taken as the cipher equivalent of the first trigraph:

Using the conversion alphabet in (1), above, "N–O–T" is converted into "12–4–19"; then the foregoing keys are applied—

$$(1\times\underline{12})+(2\times\underline{4})+(1\times\underline{19})=12+8+19=13+(1\times26)=\text{Z}$$
$$(5\times\underline{12})+(11\times\underline{4})+(3\times\underline{19})=60+44+57=5+(6\times26)=\text{D}$$
$$(2\times\underline{12})+(4\times\underline{4})+(13\times\underline{19})=24+16+247=1+(11\times26)=\text{J}$$

Thus, $\overline{\text{NOT}}_p$ is enciphered as $\overline{\text{ZDJ}}_c$. The complete encipherment would read ZDJ XMH HQH YMA DOI.

---

[43] Using "mod 26 arithmetic", one considers as the sum or product of two numbers, the number from 0–25 which is obtained by subtracting 26 (or a multiple of 26) from the ordinary arithmetical sum or product of the numbers.

[44] Encipherment of polygraphs containing n letters requires the use of n^2 keys. Thus, 9 keys are necessary for trigraphic encipherment; digraphic encipherment requires only 4 keys, whereas tetragraphic and pentagraphic encipherment necessitate the use of 16 and 25 keys, respectively. The numbers selected for use as keys must be chosen according to rather definite rules which evolve from the solution of simultaneous linear equations; otherwise, cryptographic ambiguity may result when decipherment is attempted.

(4) A large number of sets of enciphering and deciphering [45] keys can be constructed. It is even possible to construct keys which yield reciprocal encipherment, and it is this possibility which makes practicable the construction of a machine or device to accomplish the enciphering and deciphering.

i. Attention is called here to the applications of Table 13 ("Four-square individual frequencies") of Appendix 2; this table has been reproduced here for convenience. If the cryptanalyst has at hand a fairly large volume of cipher digraphs produced by encipherment with a normal four-square, he may use Table 13 as an aid in placing the initial letters and final letters

(Table 13, Appendix 2)

[Based on a count of 5,000 digraphs]

P_1									C_1
A	B	C	D	E	244	225	375	394	197
F	G	H	I J	K	125	98	193	271	95
L	M	N	O	P	229	199	188	350	251
Q	R	S	T	U	148	162	258	427	295
V	W	X	Y	Z	42	12	34	91	97
212	317	358	308	249	A	B	C	D	E
120	108	216	256	85	F	G	H	I J	K
216	140	152	435	269	L	M	N	O	P
206	121	306	364	284	Q	R	S	T	U
38	29	21	147	43	V	W	X	Y	Z
C_2									P_2

of the cipher digraphs into the appropriate cells of the cipher component sections *on the basis of their uniliteral frequencies.* Thus, if a distribution made of the initial letters of cipher pairs in a particular example shows Q_c, I_c, and C_c to be the letters of *predominantly* high frequency (listed in descending order of frequency), and if the distribution of the final letters shows F_c, Q_c, and P_c as the letters of *predominantly* high frequency (in descending order of frequency), these letters may be tentatively placed into a skeleton four-square matrix as follows (Fig. 70), based on the locations of the highest frequencies as given in Table 13:

A	B	C	D	E			C	I	
F	G	H	I	K					
L	M	N	O	P					
Q	R	S	T	U				Q	
V	W	X	Y	Z					
		P			A	B	C	D	E
					F	G	H	I	K
			F		L	M	N	O	P
			Q		Q	R	S	T	U
					V	W	X	Y	Z

Figure 70.

[45] The deciphering keys which apply to the foregoing sample encipherment are (19, 24, 9); (23, 1, 12); (14, 0, 19). The interested student may wish to decipher the cryptogram ZDJXM HHQHY MADOI and establish for himself that it deciphers as NOTHING TO REPORT using these latter keys. In so doing he should remember that, in the final mathematical operation prior to converting the intermediate numbers into plaintext letters, he must subtract 26 (or a sufficient multiple of 26) to arrive at numbers within the range of Ø to 25.

j. In attempting to diagnose the underlying cryptosystem in any particular polygraphic cipher, the student may gain some assistance from the following recapitulation:

(1) In digraphic ciphers the majority of repetitions will be an even number of letters apart and these repetitions should for the most part begin on the first letters of pairs and end on the last letters of pairs. The majority of repetitions in trigraphic ciphers will be some multiple of three letters apart and these repetitions should for the most part begin on the first letters of trigraphs and end on the last letters of trigraphs.

(2) Digraphic ciphers may be revealed as such by the digraphic phi test, with additional support being given by the digraphic blank-expectation test; the presence of a null letter at the beginning of the cipher text might be disclosed by applying the two foregoing tests to a distribution of the digraphs which are formed when the first letter of the text is omitted.

(3) If either the uniliteral frequency distribution for the initial letters or for the final letters of the digraphs in a cryptogram exhibits monoalphabeticity, the cryptogram is probably a pseudo-digraphic cipher involving a large table of the type in Figs. 47*a*, 47*b*, or 48. If both of the foregoing uniliteral frequency distributions reflect monoalphabeticity, the cryptogram may involve the use of a table of the type in Fig. 69.

(4) If the "decipherment" of a cryptogram by means of a four-square matrix containing four normal alphabets yields two monoalphabetic substitutions—one for the initial letters and one for the final letters of the pseudo-decipherment—the cryptogram may be assumed to be an inverse four-square cipher.

(5) If an ocular inspection or statistical evaluation of the cipher text of a cryptogram reveals a large number of "transparencies", the cryptogram probably involves a two-square system.

(6) If a cryptogram contains several cipher doublets, all of which are broken up when the cipher text is divided into digraphs, the cryptogram may well involve normal Playfair encipherment.

(7) If the cipher text of a cryptogram exhibits any invariable affinity of one of the letters J, K, Q, X, or Z for vowels (or, for that matter, for another cluster of 5 or 6 letters), the cryptogram probably is in a small-matrix system employing sections consisting of more than 25 letters.

k. If a particular four-square cryptogram involves the use of a matrix in which either the plain component sections or the cipher component sections are normal alphabets, the matrix will be recovered through cryptanalysis in its *original* form, even when the components which are mixed have been derived by a transposition method or by no method at all. In Playfair cipher solution, the matrix can be recovered in its *original* form as long as the original matrix has been mixed in some systematic manner. However, in the case of two-square solution, there is no guarantee that the matrix can be recovered in its original form unless the original matrix has been keyword-mixed; if the original has been transposition-mixed, for example, the matrix which has been recovered through cryptanalysis—while being cryptographically *equivalent* to the original—will undoubtedly involve a permutation of the rows and columns of the original.

l. When four-square systems are encountered in which the matrix consists of four differently mixed sections, reconstruction of the matrix is accomplished in a manner similar to that used in the analysis of two-square ciphers. If the sections are composed of keyword-mixed sequences, the original matrix may be recovered; this is done by rearranging the rows and columns of each section on the basis of VWXYZ or such similar sequences found in keyword-mixed cases. Otherwise, the reconstructed matrix will in all probability be a permutation of both the rows and the columns of the original matrix, and there may be no way of recovering or of proving the original matrix.

m. In passing, it might be well to mention that any two-square system can be solved as a four-square system in which the matrix is composed of four mixed sections; upon the realization, from phenomena in the matrix reconstruction, that a two-square matrix is involved, the proper conversion can then easily be made.

CHAPTER X

CRYPTOSYSTEMS EMPLOYING IRREGULAR-LENGTH CIPHERTEXT UNITS

74. Preliminary observations.—*a*. The cipher alphabets of nearly all of the various cryptosystems treated thus far in this text have involved cipher units of a constant length.[1] That is, the ciphertext units have been (prior to regrouping into fives for transmission) either single characters, or pairs of characters, or three-character groupings, or, in the case of the Baconian and Baudot alphabets, 5-element ciphertext units; however, within a given cryptosystem the lengths of the ciphertext units have been *consistent*, and it is this consistency that has been of most importance to the cryptanalyst.

b. There is no reason why a cryptographer could not vary the size of the cipher units in a particular cryptosystem, as long as no cryptographic ambiguity in deciphering would result thereby. Furthermore, if the size of the cryptographic units *is* varied within a particular cryptosystem, obstacles are put in the way of cryptanalytic attack on the system—varying the length of the ciphertext groupings complicates the cryptanalyst's preliminary task of dividing the cipher text into the proper units for study. In this connection, the student should refer back to par. 63 and read again the remarks on the use of nulls which *differ in size* from the real cryptographic units. The example contained therein makes it clear that, until such nulls are identified and isolated by the cryptanalyst, he is unable to divide the cipher text properly and make appropriate frequency distributions. However, nulls may sometimes be recognized as such because they do not behave like units which represent actual plaintext elements. For example, in the three almost-identical ciphertext passages below,

```
(a)  ...181Ø5  11343  71129  3219Ø  23231  52937...
(b)  ...18151  Ø1343  71129  32192  32Ø31  52937...
(c)  ...18151  13437  1Ø129  32192  3Ø231  52937...
```

the behavior of the digit Ø is characteristic of a null, and when this is recognized and eliminated, the remaining cryptographic text may be broken up into its real units and solved quite readily.

c. Since it has been indicated above that there are weaknesses in a scheme in which all cipher elements do not behave like equivalents for plaintext elements, it would be logical then

[1] The only exceptions have been in the digraphic systems using the matrices illustrated in Figs. 57*a* and 57*b* in which a plaintext digraph may be represented by a ciphertext digraph, trigraph, or tetragraph, depending upon the identity of the plaintext digraph.

shortest possible cipher text for a given plain text—but, of course, greatly limits the number of internal arrangements which may be used. Fig. 75, below, which is split into two separate

	5	2	9	7	6
5	A	B	C	D	F
2	G	H	I	J	K
9	L	M	P	Q	S
7	U	V	W	X	Y
6	Z	.	(	)	*

FIGURE 75a.

Ø	1	8	3	4
E	T	N	R	O

FIGURE 75b.

parts—one providing the monome equivalents and the other providing the dinome equivalents—illustrates another scheme for drawing up a monome-dinome cipher alphabet. In this alphabet, the digits which are used for the initial *and final* elements of dinomes are completely distinct from the digits used as monomes.

c. Most of the foregoing matrices contain a period for punctuation, and the matrix in Fig. 72, containing the single digits Ø–9, provides a means for encrypting numbers without first spelling them out. The matrices in Figs. 71, 73, and 75 contain another character, symbolized by an asterisk, which may be used for punctuation or as a special indicator.[3] The matrix in Fig. 74 uses only nine of the single digits as coordinates, the digit 6 being omitted; this single digit might be employed as a word separator, a stop, or a null. The matrix in Fig. 76, below, illustrates a scheme by which certain high-frequency plaintext digraphs and trigraphs may be represented in the matrix, as well as the single letters and digits. The symbol

	Ø	1	8	3	4	5	2	9	7	6
-	R	E	T	A	I	N	//////	//////	//////	//////
2	B	C	D	F	G	H	J	K	L	M
9	O	P	Q	S	U	V	W	X	Y	Z
7	.	,	TH	IN	ST	ED	ION	ING	*	#
6	Ø	1	2	3	4	5	6	7	8	9

FIGURE 76.

in this latter matrix could be used as a "repetition indicator" for checking numbers, as in the ciphertext passage 69 65 68 <u>76</u> 69 65 68, meaning the number 752; the symbol * might be used as an indicator meaning "the immediately preceding plaintext letter is repeated" (thus AA patterns would be suppressed in the cipher text). In all of the foregoing matrices the order of inscription of the letters within the matrix, and the particular arrangement of the row- and column-coordinates are both subject to variation.

[3] For example, this special character may be put to use as an indicator to show that plaintext numbers begin or end, thus obviating the necessity of including digits within the cipher matrix. In this usage digits in the plain text might be *tripled* and inserted in the cipher text with the appropriate indicator before and after the plaintext digits. Thus, using the matrix in Fig. 71, the plaintext fragment "..HILL 865.." would be encrypted as the cipher sequence 75 2 77 77 66 888 666 555 66 (prior to regrouping into five-character groups).

d. By prearranged convention it is possible to employ ordinary commutative bipartite matrices (such as those already described in Chapters VII and VIII) in a manner which yields monome-dinome encipherment. For example, using the matrix illustrated in Fig. 77, the plaintext word EIGHT could be encrypted as 10 29 7 8 49 (and then, of course, regrouped into five-

	6	7	8	9	Ø
1	A	B	C	D	E
2	F	G	H	I	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

FIGURE 77.

character groups). That is, the normal bipartite enciphering conventions would be used, with the exception that the row indicator for the cipher equivalent for a particular plaintext letter would *not* be employed when this row indicator is the same as that for the immediately preceding letter of the plain text.[4] As may be noted, no cryptographic ambiguity in decipherment may arise.

e. Of course, as an extension of the foregoing ideas, there could also be *monome-dinome-trinome* systems, incorporating matrices of the types illustrated in Figs. 78–82, below. In Fig.

	Ø	1	8	3	4
–	S	T	O	N	E
5	A	B	C	D	F
2	G	H	I	K	L
9	M	P	Q	R	U
7	V	W	X	Y	Z
6Ø	Ø	1	2	3	4
61	5	6	7	8	9

FIGURE 78.

	Ø	1	8	3	4
Ø	A	B	C	D	F
1	G	H	I	K	L
8	M	P	Q	S	U
3	V	W	X	Y	Z
45	Ø	1	2	3	4
42	5	6	7	8	9

FIGURE 79*a*.

5	2	9	7	6
T	E	N	O	R

FIGURE 79*b*.

	Ø	1	8	3	4	5
–	B	R	A	N	C	H
2	D	E	F	G	I	J
9	K	L	M	O	P	Q
7	S	T	U	V	W	X
29	Y	Z	Ø	1	2	3
27	4	5	6	7	8	9

FIGURE 80.

	Ø	1	8	3	4	5	2	9	7	6
–	R	E	L	A	T	I	O	N		
7	B	C	D	F	G	H	J	K	M	
6	P	Q	S	U	V	W	X	Y	Z	.
76	Ø	1	2	3	4	5	6	7	8	9

FIGURE 81.

	Ø	1	8	3	4	5	2	9
–	E	T	N	R	O	A	I	S
7	B	C	D	F	G	H	J	K
6	L	M	P	Q	U	V		
62	W	X	Y	Z	Ø	1	2	3
69	4	5	6	7	8	9	.	,

FIGURE 82.

	Ø	1	8	3	4	5
2	A	B	C	D	E	F
9	G	H	I	J	K	L
7	M	N	O	P	Q	R
62	S	T	U	V	W	X
69	Y	Z	Ø	1	2	3
67	4	5	6	7	8	9

FIGURE 83.

83 there is a matrix which may be used for *dinome-trinome* encipherment. Encipherment with this latter matrix is commutative; for example, E_p=24 or 42, and T_p=621 or 162.

[4] A variation of this method could make use of a convention by which the *column* indicator is dropped if it is the same as that for the preceding plaintext letter.

f. Literal versions of the preceding types of alphabets with irregular-length cipher units are also possible. Several types are illustrated in Figs. 84–88, including among them matrices permitting the use of variants in encryption. Furthermore, any of the commutative variant

	B	L	A	C	K
–	A	L	I	G	N
W	B	C	D	E	F
H	H	J	K	M	O
I	P	Q	R	S	T
T	U	V	W	X	Y
E	Z	.	,	(	)

FIGURE 84.

			V	W	X	Y	Z
			Q	R	S	T	U
		–	E	T	N	R	O
L	F	A	A	B	C	D	F
M	G	B	G	H	I	J	K
N	H	C	L	M	P	Q	S
O	I	D	U	V	W	X	Y
P	K	E	Z	.	,	(	)

FIGURE 85.

		L	M	N	O	P
		F	G	H	I	K
F	A	B	D	E	F	G
G	B	I	J	K	L	N
H	C	O	P	Q	S	T
I	D	U	V	W	X	Y
K	E	Z	.	,	(	)

FIGURE 86*a*.

M A R C H_p
Q R S T U_e
V W X Y Z

FIGURE 86*b*.

		N	O	P	Q	R
		I	J	K	L	M
E	A	A	B	C	D	E
F	B	F	H	I	K	L
G	C	M	N	Q	S	T
H	D	V	W	X	Y	Z

FIGURE 87*a*.

G R O U P_p
S T U V W_e
X Y Z

FIGURE 87*b*.

		Q	R	S	T	U	V	W	X	Y	Z
		G	H	I	J	K	L	M	N	O	P
	–	E	T	N	R	O	A	I	S	D	L
D	A	B	G	K	Q	W	Z	2	5	8	.
E	B	C	H	M	U	X	Ø	3	6	9	,
F	C	F	J	P	V	Y	1	4	7	(	)

FIGURE 88.

matrices treated in par. 58*a* (i. e., Figs. 27, 28, and 31) may be used in connection with the convention described in subpar. *d*, above, to provide cipher alphabets with irregular-length ciphertext units.

76. General remarks on analysis.—*a.* The first step in the analysis of any cryptogram encrypted in a system with irregular-length cipher units involves dividing the cryptogram into the proper, vari-sized cipher units—that is, reducing the cryptogram to monoalphabetic terms. After this has been done, solution proceeds along the straightforward lines which have been described in earlier chapters of the text. Thus, in this chapter, attention will be focused on this first step of dividing the text into its proper monoalphabetic units. In order to simplify somewhat the general treatment contained in this paragraph, all remarks will be directed at monome-dinome systems; most of the principles and methods outlined herein are general enough that they may be modified and applied in the solution of other types of systems with irregular-length ciphertext units.

b. A *cryptographer*, in his process of deciphering a particular monome-dinome cryptogram, would begin by considering whether or not the first digit of the cipher text were among those digits which can start a dinome—that is, whether it were a *row coordinate* or not. If it were, he would treat it along with the next digit of the text as a dinome, and then proceed to consider whether or not the following digit were a row coordinate, etc. If the first digit of the message were not a row coordinate, he would treat it as a monome, and then proceed to consider whether or not the next digit were a row coordinate, etc. One may now see that the cryptographic process of dividing the cipher text into its proper units is based solely on a knowledge of the digits which are the row coordinates of the pertinent matrix. Thus, it may further be seen that the *cryptanalytic* attack on a monome-dinome cipher would first involve an attempt to determine the identity of the row coordinates.

c. If a given cryptogram involves a matrix in which the high-frequency plaintext elements are evenly distributed throughout the various rows, it may be expected that the particular digits occurring with the greatest frequency in a uniliteral frequency distribution made on the cipher text are those which are row coordinates of the pertinent matrix. This may be explained by the fact that the digits used as row coordinates occur in the cipher equivalents for more plaintext letters than do those digits which are used as monomes. However, one must remember that a monome-dinome matrix may involve two, three, four, or more row coordinates and, although in a particular instance it may be that the most frequent cipher digits *are* those digits which have been used as row coordinates, a study of the uniliteral frequency distribution may not make it obvious as to just *how many* coordinates are involved; it may be necessary to make several trials, one considering only the *two* most frequent cipher digits as row coordinates, one considering the *three* most frequent, etc.

d. If trials of the type just mentioned do not yield reduced, monoalphabetic text which will succumb to the principles of plaintext recovery treated in the earlier chapters of this text, it may then be assumed that the cryptogram involves a matrix in which several of the high-frequency letters are arranged together in the top row or in which one or more columns are composed solely of high-frequency letters. Such matrices are likely to produce cipher text in which some of the digits which have been used as monomes occur more frequently than some of those used as row coordinates. Thus, the easy mode of entry via the uniliteral frequency distribution may not be used, and other approaches of a less clear-cut nature must be taken.

e. In an attempt to identify at least one or two probable row coordinates, the analyst should carefully scrutinize the cryptogram itself in order to find passages exhibiting *bipartite characteristics*, such as appear in the sequence 8043818741, wherein the digits 8 and 4 "act" like digits which have been used as row coordinates, being spaced off at intervals of two. A slightly more objective approach involves first making a biliteral [5] distribution of the cipher text, and then considering as a probable row coordinate the initial digit of the particular dinome which the distribution shows to be the most frequent. Of course, this approach is most likely to be valid when the particular dinome occurs with a much greater frequency than the remaining dinomes. While still on the subject of distributions, it is pointed out that the previously-mentioned "bipartite characteristics" manifested in a cryptogram might be disclosed by making a biliteral distribution of *alternate* digits of the cipher text,[6] that is, in the sequence 123456 one would consider the dinomes 13, 24, 35, 46. In such a distribution, one may expect that the most frequent dinomes will be those comprising two digits which were both row coordinates of the pertinent enciphering matrix.

f. If the cipher text of a given monome-dinome cryptogram begins with a doubled digit, this digit is most probably one of the row coordinates of the pertinent matrix; otherwise, the doublet would have to be considered as comprising two monomes and the first word of the underlying plain text would have to begin with a doublet (a very rare contingency in the English language). Similarly, if the cipher text is seen to contain any digit repeated consecutively four

[5] The use of the term *"biliteral"* in connection with *digit* cipher text may not be in conformance with the strictest rules of semantics, but the author feels that it is unnecessary to give a new name to an already-familiar type of distribution merely because it is being applied to a different kind of text. However, some who prefer to be purists in this matter term a digraphic distribution which is made on digit text as a "dinome distribution" or "dinomic distribution", and a biliteral distribution made on digit text, a "running dinome distribution".

[6] In the vernacular such a distribution is termed an "A–A" (pronounced "ay-dit-ay") distribution.

or more times, the particular digit may be assumed to be a row coordinate; otherwise, such a sequence of repeated digits would have to represent *at least* a threefold repetition of some one plaintext letter (another rare event in English, although not as rare as that mentioned in the preceding sentence).

g. On occasion it may be found that much time has been spent in the attempt to identify the row coordinates, yet apparently with not all of the coordinates being identified. In such a case, it may be found useful to consider those digits which are *least likely* to be row coordinates, specifically, those which occur least frequently in the cryptogram. The analyst may go through the troublesome cryptogram and place a slant bar (virgule) directly after each such digit as it occurs in the message. These marks may then be taken as an indication of places in the cryptogram where one bona fide cipher unit ends and the next begins. The analyst must then study the digits which directly follow these slant bars with a view to discovering new possibilities for row coordinates—possibilities which, although previously latent, have been made patent by this latest step.

h. In the foregoing subparagraphs, *a* to *g*, the secondary step of testing for the corroboration or invalidation of any particular trial decomposition has been passed over quite briefly. Actually, this step is best described with specific examples of solution, and for this reason is treated in two subsequent paragraphs, 77 and 78, with such examples. However, a few methods which can be applied for the rejection of incorrect hypotheses will be mentioned here, because they are rather basic and simple. If the cryptanalyst finds, after having divided a monome-dinome cipher on the basis of a particular hypothesis, that a long repetition in the cryptogram is not broken up in the same way on each of its occurrences, he may well reject as incorrect the hypothesis on which the division is based. Likewise, the analyst may reject any hypothesis which requires him to make the last digit of a cryptogram a monome when this particular digit has to be considered as a row coordinate as part of the basic assumption.[7] The presence of an inordinate number of *consecutive* monomes may cause one to suspect that a particular decomposition is incorrect; however, probably only continued exposure to traffic of a certain type or involving one kind of enciphering matrix would provide one with a sound basis for knowing *just how many are too many.*

i. There is one practical, straightforward measure for determining the relative goodness of an assumed decomposition which deserves particular mention. It involves considering the ratio of the *number of monomes* produced in a particular decomposition to the *number of remaining cipher units.* In the case of monome-dinome ciphers, for example, in which an assumption of only two row coordinates is made, there can be no more than eight different plaintext letters represented by monomes and the total frequency of those monomes can not exceed the frequency expected of the eight most frequent letters in the language.[8] Since in English the eight most frequent letters occur with a total relative frequency of 66%, any trial decomposition giving rise to a ratio of monomes to dinomes which is considerably more than 66 to 34 (=1.9) may be considered incorrect. Likewise, since an assumption of three row coordinates limits to seven the number of different plaintext letters which may have monome equivalents, and since the seven most frequent letters in English occur with a total relative frequency of 60%, any such assumption giving rise to a ratio of monomes to dinomes which is considerably more than 60 to 40 (=1.5) may be considered invalid. The author does, however, hasten to point out that

[7] However, the possibility of a final null or nulls must not be ignored; the presence of nulls at the end of the cipher text would invalidate this reasoning.

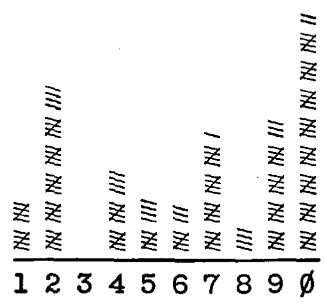
[8] The only exception to this statement would be a case wherein a word separator is included as part of the cryptosystem, and that this separator is represented by a monome. This usage, however, seems rather unlikely.

a ratio which is *smaller* in any instance than the pertinent ratio, above, *does not disprove* the particular trial decomposition since the plaintext letters represented by monomes may not necessarily be the letters of highest frequency. The examples in the next two paragraphs will serve to clarify the foregoing considerations.

77. Analysis of simple examples.—*a.* The following cryptogram, suspected to be a monome-dinome cipher, is available for study:

	5	10	15	20	25	30
A	2 4 0 9 0	1 5 7 0 9	0 8 1 2 1	0 2 0 9 2	9 2 4 0 5	5 6 0 0 1
B	2 7 0 7 2	9 0 4 8 2	4 7 6 0 7	0 9 0 2 2	1 0 2 0 9	2 9 7 2 4
C	0 7 2 9 2	9 1 2 5 7	5 2 9 6 1	0 9 0 4 2	7 2 0 0 2	0 7 2 4 7
D	5 0 5 7 0	9 6 0 8 1	7 2 4 0 9	2 9 0 4 0	4 0 9 7 1	2 4 0 9 7
E	2 9 1 2 8	7 6 0 9 0	4 0 7 5 0	6 5 2 9 7	0 9 0 6 7	2 0 9 0 2
F	0 9 0 4 0	7 4 0 7 6				

Cursory examination of the cipher text reveals nothing more significant than the fact that the digit 3 is absent; however, the significance of this escapes us for the moment. A uniliteral frequency distribution of the text is then made, as is illustrated below:



b. The uniliteral frequency distribution shows four marked peaks (2, 7, 9, and Ø) and one pronounced trough (8). A biliteral frequency distribution is made, as shown below, to assist in

	1	2	3	4	5	6	7	8	9	Ø
1	–	5	–	–	1	–	1	–	–	3
2	2	1	–	7	1	–	2	1	9	6
3	–	–	–	–	–	–	–	–	–	–
4	–	1	–	–	–	–	2	1	–	10
5	–	2	–	–	1	1	3	–	–	2
6	1	–	–	–	1	–	1	–	–	4
7	1	8	–	1	3	3	–	–	–	5
8	2	1	–	–	–	–	1	–	–	–
9	2	5	–	–	–	2	4	–	–	10
Ø	2	5	–	6	2	2	7	2	14	2

further evaluation of the properties of the cipher text. It is noted that the 2 and Ø rows, representing the two highest-frequency digits in the cipher text, have the most liberal combinations with the remaining digits; this would indicate that 2 and Ø are likely row coordinates of the cipher matrix. Since the 7 and 9 rows show less affinity of these digits for other digits, 7 and 9 are less likely to represent row coordinates of the matrix; consequently the assumption is made that the matrix involved only two numbered row coordinates, 2 and Ø.

c. The cryptogram is now divided accordingly, and the assumption of 2 and Ø as row coordinates is borne out by the bipartite character of the following passages in the cipher text:

(1) .../21/02/09/29/24/05/... (at A14)
(2) .../07/09/02/21/02/09/29/... (at B14)
(3) .../09/04/27/20/02/07/24/... (at C16)
(4) .../24/09/29/04/04/09/... (at D12)

A frequency distribution of the decomposed text is made, as illustrated below:

	1	2	3	4	5	6	7	8	9	Ø
–	7	////	–	1	6	6	12	1	1	////
2	2	–	–	7	1	–	2	1	9	2
Ø	1	5	–	6	2	2	7	2	13	1

The percentage of monomes, 35%, does not exceed the threshold for the sum of the frequencies of the eight highest-frequency plaintext letters; furthermore, since the eight monomes have a much lower frequency than the sum of the eight highest-frequency letters in English, this is an indication that some of the monomes represent plaintext letters of lower frequency.

d. The decomposed text may now be solved, and the message is found to begin with the words "SABOTAGE PLANS..." The original matrix is reconstructed, and is discovered to be based upon the key word VERMOUTH, as follows:

	9	1	6	4	5	8	7	3	Ø	2
–	V	E	R	M	O	U	T	H	////	////
Ø	A	B	C	D	F	G	I	J	K	L
2	N	P	Q	S	W	X	Y	Z	.	,

The reason for the absence of the digit 3 in the cipher text may now be seen: the digit 3 forms a part of only the letters H, J, and Z, and these letters did not occur in the plaintext message.

e. Solution of certain other cases of mixed-length systems progresses as easily as did the solution of the foregoing example.

(1) For instance, in the case of a cryptogram produced by a matrix where the digits used for both the initial and final digits of dinomes are completely distinct from the monome digits (e. g., Fig. 75), it may be seen that "eliminating" from the cipher text those particular digits which were used as monomes in the original enciphering alphabet will leave the remainder of the cryptogram broken up into units all of which contain an even number of digits. (This would not be true in the case of other types of matrices, such as Figs. 71–74, since eliminating the digits which were used as monomes in the pertinent alphabet would remove not only actual cipher monomes but also the final digits of many cipher dinomes.) In view of this fact, if one is confronted with a cryptogram which he assumes to have been produced by a matrix such as that

in Fig. 75, he may use a mechanical method by means of which he will quickly be able to determine which digits are row coordinates and which are not; or, if his basic assumption concerning the type of matrix involved is incorrect, the error will quickly become known to him. He need only make successive trials each of which involves considering a different one of the 10 digits as being one of those which is a monome in the pertinent alphabet; "eliminating" the particular digit from the cryptogram in each trial will inevitably lead to other digits which must also be eliminated throughout the cryptogram in order to maintain the stipulation that all the cipher units which remain must contain an even number of digits. For example, if one assumes that "Ø" is a digit which was a monome, then he must further assume from a sequence of cipher digits such as Ø5Ø35Ø that "5" is also a digit which was a monome; and then likewise "3". Any particular one of the ten trials which is based on an incorrect initial assumption may be expected to end up with all ten digits being considered as digits which were monomes.

(2) In the case of a monome-dinome system in which the row coordinates of the enciphering matrix are distinct from the column coordinates (as in Figs. 73 and 74), solution is expedited by capitalizing on the fact that the digits within the family comprising the row coordinates do not (and cannot) contact themselves or any other digits within the family; using Fig. 73 as an example, it is obvious that the digits 7, 8, and 9 can never be followed by a 7, 8, or 9. (This causal avoidance among certain digits exhibits itself in either a digraphic or a biliteral distribution of the cipher text.) A cryptogram enciphered by such a system may be expected to contain far fewer cipher doublets than would a cryptogram produced by a matrix without the foregoing limitation, and the doublets which do occur will themselves involve but a limited number of the 10 different digits. When solving such a cryptogram, the cryptanalyst need only consider as possible candidates for row coordinates those particular digits which do not appear in cipher doublets. Furthermore, he may with certainty go through the cryptogram placing a slant bar (to indicate the end of a valid cipher unit) after every occurrence of any digit which has appeared in a cipher doublet.

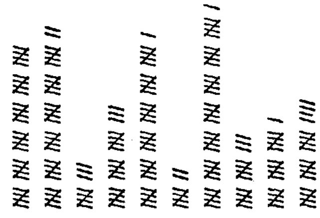
(3) The system described in subpar. 75*d* and the accompanying Fig. 77 (employing a commutative bipartite matrix) is another system which yields cipher text in which a certain family of digits—namely, the row coordinate digits—cannot contact any other digits in the same family. If the cryptanalyst is confronted by a cryptogram in this system, he knows that the first digit of the cryptogram must be a row coordinate. Then he has only to go through the cryptogram noting the digits which follow this row coordinate digit wherever it occurs in the cryptogram and, in this way, he may be able to identify all the column coordinate digits. Of course, by the process of elimination, he will then know which digits are row coordinates besides the initial digit of the cryptogram, and it will then be possible for him to divide the text into its proper irregular-length ciphertext units.

78. Analysis of more complicated examples.—*a.* In some cases, the rather simple methods of analysis applied in the preceding paragraph will not bear fruit, either because of the complexity inherent in the *number* of plaintext elements in the cipher matrix, or because of certain unpredictable aberrations caused by the particular *arrangement* of plaintext elements in the matrix. For instance, if a specific matrix contained only the highest-frequency letters in the top row, and if the matrix contained a fairly large number of plaintext elements (and therefore embodied 3 or 4, or more, row coordinates), and if the elements in the dinome rows were balanced from the frequency standpoint, so that the rows would be used with approximately equal frequency, and furthermore if certain of the columns were composed of heavier elements than others (thus producing peaks that might incorrectly be identified as row coordinates)—all these conditions would yield a cryptosystem that might pose considerable difficulties in the way of straightforward analysis. A case will now be studied that will illustrate typical techniques that would be necessary in more difficult circumstances.

b. The following cryptogram has been intercepted on an enemy net known to be passing monome-dinome traffic:

```
62719  44081  21204  71270  55042  12627

09637  06212  24712  91724  21058  12727

07055  58719  55721  04109  52847  71297

23571  82123  94578  77571  80581  97654

74572  05191  77194  52958  70012  12251

69051  15724  71389  47316  79035  47359

54742  78271  72327  05504  58255  55918
```

The uniliteral frequency distribution for the cryptogram is shown below:



c. From the appearance of the uniliteral frequency distribution, it is to be expected that from among the four peaks (1, 2, 5, and 7) some row coordinates must be represented, and since there is not much variance in frequency among these peaks, *perhaps all four* represent row coordinates. In an attempt to obtain as much information as possible from a study of the frequency characteristics of the cipher text, a biliteral distribution is made and is shown below.

	1	2	3	4	5	6	7	8	9	0
1	1	11	1	–	1	2	3	3	5	3
2	7	2	3	3	2	1	9	1	3	2
3	1	1	–	–	3	–	1	1	1	–
4	1	3	–	1	9	–	6	5	2	1
5	3	2	–	3	9	–	6	5	2	2
6	–	3	1	–	1	–	1	–	1	–
7	10	7	2	2	1	1	3	2	1	7
8	3	3	–	1	–	–	3	–	1	1
9	3	–	–	4	4	1	2	–	–	2
0	1	–	1	4	7	1	1	1	2	1

Examination of this latter distribution adds support to the impressions gained from the uniliteral frequency distribution, namely, that the row coordinates for the cipher matrix are very likely to be found among the digits 1, 2, 5, and 7. Furthermore the digit 7, because of its high frequency and because of satisfactory combinative qualities in the biliteral distribution, is selected as a definite row coordinate. This will reduce the number of trials that must subsequently be considered.

d. If *all* of the row coordinates of the cipher matrix are found among the various combinations of 7 with 1, 2, and 5, then it is clear that:

(1) if there are but two coordinates of the matrix, these must be either 7 and 1, 7 and 2, or 7 and 5, (three cases);

(2) if there are three coordinates of the matrix, these must be either 7–1–2, 7–1–5, or 7–2–5 (three cases); or

(3) if the matrix has four numbered coordinates, this must entail the combination of 7–1–2–5 (only one case).

e. On the basis of each of the foregoing seven hypotheses, the cipher text is divided and the resulting frequency distributions are shown below:

	1	2	3	4	5	6	7	8	9	Ø
–	////	17	6	16	31	4	////	10	13	9
1	1	8	–	–	–	2	2	1	2	3
7	9	7	2	2	–	1	2	2	1	7

Case I

	1	2	3	4	5	6	7	8	9	Ø
–	18	////	5	16	30	5	////	11	12	15
2	6	2	1	–	1	1	8	1	3	1
7	6	6	2	2	–	1	3	1	1	3

Case II

	1	2	3	4	5	6	7	8	9	Ø
–	21	25	6	13	////	6	////	7	14	12
5	3	2	–	3	6	–	5	5	1	–
7	6	5	2	2	–	1	3	1	1	7

Case III

	1	2	3	4	5	6	7	8	9	Ø
–	////////	////////	4	15	29	3	////	10	10	13
1	1	6	–	–	–	2	3	1	3	2
2	4	1	2	1	2	1	6	1	2	–
7	7	5	2	2	–	1	2	1	1	4

Case IV

	1	2	3	4	5	6	7	8	9	Ø
–	////	16	6	13	////	5	////	4	13	9
1	–	9	–	–	1	1	2	3	1	3
5	3	2	–	3	6	–	4	5	1	–
7	7	5	2	2	–	1	2	1	1	7

Case V

	1	2	3	4	5	6	7	8	9	Ø
–	17	////	5	12	////	5	////	8	11	14
2	6	2	1	1	1	1	8	–	2	2
5	3	2	–	3	5	–	5	5	2	–
7	4	4	2	2	–	1	3	–	1	3

Case VI

	1	2	3	4	5	6	7	8	9	Ø
–	////////	////////	4	12	////	3	////	5	10	12
1	–	7	–	–	1	2	3	3	2	2
2	4	1	2	1	2	1	6	–	1	1
5	2	2	–	3	5	–	4	5	2	–
7	5	4	2	2	–	1	2	–	1	4

Case VII

f. In order to be able to evaluate the relative merits of the seven hypotheses and choose the case which is most likely to be correct, it is possible to resort to a method wherein *group frequencies* of the high-frequency elements from each of the decompositions are studied. In the following table drawn up for this purpose, the column of figures under "x" denotes the cumulative six highest-frequency ciphertext units; under "N", we have the actual frequencies of the first, the first two, the first three . . . , the first six highest-frequency ciphertext units for each hypothesis (compare with the distributions in subpar. *e*); in the adjoining column to the right of each "N" column, the various cumulative frequency values are expressed as percentages of the total number of ciphertext units which remain after the particular trial decomposition. The column labelled "P" gives the cumulative *theoretical* frequencies of the six most frequent letters in English plain text (ETNROA), in cumulative relative order of frequency (i. e., the frequencies of E_p; of E_p and T_p; of E_p, T_p, and N_p; and so on). The following elaboration will serve to clarify the foregoing details.

	I		II		III		IV		V		VI		VII		P
x	N	$\frac{N}{158}$	N	$\frac{N}{161}$	N	$\frac{N}{157}$	N	$\frac{N}{147}$	N	$\frac{N}{138}$	N	$\frac{N}{139}$	N	$\frac{N}{129}$	
1	31	19.6	30	18.6	25	15.9	29	19.7	16	11.6	17	12.2	12	9.3	13.0
2	48	30.4	48	29.8	46	29.3	44	29.9	29	21.0	31	22.3	24	18.6	22.2
3	64	45.0	64	39.8	60	38.2	57	38.8	42	30.4	43	30.9	34	26.6	30.2
4	77	48.7	79	49.1	73	46.5	67	45.6	51	37.0	54	38.8	41	31.8	37.8
5	87	55.1	91	56.5	85	54.1	77	52.4	60	43.5	62	44.6	47	36.4	45.3
6	96	60.8	102	63.4	92	58.6	84	57.1	67	48.6	70	50.4	52	40.3	52.7

g. It is noted that in Case I, the most frequent ciphertext unit has a percentile frequency of 19.6%; the highest two units, a percentile frequency of 30.4%; the highest three, a percentile frequency of 45.0%. When these percentages are compared with the percentile frequency of the highest-frequency letter in English plain text (13.0%), of the highest two letters (22.2%), and of the highest three letters (30.2%), it is clear that Case I does not conform to the characteristics expected of a simple monoalphabetic substitution; therefore Case I is not the correct division of the cipher text. Similarly, Cases II, III, and IV can also be rejected because the cumulative values are *much higher* than the corresponding expectations for plain text. Case VII, on the other hand, demonstrates values *much lower* than the corresponding expectations for plain text; therefore this case too is rejected. This leaves only Cases V and VI, both of which show a close correspondence with plaintext expectations.

h. If there were nothing else in the manifestations of the decomposed cipher text in Case V and Case VI, these two cases would have to be tried in turn, making some tentative plaintext assumptions; of course, only the correct case would consistently yield plain text. However, there is an additional bit of reasoning which may be applied here as a means of deciding which of these two remaining cases is more likely to be correct and ought to be worked on first—namely, it may be reasoned that cipher text which has been decomposed according to an incorrect hypothesis will be likely to contain a larger ratio of monomes to dinomes than would the same text if it had been decomposed according to the correct hypothesis.[9] Case V has a monome-dinome ratio of .916 whereas Case VI has a corresponding ratio of 1.043; thus Case V is indicated as the case which is more likely to be correct.

i. The cipher text is now divided according to the hypothesis of row coordinates of 1, 5, and 7, and the plain text is quickly recovered, facilitated by the pattern of the first word, RECONNAISSANCE. The cipher matrix is reconstructed as follows:

	1	7	5	Ø	2	8	4	9	6	3
–	/	/	/	A	E	I	N	O	R	T
1	B	F	J	M	S	W	Z	1	4	7
7	C	G	K	P	U	X	.	2	5	8
5	D	H	L	Q	V	Y	Ø	3	6	9

The reason for the high frequency of the cipher digit 2 is now seen: the combined frequencies of E_p, S_p, and U_p contribute to an inordinate peak for that column coordinate.

j. In retrospect, several important points may be noted in the solution of this particular cryptogram. First of all, the four consecutive 5's in the last two groups of the cryptogram make it a very strong probability that 5 is a row coordinate; otherwise the four 5's would mean a threefold (or even fourfold) repetition of a monome letter, a comparatively rare contingency. Secondly, the digit 1 could have been selected as a row coordinate with considerable certainty, based on the fact that, since the dinome 12 was the highest-frequency element in the biliteral distribution, it may be assumed that at least a number of 12's were causal and therefore 1 must be a row coordinate. In other words, the correct set of coordinates might have been established at the very beginning of the analysis, but for pedagogical reasons it was felt necessary to proceed along the general lines of the solution as given. It is to be noted that, since at the start of solution we did not know exactly how many numbered row coordinates there were in this particular case, we could not apply the ratio of monomes to dinomes at once as the deciding criterion.

k. If mixed-length systems were encountered in actual practice, after the type of matrix became known through solution of several days' traffic, solution of subsequent days' messages would be facilitated because by this time the analyst would be familiar with the general type of matrix used. This knowledge would be of great assistance in making assumptions as to the nature of subsequent matrices. In some cases, the internal arrangement of the matrix might remain fixed, with only the coordinates being changed periodically; in other cases, the internal arrangement and the coordinates of the matrix might change, with only the *size* of the matrix remaining fixed. If it were known, for instance, that the enemy were using a monome-dinome system with

[9] This intuitive reasoning has been borne out empirically with reasonable success. 30 monome-dinome ciphers of an average length of 100 digits were decomposed in all possible ways based on the proper hypothesis of two, three, or four row coordinates—whichever correctly applied. In the case of approximately one-half of these ciphers, the correct decomposition yielded a monome-to-dinome ratio which was lower than the monome-to-dinome ratio yielded by any of the other, *in*correct decompositions. Admittedly, this 50–50 chance is of small note in connection with subpar. *h*, above, where there are only two cases from which to choose anyway, with the concomitant 50–50 chance of either choice being the right one. However, when there are *more* than two from which the analyst must make his choice, the foregoing reasoning should be quite helpful.

only *two* numbered row coordinates, then there would only be $\frac{10 \times 9}{1 \times 2}$ or 45 exhaustive trials (if these had to be made) which would be necessary to guarantee reaching the correct decomposition of the cipher text; if there were *three* numbered coordinates, then there would be a maximum of $\frac{10 \times 9 \times 8}{1 \times 2 \times 3}$ or 120 trials necessary to insure reaching the proper scheme for the decomposition of the cipher text.[10] Such trials, although laborious (and ordinarily unnecessary) when made by manual methods, would be by no means prohibitive if there were available machine processes for assistance.[11] Exhaustive trials would rarely be necessary, except in very difficult cases; in the majority of instances, straightforward methods of cryptanalysis would reduce the large number of theoretical trials to but a few, from which the correct selection could be made.

l. If the exact composition of the internal arrangement of the matrix were known, this knowledge would be useful in determining how the letters of assumed cribs would be enciphered as monomes or dinomes. In any case, if a word of pronounced idiomorphic pattern is assumed, no matter *how* the letters of the word are encrypted as monomes or dinomes, the idiomorphism must be patent in the cipher text; for example, the word ARTILLERY in a monome-dinome system *must* have a consecutively repeated monome or dinome representing L_p, closely flanked on both sides by some particular monome or dinome representing R_p. If unenciphered numbers were to appear in the encrypted text, bracketed by an indicator to signal that numbers begin and end, the recognition of these plaintext numbers would enable the analyst to identify the indicator, and thus, lead to the establishment of one row coordinate.

---

[10] The number of combinations of N different things taken *r* at a time is given by the form ${}_NC_r = \frac{N!}{r!(N-r)!}$; thus for the assumption of 3 numbered rows in a monome-dinome matrix, ${}_{10}C_3 = \frac{10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{3\cdot2\cdot1\ (7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1)} = \frac{10\cdot9\cdot8}{3\cdot2\cdot1} = 120$. The notation N! is read as "N factorial."

[11] If exhaustive trials *were* to be made by machine, an approach via the monome-to-dinome ratio would probably be as successful as any other approach and not as involved as some. As has been briefly mentioned in a preceding footnote, such an exhaustive trial procedure has been applied to 30 cryptograms of an average length of approximately 100 digits, and using the lowest monome-to-dinome ratio as the final selection criterion produced results which were quite satisfactory, *viz.*, in the case of *13* of the cryptograms tested, the procedure yielded the correct row coordinates for the underlying matrices.

Furthermore, when a study was made of those instances wherein this testing procedure failed, it was found that all but four of the unsuccessful instances involved an enciphering matrix which contained the high-frequency letters of English in the top row, that is, which provided monome equivalents for these high-frequency letters. Stated conversely, this testing procedure was quite successful when applied to cryptograms involving matrices in which the high-frequency plaintext elements were evenly distributed throughout the various rows.

The evaluation of the trial testing was carried one step further because, in the case of a cryptogram involving a matrix throughout which the high-frequency elements are *evenly* distributed, one assumes that the correct row coordinates generally can be picked out merely from a study of the uniliteral frequency distribution made on the cryptogram (see subpar. 76c). With this in mind, a further look at the results of the testing brought out that this machine process disclosed the correct row coordinates in five instances where the uniliteral frequency distribution would have led the analyst astray, and "overlooked" only two instances in which the uniliteral frequency distribution would have revealed the correct coordinates.

m. It must be pointed out that mixed-length systems, even more so than other types of systems treated in this text, often present unusual problems for the cryptanalyst. Each case is a distinctly special case,[12] but continued practice in the solution of these types of systems should, as in other situations, cultivate skill and develop abilities in this field.

n. The student may have noted that no mention has been made concerning the possible use of the ϕ test as a means for determining whether or not a particular trial decomposition represents the proper reduction of a cryptogram to monoalphabetic terms. The ϕ test has been ignored throughout this chapter because, when dealing with cipher alphabets which include plaintext elements other than single letters (e. g., such elements as syllables, numbers, indicators, etc.), the value of ϕ_p can only be loosely approximated; furthermore, computation of the value of ϕ_r in a mixed-length cipher is also a rather tenuous matter. For this reason, it has been considered best to describe only methods of solution which do not depend at all on the use of the ϕ test, and thus keep from establishing in the mind of the student any doubt as to the usefulness of this test when applied in other instances, such as those described in earlier chapters of this text.

79. Further remarks on cryptosystems employing irregular-length ciphertext units.—*a.* The subject of the diagnosis or identification of mixed-length cipher systems has not been discussed. This problem can sometimes be extremely difficult in complex cases; however, the general statement can be made that one takes advantage of any phenomena of repetitions that are present in a cryptogram to arrive at the conclusion that a mixed-length system has been encountered. If the repetitions present are separated by numbers of letters without a constant factor, or if the interval between repetitions is a prime number, and if the possibility of a null or nulls (of a different size than the real cryptographic units) has been considered and ruled out, then in all probability the cryptogram involves some sort of mixed-length cipher units. As to exactly *which* kind of mixed-length system is involved, this question can be answered only by detailed analysis, sometimes to the point of actual plaintext recoveries in order to be certain about one's conclusions.[13]

b. It is not imperative that a mixed-length cipher system be produced through the medium of a matrix with row and column coordinates. For example, in one cryptogram that was submitted for solution, the cipher text began as follows:

```
QKT2Q  3KB3K  QKTQK  T3QKT  2KB3Q  KTQR2
KKT2K  KT2KB  3QKTQ  BQRK3  KQ2QK  T2QR2....
```

The entire cryptogram, containing 490 characters, consisted only of the seven symbols B, K, Q, R, T, 2, and 3. When this cryptogram was solved, the following alphabet was recovered:

A = K3	G = KR2	N = Q2	U = Q
B = KR3	H = Q3	O = QR2	V = QB2
C = QB3	IJ = QKT3	P = QR	W = K
D = KB2	KQ = K2	R = QKT	X = KB
E = KB3	L = KKT3	S = QB	Y = KKT
F = KKT2	M = QR3	T = QKT2	Z = KR

[12] And, as one cryptowag has pointed out, some cases are more special than others.

[13] Cf. the discussion of diagnosis in subpar. 69*f*.

To the reader who is a devotee of the royal game, it will be apparent that the foregoing alphabet is based upon chess notation.[14] If however the digits 1–7 had been used in lieu of the symbols above, the cryptogram could still have been correctly divided into its component ciphertext groupings of 1, 2, 3, and 4 digits, based upon an interpretation of the characteristics present in the cipher text, and of the phenomena in a triliteral distribution showing one prefix and one suffix.[15]

c. The concept of irregular-length cryptographic units can be applied to many varieties of systems, both code and cipher. For example, in Fig. 89, below, there is illustrated a four-square matrix in which plaintext digraphs are represented by ciphertext dinomes, trinomes, or tetranomes. The positioning of the monomes in the ciphertext portions of the matrix was governed

A	B	C	D	E	10	12	5	3	13
F	G	H	I	K	14	15	16	8	17
L	M	N	O	P	18	19	40	6	Ø
Q	R	S	T	U	42	43	9	2	7
V	W	X	Y	Z	45	46	47	48	49
10	6	5	7	12	A	B	C	D	E
13	14	15	16	17	F	G	H	I	K
18	19	40	2	Ø	L	M	N	O	P
42	43	8	3	9	Q	R	S	T	U
45	46	47	48	49	V	W	X	Y	Z

FIGURE 89.

by the frequencies of individual components of four-square cipher digraphs,[16] thus permitting optimum compression of the cipher text, i. e., allowing the most liberal use of ciphertext dinomes and trinomes rather than the maximum cipher length of tetranomes; for example, the word REGIMENTAL would be encrypted RE GI ME NT AL.
76 814 06 68 1018

d. The matrix for another mixed-length cipher system, employing dinomes and trinomes for the encryption of plaintext digraphs, is shown in Figs. 90*a* and *b*. Using this matrix, the word DIVISION is encrypted as 07 883 32 746. It is noted that consonant-vowel digraphs involving eight high-frequency consonants with five vowels are represented by dinomes, and all other plaintext digraphs are represented by trinomes. In those rare cases where, as in the example MU ZZ LE, an "impossible" digraph appears in the plain text, the insertion of the letter K_p in the plain text at that point in question, similar to the normal Playfair doublet convention, enables the encryption of the word, as MU ZK ZL E. A better variation of the foregoing system

[14] The chess-playing reader might be interested in recovering the key word for this alphabet.

[15] The interested student could make up a cryptogram using seven characters in this fashion, so he could see for himself the methods of attack on such a system.

[16] See Appendix 2, Table 13, "Four-square individual frequencies."

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425
B	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451
C		452	453	454		455	456	457		458	459	460	461	462		463	464	465	466	467		468	469	470	471	472
D		473	474	475		476	477	478		479	480	481	482	483		484	485	486	487	488		489	490	491	492	493
E	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519
F	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545
G	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571
H		572	573	574		575	576	577		578	579	580	581	582		583	584	585	586	587		588	589	590	591	592
I	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618
J	619	620	621	622	623	624	625	626	627		628	629	630	631	632	633	634	635	636	637	638					
K	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664
L		665	666	667		668	669	670		671	672	673	674	675		676	677	678	679	680		681	682	683	684	685
M	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711
N		712	713	714		715	716	717		718	719	720	721	722		723	724	725	726	727		728	729	730	731	732
O	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758
P	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784
Q																					785					
R		786	787	788		789	790	791		792	793	794	795	796		797	798	799	800	801		802	803	804	805	806
S		807	808	809		810	811	812		813	814	815	816	817		818	819	820	821	822		823	824	825	826	827
T		828	829	830		831	832	833		834	835	836	837	838		839	840	841	842	843		844	845	846	847	848
U	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874
V	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900
W	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926
X	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952
Y	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978
Z	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999					

FIGURE 90a.

	A	E	I	O	U
C	00	01	02	03	04
D	05	06	07	08	09
H	10	11	12	13	14
L	15	16	17	18	19
N	20	21	22	23	24
R	25	26	27	28	29
S	30	31	32	33	34
T	35	36	37	38	39

FIGURE 90b.

might incorporate a dinome matrix for the 40 highest-frequency digraphs (comprising 42% of English plain text) such as that illustrated in Fig. 91, and a trinome matrix modified in suitable

	Ø	1	2	3	4	5	6	7	8	9
Ø	AN	AR	AS	AT	CO	DE	EA	ED	EE	EN
1	ER	ES	ET	FI	FO	HI	IN	IO	IS	LE
2	MA	ND	NE	NT	ON	OR	OU	RA	RE	RT
3	SE	SI	ST	TE	TH	TI	TO	TW	TY	VE

FIGURE 91.

fashion for the remaining digraphs (with perhaps the matrix coordinates arranged in a mixed sequence). Such a scheme would yield a greater condensing property for the cipher text, but would not be as easy to use as the system described above since the easy mnemonic feature of the matrix in Fig. 90*b* would be lost.

e. Another idea for a cryptosystem having irregular-length ciphertext groupings employs the diagram in Figs. 92*a* and *b*. This scheme incorporates Playfair digraphic encipherment (with biliteral cipher equivalents) and monographic encipherment (with uniliteral cipher equivalents). In order to disturb the regularity of usual digraphic encipherment (produced by the Playfair-type matrix in Fig. 92*a*), certain selected medium-frequency consonants are enciphered

A	C	D	E	F
H	I	K	L	N
O	P	Q	R	S
T	U	X	Y	Z

FIGURE 92*a*.

B	G	M	V	W
W	V	M	G	B

FIGURE 92*b*.

monographically and uniliterally by the reciprocal alphabet shown in Fig. 92*b*. Using Fig. 92, as an example, the phrase "BRIGADE OF ENEMY INFANTRY MOVING . . ." would be broken up and enciphered as follows:

```
B RI G AD EO FE NE M YI NF AN TR YX M OX V IN G
W PL V CE AR AF LF M UL SN FH YO ZY M QT G KH V
```

The cipher text, regrouped into fives, WPLVC EARAF LFMUL SNFHY OZYMQ TGKHV, reveals no indication of the uniliteral-biliteral encipherment involved. Since the letters BGMVW represent 8.2% of normal plain text, there is approximately 8% interruption of the regularity of normal digraphic text. Furthermore, since it is expected that about half the time these letters will occur as singles in the plain text, and about half the time an interruptor letter (such as X_p in the example above) will have to be used, this scheme is accomplished by adding only about 4% to the length of the original plain text. Other variations of the basic idea are found in Figs. 93 and 94; in Fig. 93, the Playfair matrix is a 6 x 4 rectangle omitting S and Y, and these two letters

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
T	U	V	W	X	Z

S	Y
Y	S

FIGURE 93.

A	B	C	D	E
F	G	H	I	K
L	M	N	O	P
Q	R	S	T	U
V	W	X	Y	Z

E_p
J_c

FIGURE 94.

form a reciprocal monographic encipherment convention; in Fig. 94, the Playfair matrix is the normal 5 x 5, but with the convention that, unless E_p is the second member of a digraph in the process of encryption, E_p is represented monographically by J_c. In the foregoing two figures, the SY of Fig. 93 could be replaced of course by any other two letters whose combined frequency is in the neighborhood of 6–10%, and the monographic E_p of Fig. 94 could be replaced by any other high- or medium-frequency letter. Instead of Playfair matrices, the digraphic portions of the enciphering schemes of this subparagraph could be accomplished by the use of any other small-matrix digraphic methods.

f. The Morse code, consisting as it does of irregular-length units composed of dots and dashes, lends itself to interesting cryptographic treatment. For example, the dots and dashes (and, of necessity, the *spaces* between Morse characters) might be encrypted by means of the table illustrated in Fig. 95, wherein each of the three elements has approximately the same number of variants. A better idea, however, is to employ variants in the proportions of dots

dot:	A B C D E F G H I
dash:	J K L M N O P Q R
space:	S T U V W X Y Z

FIGURE 95.

dot:	H Y D R A U L I C B E
dash:	F G J K M N O P
space:	Q S T V W X Z

FIGURE 96.

42.4%), dashes (29.1%), and spaces (28.4%) of the letters comprising normal English plain text; such a scheme for variants is shown in Fig. 96. Thus, using the example of Fig. 96, the word ENEMY (which in Morse code is . —. . —— —.——) might be encrypted as RS MDW CQ NFV PIKGZ, which would then be regrouped in fives for transmission. Other ideas for the encryption in digit form of Morse code systems might incorporate alphabets such as those illustrated in Figs. 97 and 98 below:[17]

dot:	1 2 3 4
dash:	5 6 7
space:	8 9 0

FIGURE 97.

dot:	1 3 5 7 9
dash:	2 4 6 8
space:	0

FIGURE 98.

g. Space does not permit detailed examples of analysis of some of the foregoing systems. Admittedly, some of them would pose considerable difficulty in the way of solution; however, if these systems were used in actual practice, then *operational* cryptanalytic methods and entries would make possible successful solution.

[17] Further ideas of cryptosystems based on the Morse code will be treated in *Military Cryptanalytics, Part IV*.

CHAPTER XI

MISCELLANEOUS MONOALPHABETIC SYSTEMS; CONCLUDING REMARKS

80. Cryptosystems employing syllabary squares and code charts.—*a*. The various cryptosystems treated in the preceding chapters of this text have in the main fallen into *either* the multiliteral category *or* the polygraphic category. This and the next few subparagraphs will treat of systems which represent a merger of these two categories—namely, biliteral systems which have as plaintext elements not only single letters and digits, but also certain polygraphs selected for the condensation in cipher text that their usage may permit. In addition, treatment will be made of biliteral systems which involve, as plaintext units, a selection of frequent words (that is, which occur frequently in the type of traffic for which the particular cryptosystem is intended) and perhaps some common phrases, such as "reference your message number", "request acknowledgement", "nothing to report", etc. Systems which embrace digraphs, trigraphs and other polygraphs as plaintext elements in addition to single letters and digits are called *syllabary systems* because the additional inclusion of these polygraphs permits the encrytion of plain text in a syllabic or quasi-syllabic fashion; most systems of this type involve bipartite matrices in the cryptographic scheme, and these matrices are called *syllabary squares.* When the matrix in this general type of system *also* incorporates words among the plaintext elements, the matrix is termed a *code chart.*

b. The category of systems embodying syllabary squares and code charts as the cryptographic vehicle actually constitutes a transition between *cipher* and *code* systems,[1] since a syllabary square or a code chart may be regarded equally properly as either a special type of cipher or a primitive code. However, because syllabary systems follow very closely on the ideas of bipartite matrices, these systems are included in this particular text instead of being reserved for treatment in a subsequent text.

c. A sample syllabary square is illustrated in Fig. 99, below:

	1	2	3	4	5	6	7	8	9	Ø
1	A	1	AL	AN	AND	AR	ARE	AS	AT	ATE
2	ATI	B	2	BE	C	3	CA	CE	CO	COM
3	D	4	DA	DE	E	5	EA	ED	EN	ENT
4	ER	ERE	ERS	ES	EST	F	6	G	7	H
5	8	HAS	HE	I	9	IN	ING	ION	IS	IT
6	IVE	J	Ø	K	L	LA	LE	M	ME	N
7	ND	NE	NT	O	OF	ON	OR	OU	P	Q
8	R	RA	RE	RED	RES	RI	RO	S	SE	SH
9	ST	STO	T	TE	TED	TER	TH	THE	THI	THR
Ø	TI	TO	U	V	VE	W	WE	X	Y	Z

FIGURE 99.

[1] See the distinction between the terms *cipher* and *code* as treated in subpar. 11*d*.

It will be noted that the square contains the 26 letters, the 10 digits, and 64 digraphs and trigraphs chosen both on the basis of frequency considerations and the combinative potentialities of the particular polygraphs. The internal arrangement of the square is such as to permit the easy finding of the plaintext elements to be enciphered. Other matrices, of larger dimensions, may contain not only a larger number of different plaintext elements within the matrix, but may also duplicate some of the more frequent plaintext elements and thus incorporate plaintext variants within the matrix. Furthermore, when letters are used as coordinates, variant cipher equivalents may be incorporated into the scheme.

d. Typical of the many ideas that have been employed in the past for code charts is the chart which is shown in Fig. 100, below, and which has been used as a standard tactical cryptosystem

	C,D	E,H	F,I	J,K	T,L	M,O	U,V	Y,G	Z,N	P,Q	X,R	W,S	B,A
M,H	ØØØ Action, ive, ivity, s	Ø2 Addition, al	15 Advance, d, ing, s	45 After	A Aggressor, ive (ly), s	AD Air	Spell/fig. Begins	AL Airborne	AM Aircraft/ Airplane, s	AN Ammunition	AND Antiaircraft	AR Antitank	ARE Area (of)
T,Q	ØØ Armor, ed	Ø3 Arrive, al, d, ing, s	16 Artillery	5Ø Assemble, d, ing, s	AS Attack, ed, ing, s	AT Attempt, ed, ing, s	B Azimuth (in degrees)	BA Battalion, s	BE Battery, ies	BY Begin/start, ed, ings, s	C Bomb, ed, er, ing, s	CA Bridge, d, ing, s	CAN Capture, d, ing, s
K,Z	Ø Casualty, ies,	Ø4 Command er, ing, s	17 Communicate, d, ing, ion, s	55 Company, ies	CE Complete, d, ing, ion, s	CH Concentrate, d, ing, ion, s	CO Contact, ed, ing, s	D Coordinate, d, ing, ion, s	DA Corps	DAY Counterattack, ed, ing, s	DE Cross, ed, es, ing	DI Defend/defense, s (of)	DO Delay, ed, ing, s
O,L	1 Destroy, ed, ing, s	Ø5 Detach, ed, ment (of), s	18 Dispose, al, d, ition, s	E Division, s	EA Dump, s	ED East (of)	EE Encounter, ed, ing, s	EN Enemy' s	ENT Engineer, s	ER Enlisted Man/Men	ERS Equip, ment, ped, ping	ES Escape, d, ing, s	EST Estimate, d, ing, s (at)
R,X	2 Expect, ed, ing, s (at)	Ø6 Fight, er, ing, s	19 Fire, d, ing, s	ET Flank, s	F Force, d, ing, s	FO Forward	FOR Friend, ly	G From	H Front, al, s	HA Fuel, s	HE Gun, s	I Has/have	IL Headquarters
S,P	3 Heavy, ily	Ø7 Hill, s (No.)	2Ø Hold, ing, s/held	IN Hostile, ity, ities	ING Hour, s	ION How	IS Identify, ied, ies, ing, ication	IT Immediate, ly	IVE Infantry	J Inform, ation, ed, ing, s	K Install, ation, ed, ing, s	L Junction, s (of)	LA Land, ed, ing, s
W,N	4 Large	Ø8 Left (of)	21 Line, s (of)	LE Locate, d, ing, ion, s	LI Machine gun, s (nest)	LO Main	LY Map, ped, ping, s	M Mechanize, d	MA Message, nger, s	ME Mile, s (from), (to)	MENT Mine, d, ing, s	MI Mission, s	MY Morning
A,B	5 Mortar, s	Ø9 Move, d, ing, ment, s	22 Near	N Night	NA No/not/nothing/negat	ND North (of)	NE Number, s, (of)	NI Objective, s	NO Observe, ation, d, ing, s	NOT Occupy, ied, ies, ing	NT Officer, s	O Operate, d, ing, ion, s	OF Order, ed, ing, s
C,E	6 Over	1Ø Patrol, led, ling, s	23 Penetrate, d, ing, ion, s	ON Plan, ned, ning, s (to)	OR Platoon, s	OU Point, ed, ing, s	OUR Position, s	P Post, ed, ing, s	PE Prepare, d, ation, ing, s	Q Prisoner, s	QU Proceed, ed, ing, s, ure	R Radio, ed, s	RA Railway/ Railroad, s
I,G	7 Ready (for) (to)	11 Rear	25 Receive, d, ing, s/receipt	RE Reconnaissance	RED Refer, ence, red, ring, s (to)	RES Regiment, al, s	RI Reinforce, d, ing, ment, s	RO Replace, d, ing, ment, s	RS Report, ed, ing, s	RT Request, ed, ing, s	S Require, d, ing, isition, s	SA Reserve, d, ing, s	SE Ridge, s
D,J	8 Right (of)	12 River/ Stream	3Ø Road, s/ Route, s	SH Scout, ing, s	SI Section, s/ Sector, s	SO Send, ing, s/sent (to)	ST Shell, ed, ing, s	T Small/ Small arms	TA South (of)	TE Squad, s	TED Strength, s (of)/strong	TER Stop, ped, ping, s	TH Supply, ies (of)
F,V	9 Support, ed, ing, s	13 Tank, s	35 Target, s	TI Today	TION Tomorrow	TO Tonight	TR Troop, s	U Truck, s/ Vehicle, s	UN Unit, s (of)	US Until	V Urgent, cy, ly	W Vicinity (of)	WE Water
U,Y	Ø1 West (of)	14 What/who	4Ø When	X Where	Y Will	Z With	Spell/fig. Ends	Period . Withdraw, al, ing, s	Comma , Woods	Colon : Yard, s (from), (to)	Smcln ; Yesterday	Dash — You, r	Paren ( ) Zone, s (of)

FIGURE 100.

for ground forces by AGGRESSOR, the maneuver enemy in U. S. joint maneuvers and training exercises. This chart provides 2-letter equivalents for letters, numbers, syllables, and a selection of words which occur frequently in low-echelon [2] messages. A particular plaintext value may be designated by a combination of one of the two row coordinates and one of the two column coordinates of the cell containing the plaintext value; thus each plaintext element has four variant equivalents and, for example, the word ARTILLERY contained in the chart may be encrypted *in toto* as TF, TI, QF, or QI. When a complete word contained in the chart is to be encrypted

[2] The term *low-echelon* as applied to a cryptographic system means that the system is designed for use at the lower organizational levels such as (in the army) at the regimental level and below. The term *low-grade* as applied to cryptosystems means that the inherent security afforded by the system is low. Cf. the terms *medium-grade*, and *high-echelon* and *-grade*, in the glossary.

in a message, no designator is necessary to indicate this *lower-case* meaning. However, when *upper-case* meanings (i. e., letters, numbers, and syllables) are to be encrypted, it is necessary first to encrypt the designator "Spell/fig. Begins", followed by the cipher equivalents of the particular upper-case meanings; when the spelling is completed, the designator "Spell/fig. *Ends*" is encrypted, to show the return to lower-case meanings. The coordinates of the chart, as used by AGGRESSOR, were random sequences and were changed daily; the inside of the chart remained unchanged.

e. For the most part, the steps used in the recovery of plain text from messages involving syllabary squares differ from those used in the solution of previously-discussed multiliteral systems only in that a larger number of plaintext elements may have to be considered. The cryptanalyst must accordingly modify his interpretation of the frequency characteristics and idiomorphic patterns occurring in such messages. By a careful study of the behavior of frequently recurring cipher units, the analyst is led to conclude that certain units, because of the general characteristics they exhibit, must be representative of numbers, others of punctuation, others of single letters, and so on. This classification is based upon a knowledge of the *general behavior* of the various classes of plaintext elements. For example, cipher units representing digits may be expected to appear in clusters (as in dates and time, and the designations of topographical features, such as hills, road junctions, etc.); whereas those which represent punctuation may be expected to appear at varying intervals throughout the message text (the particular intervals being dependent upon the particular punctuation mark). When this classification has proceeded upon a solid foundation far enough, each set of cipher units is underlined throughout the text in some distinctive manner by means of colored pencils. Subsequent to this, the individual members of each class of cipher units are subjected to closer scrutiny, and based upon a knowledge of the *specific behavior* of the various elements in each class, specific units are identified as having specific plaintext meanings. For example, among those cipher units which the analyst has decided constitute the class which represents plaintext digits, the particular cipher unit representing plaintext "Ø" may be expected to be readily recognizable on the basis that (1) it is one of the three units which appear as the first unit in those clusters which are suspected of representing four-digit time designations *and* (2) it is one of the two cipher units which, with any noteworthy frequency, occur doubled at the end of the same four-unit clusters.

f. When working on messages involving code charts, the cryptanalyst usually starts by attempting to *isolate* sequences of cipher units which represent plaintext letters, syllables, numbers and punctuation. Subsequent to this he proceeds to classify and identify these particular cipher units in the manner described in the foregoing subparagraph; the recovery of *word* meanings is usually accomplished much later. The isolating of the ciphertext units which represent syllabary portions may be readily accomplished in those cases wherein the underlying code chart has only one "Spell/fig. Begins" group and one "Spell/fig. Ends" group, since the recognition of these designators automatically permits one to divide the cipher text into word values and non-word values; the recognition of these designators is made on the basis of their high frequency and their alternating placements throughout the cipher text.

g. As plaintext meanings are recovered in a syllabary square system or code chart system, these meanings should be entered into a skeleton matrix in a manner similar to that used in the solution of the bipartite systems previously described (Chapters VII and VIII). This is done in order to uncover and exploit as early as possible any evidences of systematic construction arising from the arrangement which was used in the underlying matrix. It may be assumed that each syllabary square and code chart *will* normally have had its internal elements arranged in some type of systematic fashion in order to permit the ready finding of plaintext elements during the encryption of a message.

h. Even when only a *single* message encrypted in one of these systems is available, if the internal construction of the underlying matrix is known there may be special approaches to solution, based on the nature of the plaintext elements constituting each row and each column of the particular matrix. For instance, if the words REFERENCE and YOUR and MESSAGE are known to be in the same row of a particular code chart, then it would be quite possible that the ciphertext sequence LA LH LT at the beginning of a message represents the stereotype REFERENCE YOUR MESSAGE; if but a few other similarly identifiable sequences were also available to the cryptanalyst, he could possibly recover the arrangement of the outside coordinates after a relatively few steps.

i. When there are other special circumstances involved, for instance, when isologs or messages with isologous syllabary portions (i. e., spelled-out portions encrypted "off-the-cut", such as IN TER CE P TO R and I NT ER CE P T OR) are present in the cipher text, solution is considerably facilitated. For example, suppose that the enemy is known to be using the code chart of Fig. 100 (the coordinates being as yet unrecovered), and that the following sequences

(1)	. . . PR	XS	PS	AW	NP	DQ	IZ	. . .
(2)	. . . SR	RW	WM	NG	RJ	LP	IX	. . .
(3)	. . . PX	XW	NO	WY	XJ	OK	GN	. . .
(4)	. . . SX	RB	AW	WP	JS	GX	. . .	
(5)	. . . PX	XA	BW	NQ	DY	OX	. . .	

from certain proforma messages are assumed to represent different encipherments of the word KILOMETERS. First of all, the initial digraph in each sequence must represent K_p, since there are no plaintext polygraphs beginning with K in the chart. Then the $\overline{AW}_c$ in (1) and (4) is noted; this *must* represent O_p, since, from (4), $\overline{AW}_c$ can be seen to represent either L_p, or O_p, and the position of $\overline{AW}_c$ in (1) confirms the identification as O_p. The values for I_p, IL_p, L_p, and LO_p quickly follow, and the variant coordinates for these plaintext values are recorded on the edges of the chart. The endings of the sequences are now examined, and it is noted that $\overline{IZ}_c$ and $\overline{GN}_c$ must represent $\overline{RS}_p$ (since either $\overline{ERS}_p$ or S_p would have digraph equivalents ending in R or X from the already-recovered column coordinates). The recovery of the rest of the text follows easily, with but little experimentation. (The student might continue the solution and profit from the exercise.)

81. Cryptosystems employing characters other than letters or figures.—*a.* In practical cryptography today, the use of characters other than the letters of bona fide alphabets (including recognized Morse and Baudot alphabets) or the 10 digits is comparatively rare. When so-called symbol ciphers, that is, ciphers employing peculiar symbols, signs of punctuation, diacritical marks, figures of "dancing men", and so on are encountered in practical work nowadays, they are almost certain to be simple monoalphabetic ciphers. They are adequately described in romantic tales,[3] in popular books on cryptography, and in the more common types of magazine articles. No further space need be given ciphers of this type in this text, not only because of their simplicity but also because they are encountered in military cryptography only in sporadic

[3] The most famous: Edgar Allan Poe's *The Gold Bug;* Sir Arthur Conan Doyle's *The Adventure of the Dancing Men;* Jules Verne's *A Journey to the Center of the Earth.*

instances, principally in censorship activities. Even in the latter cases, it is usually found that such ciphers are employed in "intimate" correspondence for the exchange of sentiments that appear less decorous when set forth in plain language. They are very seldom used by authentic enemy agents. When such a cipher is encountered nowadays it may practically always be regarded as the work of the veriest tyro, when it is not that of a crank or a mentally-deranged person.

b. The usual preliminary procedure in handling such cases, where the symbols may be somewhat confusing to the mind because of their unfamiliar appearance to the eye, is to substitute letters for them consistently throughout the message and then treat the resulting text in the manner in which an ordinary cryptogram composed of letters is treated. This procedure also facilitates the construction of the necessary frequency distributions, which would be tedious to construct by using symbols.

c. A final word must be said on the subject of symbol ciphers by way of caution. When symbols are used to replace letters, syllables, and entire words, then the systems approach code methods in principle, and can become difficult of solution. The logical extension of the use of symbols in such a form of writing is the employment of arbitrary characters for a specially developed "shorthand" system bearing little or no resemblance to well-known and therefore nonsecret, systems of shorthand, such as Gregg, Pitman, etc.[4] Unless a considerable amount of text is available for analysis, a privately-devised shorthand may be very difficult to solve. Fortunately, such systems are rarely encountered in military cryptography. They fall under the heading of cryptographic curiosities, of interest to the cryptanalyst in his leisure moments.[5]

82. Special remarks concerning the initial classification of cryptograms.—*a*. The student should by this time have a good conception of the basic nature of monoalphabetic substitution and of the many variations which may be played upon this simple tune. The first step of all, naturally, is to be able to classify a cryptogram properly and place it in either the transposition or the substitution class. The tests for this classification have been given and as a rule the student will encounter no difficulty in this respect.

b. There are, however, certain kinds of cryptograms whose class cannot be determined in the usual manner, as outlined in par. 25 of this text. First of all there is the type of code message which employs bona fide dictionary words as code groups. Naturally, a frequency distribution of such a message will approximate that for normal plain text. The appearance of the message, however, gives clear indications of what is involved. The study of such cases will be taken up in its proper place. At the moment it is only necessary to point out that these are *code* messages and not *cipher*, and it is for this reason that in pars. 24 and 25 the words "cipher" and "cipher messages" are used, the word "cryptogram" being used only where technically correct.

c. Secondly, there come the unusual and borderline cases, including cryptograms whose nature and type can *not* be ascertained from frequency distributions. Here, the cryptograms are technically not ciphers but special forms of disguised secret writings which are rarely sus-

---

[4] The use of symbols for abbreviation and speed in writing goes back to the days of antiquity. Cicero's freedman and amanuensis, Tiro, is reported to have drawn up "a book like a dictionary, in which he placed before each word the notation (symbol) which should represent it, and so great was the number of notations and words that whatever could be written in Latin could be expressed in his notation." The designation "Tironian notes" is applied to this type of shorthand.

[5] An example is found in the famous Pepys Diary, which was written in shorthand, purely for his own eyes by Samuel Pepys (1633–1703). "He wrote it in Shelton's system of tachygraphy (1641), which he complicated by using foreign languages or by varieties of his own invention whenever he had to record passages least fit to be seen by his servants, or by 'all the world.' "

ceptible of being classed as transposition or substitution. These include a large share of the cases wherein the cryptographic messages are disguised and carried under an external, innocuous text which is innocent and seemingly without cryptographic content—for instance, in a message wherein specific letters are indicated in a way not open to suspicion under censorship, these letters being intended to constitute the letters of the cryptographic messages and the other letters constituting "dummies." Obviously, no amount of frequency tabulations will avail a competent, expert cryptanalyst in demonstrating or disclosing the presence of a cryptographic message, written and secreted within the "open" message, which serves but as an envelop and disguise for its authentic or real import. Certainly, such frequency tabulations can disclose the existence *neither* of substitution *nor* transposition in these cases, since both forms are absent. The next paragraph contains more about these latter cases.[6]

83. Disguised secret communications.—*a*. As was mentioned above, there is a general class of methods of secret writing in which a secret message is concealed within the text of an apparently innocuous plaintext message; also, by extension, a secret message may be concealed within otherwise bona fide media such as maps, drawings, charts, music manuscripts, bridge hands, chess problems, shopping lists, stock quotations, and so on. The addressee of such a communication, knowing where to look for the secret elements, does so and from them is able to read the message contained within its covering disguise. When the plaintext elements of the secret message are concealed by surrounding them with the plaintext elements of an innocent cover text, such a system is known as a *concealment system*. When, however, the plaintext elements of the secret message are not themselves concealed within a cover text, but instead have *code* equivalents which are actual plaintext words or phrases and which are used to form an apparently innocent message, such a system is called an *open code system*.

b. An example of a concealment system message is the communication "HAVE ESTABLISHED LOW PRIORITY", in which the secret message "help" has been concealed as the first letter of each word of the covering text. As an example of an open code, in the message "AUNT MARY LEFT FOR DETROIT ON FRIDAY", the words AUNT MARY might stand for "five troop ships", DETROIT might mean "Southampton", and FRIDAY might stand for "Monday." An often-cited case of open code is the message "A SON IS BORN", which allegedly was sent out by German-controlled radio stations all over the world in August, 1914, meaning that war was about to be declared.

c. The solution of *concealment systems* may pose considerable difficulties for the cryptanalyst, who is placed in the rather odd situation where he might have before him a simple system, *if* he can but find the system. Most of the statistical and other tools at the disposal of the cryptanalyst are of no avail to him in the attack on concealment systems. First of all, he might not even know whether or not a given piece of correspondence *does* contain a secret message; often the only reason for an examination of a particular message, other than a random sampling case, is that the originator or the addressee is on a suspect list and therefore the communication is considered for possible secret writing. The difficulty in analysis is usually not brought about by the complexity of the system, for concealment systems are almost always cryptographically simple. The difficulty of the problem arises from the lack, at the outset, of tangible cryptographic elements into which the cryptanalyst can "get his teeth". There is primarily the

[6] The subparagraph which the student has just read (82*c*) contains a hidden cryptographic message. With the hints given in par. 83 let the student see if he can uncover it.

question of determining whether or not a secret text actually exists,[7] and if it does, locating the elements which constitute it.

d. Clearly objective methods for recognizing a concealment system message as such prior to recovering the secret text itself are not available. However, the reader may find useful a list of various situations which a censorship cryptanalyst should regard with suspicion and which may be indications of concealment system messages. Such a list of situations is given below:

(1) A letter is sent airmail or special delivery, when the contents do not warrant such speed and expense; (2) there is a discrepancy between the dating and the postmark of a letter; (3) the contents of a letter do not seem to warrant the time taken to write it, yet it appears to be composed with care and exceptional neatness; (4) the subject matter is out of accord with known facts and circumstances; (5) there are undue spacings between words, or there appear to be some "carefully placed" words in the text; (6) a writer is known to use colons habitually after the salutation, and a letter is intercepted which has a dash or a semicolon after the salutation; (7) there is a pronounced irregularity in the manner of the dotting of *i*'s or the crossing of *t*'s, or there is undue shading or other abberations in the formation of letters; (8) there is a pronouncedly stilted style or forced terminology; (9) there are inconsistencies in the style of the writer, involving misspellings or other errors incompatible with the apparent education of the writer; (10) there are peculiar or excessive underlinings which are not rational with the apparent stress intended; (11) the writer purports to be a child, writing in a childish scrawl, yet he uses words which are unlikely for a child or he forces misspellings which somehow do not ring true; (12) in a map or sketch, there are unnecessary breaks in border outlines, routes, etc.; (13) in a music manuscript, there is an excruciatingly bad (even for an "amateur composer") melodic or harmonic progression, or there are implausible accents or marks of expression; (14) in a diagram of a chess problem, there is an "impossible situation"; (15) there are entirely too many references to names of people, places, objects, or items, in what purports to be a friendly letter; (16) in short, *anything* that appears "just not quite right."

e. Locating the elements constituting the secret text of a concealment system message and deriving the meaning of the secret text are practically synonymous. This phase in the solution of a concealment system message can involve a tremendous amount of time and labor, simply

[7] Success in this type of analytic work requires extraordinary patience and perseverance, keen powers of observation nurtured by unrelenting suspicion, a lively imagination, exceptional ingenuity, and organized methods of analysis—plus a firm foundation and considerable experience in the methods and practices of concealment systems.

In this connection, it is worthwhile to cite an extract from an official report prepared in 1946 by the wartime Office of Censorship:

> "Detection of concealed messages is based on the principle that there is no absolutely safe disguise for duplicity. Espionage letters have weaknesses and identifying characteristics, which modern techniques can minimize but never completely eliminate. Seasoned examiners develop an ability to relate facts and think clearly about possibilities. They develop a keen perception of, or alertness to, certain peculiarities, an attitude of suspicion toward certain indicators, and experience or training in handling certain types of materials.
>
> "The texts of letters containing concealed messages do not ring true; they lack spontaneity, and the normal emphasis which people give to certain thoughts or ideas is absent. Something comparable in social life is the stilted behavior and speech of a person who is obliged to entertain a stranger with whom he feels nothing in common; he behaves unnaturally; he desires to be polite, but in order to do so he must hide his boredom and pretend an interest he does not feel. Exactly the same is true in the writing of cover texts or open code letters—the attempt to pursue two aims simultaneously results in strain. Skill and experience may overcome the strained-text hazard to a high degree, but they can never completely dispel the distortion and dislocation of a normal emphasis inevitable in a cover letter."

because it generally requires considerable experimentation with possible systems—and the number possible is enormous. Appendix 6, "Classification guide to concealment systems", presents an extensive list, but at this point it will suffice to indicate a few such systems. The letters of the secret message might be concealed as the first, second, or third letters of the cover text; or they might be concealed as the final, penultimate, or antepenultimate letters of the words; or they might be concealed by means of a specific key into prearranged variable placements within the words of the innocent text. The secret text might be read by considering the letters which follow or precede all unnecessary breaks in cursive handwriting; or the secret text might be indicated by shaded letters or by pin pricks over significant letters, or even by elongated tails on words pointing to significant letters in the line above. In the analysis of concealed-letter systems, it is advisable to write the successive words of the cover text one below the other, in a column, aligned by their beginnings and subsequently to rewrite them columnwise aligned by their endings; this will assist in disclosing a secret text hidden in a fixed position relative to the beginnings or endings, or in diagonal routes near those locations (see Fig. 101*a*). It is also advisable to write out the cover text in rectangular arrangements of various widths, in order to disclose secret text which might have been concealed in every *n*th letter of the entire cover text (see Fig. 101*b*). In cases where physical indicators are employed, such as breaks in handwriting or as shaded letters, an examination of the letters in the immediate vicinity of such indicators would disclose the secret text.

Cover text:

```
UNCLE EZRA SEEMS DESPONDENT.
HAVE YOU HEARD THE LAST REPORT?
```

```
            U N C L E
              E Z R A
            S E E M S
D E S P O N D E N T
              H A V E
                Y O U
            H E A R D
                T H E
              L A S T
          R E P O R T
```

Secret text: NEED HELP

FIGURE 101*a*.

Cover text:

```
WHEN YOU SEE CHESTER AT
MADISON'S HOUSE TELL HIM
LOIS DEPARTED.
```

```
W H E N Y O
U S E E C H
E S T E R A
T M A D I S
O N S H O U
S E T E L L
H I M L O I
S D E P A R
T E D
```

Secret text: NEED HELP

FIGURE 101*b*.

f. Some systems involve the concealment of entire words, instead of just individual letters. Thus, for example, the secret text might consist of (1) every *n*th word of the cover text; (2) the first and last words of every line; (3) words preceding or following punctuation marks; (4) words bisected by an imaginary line running diagonally from the upper left to the lower right of the sheet of paper; or countless varieties of similar schemes. Grilles have also been used, the secret text being written through the apertures of the grille on placed positions on the sheet of paper, and then a covering letter written to surround and camouflage the secret text. In the solution of concealed-word systems, examining the text produced by counting off every *n*th word may bear fruit; if the secret text is long enough, the validity of the assumed secret text may be proved by the consistency of the decimation. In cases wherein a variable key has been used to indicate which words constitute the secret text, proof of the assumed secret text may be impossible, unless the key is short compared to the message lengths, or unless additional messages in *exactly the same key* are available for comparison to test an assumed key.

g. There have been many cases in which a secret text has first been converted into the dots, dashes, and spaces of the Morse code, or encrypted in a Baconian or a Trithemius cipher; then this converted text was concealed within an innocent text in any one of the almost infinite number of possible ways. Some of these ways in which the multiliteral elements of the preliminary conversion may be represented are by (1) the lengths of words; (2) the number of vowels or consonants in the words; (3) the number of syllables in the words; or (4) by the ways in which *t*'s are crossed or *i*'s are dotted. The solution of such systems involves experimentation with basic hypotheses concerning the manner in which multiliteral elements are denoted, followed by a recombination into monoalphabetic terms (under the assumption of a Morse, Trithemian, or Baconian system) and solving the reduced monoalphabetic text. Another method for a concealment system involves the use of a bipartite matrix employing coordinates consisting of vowels (or, for that matter, any other set of five or six letters); the secret text is first enciphered in this biliteral system, and then the vowels are surrounded by consonants to form the plain text of an innocent cover message. As in most concealment systems, *once such a subterfuge is suspected or assumed,* then and only then is solution possible.[8]

h. In addition to literal vehicles to conceal secret text, pictorial or physical vehicles have often been used for this purpose. Sketches, drawings, graphs, etc., have been used as the surrounding medium for actual letters of the secret plain text, or the secret text has been incorporated in such sketches and drawings by means of a shorthand or the multiliteral equivalents in a Morse, Baconian, or Trithemius alphabet. In Fig. 102*a* there is an actual example from World War II censorship activities, while in Fig. 102*b* there is a problem submitted to one of the authors by a World War II class of officers undergoing instruction in cryptanalysis. Solution of these examples is left to the student who is inclined to pursue such matters. As a matter of analysis, the student should be able to see why the second example was at once diagnosed as containing a plain text written backwards, enciphered with an *arbitrary* alphabet. If the student has not yet consulted the guide to concealment systems in Appendix 6, he should do so now.

i. The detailed discussion thus far has been limited to concealment systems. In cases of *open code,* unfortunately there are neither clear-cut methods of analysis nor of recognition; there is simply *no* rational way of proving that a message such as "AUNT MARY LEFT FOR DETROIT TODAY" contains a secret meaning, unless it is known for a fact that the sender has no aunt named Mary; and even then there still might exist a friend of the sender's who is affectionately called "Aunt Mary"—or, for that matter, she might be someone *else's* aunt.[9] And once having suspected or even proved that there *is* something rotten in Denmark, proof of the content of the hidden meaning is simply out of the question unless the sender is somehow convinced to mend his

---

[8] At this point the student might like to try his hand on the secret text hidden in subpar. 82*c*.

[9] In one instance, it has been related that a censor reviewed a telegram transmitted by a person on a suspect list. The telegram read "FATHER IS DECEASED." The censor, smelling a rat, changed the text to read "FATHER IS DEAD" and waited. Sure enough, several hours later came a query: "IS FATHER DEAD OR DECEASED?"

FIGURE 102*a*.

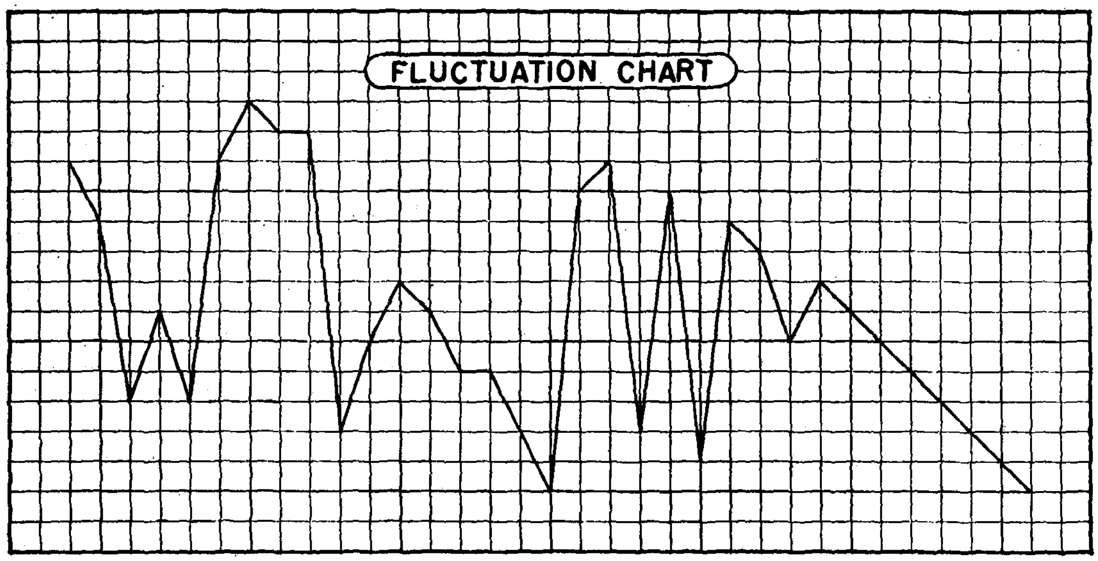


FIGURE 102*b*.

ways and thereupon volunteers the information. In many wartime instances where open codes have been used, a legal case could not be proved against a suspect without his cooperation.[10]

j. A prominent case of the use of open code in espionage communications was that of an Axis spy, Mrs. Velvalee Dickinson, who in August, 1944, was sentenced in New York to ten years' imprisonment and was fined $10,000 after pleading guilty to the charge that a series of letters she had written to an agent in Buenos Aires in the early part of 1942 contained secret messages hidden in the plain text. These messages gave information regarding the location and condition of Allied warships in Pacific ports. The two agents professed to be dealers in antique dolls and used a prearranged code giving secondary meanings to words pertaining to the sale of dolls. Mrs. Dickinson would send out letters advertising or offering to sell certain of her antique dolls to the addressee. She would write the doll's name and after the name a brief description; then she would write, as in an ordinary business letter, the price of each doll. The original cause for suspicion was the extreme variation in prices over a range of three or four letters of what was apparently the same doll or the same type of doll. A great many letters were necessary in order to build up a case sufficient to prove the use of open code. It is doubtful even then that the use of open code could have been legally proven except for the fact that, faced with so much evidence against her, she chose to confess this use.

k. In addition to concealment systems and open codes, there are still other methods for hiding the existence of secret text. The majority of these methods embrace the following:

(1) secret inks;

(2) microscopic writing, involving use of micropantographs; and

(3) photographic methods, including "microdots" (i. e., the reduction of a page of copy to a negative the size of a miniature dot, which is then affixed on a period or on the dot of an "i"), double printing, double exposure, or concealment within photographs.

The *modus operandi* and analysis of these latter methods are, however, beyond the scope of this text.

84. Concluding remarks.—*a.* The student will have by this time appreciated that monoalphabetic substitution ciphers are for the most part quite easy to solve, once the underlying principles are thoroughly understood. As in other arts, continued practice with many examples leads to facility and skill in solution, especially where the student concentrates his attention upon traffic all of the same general nature, so that the type of text which he is continually encountering becomes familiar to him and its peculiarities or characteristics of construction give clues for short cuts to solution. It is true that a knowledge of the general phraseology of messages, the kind of words used, their sequences, and so on, is of very great assistance in practical work in all fields of cryptanalysis. In operational cryptanalysis, it is of vital importance to gain a knowledge of the language habits of a particular group of correspondents, to permit the rapid exploitation of the cryptosystem involved. Thus, at least initially, all possible traffic is cryptanalyzed, even that in simple systems and that of comparatively little intelligence value. Word lists obtained empirically are of more value than "intuitive" or academic compilations; however, at the outset, reference may of course be made to these latter compilations.[11]

---

[10] Almost any element of a communication can have a code meaning; e. g., a reference to a particular kind of a flower might mean "two transports leaving tomorrow." Among the elements that have been used are the following: (1) proper nouns, place names, person's names, relatives, flowers, etc.; (2) description of bidding in a game of bridge; (3) references to particular novels or other books (4) advertisements; (5) military 24-hour clock system (permitting 1440 different prearranged meanings); (6) references to musical compositions.

[11] See in this connection the word and idiomorph lists comprising Appendix 3.

b. Some of the simpler subterfuges which the student should be on the lookout for in monoalphabetic substitution are the following:

(1) As a simple departure from monoalphabetic substitution, a message might be broken up into sections, and each section enciphered monoalphabetically with a different mixed cipher alphabet. Obviously, a single, composite frequency distribution for the whole message will not show the characteristic crest and trough appearance of a simple monoalphabetic cipher, since a given cipher unit will represent different plaintext letters in different parts of the message. But if the cryptanalyst will carefully observe the distribution *as it is being compiled*, he will note that at first it presents the characteristic crest and trough appearance of monoalphabeticity, and that after a time it begins to lose this appearance. If possible he should be on the lookout for some peculiarity of grouping of letters which serves as an indicator for the shift from one cipher alphabet to the next. If he finds such an indicator he should begin a second distribution from that point on, and proceed until another shift is encountered. By thus isolating the different portions of the text, and restricting the frequency distributions to the separate monoalphabets, the problem may be treated then as an ordinary simple monoalphabetic substitution.[12] Consideration of these remarks in connection with instances of this kind leads to the comment that it is often more advisable for the cryptanalyst to compile his own data, than to have the latter prepared by clerks, especially when studying a system *ab initio*. For observations which will certainly escape an untrained clerk can be most useful and may indeed facilitate solution. For example, in the case under consideration, if a clerk should merely hand the completed over-all uniliteral distribution to the cryptanalyst, the latter may be led astray; the appearance of the composite distribution might convince him that the cryptogram is much more complicated than it really is. While still on the subject of frequency distributions, it is pointed out that, although earlier (par. 43) the *triliteral frequency distribution* was cited primarily for its usefulness in extracting frequency data relative to the digraphs and trigraphs occurring in a simple substitution cipher, this particular type of distribution is used extensively in the manual attack on many other types of cryptograms because it provides one of the best means for systematically locating all of the repetitions which appear in a message.

(2) There have been cases where direct and reversed standard alphabets have been used alternately in a single cryptogram, the change of alphabets being made at irregular intervals, or changed at the end of every word or with each group of five letters. If the interruption takes place at too short an interval, not only will a frequency distribution be of no avail, but also it would be almost impossible for the cryptanalyst to determine when and how the change of alphabets occurs from a mere examination of the cipher text. However, if the cryptanalyst is on the alert *to try the simplest thing first*, completing the plain-component sequence on the assumption of standard alphabets will yield a solution where otherwise a solution might be out of the question.

(3) Another subterfuge that has been encountered is the encryption by means of a monoalphabetic uniliteral substitution of a message whose plain text has first been written *backwards* (or for that matter, an ordinary simple substitution cipher *sent* backwards). Ciphers of this type may successfully resist the unsystematic attempts of solution which a tyro might make;

---

[12] The cryptanalyst should be on the alert for the possibility of *related* alphabets in such a system; if related alphabets *have* been used, the reconstruction of the primary components from the solution of one portion of the message would enable the reading of the other portions of the message by means of the generatrix method treated in par. 50.

however, the experienced analyst would probably quickly recognize the weak subterfuge if he were to examine the frequencies of cipher digraphs, trigraphs, and tetragraphs, in relation to the uniliteral frequencies of their component letters.

c. Monoalphabetic substitution with variants represents an extension of the basic principle, with the intention of masking the characteristic frequencies resulting from a strict monoalphabeticity, by means of which solutions are rather readily obtained. Some of the subterfuges applied in the establishment of variant or multiple values are simple and more or less fail to serve the purpose for which they are intended; others, on the contrary, may interpose serious difficulties to a straightforward solution. But in no case may the problem be considered of more than ordinary difficulty. Furthermore, it should be recognized that where these subterfuges are really adequate to the purpose, the complications introduced are such that the practical manipulation of the system becomes as difficult for the cryptographer as for the cryptanalyst.

(1) A few words may be added here in regard to a method which often suggests itself to laymen, but which is very old indeed in the art. This consists in using a book possessed by all the correspondents and indicating the letters of the message by means of numbers referring to specific letters in the book. One way consists in selecting a certain page and then giving the line number and position of the letter in the line, the page number being shown by a single initial indicator. Another way is to use the entire book, giving the cipher equivalents in groups of three numbers representing page, line, and number of letter (for example, 75–8–10 means page 75, 8th line, 10th letter in the line). Such systems are, however, extremely cumbersome to use and, when the enciphering is done carelessly, can be solved. The basis for solution in such cases rests upon the use of adjacent letters on the same line, the accidental repetitions of certain letters, and the occurrence of unenciphered words in the messages, when laziness or fatigue intervenes in the enciphering.[13]

(2) It may also be indicated that human nature and the fallibility of cipher clerks is such that it is rather rare for an encipherer to make full use of the complement of variants placed at his disposal. The result is that in most cases certain of the equivalents will be used so much more often than others that diversities in frequencies will soon manifest themselves, affording important data for attack by the cryptanalyst.

d. There is one additional aspect of cryptography within the realm of monoalphabetic substitution ciphers that should be discussed at this point—the aspect involving *repetitive* monoalphabetic substitution.

(1) Suppose a message undergoes a primary encipherment by means of a single mixed, nonreciprocal cipher alphabet, and this primary cipher text then undergoes a secondary encipherment by means of the same or a *different* mixed alphabet. The resulting cryptogram is still monoalphabetic in character, and presents very little, if any, augmentation in the degree of security (depending upon the type of alphabet employed).[14] Here an entirely illusory increase in

---

[13] In 1915 the German Government conspired with a group of Hindu revolutionaries to stir up a rebellion in India, the purpose being to cause the withdrawal of British troops from the Western Front. Hindu conspirators in the United States were given money to purchase arms and ammunition and to transport them to India. For communication with their superiors in Berlin the conspirators used, among others, the system described in this subparagraph. A 7-page typewritten letter, built up from page, line, and letter-number references to a book known only to the communicants, was intercepted by the British and turned over to the United States Government for use in connection with the prosecution of the Hindus for violating our neutrality. The author [W. F. F.] solved this message without the book in question, by taking full advantage of the clues referred to.

[14] The only possible slight increase in security lies in the fact that the key words for the primary and secondary encipherments might be made more difficult to recover or even impossible to recover.

security is involved and an ineffectual complexity is introduced; the process may indeed be repeated indefinitely without producing the desirable result of added security. Similarly, the same illusory increase in security is present in the case of repetitive multiliteral encipherments involving regular-length ciphertext units, *as long as the repetitive encipherments are made "on the cut"*.

(2) In the case of repetitive polygraphic encipherment made on the cut, a moderate increase in security is achieved over the degree of security normally provided by a single polygraphic encipherment. For instance, in the case of repetitive digraphic encipherment using, let us say, a four-square system for the first encipherment and a modified Playfair system for the second step, the final encipherment is still monoalphabetic digraphic in character, except that the cryptosystem might have to be resolved as involving a more-or-less random square table, instead of being recovered in its primary and secondary steps; all the repetitive encipherment has accomplished is that it has added to the difficulty of reconstruction of the matrices used—but this, in the case of a digraphic system, is a reasonably fair increase in security, since we expect solution to be expedited through an early recovery of the matrix.

(3) When, however, successive multiliteral or polygraphic encipherments are made "off the cut" for the second step, the increase in security can be considerable, since the end result no longer exhibits the phenomena of monoalphabeticity and the *cryptanalytic* complexity of the system has been thereby materially enhanced.[15] For example, using the two-square matrix illustrated in Fig. 55 on p. 138, the message REENFORCEMENTS NEEDED undergoes the following encipherments:

	RE	EN	FO	RC	EM	EN	TS	NE	ED	ED
Cipher I	IL	DP	UM	CF	KT	DP	GI	UL	DF	DF
Cipher II	OC	OT	MC	MR	TD	QF	TO	OC	AH	

The first encipherment, IL DP UM. . . . , is subjected to a second encipherment by considering the digraphs "off the cut", resulting in the encryptment OC OT MC. . . . In the final cryptogram, the first and last letters of the primary encipherment may be retained as is, or they may be combined

[15] A rather ingenious idea proposed by Charles Eyraud in his excellent work, *Précis de Cryptographie Moderne*, Paris, 1953, pp. 224–225, involves a repetitive encipherment using two different monome-dinome matrices. In Eyraud's example, using the two matrices illustrated, the plain text "ECRITURES SECRETES" is first enciphered

Matrix I

	1	2	3	4	5	6	7	8	9
–	////	////	E	S	A	N	T	I	R
1	U	D	L	F	V	Q	M	P	C
2	H	G	O	B	X	W	J	Z	K

Matrix II

	1	2	3	4	5	6	7	8	9
–	///	P	///	B	G	M	U	Z	A
1	C	H	N	V	R	D	J	O	W
3	I	E	K	Q	X	S	F	L	T
	Y		Y						

E	C	R	I	T	U	R	E	S	S	E	C	R	E	T	E	S
3	19	9	8	7	11	9	3	4	4	3	19	9	3	7	3	4

31	9	9	8	7	11	9	34	4	31	9	9	37	34
I	A	A	Z	U	C	A	Q	B	I	A	A	F	Q

with Matrix I, then the digits are recombined into letters using Matrix II with the resulting cipher text IAAZU . . . . (It is interesting to note that the 17 letters of the plain text are encrypted by only *14* letters in the final cipher!) The letter Y_p is eliminated from Matrix I, and is included in Matrix II to take care of a final 1 or 3 in the first encipherment which otherwise could not have been encrypted as a single element.

for the second encryption, for added security; thus the final cryptogram may read either IOCOT .....OCAHF, or ROCOT.....OCAHG. When this sort of secondary encipherment is applied in a repetitive multiliteral cipher, the system is called a *fractionating system.* The cryptanalysis of these systems, which is often quite complex, will be treated in subsequent texts.

e. If the cryptanalyst is fortunate enough to have a pair of isologs, one message of which is in a monoalphabetic substitution system and the other in a transposition system, it may be possible for him to make exact identifications of the elements in the substitution cipher based on the plaintext letter frequencies present in the transposition cipher. Then, having the plain text, the solution of the transposition is greatly facilitated.

f. As has already been stated in subpar. 2*c*, mathematics and mathematical methods have an important place in the art of cryptanalysis. This text has included only those introductory statistical and mathematical applications which apply to monoalphabetic systems. If it appears to the student that there has been a rather extensive treatment of too-specialized techniques, let him be reassured that these have been included as being in the nature of collateral information, rather than being an absolute necessity in the solution of the particular problems to which they were applied. As a final word of caution to the student the following extract from a report by C. H. O'D. Alexander is included:

"There is a considerable danger that a learner, when he realizes that statistical methods can be of some use, will attempt to use them where they are quite inappropriate. If he does this a few times and finds it gets him nowhere, he then gives the whole thing up as a waste of time and does not use such methods where he might. There is also the worse danger of doing statistical tests for their own sake so that they are used as a method of passing the time and avoiding real thought about the problem to be solved."

g. The general problem of cryptanalytic diagnosis has been discussed briefly in various chapters of this text. The problem is far from simple, since many variations and conventions may be encountered in the various systems treated in this text; furthermore, the problem is made even harder by the fact that certain systems, themselves quite simple, may be combined to produce a system much more difficult to diagnose. The lack of *precise* diagnostic tests, such as those available in the natural sciences,[16] is brought about by the fact that variations and conventions introduced into otherwise conventional systems may change radically the

[16] The author feels that it is of value to pursue further a discussion of how the science of cryptanalytics compares with some branch of one of the natural sciences, when the diagnostic procedures involved in each are considered. In that branch of biology called taxonomic botany, for example, the first steps in the classificatory process are based upon observation of externally quite marked differences; as the process continues, the observational details become finer and finer, involving more and more difficulties as the work progresses. Towards the end of the work the botanical taxonomist may have to dissect the specimen and study internal characteristics. The whole process is largely a matter of painstaking, accurate observation of data and drawing proper conclusions therefrom. Except for the fact that the botanical taxonomist depends almost entirely upon ocular observation of characteristics while the cryptanalyst in addition to observation must use some statistics, the steps taken by the former are quite similar to those taken by the latter. It is only at the very end of the work that a significant dissimilarity between the two sciences arises. If the botanist makes a mistake in observation or deduction, he merely fails to identify the specimen correctly; he has an "answer"—but the answer is wrong. He may not be cognizant of the error; however, other more skillful botanists will find him out. But if the cryptanalyst makes a mistake in observation or deduction, he fails to get any "answer" at all; he needs nobody to tell him he has failed. Further, there is one additional important point of difference. The botanist is studying a bit of Nature—and she does not consciously interpose obstacles, pitfalls, and dissimulations in the path of those trying to solve her mysteries. The cryptanalyst, on the other hand, is studying a piece of writing prepared with the express purpose of preventing its being read by any persons for whom it is not intended. The obstacles, pitfalls, and dissimulations are here consciously interposed by the one who encrypted the message. These, of course, are what make cryptanalytics different and difficult.

appearance and manifestations expected in the cipher text produced by the known systems, yielding "hitherto-unencountered phenomena." Each cryptosystem is then actually an individual and unique case in diagnosis.[17]

(1) For example, encrypted text which is made up of four-letter groups having the pattern consonant-vowel-consonant-consonant does not *necessarily* involve a code system, even, though this grouping is a frequent one in four-letter code systems; the basic system might still be a cipher system, with the apparent characteristics of a code system. Upon closer examination, it might be possible to disprove a code system, based on the nonappearance of certain other characteristics that should be present in a code system.

(2) If a cryptogram or a set of cryptograms contain only the letters A through O in the cipher text, all that can be said initially is that only 15 letters are present in the encrypted text, and that the system must be one of substitution, either cipher or code. If a cipher, then the system must of course be a multiliteral system (including perhaps a mixed-length system), not excluding, for example, a digraphic system or a code chart. For instance, in the biliteral matrix of Fig. 103, below, the ciphertext units consist only of pairs of consonants, and the plaintext elements include the 26 letters and the 374 most frequent digraphs; thus the system is essentially a digraphic system. Such a system would not be at once recognized as a digraphic system; and if the vowels were used as nulls, the diagnosis of the cryptosystem would be considerably impeded.[18]

	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Z
B	A	AA	AB	AC	AD	AE	AF	AG	AH	AI	AK	AL	AM	AN	AO	AP	AR	AS	AT	AU
C	AV	AW	AY	B	BA	BE	BI	BL	BO	BR	BT	BU	BY	C	CA	CC	CE	CH	CI	CK
D	CL	CO	CR	CT	CU	CY	D	DA	DB	DC	DD	DE	DF	DG	DH	DI	DL	DM	DN	DO
F	DP	DQ	DR	DS	DT	DU	DV	DW	DY	E	EA	EB	EC	ED	EE	EF	EG	EH	EI	EJ
G	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW	EX	EY	EZ	F	FA	FC	FE	FF
H	FI	FL	FO	FR	FS	FT	FU	FY	G	GA	GC	GE	GF	GG	GH	GI	GL	GN	GO	GP
J	GR	GS	GT	GU	GW	H	HA	HB	HC	HD	HE	HF	HI	HL	HM	HN	HO	HR	HS	HT
K	HU	HY	I	IA	IB	IC	ID	IE	IF	IG	IK	IL	IM	IN	IO	IP	IR	IS	IT	IV
L	IX	IZ	J	JA	JE	JO	JU	K	KA	KE	KI	KS	L	LA	LB	LC	LD	LE	LF	LG
M	LI	LL	LM	LN	LO	LP	LR	LS	LT	LU	LV	LW	LY	M	MA	MB	MC	ME	MI	MM
N	MO	MP	MR	MS	MT	MU	MY	N	NA	NB	NC	ND	NE	NF	NG	NH	NI	NK	NL	NM
P	NN	NO	NP	NR	NS	NT	NU	NV	NW	NY	O	OA	OB	OC	OD	OE	OF	OG	OH	OI
Q	OK	OL	OM	ON	OO	OP	OR	OS	OT	OU	OV	OW	OX	OY	P	PA	PE	PF	PH	PI
R	PL	PM	PN	PO	PP	PR	PS	PT	PU	PY	Q	QU	R	RA	RB	RC	RD	RE	RF	RG
S	RH	RI	RL	RM	RN	RO	RP	RR	RS	RT	RU	RV	RW	RY	S	SA	SB	SC	SD	SE
T	SF	SG	SH	SI	SK	SL	SM	SN	SO	SP	SR	SS	ST	SU	SW	SY	T	TA	TB	TC
V	TD	TE	TF	TG	TH	TI	TL	TM	TN	TO	TP	TR	TS	TT	TU	TW	TY	TZ	U	UA
W	UB	UC	UD	UE	UG	UI	UL	UM	UN	UP	UR	US	UT	V	VA	VE	VI	VO	W	WA
X	WE	WH	WI	WL	WN	WO	WR	WY	X	XA	XC	XE	XF	XI	XN	XP	XT	Y	YA	YB
Z	YC	YD	YE	YF	YG	YH	YI	YL	YM	YN	YO	YP	YR	YS	YT	YW	Z	ZA	ZE	ZI

FIGURE 103.

[17] Baudouin (*op. cit.*, Chapter XIV) drew up a sort of check list of the classificatory procedures which an analyst might follow when attempting to diagnose the cryptosystem underlying a particular cryptogram or cryptograms. However, the science of cryptanalytics, being what it is, does not lend itself to successful completion of such diagnostic "check lists." Thus, the one compiled by Baudouin is far from satisfactory and is of no more than academic interest to the present-day practicing cryptanalyst.

[18] For a discussion of how such a system would be attacked from scratch, see par. 10 of Appendix 7.

h. The often extensive and elaborate treatment of the many varieties of cryptosystems within the scope of this text has not been given solely for the sake of the analysis of the particular systems involved, but rather to illustrate the general cryptanalytic techniques which are applied to various problems. In being guided along the lines of "thinking cryptanalytically", the student has been put in a position to analyze successfully many possible variations and modifications of the cryptosystems treated in this text and in the accompanying course of problems. The cryptosystems in this text and accompanying course have been solved for the most part from one or two messages. Naturally, there is a certain amount of artificiality in the examples and messages employed herein. The texts of messages have been manipulated, especially in connection with the accompanying problems, in order to illustrate pedagogical principles and the application of cryptanalytic techniques. In actual practice, instead of the one or two messages, five might be required; or for that matter, fifty or more might be necessary in order to effect a solution. In operational practice, there is frequently a high incidence of garbles which would have a pronounced impact on not only a facile identification of the cryptosystem but also on its subsequent solution. *Speed* is an essential criterion in operational practice; a cryptosystem must be broken and messages read as soon as possible, to be of maximum use to a field commander—messages read six or twelve months after they were sent are hardly of more than historical interest. Nevertheless, when a system is cryptanalyzed for the first time, no matter *when* it is broken it helps maintain cryptologic continuity which is of extreme importance in successful operational practice.[19]

i. The student should now study, if he has not already done so, the various appendices to this text. Through them, he may gain an insight into further aspects of cryptology and topics related to the art of cryptanalysis. Practice on many different ciphers of the types covered in this text will tend to sharpen the wits and give to the student confidence and facility in the cryptanalysis of unknown examples. It is for this reason that a course of problems (Appendix 9) is a necessary adjunct to the study of this text; as was previously mentioned, one month's actual practice in solution is worth a whole year's mere reading of theoretical principles.

j. It may be of assistance to indicate, by means of a graphic outline, the relationship existing among the various cryptographic systems thus far considered. The outline will be augmented with each succeeding text as the different cryptosystems are encountered, and will constitute what is termed a "synoptic chart of cryptography". The synoptic chart for this text (Chart 9) is appended at the end of this chapter. Looking at this chart the student may see that, although it is essentially dichotomous in form, at several levels there appears a sort of cryptographic *tertium quid*—some category (or categories) of cryptosystems which properly belongs at the particular level shown, but which does not directly fit into either of the two *primary* subdivisions already appearing at that level. However, if the student will study the synoptic chart attentively, it will assist him in fixing in mind the manner in which the various systems covered thus far are related to one another, and this will be of benefit in clearing away some of the mental fog or haziness from which he is at first apt to suffer.

k. There remain five more volumes to this series of basic texts on the art of cryptanalysis. *Military Cryptanalytics, Part II,* will deal mainly with periodic polyalphabetic substitution ciphers, including periodic numerical systems, together with an introduction to transposition solution; *Part III* will deal with varieties of aperiodic substitution systems, elementary cipher devices and cryptomechanisms, and will embrace a detailed treatment of cryptomathematics and diagnostic tests in cryptanalysis; *Part IV* will treat transposition and fractionating systems,

[19] See also in this connection the remarks in Appendix 7, par. 8.

and combined substitution-transposition systems; *Part V* will treat the reconstruction of codes, and the solution of enciphered code systems; and *Part VI* will treat the solution of representative machine cipher systems. In addition, throughout the five remaining texts there will be interpolated statistical techniques applicable to the systems treated, and information on the application of analytical machines in cryptanalytic problems. The security classification of each succeeding text will vary according to the information contained therein. It is not intended that the student study all six texts; life is too short to become an expert cryptanalyst in all fields of the art. *Parts I* and *II* embrace most of the necessary fundamentals of cryptanalysis; the succeeding four volumes will impart knowledge on more specific advanced categories of systems with which the cryptanalyst may be faced.

VUIFH HAFWN JMVDJ JWHIZ JWNRJ M

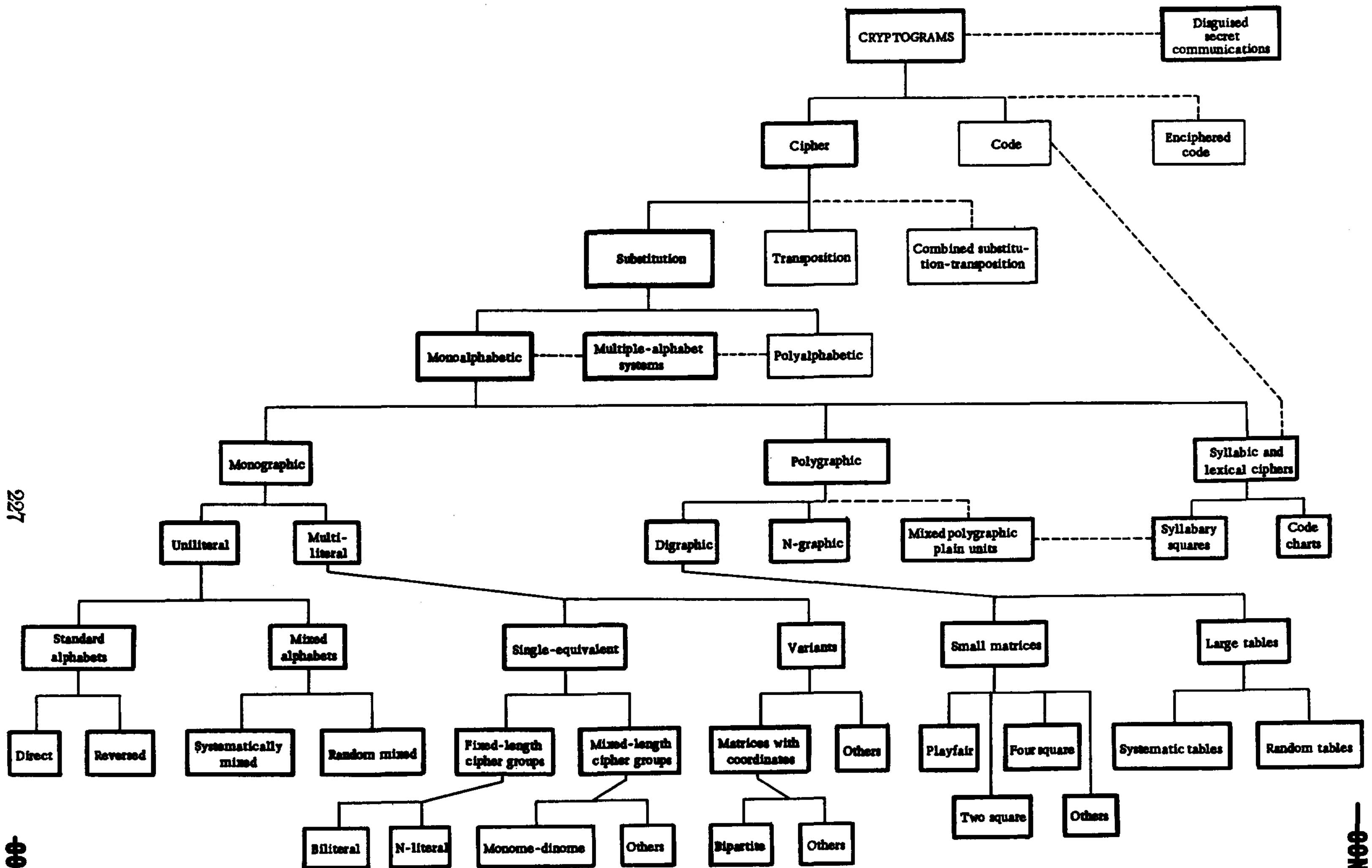


CHART 9. Synoptic chart of cryptography for *Military Cryptanalytics, Part I.*

1

2

3

4

APPENDICES

5

6

7

8

9

INDEX

APPENDIX 1

GLOSSARY FOR MILITARY CRYPTANALYTICS, PART I

This glossary is limited in scope to cryptologic terms actually appearing in this text, terms likely to be encountered in other cryptologic literature of approximately the same level as this text, and a few other terms considered necessary to complement or to clarify certain definitions.

GLOSSARY FOR MILITARY CRYPTANALYTICS, PART I

accidental repetition. A *repetition* produced fortuitously, and not by the encipherment of identical plaintext characters by identical keying elements. Cf. CAUSAL REPETITION.

additive, n. A single digit, a numerical group, or a series of digits which for the purpose of encipherment, is added to a numerical cipher unit, code group, or plain text, usually by *cryptographic arithmetic.*

additive book. A book comprising a group of additive tables.

additive system. A cryptosystem in which encipherment is accomplished through the application of additives.

additive table. A tabular arrangement of additives.

addressee, n. The office, headquarters, activity, or individual to whom a message is directed by the originator.

ADFGVX system. A German high-command cipher system used in World War I. Essentially, a biliteral substitution system employing a 6 x 6 square, to which a columnar transposition was subsequently applied.

anagram, n. Plain language reconstructed from a transposition cipher by restoring the letters of the cipher text to their original order.—v. t. To cryptanalyze a transposition cipher in whole or in part by combining one series of characters with another series from the same message to produce plain text, plain code, or intermediate cipher text.

appliqué unit, teleprinter. A special cipher attachment used in connection with a teleprinter to encrypt and decrypt teleprinter messages.

artificial word. A group of letters having no real meaning, constructed by the systematic arrangement of vowels and consonants so as to give the appearance and pronounceability of a bona fide word.

Baconian cipher. A multiliteral cipher system invented by Sir Francis Bacon (1561–1626) in which the cipher units are composed of arrangements of five elements, each of which may be chosen from one of two categories.

baud, n. The unit impulse of the code employed. Normally the impulse of shortest duration which can appear alone in a given telegraphic system, e. g., the dot in the Morse code, the impulse of teleprinter systems.

Baudot alphabet. A five-unit code applied to teleprinter systems by Jean Maurice Emile Baudot (1845–1903). It employs a 32-element alphabet designed particularly for telecommunications wherein each symbol intended for transmission is represented by a unique arrangement of five *mark* or *space impulses,* q. v.

biliteral, adj. Of or pertaining only to cryptosystems, cipher alphabets, and frequency distributions which involve cipher units of two letters or characters. *See* the more inclusive term DIGRAPHIC; *see also* BILITERAL FREQUENCY DISTRIBUTION.

biliteral alphabet. A cipher alphabet having a cipher component composed of two-character units.

biliteral frequency distribution. A frequency distribution of pairs formed by combining successive letters or characters. Thus, a biliteral distribution of ABCDEF would list the following pairs: AB, BC, CD, DE, EF. Cf. DIGRAPHIC FREQUENCY DISTRIBUTION.

bipartite alphabet. A multiliteral alphabet in which the cipher units may be divided into two separate parts whose functions are clearly defined, e. g., row indicators and column indicators of a matrix.

bipartite system. A substitution system involving the use of a bipartite alphabet.

blank expectation test. *See* LAMBDA TEST.

book cipher. A cipher system, utilizing any agreed-upon book, in which the cipher identifies a plain element present in the book.

bust message. A message or set of related messages containing an error in encipherment or violating standard cryptographic security practices so as to jeopardize the security of the message or the system and thus be of potential value to the cryptanalyst.

Caesar's cipher. An ancient form of simple substitution cipher in which each plaintext letter was replaced by the letter three places to the right of it in the normal alphabet; attributed to Julius Caesar.

callsign, adj. Of or pertaining to a call sign or call signs; as, the *callsign* generation.

call sign, n. Any combination of letters, numbers, or a combination of both, used as the identification for a communications facility, command, authority, activity, or unit; used for establishing and maintaining communications. In U. S. military practice used also for the purpose of identifying message originators and addressees.

causal repetition. A repetition produced by the encipherment of identical plaintext characters by identical keying elements. Cf. ACCIDENTAL REPETITION.

cell, n. An individual small square on cross-section paper, grilles, etc.

chadded tape. Perforated teleprinter tape; also known as fully-chadded tape, punched tape, and chad tape. Cf. CHADLESS TAPE.

chadless tape. A tape used in printing telegraphy/teleprinter operation. The perforations are not completely severed from the tape, thereby permitting the characters representing the perforations in the tape to be printed on the same tape.

characteristic frequency. *See* NORMAL FREQUENCY.

chi-square (χ^2) table. A mathematical table listing the probabilities of occurrence by chance of a chi-square value higher than those observed in a given case; an adjunct to the *chi-square test.*

chi-square (χ^2) test. A mathematical means for determining the relative likelihood that two distributions derive from the same source. For example, the test can be used to aid in the determination of whether a distribution is more likely to be random than not; in this usage, the observed distribution is compared with a theoretical distribution representing that which is expected for random. The end result of the test is a value representing the discrepancy between the two distributions which have been compared. This value, called a "chi-square value" may be interpreted as it is, or it may be interpreted through the use of a *chi-square table.*

chi (χ) test. A test applied to the distributions of the elements of two cipher texts either to determine whether the distributions are the result of encipherment by identical cipher alphabets, or to determine whether the underlying cipher alphabets are related. Also called the *cross-product sum.*

cifax, n. Enciphered facsimile. The process of converting a plane image into an unintelligible image or series of electrical impulses and of reconverting it or them into intelligibility through the use of a key.—adj. Using or pertaining to cifax.

cipher, n. 1. A cipher system. 2. A cryptogram produced by a cipher system.—adj. Pertaining to that which enciphers or is enciphered. *See also* CIPHERTEXT.

cipher alphabet. An ordered arrangement of the letters (or other conventional signs, or both) of a written language and of the characters which replace them in a cryptographic process of substitution. Also called a *substitution alphabet.*

cipher clerk. A clerk who enciphers and deciphers messages.

cipher component. The sequence of a cipher alphabet containing the symbols which replace the plaintext symbols in the process of substitution.

cipher device. A relatively simple mechanical contrivance for encipherment and decipherment, usually hand-operated or manipulated by the fingers, such as sliding strips or rotating disks.

cipher disk. A cipher device consisting of two or more concentric disks, each bearing on its periphery one component of a cipher alphabet.

cipher machine. A relatively complex apparatus or mechanism for encipherment and decipherment, usually equipped with a keyboard and often requiring an external power source.

cipher square. An orderly arrangement or collection of sequences set forth in a rectangular form, commonly a square (e. g., a Vigenère square), and employed in a cipher system.

cipher system. Any cryptosystem in which cryptographic treatment is applied to plaintext units of regular length, usually monographic or digraphic. Cf. CODE SYSTEM.

ciphertext, adj. Of or pertaining to the encrypted text produced by a cipher system or to the elements which comprise such text; as the *ciphertext* distribution. Often shortened to *cipher.*

cipher text. The text of a cryptogram which has been produced by means of a cipher system.

ciphony, n. Enciphered telephony. The process of converting vocal communications into unintelligibility and of reconverting them into intelligibility through cryptographic treatment.—adj. Using or pertaining to ciphony.

citrol, n. The process of converting control and telemetering signals, such as those used in missile guidance, into unintelligibility and reconverting them into intelligibility through cryptographic treatment.—adj. Using or pertaining to citrol.

civision, n. Enciphered television. A system of converting television signals into unintelligible signals and vice versa, in accordance with certain predetermined cryptographic procedures.—adj. Using or pertaining to civision.

clear text. *Plain text,* q. v.

code, n. 1. A *code system,* q. v. 2. A *code book,* q. v. 3. A system of signals used in electrical or electronic communication.—adj. Pertaining to that which encodes or is encoded.

code book. A book or document used in a code system, arranged in systematic form, containing units of plain text of varying length (letters, syllables, words, phrases, or sentences) each accompanied by one or more arbitrary groups of symbols used as equivalents in messages.—adj. Codebook.

code chart. A chart in the form of a matrix containing letters, syllables, numbers, words, and occasionally, phrases. The matrix has row and column coordinates for the purpose of designating the plaintext elements within.

code clerk. A clerk who encodes and decodes messages.

code group. A group of letters or numbers, or a combination of both, assigned (in a code system) to represent a plaintext element.

code message. A cryptogram produced by encodement.

code system. A cryptosystem in which arbitrary groups of symbols represent plaintext units of varying length, usually syllables, whole words, phrases, and sentences.

code text. The text of a cryptogram which has been produced by means of a code system.

coincidence, n. A recurrence of textual elements (single letters, digits, digraphs, etc.) occurring within a message or between messages.

coincidence test. The kappa test, a statistical test applied to two ciphertext messages to determine whether or not they both involve encipherment by the same sequence of cipher alphabets.

columnar transposition. A method of transposition in which the cipher text is obtained by inscribing the plain text into a matrix in any way except vertically and then transcribing the columns of the matrix.

column coordinate. A symbol normally at the top of a matrix or cryptographic table, identifying a specific column of cells, used in conjunction with a row coordinate to specify an individual cell in the matrix or table. Also called *column indicator.*

column indicator. *See* COLUMN COORDINATE.

communication intelligence. Information derived from the study of intercepted communications. Abbr. COMINT.

communication security. The protection resulting from all measures designed to deny to unauthorized persons information of value which might be derived from a study of communications. *Cryptosecurity* and *transmission security* are the components of communication security. Abbr. COMSEC.

commutative, adj. As applied to cipher matrices, so constructed as to permit coordinates to be read in either row-column or column-row order without cryptographic ambiguity.

component, n. 1. One of the two sequences (plain and cipher) which compose a cipher alphabet. 2. An independent or semi-independent part of a machine or device.

compromise, n. The availability of classified material to unauthorized persons through loss, theft, capture, recovery by salvage, defections of individuals, unauthorized viewing, or any other physical means.

computer, n. A machine for executing prescribed programs, especially a high speed automatically sequenced machine.

concealment system. A method of secret communication so designed as to convey a secret message without its presence being suspected by others than the addressee. In its most usual form, the plaintext elements are concealed by combining them with extraneous plaintext elements in such a way that the end result is an intelligible and apparently innocent message. Cf. OPEN CODE.

continuity, n. Identity with respect to a series of changes. In cryptanalytic procedure, the maintenance of continuity involves keeping current a systematic record of changes in such variable elements as indicators, keys, discriminants, code books, etc., on a given cryptochannel. In traffic analysis, the maintenance of continuity involves the tracing of changes in call signs, frequencies, schedules, or other variable elements assigned to a given radio station, link, or net.

crest, n. In its cryptologic application, a point of high relative frequency in a frequency distribution. Also called a *peak*.

crib, n. 1. Plain text assumed or known to be present in a cryptogram. 2. Keys known or assumed to have been used in a cryptogram.—v. t. 1. To fit assumed or known plain text or keys into the proper position in an encrypted message. 2. In traffic analysis, to equate an unknown element, particularly call signs and addresses, to one that is already known, especially applicable in case of compromise.

cross-product sum. *See* CHI TEST.

crypt-, crypto-. In general, a combining form meaning "hidden," "covered," or "secret." Used as a prefix in compound words, crypt-, crypto-, pertains to *cryptologic*, *cryptographic*, or *cryptanalytic*, depending upon the use of the particular word as defined.

cryptanalysis, n. The analysis of encrypted messages; the steps or processes involved in converting encrypted messages into plain text without initial knowledge of the key employed in the encryption. Abbr. C/A.

cryptanalyst, n. A person versed in the art of cryptanalysis.

cryptanalytic, adj. Of, pertaining to, or used in cryptanalytics.

cryptanalytics, n. That branch of cryptology which deals with the principles, methods, and means employed in the solution or analysis of cryptosystems.

cryptanalyze, v. t. To solve by cryptanalysis.

cryptochannel, n. A complete system for encrypted communications between two or more holders.

cryptogram, n. A communication in visible writing which conveys no intelligible meaning in any known language, or which conveys some meaning other than the real meaning.

cryptographer, n. One who encrypts or decrypts messages or has a part in making a cryptographic system.

cryptographic, adj. Of, pertaining to, or concerned with cryptography.

cryptographic ambiguity. Uncertainty as to the method of decryption or as to the meaning intended after decryption; created by a fault in the structure of a cryptosystem.

cryptographic arithmetic. The method of modular arithmetic used in cryptographic procedures which involves no carrying in addition and no borrowing in subtraction.

cryptographic security. *See* CRYPTOSECURITY.

cryptographic system. *See* CRYPTOSYSTEM.

cryptographic text. Encrypted text; the text of a cryptogram.

cryptography, n. That branch of cryptology which treats of the means, methods, and apparatus for converting or transforming plaintext messages into cryptograms, and for reconverting the cryptograms into their original plaintext form by a simple reversal of the steps used in their transformation.

cryptolinguistics, n. The study of those characteristics of languages which have some particular application in cryptology (e. g., frequency data, word patterns, unusual or impossible letter combinations, etc.).

cryptologic, adj. Of, pertaining to, or concerned with cryptology.

cryptology, n. That branch of knowledge which treats of hidden, disguised, or encrypted communications. It embraces all means and methods of producing communication intelligence and maintaining communication security; for example, cryptology includes cryptography, cryptanalytics, traffic analysis, interception, specialized linguistic processing, secret inks, etc.

cryptomaterial, n. All documents, devices, and machines employed in encrypting and decrypting messages.

cryptomathematician, n. One versed in cryptomathematics.

cryptomathematics, n. Those portions of mathematics and those mathematical methods which have cryptologic applications.

cryptoperiod, n. The specific length of time throughout which there is no change in cryptographic procedure (keys, codes, etc.).

cryptosecurity, n. That component of communicacation security which results from the provision of technically sound cryptographic systems and from their proper use.

cryptosystem, n. The associated items of cryptomaterial and the methods and rules by which these items are used as a unit to provide a single means of encryption and decryption. A cryptosystem embraces the *general cryptosystem* and the *specific keys* essential to the employment of the general cryptosystem.

cyclic, adj. Periodic; continuing or repeating so that the first term of a series follows the last; characterized by a ring or closed-chain formation.

cyclic permutation. Any rearrangement of a sequence of elements which merely involves shifting all the elements of common distance to the right or left of their initial positions in the sequence, the relative order remaining undisturbed; such a rearrangement requires that one consider the basic sequence as being circular in nature so that, for example, shifting that element which occupies the left-most position in the sequence one place to the left places this element in the right-most position.

daily keying element. That part of the specific-key that changes at predetermined intervals, usually daily.

decimated alphabet. An alphabet produced by *decimation,* q. v.

decimation, n. The process of selecting members of a series by counting off at an arbitrary interval, the original series being treated as cyclic; or the result of the foregoing process.

decimation-mixed sequence. A mixed sequence produced by *decimation,* q. v.

decipher, v. t. To convert an enciphered message into its equivalent plain text by a reversal of the cryptographic process used in the encipherment. (This does not include solution by cryptanalysis.)

deciphering alphabet. A cipher alphabet in which the sequence of symbols in the cipher component is arranged in normal order for convenience in decipherment.

decipherment, n. 1. The process of deciphering. 2. The plain text of a deciphered cryptogram. 3. In an enciphered code system, the code text resulting from the removal of the encipherment.

decode, n. 1. That section of a code book in which the code groups are in alphabetical, numerical, or other systematic order. 2. The decoded, but not translated, version of a code message.—v. t. To convert an encoded message into its plain text by means of a code book. (This does not include solution by cryptanalysis.)

decodement, n. 1. The process of decoding. 2. The decoded, but not translated, version of a cryptogram.

decrypt, n. A decrypted, but not translated, message.—v. t. To transform an encrypted communication into an intelligible one by a reversal of the cryptographic process used in encryption. (This does not include solution by cryptanalysis.)

decryption, n. The act of decrypting.

degarble, v. t. To make emendations in a garbled text.

delta I. C. Index of coincidence applied to a small sample. *See* INDEX OF COINCIDENCE.

derived numerical key. A key produced by assigning numerical values to a selected literal key.

diagnosis, n. In cryptanalysis, a systematic examination of cryptograms with a view to discovering the general system underlying these cryptograms.

digraph, n. A pair of letters.

digraphic, adj. Of or pertaining to any combination of two characters.

digraphic frequency distribution. A frequency distribution of successive pairs of letters or characters. A digraphic distribution of ABCDEF would list the pairs: AB, CD, EF. Cf. BILITERAL FREQUENCY DISTRIBUTION.

digraphic idiomorph. A plaintext or cipher sequence which contains or shows a pattern in its construction as regards the number and position of repeated digraphs.

digraphic substitution. Encipherment by substitution methods in which the plaintext units are pairs of characters and their cipher equivalents usually consist of two characters.

dinome, n. A pair of digits.

direct standard cipher alphabet. A cipher alphabet in which both the plain and cipher components are the normal sequence, the two components being juxtaposed in any of the non-crashing placements. Cf. REVERSED STANDARD CIPHER ALPHABET.

discriminant, n. A group of symbols indicating the specific cryptosystem used in encrypting a given message. Also called *system indicator.*

distribution, n. *See* FREQUENCY DISTRIBUTION.

doublet, n. A digraph or dinome in which a letter or a digit is repeated (e. g., LL, EE, 22, 66, etc.).

double transposition. A cryptosystem in which the characters of a first or primary transposition are subjected to a second transposition.

encipher, v. t. To convert a plaintext message into unintelligible language or signals by means of a cipher system.

enciphered code. A cryptographic system in which a cipher system is applied to encoded text.

enciphering alphabet. A cipher alphabet in which the sequence of letters in the plain component is arranged in normal order for convenience in encipherment.

encipherment, n. 1. The process of enciphering. 2. Text which has been enciphered.

encode, n. That section of a code book in which the plaintext equivalents of the code groups are in alphabetical, numerical, or other systematic order.—v. t. To convert a plaintext message into unintelligible language by means of a code book.

encodement, n. 1. The act or process of encrypting plain text with a code system. 2. The text produced by encoding plain text.

encrypt, v. t. To convert a plaintext message into unintelligible language or signals by means of a cryptosystem.

encrypted text. The text produced by the application of a cryptosystem to a plaintext message.

encryption, n. 1. The act of encrypting. 2. Encrypted text.

external text. In concealment systems, the apparently innocent enveloping text within which a secret message is hidden.

flat, adj. As a characteristic of a frequency distribution, implies statistically not rough. Cf. SMOOTHNESS.

four-level dinome cipher. A biliteral substitution cipher system employing four cipher sequences composed of two-digit numbers, by means of which all or nearly all of the plaintext letters are provided with four two-digit variant equivalents.

four-square matrix system. A digraphic substitution system employing a matrix which usually consists of four 5 x 5 squares in which the letters of 25-element alphabets (usually combining I and J) are inserted according to any prearranged order.

fractionation, n. A cryptographic system in which plaintext units are represented by two or more cipher symbols which in turn are dissociated and subjected to further encipherment by substitution or transposition or both.

frequency, n. In cryptology, the number of actual occurrences of a textual element within a given text. Cf. RELATIVE FREQUENCY.

frequency distribution. A tabulation of the frequency of occurrence of plaintext, ciphertext, or codetext units in a message or a group of messages. A frequency count.

frequential matrix. A type of cipher matrix providing variants. A matrix in which the number of different cipher values available to represent any given plaintext letter closely approximates its relative plaintext frequency.

garble, n. An error in transmission, reception, encryption, or decryption which renders incorrect or undecryptable a message or transmission or a portion thereof.—v. t. To make an error in transmission, reception, encryption, or decryption of a message.

general cryptosystem. The basic invariable method of encryption of a cryptosystem, excluding the specific keys essential to its employment.

general solution. A solution dependent on exploiting the inherent weaknesses of the cryptographic system arising from its own mechanics, without the presence of any specialized circumstances.

general system. *See* GENERAL CRYPTOSYSTEM.

generatrix, n. 1. One decipherment or encipherment out of a set of decipherments or encipherments of the same text, the set being exhaustive on a given hypothesis or given cryptographic principle. The elements of a generatrix are at a constant alphabetic (normal or cipher) interval from those of another generatrix of the set (e. g., as in a strip system). 2. In connection with the method of completing the plain component sequence, any one of the rows, each of which represents a trial "decipherment" of the original cryptogram.

Grandpré cipher. A type of substitution system providing dinome variants. This system employs a cipher square in which are inscribed ten 10-letter words containing all the letters of the alphabet in their approximate plaintext frequencies. These ten words are further linked together by a 10-letter word which appears vertically in the first column as a mnemonic feature for the inscription of the words in the rows.

grid, n. In a transposition system, a form or matrix over which a grille is placed for the purpose of enciphering or deciphering.

grille, n. 1. A sheet of paper, cardboard, thin metal, plastic, or like material in which perforations have been made for the uncovering of spaces in which textual units or key may be written or read on a grid. 2. A matrix in which certain squares are blocked out or otherwise marked so as not to be used. Also called a *stencil.*

group, n. 1. A number of digits, letters, or characters forming a unit for transmission or for cryptographic treatment. 2. In radio, one or more links whose stations work together as a communication entity under a common operating control.

high-echelon, adj. Pertaining to organizational units at the army divisional level or higher, or their equivalents in the other Services.

high-grade, adj. Pertaining to a cryptosystem which offers a maximum of resistance to cryptanalysis; for example: (1) complex cipher machines, (2) one-time systems, (3) two-part codes enciphered with an additive book. Cf. LOW-GRADE and MEDIUM-GRADE.

Hill's algebraic encipherment. A polygraphic system for the encipherment of polygraphs of any order, involving algebraic treatment for the transformation of a plaintext polygraph into its ciphertext polygraphic equivalent, and vice versa. Invented by Professor Lester S. Hill of Hunter College.

horizontal two-square matrix system. A digraphic substitution system employing a matrix which normally consists of two 5 x 5 squares placed side by side.

identification, n. The determination of the plaintext value of a cipher element or code group.

identify, v. t. To determine the plaintext value of a cipher element or code group.

idiomorph, n. A plaintext or cipher sequence which contains or shows a pattern in its construction as regards the number and positions of repeated elements.

idiomorphic, adj. Exhibiting the phenomenon of idiomorphism.

idiomorphism, n. In a plaintext or cipher sequence, the phenomenon of showing a pattern as regards the number and positions of repeated letters.

index of coincidence. The ratio of the observed number of coincidences in a given body of text or keys to the number of coincidences expected in a sample of random text of the same size. Commonly known as I. C. *See also* DELTA I. C.

indicator, n. In cryptography, an element inserted within the text or heading of a message which serves as a guide to the selection or derivation and application of the correct system and key for the prompt decryption of the message. *See also* the more precise terms DISCRIMINANT and MESSAGE INDICATOR.

inscription, n. 1. In a transposition system, the process of writing a message into a matrix. 2. The process of writing a series of numbers, letters, or coded meanings into a code chart or table.

intelligence, n. The product resulting from the collecting and processing of information concerning actual and potential situations and conditions relating to foreign activities and to foreign or enemy-held areas. The processing includes the evaluation and collation of the information obtained from all available sources, and the analysis, synthesis and interpretation thereof for subsequent presentation and dissemination.

intercept, n. A copy of a message obtained by interception.—v. t. To engage in interception.

interception, n. The process of gaining possession of communications intended for others without obtaining the consent of the addressees and ordinarily without delaying or preventing the transmission of the communications to those addressees.

internal text. In concealment systems, the secret text which is enveloped by open or apparently innocent text.

international Morse code. A widely-used code in which letters and numbers are represented by specific groupings of dots, dashes, or combinations of both. The international Morse code is used especially in radio telegraphy.

international teleprinter code. *See* BAUDOT ALPHABET.

interrupted-key columnar transposition. A columnar transposition system in which the plaintext elements are inscribed in a matrix in rows of irregular length as determined by a numerical key.

interval, n. A distance between two points or occurrences, especially between recurrent conditions or states. The number of units between a letter, digraph, code group, etc., and the recurrence of the same letter, digraph, code group, etc., counting either the first or second occurrence but not both. Frequently called *cryptanalyst's interval.*

intuitive method. A method of solution making use of probable words, probable keys, the supposed psychology of the encipherer, the reports of espionage services, and all other factors derivable from a given situation.

inverse four-square matrix system. A four-square matrix system in which the cipher sections contain normal alphabets while the plain component sections contain mixed alphabets.

invisible ink. Any of several chemicals used for writing or printing which has the property either of being initially invisible to the naked eye or of becoming so after a short time.

invisible writing. Writing not visible to the naked eye. The characters composing such writing may be microscopic or inscribed with invisible ink.

isolog, n. A cryptogram in which the plain text is identical or nearly identical with that of a message encrypted in another system, key, code, etc.

isologous, adj. Pertaining to or having the nature of an isolog.

Jefferson cipher. A polyalphabetic substitution system invented by Thomas Jefferson and independently at a later date by the French cryptographer Bazeries. It provided for encipherment by means of a manually operated device involving a number of revolvable disks, each bearing a mixed alphabet on its periphery.

kappa plain constant. A mathematical constant employed in coincidence tests such as the phi test, to denote the probability of a coincidence of a given plaintext element or unit. It is the sum of the squares of the probabilities of occurrence of the different textual elements or units as they are employed in writing the text; for example, in English telegraphic plain text, the monographic and digraphic plain constants are .0667 and .0069 respectively.

kappa random constant. A mathematical constant employed in coincidence tests such as the phi test to denote the probability of coincidence of a given textual element in random text. It is merely the reciprocal of the total number of characters used in writing the text. If a 26-letter alphabet were employed, for instance, the constant denoting the probability of coincidence of various textual elements would be derived as follows:

a. single letters	1/26=.0385
b. digraphs	1/676=.00148
c. trigraphs	1/17,576=.000057

kappa test. *See* COINCIDENCE TEST.

key, n. 1. In cryptography, a symbol or sequence of symbols applied to successive textual elements of a message to control their encryption or decryption. 2. A *specific key*.

key book. A book containing key text, or plain text forming specific keys.

keyed columnar transposition. A transposition system in which the columns of a matrix are taken off in the order determined by the specific key, which is often a derived numerical key.

key phrase. An arbitrarily selected phrase used as a key or from which a key is derived.

key recovery. The cryptanalytic reconstruction of a key.

key text. Text from which a key is derived.

keyword, adj. Of or pertaining to a key word or key words; as, the *keyword* recovery.

key word. An arbitrarily selected word used as a key *per se*, or from which a key is derived.

keyword-mixed alphabet. An alphabet constructed by writing a prearranged key word or key phrase (repeated letters, if present, being omitted after their first occurrence), and then completing the sequence from the unused letters of the alphabet in their normal sequence.

lambda (Λ) test. A test for monoalphabeticity in a message, based on a comparison of the observed number of blanks in its frequency distribution with the theoretically expected number of blanks both in (a) a normal plaintext message of equal length and (b) a random assortment of an equal number of letters. Also called the *blank-expectation test.*

latent repetition. A plaintext repetition not apparent in cipher text but susceptible of being made patent as a result of analysis.

Latin square. A cipher square in which no row or column contains a repeated symbol.

lexical, adj. Of, pertaining to, or connected with words. In its cryptologic sense, the word is used to characterize those cryptographic methods (chiefly codes) which deal with plaintext elements comprising complete words, phrases, and sentences.

literal key. A key composed of a sequence of letters. Cf. NUMERICAL KEY.

logarithmic weights. Numerical weights assigned to units of text, which weights are actually logarithms of the probabilities of the textual units, and which are used to evaluate the results of certain cryptanalytic operations.

low-echelon, adj. Pertaining to organizational units below the level of the army division or its equivalent in the other Services.

low-grade, adj. Pertaining to a cryptosystem which offers only slight resistance to cryptanalysis; for example: (1) Playfair ciphers, (2) single transposition, (3) unenciphered one-part codes. Cf. MEDIUM-GRADE and HIGH-GRADE.

mark, n. *Mark impulse*, q. v.

mark impulse. One of the two types of impulses used in teleprinter transmission; normally, that impulse during which current flows through the teleprinter receiving magnet. The other type of impulse is the *space impulse*, q. v.

matrix, n. A geometric form or pattern. In transposition systems, the figure or diagram in which the various steps of the transposition are effected; in substitution systems, the figure or diagram containing the sequence or sequences of plaintext or cipher symbols.

medium-grade, adj. Pertaining to a cryptosystem which offers considerable resistance to cryptanalysis; for example: (1) strip ciphers, (2) double transposition, (3) unenciphered two-part codes. Cf. LOW-GRADE and HIGH-GRADE.

message, n. Any thought or idea expressed in plain or secret language, prepared in a form suitable for transmission by any means of communication.

message indicator. A group of letters or numbers placed within an encrypted message to designate the keying elements applicable to that message.

message keying element. That part of the key which changes with every message.

mixed cipher alphabet. A cipher alphabet in which the sequence of letters or characters in one or both of the components is not the normal sequence.

mixed-length system. A cryptosystem in which the units of cipher text or code text are of irregular or nonconstant length, as for example, a monome-dinome system, or a code system employing both 4-letter and 5-letter groups.

mnemonic key. A key so constructed as to be easily remembered.

modulo, adv. With respect to a modulus, q. v. (Abbr. *mod;* e. g., mod 10, mod 26, etc.)

modulus, n. Scale or basis of arithmetic; the number *n* is called the modulus when all numbers which differ from each other by *n* or a multiple of *n* are considered equivalent.

monoalphabeticity, n. A characteristic of encrypted text which indicates that it has been produced by means of a single cipher alphabet or an unenciphered code system using a single code book. It is normally disclosed by frequency distributions which display "roughness," or pronounced variation in relative frequencies.

monoalphabetic substitution. A type of substitution employing a single cipher alphabet by means of which each cipher equivalent, composed of one or more elements, invariably represents one particular plaintext unit, wherever it occurs throughout any given message.

monographic, adj. Of or pertaining to any units comprising single characters.

monographic substitution. Encipherment by substitution methods in which the plaintext units are single characters and their cipher equivalents usually consist of single characters.

monome, n. A single digit. A contraction of mononome.

monome-dinome system. A substitution system in which certain plaintext elements have single-digit cipher equivalents, while others are represented by pairs of digits.

Morse codes. Various communication codes, of special and limited usage, in which letters and numbers are represented by specific groupings of dots, dashes, or combinations of both.

multiliteral, adj. Of or pertaining only to cryptosystems, cipher alphabets, and frequency distributions which involve cipher units of two or more letters or characters. *See also* POLYGRAPHIC.

multiliteral cipher alphabet. A cipher alphabet in which one plaintext letter is represented by cipher units of two or more elements.

multiliteral system. A substitution system involving one or more multiliteral cipher alphabets.

multiple-alphabet system. A type of substitution in which successive lengthy portions of a message are each monoalphabetically enciphered by a different alphabet; monoalphabetic encipherment by sections.

noncarrying sum. A sum produced in cryptographic (mod 10) arithmetic.

noncommutative, adj. As applied to bipartite matrices, so constructed that row and column coordinates must be read in a certain prescribed order, for example, in a row-column order.

noncrashing, adj. A term used to describe that feature of the structure of certain cryptosystems which does not permit a plaintext unit to be represented in the cipher text by the same unit.

normal alphabet. The conventional sequence of letters which form the elements of written language and are used to represent approximately the sounds of the spoken language. The direct standard alphabet beginning with "A" and ending with "Z"

normal frequency. The standard frequency of a textual unit or letter relative to other textual units or letters, as disclosed by the statistical study of a large volume of homogeneous text. Also called *characteristic frequency.*

normal sequence. The normal alphabetical sequence of those letters which are used in the written text of any particular language, or any cyclic permutation thereof.

normal uniliteral frequency distribution. A distribution showing the standard relative frequency of single plaintext symbols as disclosed by statistical study of a large volume of text.

null, n. In cryptography, a symbol or unit of encrypted text having no plaintext significance.

numerical key. A key composed of a sequence of numbers. Cf. LITERAL KEY.

numerically-keyed columnar transposition. A columnar transposition system in which the columns of a matrix are taken off in the order determined by a numerical key.

off the cut. As applied to the division of cipher text into polygraphs, beginning elsewhere than with the initial character of a bona fide polygraph.

one-part code. A code in which the plaintext elements are arranged in alphabetical, numerical, or other systematic order accompanied by their code groups also arranged in alphabetical, numerical, or other systematic order.

one-time pad. A form of key book used in a one-time system, so designed as to permit the destruction of each page of key as soon as it has been used.

one-time system. A cryptosystem in which the key, normally of a random nature, is used only once.

on the cut. As applied to the division of text into polygraphs, beginning with the first textual character.

open code. A system of disguised secret writing in which units of plain text are used as the code equivalents for letters, numbers, words, phrases, or sentences. The code equivalents, themselves, usually words or phrases, can be combined to form the intelligible text of apparently innocent messages. Cf. CONCEALMENT SYSTEM.

originator, n. The command by whose authority a message is sent. The originator is responsible for the functions of the drafter and releasing officer.

padding, n. Extraneous text added to a message for the purpose of concealing its length and beginning or ending or both.

paraphrase, v. t. To change the phraseology of a message without changing its meaning.

partially-polygraphic system. Any polygraphic substitution system in which the encipherment of certain members of the polygraphs show group relationships; small matrix systems, such as the four-square, two-square and Playfair systems involve such group relationships and are considered to be partially-digraphic systems.

partition, n. Resolution of an integer into a set of integers (e. g., representation of the integer 6 as 1 and 5, 2 and 4, or 3 and 3.

patent repetition. A repetition which is externally visible in encrypted text. Cf. LATENT REPETITION.

pentagraph, n. A set of five letters.

pentanome, n. A set of five digits.

periodic substitution. Periodic polyalphabetic substitution. A method of encipherment involving the cyclic use of two or more alphabets. Also called *repeating key method*.

permutation table. A table designed for the systematic construction of code groups. It may also be used to correct garbles in groups of code text.

phi (φ) test. A test applied to a frequency distribution to determine whether it is monoalphabetic or not. *See also* KAPPA PLAIN CONSTANT and KAPPA RANDOM CONSTANT.

physical security. That component of security which results from all physical measures necessary to safeguard classified equipment and material from access by unauthorized persons.

plain, adj. Of or pertaining to that which is unencrypted. *See also* PLAINTEXT.

plain code. Unenciphered code.

plain component. The sequence of plaintext symbols in a cipher alphabet.

plain component equivalents. In connection with the method of completing the plain component sequence, the plaintext equivalents for cipher units derived from an arbitrary juxtaposition of the components of a cipher alphabet.

plain language. *Plain text*, q. v.

plaintext, adj. Of or pertaining to that which conveys an intelligible meaning in the language in which it is written with no hidden meaning; as the *plaintext* equivalents. Often shortened to *plain*.

plain text. 1. Normal text or language which, with no hidden or secret meaning, conveys knowledge. 2. The intelligible text underlying a cryptogram.

Playfair system. A type of digraphic substitution using a single matrix normally of 25 cells.

Poisson table. Table of the Poisson distribution. A special type of mathematical table containing probability data applicable to the phenomena of repetitions expected to obtain in samples of random text; used in cryptanalysis to determine whether or not the repetitions observed in a given sample of cryptographic text are causal or random repetitions.

polyalphabetic substitution. A type of substitution in which the successive plaintext elements of a message, usually single letters, are enciphered by a succession of different alphabets.

polygraphic, adj. Of, pertaining to, or connected with any groupings comprising two or more letters or characters.

polygraphic substitution. Encipherment by substitution methods in which the plaintext units are regular length groupings of more than one element.

probable word. A word assumed or known to be present in the underlying plain text of a cryptogram. A *crib*.

probable-word method. The method of solution involving the trial of plain text assumed to be present in a cryptogram.

proforma message. A message in standardized form, designed to convey intelligence by conventions of arrangement and abbreviation.

pseudo-code system. A cipher system which produces a cryptogram whose groups resemble those of a code system.

pseudo-polygraphic system. A polygraphic substitution system in which at least one of the letters in each polygraph is enciphered monoalphabetically.

quinqueliteral alphabet. A cipher alphabet in which each plaintext letter is represented by a 5-character equivalent.

random, adj. 1. In mathematics, pertaining to unsystematic or chance variations from an expected norm. 2. In cryptanalysis, pertaining to any situation in which a statistical analysis will show variations from a calculated expected norm which variations are indistinguishable from those due to chance.

random text. Text which appears to have been produced by chance or accident, having no discernible patterns or limitations.

rapid analytical machinery. Any high-speed cryptanalytic machinery, usually electronic or photoelectric in nature. Abbr. RAM.

raw traffic. Intercepted traffic showing no evidence of processing for communication intelligence purposes beyond sorting by clear address elements, elimination of unwanted messages, and the inclusion of an arbitrary traffic designator.

read, v. t. 1. To decrypt, especially as the result of successful cryptanalytic investigation.—v. i. To yield intelligible plain text when decrypted.

readable, adj. Pertaining to those code and cipher systems in which sufficient plaintext values or keys have been recovered to permit the reading of messages encrypted in these systems.

reciprocal cipher alphabet. A cipher alphabet in which either of the two sequences may serve as plain or cipher since the equivalents exhibit reciprocity.

reciprocity, n. As used in cryptology, interchangeability of plain-cipher relationships (e. g., $A_p=B_c$ and $B_p=A_c$).

reconstruction matrix. A skeleton matrix employed in the solution of cryptosystems involving a substitution matrix. It aids in the correct relative placement of plaintext or ciphertext values as recovered, and thus often affords clues as to the internal arrangement of the original matrix.

related alphabets. Any of the several secondary cipher alphabets which are produced by sliding any given pair of primary components against each other.

relative code. Code text from which an encipherment has been removed in relative terms, but not reduced to plain-code text, so that the groups differ from the actual original plain code by an interval constant for every group, thus the difference between two relative code groups is the same as that between their plain-code equivalents.

repeating-key method. *See* PERIODIC SUBSTITUTION.

repetitive encipherment. A type of encipherment in which the primary cipher text of a cryptogram is subjected to further encipherment with either the same or a different system. Double transposition is a frequently-encountered example of repetitive encipherment.

reversed standard cipher alphabet. A cipher alphabet in which both the plain and cipher components are the normal sequence, the cipher component being reversed in direction from the plain component.

reversibility, n. That characteristic of the relationship between a plaintext digraph and its cipher digraph equivalent which permits the elements of each to be reversed without disrupting the equivalency (e. g., $AB_p=CD_c$ and $BA_p=DC_c$).

revolving grille. A type of grille in which the apertures are so distributed that when the grille is turned successively through four angles of 90 degrees and set in position on the grid, all the cells on the grid are disclosed only once. Also called *rotating grille.*

rotating grille. *See* REVOLVING GRILLE.

rotor, n. A disk designed to rotate within a cipher machine and which controls the action of some other machine component or produces a variation in some textual or keying element.

roughness, n. A pronounced variation in relative frequencies of the elements considered in a frequency distribution. Cf. SMOOTHNESS.

route transposition. A method of transposition in which the ciphertext equivalent of a message is obtained by transcribing, according to any prearranged route, the letters inscribed in the cells of a matrix into which the message was inscribed earlier according to some prearranged route.

row coordinate. A symbol normally at the side of a matrix, or cryptographic table, identifying a specific row of cells, used in conjunction with a column coordinate to specify an individual cell in the matrix or table. Also called *row indicator.*

row indicator. *See* ROW COORDINATE.

secret ink. Any of several chemicals used for writing or printing which has the property of being initially invisible to the naked eye or of becoming so after a short time. Also called *invisible ink* or *sympathetic ink.*

secret language. Text which conveys no intelligible meaning in any language or which conveys an intelligible meaning which is not the real, hidden meaning.

secret writing. 1. Visible writing in secret language. 2. Invisible writing.

separator, n. *See* WORD SEPARATOR.

sequence, n. An ordered arrangement of symbols (letters, digits, etc.) having continuity. Specifically, the members of a component of a cipher alphabet in order; the symbols in a row, column, or diagonal of a cipher square in order; key letters or key figures in order.

setting, n. The arrangement and alignment of the variable elements of a cryptographic device or machine at any moment during its operation.

sigma (σ), n. A symbol for the standard deviation.

sigmage, n. As used in cryptomathematics, a measure of the deviation from the normal, expressed in terms of numbers of sigmas (σ).

simple substitution. Monoalphabetic uniliteral substitution.

simple transposition. *See* SINGLE TRANSPOSITION.

single transposition. A transposition in which only one inscription and one transcription are effected.

sliding strip. A strip of cardboard or similar material which bears a sequence and which can be slid against other such strips to various juxtapositions.

smoothness, n. The lack of pronounced variation in relative frequencies of the elements considered in a frequency distribution. Cf. ROUGHNESS.

solution, n. In its cryptanalytic application, the process or result of solving a cryptogram or cryptosystem by cryptanalysis.

solve, v. t. To cryptanalyze. To find the plain text of encrypted communications by cryptanalytic processes, or to recover by analysis the keys and the principles of their application.

space impulse. One of the two types of impulses used in teleprinter transmission; normally, that impulse during which no current flows through the teleprinter receiving magnet. The other type of impulse is the *mark impulse,* q. v.

special solution. A solution which depends on circumstances which are not caused by the inherent principles of the particular cryptosystem. For example, solution of a periodic system by exploiting a pair of isologs which have been produced by identical sliding components but which use two different repeating keys; solution of a double transposition system by simultaneously anagramming the corresponding elements of several crytograms which are of identical length and which all use the same specific key; etc.

specific key. An element which is used with a specific cryptosystem to determine the encipherment of a message and which includes both the *message keying element* and the *daily keying element.* It may consist of a letter, number, word, phrase, sentence, a special document, book, or table, etc., usually of a variable nature and easily changeable at the will of the correspondents, or prearranged for them or for their agents by higher authority.

square, n. *See* MATRIX.

standard cipher alphabet. A cipher alphabet in which the sequence of letters in the plain component is the normal, and in the cipher component is the same as the normal, but either reversed in direction or shifted from its normal point of coincidence with the plain component.

standard uniliteral frequency distribution. *See* NORMAL UNILITERAL FREQUENCY DISTRIBUTION.

stereotype, n. A word, number, phrase, abbreviation, etc., which as a result of language habits, has a high probability of occurrence, especially at the beginning or ending of a message.

stereotyped messages. Related encrypted messages which are recognizable as such because of distinctive characteristics of the underlying plain text.

strip-cipher device. A cipher device employing sliding alphabet strips.

substitution alphabet. *See* CIPHER ALPHABET.

substitution cipher. 1. A cipher system in which the elements of the plain text are replaced by other elements. 2. A cryptogram produced by enciphering a plaintext message with a substitution system.

substitution system. A system in which the elements of the plain or code text are replaced by other elements.

sum check. A digit of a textual group which is the sum (mod 10) of the other digits in the group—v. i. To exhibit the property of a sum check.

sum-checking digit. A preselected digit (normally the final digit) in a code or cipher group which is the noncarrying sum of the other digits in the group.

summing-trinome system. A substitution system in which each plaintext letter is assigned a unique numerical value of 0 to 27. This value is then expressed as a trinome, the digits of which sum to the designated value of the letter.

superencipherment, n. A form of superencryption in which the final step involves encipherment.

superencryption, n. A further encryption of the text of a cryptogram for increased security. Enciphered code is a frequently encountered example of superencryption.

switch group. A group used within a message to indicate that the following textual elements are encrypted in a different manner.

syllabary, n. In a code book, a list of individual letters, combinations of letters, or syllables, accompanied by their equivalent code groups, usually provided for spelling out words or proper names not present in the vocabulary of a code; a spelling table.

syllabary square. A cipher matrix containing individual letters, digits, syllables, frequent digraphs, trigraphs, etc., which are encrypted by the row and column coordinates of the matrix.

syllabic, adj. Of, pertaining to, or denoting syllables.

system. *See* CRYPTOSYSTEM.

systematically-mixed cipher alphabet. A cipher alphabet in which the component that is mixed has been disarranged by systematic procedure.

system indicator. *See* DISCRIMINANT.

telecommunications, n. Any transmission, emission, or reception of signs, signals, writing, images and sounds, or intelligence of any nature by wire, radio, visual, electronic, or other means.

teleprinter, n. An electrically-operated instrument used in the transmission and reception-printing of messages by proper sensing and interpretation of electrical signals. Also called *teletypewriter, radioprinter.* A specific variety of teleprinter is the Teletype, a trademarked machine manufactured by the Teletype Corporation.

teletypewriter, n. A *teleprinter,* q. v.

tetragraph, n. A set of four letters.

tetranome, n. A set of four digits.

text, n. The part of a message containing the basic information which the originator desires to be communicated.

traffic, n. All transmitted and received communications. Abbr. tfc.

traffic analysis. That branch of cryptology which deals with the study of the external characteristics of signal communications and related materials for the purpose of obtaining information concerning the organization and operation of a communication system. Abbr. T/A.

traffic intercept. A copy of a communication obtained through interception.

transcription, n. 1. In a transposition system, the process of removing the text from a matrix or grid by a method or route different from that used in the inscription. 2. A written copy of a previously recorded radio transmission; also the process of preparing such copy from tapes or records.

transmission security. That component of communication security which results from all measures designed to protect transmissions from interception, traffic analysis, and imitative deception.

transparency, direct. That characteristic of cipher text which indicates that certain plaintext elements may have been self-enciphered.

transparency, inverse. In a digraphic system, that characteristic of cipher text which indicates that certain cipher digraphs may be merely reversals of the corresponding plaintext digraphs.

transposition cipher. 1. A transposition system. 2. A cryptogram produced by enciphering a message with a transposition system.

transposition-mixed cipher alphabet. A cipher alphabet in which at least one component (plain or cipher) has been constructed by applying a form of transposition to either a standard or a mixed sequence.

transposition system. A cryptosystem in which the elements of plain text, whether individual letters, groups of letters, syllables, words, phrases, sentences, or code groups or their components undergo some change in their relative positions without a change in their identities.

trigraph, n. A set of three letters.

trigraphic, adj. Of or pertaining to any three-character group.

trigraphic frequency distribution. A frequency distribution of successive trigraphs. A trigraphic frequency distribution of ABCDEF would consider only the trigraphs ABC and DEF. Cf. TRILITERAL FREQUENCY DISTRIBUTION.

trigraphic substitution system. A substitution system in which the plaintext units are composed of three elements.

triliteral, adj. Of, or pertaining only to cryptosystems, cipher alphabets, and frequency distributions which involve cipher units of three letters or characters. *See* the more inclusive term TRIGRAPHIC; *See also* TRILITERAL FREQUENCY DISTRIBUTION.

triliteral frequency distribution. A distribution of the characters in the text of a message in sets of three, which will show: (a) each character with its two preceding characters; or (b) each character with its two succeeding characters; or in its most usual form, (c) each character with one preceding and one succeeding character. A triliteral frequency distribution of ABCDEF would consider the groups ABC, BCD, CDE, DEF.

trinome, n. A set of three digits.

trinome-digraphic system. A substitution system in which plaintext digraphs are represented by 3-digit cipher elements.

tripartite alphabet. A multiliteral alphabet in which the cipher units may be divided into three separate parts whose functions are clearly defined, *viz.,* page, row, and column indicators of a dictionary system.

triplet, n. A group of three like symbols.

trough, n. In its cryptologic application, a point of low relative frequency in a frequency distribution.

true polygraphic system. Any polygraphic substitution system in which the individual elements of the cipher units display no evidence of monoalphabeticity, nor evidence of relationships within any group; that is, in a true polygraphic system, changing one letter in any plaintext polygraph affects the equivalent ciphertext unit in its entirety. Cf. PARTIALLY POLYGRAPHIC SYSTEM and PSEUDO-POLYGRAPHIC SYSTEM.

two-element differential. The characteristic incorporated in certain codes in which the groups differ from one another by a minimum of two elements, either in identity or the positions occupied. When the elements are letters, the characteristic is called a *two-letter differential;* when the elements are digits, it is called a *two-digit differential.*

two-part code. A randomized code, consisting of an encoding section in which the plaintext groups are arranged in an alphabetical or other systematic order accompanied by their code groups arranged in a nonalphabetical or random order; and a decoding section, in which the code groups are arranged in alphabetical or numerical order and are accompanied by their meanings as given in the encoding section.

two-square matrix system. A digraphic substitution system which normally employs a matrix consisting of two 5 x 5 squares arranged either horizontally or vertically.

uniliteral, adj. Of, or pertaining only to cryptosystems, cipher alphabets and frequency distributions which involve cipher units of single letters or characters. *See* MONOGRAPHIC; *see also* UNILITERAL FREQUENCY DISTRIBUTION.

uniliteral frequency distribution. A simple tabulation showing the frequency of individual characters of a text.

uniliteral substitution. A cryptographic process in which the individual letters of a message text are replaced by single-letter cipher equivalents.

variant, n. 1. One of two or more cipher or code symbols which have the same plain equivalent; also called *variant value.* 2. One of several plaintext meanings which may be represented by a single code group.

variant system. A substitution system in which some or all plaintext letters may be represented by more than one cipher equivalent.

variant value. *See* VARIANT.

vertical two-square matrix system. A digraphic substitution system employing a matrix which normally consists of two 5 x 5 squares arranged vertically.

Vigenère square. The cipher square commonly attributed in cryptographic literature to the French cryptographer Blaise de Vigenère (1523–1596), having the normal sequence at the top (or bottom) and at the left (or right), with cyclic permutations of the normal sequence forming the successive rows (or columns) within the square.

visible writing. Writing in which the characters are inscribed with ordinary writing materials and can be seen with the naked eye. Cf. INVISIBLE WRITING.

Wheatstone cipher device. A cipher device consisting essentially of two rings mounted concentrically in a single plane, the outer (and larger) ring being the plain component of the device and comprising 27 equisized divisions, the inner (and smaller) ring being the cipher component, comprising 26 smaller divisions. The device incorporates two hands (similar to those on a clock) pivoted at the center of the device—the larger hand serving the outer ring and the smaller hand the inner—so geared together that for each complete revolution of the larger, the smaller turns through one complete revolution plus one twenty-sixth.

word pattern. The characteristic arrangement of repeated letters in a word which tends to make it readily identifiable when enciphered monoalphabetically. *See* IDIOMORPHISM.

word separator. A unit of one or more characters employed in certain cryptosystems to indicate the space between words. It may be enciphered or unenciphered. Also called a *word spacer.*

word transposition. A cryptosystem in which whole words are transposed according to a certain prearranged route or pattern.

APPENDIX 2

LETTER FREQUENCY DATA—ENGLISH

LETTER FREQUENCY DATA—ENGLISH

Table No. | Pages

* * * * *

SPECIAL-PURPOSE DATA

TABLE 1-A.—*Absolute frequencies of letters appearing in five sets of Governmental plaintext telegrams, each set containing 10,000 letters, arranged alphabetically*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

~~CONFIDENTIAL~~

TABLE 1–B.—*Absolute frequencies of letters appearing in five sets of Governmental plaintext telegrams, each set containing 10,000 letters, arranged according to frequency*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
T	936	T	879	T	894	T	958	T	928
N	786	N	794	N	815	N	800	R	786
R	760	A	783	O	791	O	756	N	780
I	742	O	770	I	787	A	740	O	762
A	738	I	750	R	762	R	735	A	741
O	685	R	745	A	681	I	700	I	697
S	658	S	583	S	585	S	628	S	604
D	387	D	413	D	423	D	451	D	448
L	365	L	393	H	335	L	386	H	349
C	319	H	351	L	333	H	349	L	344
H	310	C	300	P	317	C	326	C	301
U	270	F	287	U	312	F	287	F	281
F	253	P	272	F	308	M	249	M	268
M	242	M	240	C	288	U	247	P	260
P	241	U	233	M	238	P	245	U	238
Y	191	G	175	Y	179	Y	213	Y	229
G	166	V	173	G	161	G	167	W	182
W	166	W	163	V	142	V	133	V	155
V	163	Y	155	W	136	W	133	G	150
B	104	B	103	B	98	B	83	B	99
X	43	X	50	Q	45	X	53	X	41
Q	40	K	38	X	44	Q	38	K	31
K	36	Q	22	K	22	K	21	Q	30
J	18	J	17	J	10	J	21	J	16
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

TABLE 1–C.—*Absolute frequencies of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants appearing in five sets of Governmental plaintext telegrams, each set containing 10,000 letters*

Set No.	Vowels	High-Frequency Consonants	Medium-Frequency Consonants	Low-Frequency Consonants
1	3, 993	3, 527	2, 329	151
2	3, 985	3, 414	2, 457	144
3	4, 042	3, 479	2, 356	123
4	3, 926	3, 572	2, 358	144
5	3, 942	3, 546	2, 389	123
Total [1]	19, 888	17, 538	11, 889	685

[1] Grand total, 50,000.

TABLE 2–A.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters, arranged alphabetically*

A	3,683	G	819	L	1,821	Q	175	V	766
B	487	H	1,694	M	1,237	R	3,788	W	780
C	1,534	I	3,676	N	3,975	S	3,058	X	231
D	2,122	J	82	O	3,764	T	4,595	Y	967
E	6,498	K	148	P	1,335	U	1,300	Z	49
F	1,416								

TABLE 2–B.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters, arranged according to frequency*

E	6,498	I	3,676	C	1,534	Y	967	X	231
T	4,595	S	3,058	F	1,416	G	819	Q	175
N	3,975	D	2,122	P	1,335	W	780	K	148
R	3,788	L	1,821	U	1,300	V	766	J	82
O	3,764	H	1,694	M	1,237	B	487	Z	49
A	3,683								

TABLE 2–C.—*Absolute frequencies of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants appearing in the combined five sets of messages totalling 50,000 letters*

Vowels	19,888
High-frequency consonants (D, N, R, S, and T)	17,538
Medium-frequency consonants (B, C, F, G, H, L, M, P, V, and W)	11,889
Low-frequency consonants (J, K, Q, X, and Z)	685
Total	50,000

TABLE 2–D.—*Absolute frequencies of letters as initial letters of 10,000 words found in Governmental plaintext telegrams*

(1) ARRANGED ALPHABETICALLY

A	905	G	109	L	196	Q	30	V	77
B	287	H	272	M	384	R	611	W	320
C	664	I	344	N	441	S	965	X	4
D	525	J	44	O	646	T	1,253	Y	88
E	390	K	23	P	433	U	122	Z	12
F	855								
								Total	10,000

(2) ARRANGED ACCORDING TO FREQUENCY

T	1,253	R	611	M	384	L	196	J	44
S	965	D	525	I	344	U	122	Q	30
A	905	N	441	W	320	G	109	K	23
F	855	P	433	B	287	Y	88	Z	12
C	664	E	390	H	272	V	77	X	4
O	646								
								Total	10,000

~~CONFIDENTIAL~~

TABLE 2-E.—*Absolute frequencies of letters as final letters of 10,000 words found in Governmental plaintext telegrams*

(1) ARRANGED ALPHABETICALLY

A	269	G	225	L	354	Q	8	V	4
B	22	H	450	M	154	R	769	W	45
C	86	I	22	N	872	S	962	X	116
D	1,002	J	6	O	575	T	1,007	Y	866
E	1,628	K	53	P	213	U	31	Z	9
F	252								
								Total	10,000

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,628	R	769	F	252	C	86	I	22
T	1,007	O	575	G	225	K	53	Z	9
D	1,002	H	450	P	213	W	45	Q	8
S	962	L	354	M	154	U	31	J	6
N	872	A	269	X	116	B	22	V	4
Y	866								
								Total	10,000

TABLE 3.—*Relative frequencies of letters appearing in 1,000 letters based upon Table 2-B*

(1) ARRANGED ALPHABETICALLY

A	73.66	G	16.38	L	36.42	Q	3.50	V	15.32
B	9.74	H	33.88	M	24.74	R	75.76	W	15.60
C	30.68	I	73.52	N	79.50	S	61.16	X	4.62
D	42.44	J	1.64	O	75.28	T	91.90	Y	19.34
E	129.96	K	2.96	P	26.70	U	26.00	Z	.98
F	28.32								
								Total	1,000.00

(2) ARRANGED ACCORDING TO FREQUENCY

E	129.96	I	73.52	C	30.68	Y	19.34	X	4.62
T	91.90	S	61.16	F	28.32	G	16.38	Q	3.50
N	79.50	D	42.44	P	26.70	W	15.60	K	2.96
R	75.76	L	36.42	U	26.00	V	15.32	J	1.64
O	75.28	H	33.88	M	24.74	B	9.74	Z	.98
A	73.66								
								Total	1,000.00

(3) VOWELS

A	73.66
E	129.96
I	73.52
O	75.28
U	26.00
Y	19.34
Total	397.76

(4) HIGH-FREQUENCY CONSONANTS

D	42.44
N	79.50
R	75.76
S	61.16
T	91.90
Total	350.76

~~CONFIDENTIAL~~

TABLE 3, Continued.—*Relative frequencies of letters appearing in 1,000 letters based upon Table 2–B*

(5) MEDIUM-FREQUENCY CONSONANTS		(6) LOW-FREQUENCY CONSONANTS	
B	9.74	J	1.64
C	30.68	K	2.96
F	28.32	Q	3.50
G	16.38	X	4.62
H	33.88	Z	.98
L	36.42	Total	13.70
M	24.74		
P	26.70		
V	15.32		
W	15.60		
Total	237.78	Total (3), (4), (5), (6)	1,000.00

TABLE 4.—*Frequency distribution for 10,000 letters of nontelegraphic English military text, as compiled by Hitt* [1]

(1) ARRANGED ALPHABETICALLY

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

TABLE 5.—*Frequency distribution for 10,000 letters of telegraphic English military text, as compiled by Hitt*[1]

(1) ARRANGED ALPHABETICALLY

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

[1] Hitt, Capt. Parker. ***Manual for the Solution of Military Ciphers.*** **Army Service Schools Press, Fort Leavenworth, Kansas, 1916.**

TABLE 6–A.—*Frequency distribution of digraphs, based on 50,000 letters of Governmental plaintext telegrams; reduced to 5,000 digraphs*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Total	Blanks
A	3	6	14	27	1	4	6	2	17	1	2	32	14	64	2	12		44	41	47	13	7	3		12		374	3
B	4				18				2	1		6	1		4			2	1	1	2				7		49	14
C	20		3	1	32	1		14	7		4	5	1	1	41			4	1	14	4		1		1		155	8
D	32	4	4	8	33	8	2	2	27	1		3	5	4	16	5	2	12	13	15	5	3	4		1		209	3
E	35	4	32	60	42	18	4	7	27	1		29	14	111	12	20	12	87	54	37	3	20	7	7	4	1	648	1
F	5		2	1	10	11	1		39			2	1		40	1		9	3	11	3		1		1		141	9
G	7		2	1	14	2	1	20	5	1		2	1	3	6	2		5	3	4	2		1				82	7
H	20	1	3	2	20	5			33			1	2	3	20	1	1	17	4	28	8		1		1		171	7
I	8	2	22	6	13	10	19				2	23	9	75	41	7		27	35	27		25		15		2	368	7
J	1				2										2						2						7	22
K	1		1		6				2			1		1					1								13	19
L	28	3	3	9	37	3	1	1	20			27	2	1	13	3		2	6	8	2	2	2		10		183	5
M	36	6	3	1	26	1		1	9				13		10	8		2	4	2	2				2		126	10
N	26	2	19	52	57	9	27	4	30	1	2	5	5	8	18	3	1	4	24	82	7	3	3		5		397	2
O	7	4	8	12	3	25	2	3	5	1	2	19	25	77	6	25		64	14	19	37	7	8	1	2		376	2
P	14	1	1	1	23	2		3	6			13	4	1	17	11		18	6	8	3	1	1		1		135	6
Q													1					1			15						17	23
R	39	2	9	17	98	6	7	3	30	1	1	5	9	7	28	13		11	31	42	5	5	4		9		382	3
S	24	3	13	5	49	12	2	26	34		1	2	3	4	15	10		5	19	63	11	1	4		1		307	4
T	28	3	6	6	71	7	1	78	45			5	6	7	50	2	1	17	19	19	5		36		41	1	454	4
U	5	3	3	3	11	1	8		5			6	5	21	1	2		31	12	12		1					130	9
V	6				57				12						1					1							77	21
W	12				22			4	13			1		2	19			1	1						1		76	16
X	2		2	1	1	1		1	2					1	1	2		1	1	7							23	13
Y	6	2	4	4	9	11	1	1	3			2	2	6	10	3		4	11	15	1		1				96	7
Z	1				2				1																		4	23
Total	370	46	154	217	657	137	82	170	374	8	14	189	123	397	373	130	17	368	304	462	130	75	77	23	99	4	5, 000	
Blanks	1	11	6	7	1	7	12	10	3	18	19	6	6	7	3	8	21	4	4	5	7	15	11	23	10	23		248

~~CONFIDENTIAL~~

TABLE 6-B.—*Frequency distribution of digraphs (naval text), based on 20,000 letters of naval text; reduced to 2,000 digraphs* [1]

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Total	Blanks
A	1	4	9	5		2	3	1	8		3	7	2	29		4		16	11	31	1	3		1	5		146	6
B	4			1	8				1			6	2		4						1				2		29	17
C	7		1		10		2	5	1		4				22	1		4	1	4							62	14
D	10	2	2	2	15	3	1	1	12			2	2		4	3		8	6	6	2	1	1		3		86	6
E	9	3	8	24	25	7	1	2	7	1	1	6	6	34	6	10	1	43	23	18	1	7	2	4	1	4	254	0
F	2		1		2	1			13			5	1		12	1		2	1	5	1				1		48	12
G	4		1	1	8	1	1	11	2			2		1	2	1		2	2	6	3					1	49	9
H	6				7	1			6				1		3	1		7	1	11	6		1				51	14
I	2	1	6	2	2	5	11					8	2	42	21	2		10	10	11		9		5			149	9
J															2												2	25
K	1	1	1		3	1			2			1		1													11	. 18
L	14	1	1		15	1			8			6			7	2		1	2	1	1	2			2		64	11
M	11	1			5				4			1	2		4	2			1						3		34	16
N	10	3	8	22	22	5	22	2	6		2	2	2	3	10	2		2	9	27	3		1				163	6
O	3	3	3	11	4	9	2		6		1	4	9	38	2	8		20	9	7	20	1	4	1	1	1	167	3
P	4				18			1	1			5			7	3		8	3	2	1						53	15
Q																					3						3	25
R	14	2	6	9	34	2	3		19		1	1	3	3	24	2		2	8	10	4		1				148	7
S	8	2	8	1	15	2		4	13			2	1	1	5	6	1	1	6	23	6		3				108	7
T	16	1	4	3	27	4	1	21	23			3	1	2	22	3		10	8	8	4		12		8	4	185	5
U	4	3	1	2	3		1		4			2	2	9		1		1	4	10							47	12
V	3				17				4						1												25	22
W	4				10			1	5						6			1									27	20
X			1			1		1	4			1								2							10	20
Y	3	1	2	1	2	3			1			3	2		2	2		1	2	2			1				28	11
Z					10																	1					11	24
Total	140	28	63	84	262	48	48	50	150	1	12	67	38	163	166	54	2	139	107	184	57	24	26	11	26	10	1, 960	
Blanks	4	12	9	13	4	10	15	15	4	25	20	7	11	15	6	8	24	8	8	8	11	19	17	22	17	22		334

[1] Fractional values have been discarded. This accounts for the discrepancy between the indicated total (1,960) and the stated total (2,000).

~~CONFIDENTIAL~~

TABLES 7–11, Inclusive

Absolute frequencies of digraphs, trigraphs, and tetragraphs and the logarithms of their assigned probabilities [1]

1. For each of the following 18 tables, the basic data were first arranged according to their absolute frequencies (F), and then the logarithms—$L_{10}(F)$ of the frequencies found.

2. The tables are designed to facilitate determination of the relative weights or probability of occurrence of sets of digraphs, trigraphs, or tetragraphs, particularly with respect to various "matching" operations. For example, are the matched digraphs RE and ET more probable than the matched digraphs RT and EF? Table 7–A shows the frequencies (F) of the digraphs to be as follows: RE=98, ET=37, RT=42, EF=18. Therefore, 98 times 37 is compared with 42 times 18, or 3,626 with 756. This arithmetic method of approach is extremely cumbersome for a large number of comparisons. By using the logarithms of the individual frequencies, the operation is greatly simplified, since the addition of the logarithms of two numbers is equivalent to the multiplication of their equivalent arithmetic values. Thus, the foregoing computation may be expressed as Log 98+Log 37, compared with Log 42+Log 18, or 0.96+0.79 versus 0.81+0.66 (see Table 7–A and explanation below). If more than one occurrence of a particular digraph is involved, it is merely necessary to multiply the logarithmic value by the number of the occurrences, *viz.*, Log X+2(Log Y)+3(Log Z), as compared with Log A+3(Log B)+2(Log C).

3. The logarithm of any given number is the power to which 10 must be raised to equal the given number. Thus, $10^2=100$, or the logarithm of 100=2. Similarly, $10^3=1{,}000$, or the logarithm of 1,000=3. The sum of logarithms is equal to the logarithm of the product of their antilogs (arithmetic numbers they represent). For example, $10^2=100$; $10^3=1000$; $10^{2+3}=100\times1000$; Log 100,000=5. Also, $10^0=1$, or Log 1=0. The Log of 0 is minus infinity $(-\infty)$.

4. In the compilation of the logarithms of the elements constituting these tables, frequencies of 1, of course, had a logarithmic value of 0.00. Digraphs which did not occur,[2] i. e., those with 0 occurrences, had a logarithmic value of minus infinity $(-\infty)$. For practical use, each of the original frequency occurrences in these tables was doubled; i. e., EN was given a frequency of 222 instead of 111, the frequency of RE became 196 instead of 98, etc. Thus, single occurrences were doubled (2×1=2), and the logarithms of those elements became 0.30 instead of 0. This is equivalent to saying Log 1+Log 2=0.00+0.30=0.30. Those elements which occurred 0 times, now were assumed to have an occurrence of 1, with an equivalent logarithmic value of 0.00.

[1] These frequency distributions are based upon data derived from 50,000 letters of U. S. Governmental plaintext telegrams, reduced to 5,000 digraphs.

[2] While in general it is possible to assign probability values to digraphs in accordance with their observed frequencies, it is not strictly correct to associate the probability "Ø" with a frequency of zero. This would be equivalent to saying: "Because a specified digraph has not occurred, it cannot occur," and would be reflected in the mathematics: "Log probability zero equals minus infinity." What may be said is: "Since a specified digraph has not occurred in the data its true probability value is unknown, except that it must be below the probability value assigned to a frequency of one." The proper way to assign a probability value to digraphs with frequencies of zero is to continue counting until they have at least one occurrence; then the true relative probability can be found.

A simple practical method of taking this difficulty into account is merely to assume that in twice the amount of data the digraph probably would have occurred at least once; that is, it has a frequency of one-half.

It should be pointed out, however, that since probabilities are multiplied (by summing logarithms) a 10% error in evaluating the digraph ZZ for example, makes the product, wherever ZZ occurs, 10% wrong, and is just as serious as a 10% error in evaluating the high-frequency digraph EN.

In practice, however, results obtained from the logarithmic method are so satisfactory that refinements are not needed.

5. In order to place all the logarithms of the initial frequencies on a comparable logarithmic basis, it was merely necessary to add 0.30 to each of them. While EN had a frequency of 111 in the original compilation, it now had a frequency of 222, or 2(111). The logarithm of 222 is 2.35. This is equivalent to saying Log 111+Log 2=2.05+0.30=2.35.

6. The frequencies as stated in terms of their actual logarithms do not readily indicate their relative size for each distribution. Therefore, the highest frequency in each group was given a value of 0.99, and the lowest a value of 0; frequencies intermediate between these extremes were evaluated in proportion to their respective frequencies. This is equivalent to expressing the frequencies in logarithms with a base other than 10. In other words, this procedure of converting the logarithms to the range from .00 to .99 consists in dividing up the original range of logarithms into 100 equal parts and assigning each one to the proper rank in the range.

7. The new base (C) used to convert each of the digraphic frequencies to the logarithmic range 0 to 0.99 is derived as follows, when 222 is the highest frequency (F):

$$\text{Let } 222 = C^{0.99}$$
$$\text{Log}_{10}\ 222 = \text{Log}_{10}\ C^{0.99}$$
$$\text{Log}_{10}\ 222 = (0.99)\ (\text{Log}_{10}\ C)$$
$$C = \text{Antilog}\ \frac{\text{Log}_{10}\ 222}{0.99} = \text{Antilog}\ \frac{2.35}{0.99}$$
$$C = 224$$

8. The formula for the computation of the logarithm to the new base (C) of any actual frequency (Y) of a series is:

$$\text{Log}_c\ Y = \frac{\text{Log}_{10}\ Y}{\text{Log}_{10}\ C}$$

It is more expeditious to use reciprocals in the conversion of a whole series of logarithmic values, as in this instance. The formula is: $(\text{Log}_{10}\ C)^{-1} \cdot (\text{Log}_{10}\ Y) = \text{Log}_c\ Y$.

9. The digraphic index chart, Table 15, on page 37, summarizes the logarithmic frequencies of all English plaintext digraphs, computed to a base of 224 so that the logarithm of the highest frequency (EN) is 0.99.

Example:

$$\text{EN} = 222$$
$$\text{Log}_{10}\ 222 = 2.35$$
$$(\text{Log}_{10}\ C)^{-1} = (\text{Log}_{10}\ 224)^{-1} = 0.421$$
$$\text{Log}_c\ 222 = 0.421 \times 2.35 = 0.99$$

10. Likewise, the trigraphs and tetragraphs have been computed to the bases L_{586} and L_{244}, respectively, so that the logarithms of the highest-frequency trigraph (ENT) and tetragraph (TION) are 0.99. Since no use is being made of the trigraphs appearing less than 100 times and tetragraphs appearing less than 50 times, the basic frequencies of the trigraphs and tetragraphs have not been doubled in computing the new bases of the logarithms.

TABLE 7-A.—*The 428 different digraphs of Table 6-A, arranged according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{224}(2F)$
EN	111	2.05	.99
RE	98	1.99	.96
ER	87	1.94	.94
NT	82	1.91	.93
TH	78	1.89	.92
ON	77	1.89	.92
IN	75	1.88	.92
TE	71	1.85	.91
AN	64	1.81	.89
OR	64	1.81	.89
ST	63	1.80	.88
ED	60	1.78	.88
NE	57	1.76	.87
VE	57	1.76	.87
ES	54	1.73	.86
ND	52	1.72	.85
TO	50	1.70	.84
SE	49	1.69	.84
	[1] 1,249		
AT	47	1.67	.83
TI	45	1.65	.82
AR	44	1.64	.82
EE	42	1.62	.81
RT	42	1.62	.81
AS	41	1.61	.80
CO	41	1.61	.80
IO	41	1.61	.80
TY	41	1.61	.80
FO	40	1.60	.80
FI	39	1.59	.80
RA	39	1.59	.80
ET	37	1.57	.79
LE	37	1.57	.79
OU	37	1.57	.79
MA	36	1.56	.78
TW	36	1.56	.78
EA	35	1.54	.78
IS	35	1.54	.78
SI	34	1.53	.77
DE	33	1.52	.77
HI	33	1.52	.77
AL	32	1.51	.76
CE	32	1.51	.76

	F	$L_{10}(F)$	$L_{224}(2F)$
DA	32	1.51	.76
EC	32	1.51	.76
RS	31	1.49	.75
UR	31	1.49	.75
NI	30	1.48	.75
RI	30	1.48	.75
EL	29	1.46	.74
HT	28	1.45	.74
LA	28	1.45	.74
RO	28	1.45	.74
TA	28	1.45	.74
	[2] 2,495		
AD	27	1.43	.73
DI	27	1.43	.73
EI	27	1.43	.73
IR	27	1.43	.73
IT	27	1.43	.73
LL	27	1.43	.73
NG	27	1.43	.73
ME	26	1.41	.72
NA	26	1.41	.72
SH	26	1.41	.72
IV	25	1.40	.72
OF	25	1.40	.72
OM	25	1.40	.72
OP	25	1.40	.72
NS	24	1.38	.71
SA	24	1.38	.71
IL	23	1.36	.70
PE	23	1.36	.70
IC	22	1.34	.69
WE	22	1.34	.69
UN	21	1.32	.68
CA	20	1.30	.67
EP	20	1.30	.67
EV	20	1.30	.67
GH	20	1.30	.67
HA	20	1.30	.67
HE	20	1.30	.67
HO	20	1.30	.67
LI	20	1.30	.67
IG	19	1.28	.67
NC	19	1.28	.67

	F	$L_{10}(F)$	$L_{224}(2F)$
OL	19	1.28	.67
OT	19	1.28	.67
SS	19	1.28	.67
TS	19	1.28	.67
TT	19	1.28	.67
WO	19	1.28	.67
BE	18	1.26	.66
EF	18	1.26	.66
NO	18	1.26	.66
PR	18	1.26	.66
AI	17	1.23	.64
HR	17	1.23	.64
PO	17	1.23	.64
RD	17	1.23	.64
TR	17	1.23	.64
DO	16	1.20	.63
DT	15	1.18	.62
IX	15	1.18	.62
QO	15	1.18	.62
SO	15	1.18	.62
YT	15	1.18	.62
AC	14	1.15	.61
AM	14	1.15	.61
CH	14	1.15	.61
CT	14	1.15	.61
EM	14	1.15	.61
GE	14	1.15	.61
OS	14	1.15	.61
PA	14	1.15	.61
AU	13	1.11	.59
DS	13	1.11	.59
IE	13	1.11	.59
LO	13	1.11	.59
MM	13	1.11	.59
PL	13	1.11	.59
RP	13	1.11	.59
SC	13	1.11	.59
WI	13	1.11	.59
	[3] 3,745		
AP	12	1.08	.58
AY	12	1.08	.58
DR	12	1.08	.58
EO	12	1.08	.58

	F	$L_{10}(F)$	$L_{224}(2F)$
EQ	12	1.08	.58
OD	12	1.08	.58
SF	12	1.08	.58
US	12	1.08	.58
UT	12	1.08	.58
VI	12	1.08	.58
WA	12	1.08	.58
FF	11	1.04	.56
FT	11	1.04	.56
PP	11	1.04	.56
RR	11	1.04	.56
SU	11	1.04	.56
UE	11	1.04	.56
YF	11	1.04	.56
YS	11	1.04	.56
FE	10	1.00	.55
IF	10	1.00	.55
LY	10	1.00	.55
MO	10	1.00	.55
SP	10	1.00	.55
YO	10	1.00	.55
FR	9	0.95	.53
IM	9	0.95	.53
LD	9	0.95	.53
MI	9	0.95	.53
NF	9	0.95	.53
RC	9	0.95	.53
RM	9	0.95	.53
RY	9	0.95	.53
YE	9	0.95	.53
DD	8	0.90	.51
DF	8	0.90	.51
HU	8	0.90	.51
IA	8	0.90	.51
LT	8	0.90	.51
MP	8	0.90	.51
NN	8	0.90	.51
OC	8	0.90	.51
OW	8	0.90	.51
PT	8	0.90	.51
UG	8	0.90	.51
AV	7	0.85	.48
BY	7	0.85	.48
CI	7	0.85	.48

[1] The 18 digraphs above this line compose 25% of the total.
[2] The 53 digraphs above this line compose 50% of the total.
[3] The 122 digraphs above this line compose 75% of the total.

TABLE 7–A, Continued.—*The 428 different digraphs of Table 6–A, arranged according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
EH------	7	0.85	.48	RU------	5	0.70	.42	GS------	3	0.48	.33	JE------	2	0.30	.25
EW------	7	0.85	.48	RV------	5	0.70	.42	HC------	3	0.48	.33	JO------	2	0.30	.25
EX------	7	0.85	.48	SD------	5	0.70	.42	HN------	3	0.48	.33	JU------	2	0.30	.25
GA------	7	0.85	.48	SR------	5	0.70	.42	LB------	3	0.48	.33	KI------	2	0.30	.25
IP------	7	0.85	.48	TL------	5	0.70	.42	LC------	3	0.48	.33	LM------	2	0.30	.25
NU------	7	0.85	.48	TU------	5	0.70	.42	LF------	3	0.48	.33	LR------	2	0.30	.25
OA------	7	0.85	.48	UA------	5	0.70	.42	LP------	3	0.48	.33	LU------	2	0.30	.25
OV------	7	0.85	.48	UI------	5	0.70	.42	MC------	3	0.48	.33	LV------	2	0.30	.25
RG------	7	0.85	.48	UM------	5	0.70	.42	NP------	3	0.48	.33	LW------	2	0.30	.25
RN------	7	0.85	.48	AF------	4	0.60	.38	NV------	3	0.48	.33	MR------	2	0.30	.25
TF------	7	0.85	.48	BA------	4	0.60	.38	NW------	3	0.48	.33	MT------	2	0.30	.25
TN------	7	0.85	.48	BO------	4	0.60	.38	OE------	3	0.48	.33	MU------	2	0.30	.25
XT------	7	0.85	.48	CK------	4	0.60	.38	OH------	3	0.48	.33	MY------	2	0.30	.25
AB------	6	0.78	.45	CR------	4	0.60	.38	PH------	3	0.48	.33	NB------	2	0.30	.25
AG------	6	0.78	.45	CU------	4	0.60	.38	PU------	3	0.48	.33	NK------	2	0.30	.25
BL------	6	0.78	.45	DB------	4	0.60	.38	RH------	3	0.48	.33	OG------	2	0.30	.25
GO------	6	0.78	.45	DC------	4	0.60	.38	SB------	3	0.48	.33	OK------	2	0.30	.25
ID------	6	0.78	.45	DN------	4	0.60	.38	SM------	3	0.48	.33	OY------	2	0.30	.25
KE------	6	0.78	.45	DW------	4	0.60	.38	TB------	3	0.48	.33	PF------	2	0.30	.25
LS------	6	0.78	.45	EB------	4	0.60	.38	UB------	3	0.48	.33	RB------	2	0.30	.25
MB------	6	0.78	.45	EG------	4	0.60	.38	UC------	3	0.48	.33	SG------	2	0.30	.25
OO------	6	0.78	.45	EY------	4	0.60	.38	UD------	3	0.48	.33	SL------	2	0.30	.25
PI------	6	0.78	.45	GT------	4	0.60	.38	YI------	3	0.48	.33	TP------	2	0.30	.25
PS------	6	0.78	.45	HS------	4	0.60	.38	YP------	3	0.48	.33	UP------	2	0.30	.25
RF------	6	0.78	.45	MS------	4	0.60	.38	AH------	2	0.30	.25	WN------	2	0.30	.25
TC------	6	0.78	.45	NH------	4	0.60	.38	AK------	2	0.30	.25	XA------	2	0.30	.25
TD------	6	0.78	.45	NR------	4	0.60	.38	AO------	2	0.30	.25	XC------	2	0.30	.25
TM------	6	0.78	.45	OB------	4	0.60	.38	BI------	2	0.30	.25	XI------	2	0.30	.25
UL------	6	0.78	.45	PM------	4	0.60	.38	BR------	2	0.30	.25	XP------	2	0.30	.25
VA------	6	0.78	.45	RW------	4	0.60	.38	BU------	2	0.30	.25	YB------	2	0.30	.25
YA------	6	0.78	.45	SN------	4	0.60	.38	DG------	2	0.30	.25	YL------	2	0.30	.25
YN------	6	0.78	.45	SW------	4	0.60	.38	DH------	2	0.30	.25	YM------	2	0.30	.25
CL------	5	0.70	.42	WH------	4	0.60	.38	DQ------	2	0.30	.25	ZE------	2	0.30	.25
DM------	5	0.70	.42	YC------	4	0.60	.38	FC------	2	0.30	.25	AE------	1	0.00	.13
DP------	5	0.70	.42	YD------	4	0.60	.38	FL------	2	0.30	.25	AJ------	1	0.00	.13
DU------	5	0.70	.42	YR------	4	0.60	.38	GC------	2	0.30	.25	BJ------	1	0.00	.13
FA------	5	0.70	.42	AA------	3	0.48	.33	GF------	2	0.30	.25	BM------	1	0.00	.13
GI------	5	0.70	.42	AW------	3	0.48	.33	GL------	2	0.30	.25	BS------	1	0.00	.13
GR------	5	0.70	.42	CC------	3	0.48	.33	GP------	2	0.30	.25	BT------	1	0.00	.13
HF------	5	0.70	.42	DL------	3	0.48	.33	GU------	2	0.30	.25	CD------	1	0.00	.13
NL------	5	0.70	.42	DV------	3	0.48	.33	HD------	2	0.30	.25	CF------	1	0.00	.13
NM------	5	0.70	.42	EU------	3	0.48	.33	HM------	2	0.30	.25	CM------	1	0.00	.13
NY------	5	0.70	.42	FS------	3	0.48	.33	IB------	2	0.30	.25	CN------	1	0.00	.13
OI------	5	0.70	.42	FU------	3	0.48	.33	IK------	2	0.30	.25	CS------	1	0.00	.13
RL------	5	0.70	.42	GN------	3	0.48	.33	IZ------	2	0.30	.25	CW------	1	0.00	.13

TABLE 7-A, Concluded.—*The 428 different digraphs of Table 6-A, arranged according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
CY	1	0.00	.13	HW	1	0.00	.13	PD	1	0.00	.13	WL	1	0.00	.13
DJ	1	0.00	.13	HY	1	0.00	.13	PN	1	0.00	.13	WR	1	0.00	.13
DY	1	0.00	.13	JA	1	0.00	.13	PV	1	0.00	.13	WS	1	0.00	.13
EJ	1	0.00	.13	KA	1	0.00	.13	PW	1	0.00	.13	WY	1	0.00	.13
EZ	1	0.00	.13	KC	1	0.00	.13	PY	1	0.00	.13	XD	1	0.00	.13
FD	1	0.00	.13	KL	1	0.00	.13	QM	1	0.00	.13	XE	1	0.00	.13
FG	1	0.00	.13	KN	1	0.00	.13	QR	1	0.00	.13	XF	1	0.00	.13
FM	1	0.00	.13	KS	1	0.00	.13	RJ	1	0.00	.13	XH	1	0.00	.13
FP	1	0.00	.13	LG	1	0.00	.13	RK	1	0.00	.13	XN	1	0.00	.13
FW	1	0.00	.13	LH	1	0.00	.13	SK	1	0.00	.13	XO	1	0.00	.13
FY	1	0.00	.13	LN	1	0.00	.13	SV	1	0.00	.13	XR	1	0.00	.13
GD	1	0.00	.13	MD	1	0.00	.13	SY	1	0.00	.13	XS	1	0.00	.13
GG	1	0.00	.13	MF	1	0.00	.13	TG	1	0.00	.13	YG	1	0.00	.13
GJ	1	0.00	.13	MH	1	0.00	.13	TQ	1	0.00	.13	YH	1	0.00	.13
GM	1	0.00	.13	NJ	1	0.00	.13	TZ	1	0.00	.13	YU	1	0.00	.13
GW	1	0.00	.13	NQ	1	0.00	.13	UF	1	0.00	.13	YW	1	0.00	.13
HB	1	0.00	.13	OJ	1	0.00	.13	UO	1	0.00	.13	ZA	1	0.00	.13
HL	1	0.00	.13	OX	1	0.00	.13	UV	1	0.00	.13	ZI	1	0.00	.13
HP	1	0.00	.13	PB	1	0.00	.13	VO	1	0.00	.13		5,000		
HQ	1	0.00	.13	PC	1	0.00	.13	VT	1	0.00	.13				

TABLE 7-B.—*The 18 digraphs composing 25% of the digraphs in Table 6-A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
AN	64	1.81	.89	ON	77	1.89	.92
				OR	64	1.81	.89
ED	60	1.78	.88	RE	98	1.99	.96
EN	111	2.05	.99				
ER	87	1.94	.94	SE	49	1.69	.84
ES	54	1.73	.86	ST	63	1.80	.88
				TE	71	1.85	.91
IN	75	1.88	.92	TH	78	1.89	.92
				TO	50	1.70	.84
ND	52	1.72	.85	VE	57	1.76	.87
NE	57	1.76	.87		1,249		
NT	82	1.91	.93				

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
AN	64	1.81	.89	ON	77	1.89	.92
				OR	64	1.81	.89
EN	111	2.05	.99	RE	98	1.99	.96
ER	87	1.94	.94				
ED	60	1.78	.88	ST	63	1.80	.88
ES	54	1.73	.86	SE	49	1.69	.84
				TH	78	1.89	.92
IN	75	1.88	.92	TE	71	1.85	.91
				TO	50	1.70	.84
NT	82	1.91	.93	VE	57	1.76	.87
NE	57	1.76	.87		1,249		
ND	52	1.72	.85				

TABLE 7–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
AL	32	1.51	.76	MA	36	1.56	.78
AN	64	1.81	.89				
AR	44	1.64	.82	ND	52	1.72	.85
AS	41	1.61	.80	NE	57	1.76	.87
AT	47	1.67	.83	NI	30	1.48	.75
				NT	82	1.91	.93
CE	32	1.51	.76				
CO	41	1.61	.80	ON	77	1.89	.92
				OR	64	1.81	.89
DA	32	1.51	.76	OU	37	1.57	.79
DE	33	1.52	.77				
				RA	39	1.59	.80
EA	35	1.54	.78	RE	98	1.99	.96
EC	32	1.51	.76	RI	30	1.48	.75
ED	60	1.78	.88	RO	28	1.45	.74
EE	42	1.62	.81	RS	31	1.49	.75
EL	29	1.46	.74	RT	42	1.62	.81
EN	111	2.05	.99				
ER	87	1.94	.94	SE	49	1.69	.84
ES	54	1.73	.86	SI	34	1.53	.77
ET	37	1.57	.79	ST	63	1.80	.88
FI	39	1.59	.80	TA	28	1.45	.74
FO	40	1.60	.80	TE	71	1.85	.91
				TH	78	1.89	.92
HI	33	1.52	.77	TI	45	1.65	.82
HT	28	1.45	.74	TO	50	1.70	.84
				TW	36	1.56	.78
IN	75	1.88	.92	TY	41	1.61	.80
IO	41	1.61	.80				
IS	35	1.54	.78	UR	31	1.49	.75
LA	28	1.45	.74	VE	57	1.76	.87
LE	37	1.57	.79		2,495		

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
AN	64	1.81	.89	MA	36	1.56	.78
AT	47	1.67	.83				
AR	44	1.64	.82	NT	82	1.91	.93
AS	41	1.61	.80	NE	57	1.76	.87
AL	32	1.51	.76	ND	52	1.72	.85
				NI	30	1.48	.75
CO	41	1.61	.80				
CE	32	1.51	.76	ON	77	1.89	.92
				OR	64	1.81	.89
DE	33	1.52	.77	OU	37	1.57	.79
DA	32	1.51	.76				
				RE	98	1.99	.96
EN	111	2.05	.99	RT	42	1.62	.81
ER	87	1.94	.94	RA	39	1.59	.80
ED	60	1.78	.88	RS	31	1.49	.75
ES	54	1.73	.86	RI	30	1.48	.75
EE	42	1.62	.81	RO	28	1.45	.74
ET	37	1.57	.79				
EA	35	1.54	.78	ST	63	1.80	.88
EC	32	1.51	.76	SE	49	1.69	.84
EL	29	1.46	.74	SI	34	1.53	.77
FO	40	1.60	.80	TH	78	1.89	.92
FI	39	1.59	.80	TE	71	1.85	.91
				TO	50	1.70	.84
HI	33	1.52	.77	TI	45	1.65	.82
HT	28	1.45	.74	TY	41	1.61	.80
				TW	36	1.56	.78
IN	75	1.88	.92	TA	28	1.45	.74
IO	41	1.61	.80				
IS	35	1.54	.78	UR	31	1.49	.75
LE	37	1.57	.79	VE	57	1.76	.87
LA	28	1.45	.74		2,495		

TABLE 7-D.—*The 122 digraphs composing 75% of the 5,000 digraphs of Table 6-A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

	F	$L_{10}(F)$	$L_{224}(2F)$
AC	14	1.15	.61
AD	27	1.43	.73
AI	17	1.23	.64
AL	32	1.51	.76
AM	14	1.15	.61
AN	64	1.81	.89
AR	44	1.64	.82
AS	41	1.61	.80
AT	47	1.67	.83
AU	13	1.11	.59
BE	18	1.26	.66
CA	20	1.30	.67
CE	32	1.51	.76
CH	14	1.15	.61
CO	41	1.61	.80
CT	14	1.15	.61
DA	32	1.51	.76
DE	33	1.52	.77
DI	27	1.43	.73
DO	16	1.20	.63
DS	13	1.11	.59
DT	15	1.18	.62
EA	35	1.54	.78
EC	32	1.51	.76
ED	60	1.78	.88
EE	42	1.62	.81
EF	18	1.26	.66
EI	27	1.43	.73
EL	29	1.46	.74
EM	14	1.15	.61
EN	111	2.05	.99
EP	20	1.30	.67
ER	87	1.94	.94
ES	54	1.73	.86
ET	37	1.57	.79
EV	20	1.30	.67
FI	39	1.59	.80
FO	40	1.60	.80
GE	14	1.15	.61
GH	20	1.30	.67
HA	20	1.30	.67
HE	20	1.30	.67
HI	33	1.52	.77
HO	20	1.30	.67
HR	17	1.23	.64
HT	28	1.45	.74
IC	22	1.34	.69
IE	13	1.11	.59
IG	19	1.28	.67
IL	23	1.36	.70
IN	75	1.88	.92
IO	41	1.61	.80
IR	27	1.43	.73
IS	35	1.54	.78
IT	27	1.43	.73
IV	25	1.40	.72
IX	15	1.18	.62
LA	28	1.45	.74
LE	37	1.57	.79
LI	20	1.30	.67
LL	27	1.43	.73
LO	13	1.11	.59
MA	36	1.56	.78
ME	26	1.41	.72
NA	26	1.41	.72
NC	19	1.28	.67
ND	52	1.72	.85
NE	57	1.76	.87
NG	27	1.43	.73
NI	30	1.48	.75
NO	18	1.26	.66
NS	24	1.38	.71
NT	82	1.91	.93
OF	25	1.40	.72
OL	19	1.28	.67
OM	25	1.40	.72
ON	77	1.89	.92
OP	25	1.40	.72
OR	64	1.81	.89
OS	14	1.15	.61
OT	19	1.28	.67
OU	37	1.57	.79
PA	14	1.15	.61
PE	23	1.36	.70
PO	17	1.23	.64
PR	18	1.26	.66
QU	15	1.18	.62
RA	39	1.59	.80
RD	17	1.23	.64
RE	98	1.99	.96
RI	30	1.48	.75
RO	28	1.45	.74
RS	31	1.49	.75
RT	42	1.62	.81
SA	24	1.38	.71
SE	49	1.69	.84
SH	26	1.41	.72
SI	34	1.53	.77
SO	15	1.18	.62
SS	19	1.28	.67
ST	63	1.80	.88
TA	28	1.45	.74
TE	71	1.85	.91
TH	78	1.89	.92
TI	45	1.65	.82
TO	50	1.70	.84
TR	17	1.23	.64
TS	19	1.28	.67
TT	19	1.28	.67
TW	36	1.56	.78
TY	41	1.61	.80
UN	21	1.32	.68
UR	31	1.49	.75
VE	57	1.76	.87
WE	22	1.34	.69
WO	19	1.28	.67
YT	15	1.18	.62
	3,745		

~~CONFIDENTIAL~~

TABLE 7–D, Concluded.—*The 122 digraphs composing 75% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their initial letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
AN	64	1.81	.89	EI	27	1.43	.73	MA	36	1.56	.78	RI	30	1.48	.75
AT	47	1.67	.83	EP	20	1.30	.67	ME	26	1.41	.72	RO	28	1.45	.74
AR	44	1.64	.82	EV	20	1.30	.67					RD	17	1.23	.64
AS	41	1.61	.80	EF	18	1.26	.66	NT	82	1.91	.93				
AL	32	1.51	.76	EM	14	1.15	.61	NE	57	1.76	.87				
AD	27	1.43	.73					ND	52	1.72	.85	ST	63	1.80	.88
AI	17	1.23	.64	FO	40	1.60	.80	NI	30	1.48	.75	SE	49	1.69	.84
AC	14	1.15	.61	FI	39	1.59	.80	NG	27	1.43	.73	SI	34	1.53	.77
AM	14	1.15	.61					NA	26	1.41	.72	SH	26	1.41	.72
AU	13	1.11	.59	GH	20	1.30	.67	NS	24	1.38	.71	SA	24	1.38	.71
				GE	14	1.15	.61	NC	19	1.28	.67	SS	19	1.28	.67
BE	18	1.26	.66					NO	18	1.26	.66	SO	15	1.18	.62
				HI	33	1.52	.77								
CO	41	1.61	.80	HT	28	1.45	.74					TH	78	1.89	.92
CE	32	1.51	.76	HA	20	1.30	.67	ON	77	1.89	.92	TE	71	1.85	.91
CA	20	1.30	.67	HE	20	1.30	.67	OR	64	1.81	.89	TO	50	1.70	.84
CH	14	1.15	.61	HO	20	1.30	.67	OU	37	1.57	.79	TI	45	1.65	.82
CT	14	1.15	.61	HR	17	1.23	.64	OF	25	1.40	.72	TY	41	1.61	.80
								OM	25	1.40	.72	TW	36	1.56	.78
				IN	75	1.88	.92	OP	25	1.40	.72	TA	28	1.45	.74
DE	33	1.52	.77	IO	41	1.61	.80	OL	19	1.28	.67	TS	19	1.28	.67
DA	32	1.51	.76	IS	35	1.54	.78	OT	19	1.28	.67	TT	19	1.28	.67
DI	27	1.43	.73	IR	27	1.43	.73	OS	14	1.15	.61	TR	17	1.23	.64
DO	16	1.20	.63	IT	27	1.43	.73								
DT	15	1.18	.62	IV	25	1.40	.72	PE	23	1.36	.70				
DS	13	1.11	.59	IL	23	1.36	.70	PR	18	1.26	.66	UR	31	1.49	.75
				IC	22	1.34	.69	PO	17	1.23	.64	UN	21	1.32	.68
EN	111	2.05	.99	IG	19	1.28	.67	PA	14	1.15	.61				
ER	87	1.94	.94	IX	15	1.18	.62					VE	57	1.76	.87
ED	60	1.78	.88	IE	13	1.11	.59								
ES	54	1.73	.86					QU	15	1.18	.62	WE	22	1.34	.69
EE	42	1.62	.81	LE	37	1.57	.79					WO	19	1.28	.67
ET	37	1.57	.79	LA	28	1.45	.74	RE	98	1.99	.96				
EA	35	1.54	.78	LL	27	1.43	.73	RT	42	1.62	.81	YT	15	1.18	.62
EC	32	1.51	.76	LI	20	1.30	.67	RA	39	1.59	.80				
EL	29	1.46	.74	LO	13	1.11	.59	RS	31	1.49	.75		3,745		

TABLE 7–E.—*All the 428 digraphs of Table 6–A, arranged first alphabetically according to their initial letters and then alphabetically according to their final letters.*

(SEE TABLE 6–A.—READ ACROSS THE ROWS)

TABLE 8.—*The 428 different digraphs of Table 6–A, arranged first alphabetically according to their initial letters and then according to their absolute frequencies under each initial letter,*[1] *accompanied by the logarithms of their assigned probabilities*

F		L_{10}(F)	L_{224} (2F)
AN	64	1.81	.89
AT	47	1.67	.83
AR	44	1.64	.82
AS	41	1.61	.80
AL	32	1.51	.76
AD	27	1.43	.73
AI	17	1.23	.64
AC	14	1.15	.61
AM	14	1.15	.61
AU	13	1.11	.59
AP	12	1.08	.58
AY	12	1.08	.58
AV	7	0.85	.48
AB	6	0.78	.45
AG	6	0.78	.45
AF	4	0.60	.38
AA	3	0.48	.33
AW	3	0.48	.33
AH	2	0.30	.25
AK	2	0.30	.25
AO	2	0.30	.25
AE	1	0.00	.13
AJ	1	0.00	.13
BE	18	1.26	.66
BY	7	0.85	.48
BL	6	0.78	.45
BA	4	0.60	.38
BO	4	0.60	.38
BI	2	0.30	.25
BR	2	0.30	.25
BU	2	0.30	.25
BJ	1	0.00	.13
BM	1	0.00	.13
BS	1	0.00	.13
BT	1	0.00	.13
CO	41	1.61	.80
CE	32	1.51	.76
CA	20	1.30	.67
CH	14	1.15	.61

F		L_{10}(F)	L_{224} (2F)
CT	14	1.15	.61
CI	7	0.85	.48
CL	5	0.70	.42
CK	4	0.60	.38
CR	4	0.60	.38
CU	4	0.60	.38
CC	3	0.48	.33
CD	1	0.00	.13
CF	1	0.00	.13
CM	1	0.00	.13
CN	1	0.00	.13
CS	1	0.00	.13
CW	1	0.00	.13
CY	1	0.00	.13
DE	33	1.52	.77
DA	32	1.51	.76
DI	27	1.43	.73
DO	16	1.20	.63
DT	15	1.18	.62
DS	13	1.11	.59
DR	12	1.08	.58
DD	8	0.90	.51
DF	8	0.90	.51
DM	5	0.70	.42
DP	5	0.70	.42
DU	5	0.70	.42
DB	4	0.60	.38
DC	4	0.60	.38
DN	4	0.60	.38
DW	4	0.60	.38
DL	3	0.48	.33
DV	3	0.48	.33
DG	2	0.30	.25
DH	2	0.30	.25
DQ	2	0.30	.25
DJ	1	0.00	.13
DY	1	0.00	.13
EN	111	2.05	.99
ER	87	1.94	.94

F		L_{10}(F)	L_{224} (2F)
ED	60	1.78	.88
ES	54	1.73	.86
EE	42	1.62	.81
ET	37	1.57	.79
EA	35	1.54	.78
EC	32	1.51	.76
EL	29	1.46	.74
EI	27	1.43	.73
EP	20	1.30	.67
EV	20	1.30	.67
EF	18	1.26	.66
EM	14	1.15	.61
EO	12	1.08	.58
EQ	12	1.08	.58
EH	7	0.85	.48
EW	7	0.85	.48
EX	7	0.85	.48
EB	4	0.60	.38
EG	4	0.60	.38
EY	4	0.60	.38
EU	3	0.48	.33
EJ	1	0.00	.13
EZ	1	0.00	.13
FO	40	1.60	.80
FI	39	1.59	.80
FF	11	1.04	.56
FT	11	1.04	.56
FE	10	1.00	.55
FR	9	0.95	.53
FA	5	0.70	.42
FS	3	0.48	.33
FU	3	0.48	.33
FC	2	0.30	.25
FL	2	0.30	.25
FD	1	0.00	.13
FG	1	0.00	.13
FM	1	0.00	.13
FP	1	0.00	.13
FW	1	0.00	.13
FY	1	0.00	.13

F		L_{10}(F)	L_{224} (2F)
GH	20	1.30	.67
GE	14	1.15	.61
GA	7	0.85	.48
GO	6	0.78	.45
GI	5	0.70	.42
GR	5	0.70	.42
GT	4	0.60	.38
GN	3	0.48	.33
GS	3	0.48	.33
GC	2	0.30	.25
GF	2	0.30	.25
GL	2	0.30	.25
GP	2	0.30	.25
GU	2	0.30	.25
GD	1	0.00	.13
GG	1	0.00	.13
GJ	1	0.00	.13
GM	1	0.00	.13
GW	1	0.00	.13
HI	33	1.52	.77
HT	28	1.45	.74
HA	20	1.30	.67
HE	20	1.30	.67
HO	20	1.30	.67
HR	17	1.23	.64
HU	8	0.90	.51
HF	5	0.70	.42
HS	4	0.60	.38
HC	3	0.48	.33
HN	3	0.48	.33
HD	2	0.30	.25
HM	2	0.30	.25
HB	1	0.00	.13
HL	1	0.00	.13
HP	1	0.00	.13
HQ	1	0.00	.13
HW	1	0.00	.13
HY	1	0.00	.13

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6–A.

TABLE 8, Continued.—*The 428 different digraphs of Table 6–A, arranged first alphabetically according to their initial letters and then according to their absolute frequencies under each initial letter,[1] accompanied by the logarithms of their assigned probabilities*

F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$
IN------	75	1.88	.92	LO------	13	1.11	.59	ND------	52	1.72	.85	OV------	7	0.85	.48
IO------	41	1.61	.80	LY------	10	1.00	.55	NI------	30	1.48	.75	OO------	6	0.78	.45
IS------	35	1.54	.78	LD------	9	0.95	.53	NG------	27	1.43	.73	OI------	5	0.70	.42
IR------	27	1.43	.73	LT------	8	0.90	.51	NA------	26	1.41	.72	OB------	4	0.60	.38
IT------	27	1.43	.73	LS------	6	0.78	.45	NS------	24	1.38	.71	OE------	3	0.48	.33
IV------	25	1.40	.72	LB------	3	0.48	.33	NC------	19	1.28	.67	OH------	3	0.48	.33
IL------	23	1.36	.70	LC------	3	0.48	.33	NO------	18	1.26	.66	OG------	2	0.30	.25
IC------	22	1.34	.69	LF------	3	0.48	.33	NF------	9	0.95	.53	OK------	2	0.30	.25
IG------	19	1.28	.67	LP------	3	0.48	.33	NN------	8	0.90	.51	OY------	2	0.30	.25
IX------	15	1.18	.62	LM------	2	0.30	.25	NU------	7	0.85	.48	OJ------	1	0.00	.13
IE------	13	1.11	.59	LR------	2	0.30	.25	NL------	5	0.70	.42	OX------	1	0.00	.13
IF------	10	1.00	.55	LU------	2	0.30	.25	NM------	5	0.70	.42				
IM------	9	0.95	.53	LV------	2	0.30	.25	NY------	5	0.70	.42	PE------	23	1.36	.70
IA------	8	0.90	.51	LW------	2	0.30	.25	NH------	4	0.60	.38	PR------	18	1.26	.66
IP------	7	0.85	.48	LG------	1	0.00	.13	NR------	4	0.60	.38	PO------	17	1.23	.64
ID------	6	0.78	.45	LH------	1	0.00	.13	NP------	3	0.48	.33	PA------	14	1.15	.61
IB------	2	0.30	.25	LN------	1	0.00	.13	NV------	3	0.48	.33	PL------	13	1.11	.59
IK------	2	0.30	.25					NW------	3	0.48	.33	PP------	11	1.04	.56
IZ------	2	0.30	.25	MA------	36	1.56	.78	NB------	2	0.30	.25	PT------	8	0.90	.51
				ME------	26	1.41	.72	NK------	2	0.30	.25	PI------	6	0.78	.45
JE------	2	0.30	.25	MM------	13	1.11	.59	NJ------	1	0.00	.13	PS------	6	0.78	.45
JO------	2	0.30	.25	MO------	10	1.00	.55	NQ------	1	0.00	.13	PM------	4	0.60	.38
JU------	2	0.30	.25	MI------	9	0.95	.53					PH------	3	0.48	.33
JA------	1	0.00	.13	MP------	8	0.90	.51					PU------	3	0.48	.33
				MB------	6	0.78	.45	ON------	77	1.89	.92	PF------	2	0.30	.25
KE------	6	0.78	.45	MS------	4	0.60	.38	OR------	64	1.81	.89	PB------	1	0.00	.13
KI------	2	0.30	.25	MC------	3	0.48	.33	OU------	37	1.57	.79	PC------	1	0.00	.13
KA------	1	0.00	.13	MR------	2	0.30	.25	OF------	25	1.40	.72	PD------	1	0.00	.13
KC------	1	0.00	.13	MT------	2	0.30	.25	OM------	25	1.40	.72	PN------	1	0.00	.13
KL------	1	0.00	.13	MU------	2	0.30	.25	OP------	25	1.40	.72	PV------	1	0.00	.13
KN------	1	0.00	.13	MY------	2	0.30	.25	OL------	19	1.28	.67	PW------	1	0.00	.13
KS------	1	0.00	.13	MD------	1	0.00	.13	OT------	19	1.28	.67	PY------	1	0.00	.13
				MF------	1	0.00	.13	OS------	14	1.15	.61				
LE------	37	1.57	.79	MH------	1	0.00	.13	OD------	12	1.08	.58				
LA------	28	1.45	.74					OC------	8	0.90	.51	QU------	15	1.18	.62
LL------	27	1.43	.73	NT------	82	1.91	.93	OW------	8	0.90	.51	QM------	1	0.00	.13
LI------	20	1.30	.67	NE------	57	1.76	.87	OA------	7	0.85	.48	QR------	1	0.00	.13

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6–A.

TABLE 8, Concluded.—*The 428 different digraphs of Table 6–A, arranged first alphabetically according to their initial letters and then according to their absolute frequencies under each initial letter,[1] accompanied by the logarithms of their assigned probabilities*

F	$L_{10}(F)$	$L_{234}(2F)$	F	$L_{10}(F)$	$L_{234}(2F)$	F	$L_{10}(F)$	$L_{234}(2F)$	F	$L_{10}(F)$	$L_{234}(2F)$
RE------ 98	1.99	.96	SR------ 5	0.70	.42	US------ 12	1.08	.58	XI------ 2	0.30	.25
RT------ 42	1.62	.81	SN------ 4	0.60	.38	UT------ 12	1.08	.58	XP------ 2	0.30	.25
RA------ 39	1.59	.80	SW------ 4	0.60	.38	UE------ 11	1.04	.56	XD------ 1	0.00	.13
RS------ 31	1.49	.75	SB------ 3	0.48	.33	UG------ 8	0.90	.51	XE------ 1	0.00	.13
RI------ 30	1.48	.75	SM------ 3	0.48	.33	UL------ 6	0.78	.45	XF------ 1	0.00	.13
RO------ 28	1.45	.74	SG------ 2	0.30	.25	UA------ 5	0.70	.42	XH------ 1	0.00	.13
RD------ 17	1.23	.64	SL------ 2	0.30	.25	UI------ 5	0.70	.42	XN------ 1	0.00	.13
RP------ 13	1.11	.59	SK------ 1	0.00	.13	UM------ 5	0.70	.42	XO------ 1	0.00	.13
RR------ 11	1.04	.56	SV------ 1	0.00	.13	UB------ 3	0.48	.33	XR------ 1	0.00	.13
RC------ 9	0.95	.53	SY------ 1	0.00	.13	UC------ 3	0.48	.33	XS------ 1	0.00	.13
RM------ 9	0.95	.53				UD------ 3	0.48	.33			
RY------ 9	0.95	.53	TH------ 78	1.89	.92	UP------ 2	0.30	.25	YT------ 15	1.18	.62
RG------ 7	0.85	.48	TE------ 71	1.85	.91	UF------ 1	0.00	.13	YF------ 11	1.04	.56
RN------ 7	0.85	.48	TO------ 50	1.70	.84	UO------ 1	0.00	.13	YS------ 11	1.04	.56
RF------ 6	0.78	.45	TI------ 45	1.65	.82	UV------ 1	0.00	.13	YO------ 10	1.00	.55
RL------ 5	0.70	.42	TY------ 41	1.61	.80				YE------ 9	0.95	.53
RU------ 5	0.70	.42	TW------ 36	1.56	.78	VE------ 57	1.76	.87	YA------ 6	0.78	.45
RV------ 5	0.70	.42	TA------ 28	1.45	.74	VI------ 12	1.08	.58	YN------ 6	0.78	.45
RW------ 4	0.60	.38	TS------ 19	1.28	.67	VA------ 6	0.78	.45	YC------ 4	0.60	.38
RH------ 3	0.48	.33	TT------ 19	1.28	.67	VO------ 1	0.00	.13	YD------ 4	0.60	.38
RB------ 2	0.30	.25	TR------ 17	1.23	.64	VT------ 1	0.00	.13	YR------ 4	0.60	.38
RJ------ 1	0.00	.13	TF------ 7	0.85	.48				YI------ 3	0.48	.33
RK------ 1	0.00	.13	TN------ 7	0.85	.48	WE------ 22	1.34	.69	YP------ 3	0.48	.33
			TC------ 6	0.78	.45	WO------ 19	1.28	.67	YB------ 2	0.30	.25
ST------ 63	1.80	.88	TD------ 6	0.78	.45	WI------ 13	1.11	.59	YL------ 2	0.30	.25
SE------ 49	1.69	.84	TM------ 6	0.78	.45	WA------ 12	1.08	.58	YM------ 2	0.30	.25
SI------ 34	1.53	.77	TL------ 5	0.70	.42	WH------ 4	0.60	.38	YG------ 1	0.00	.13
SH------ 26	1.41	.72	TU------ 5	0.70	.42	WN------ 2	0.30	.25	YH------ 1	0.00	.13
SA------ 24	1.38	.71	TB------ 3	0.48	.33	WL------ 1	0.00	.13	YU------ 1	0.00	.13
SS------ 19	1.28	.67	TP------ 2	0.30	.25	WR------ 1	0.00	.13	YW------ 1	0.00	.13
SO------ 15	1.18	.62	TG------ 1	0.00	.13	WS------ 1	0.00	.13			
SC------ 13	1.11	.59	TQ------ 1	0.00	.13	WY------ 1	0.00	.13	ZE------ 2	0.30	.25
SF------ 12	1.08	.58	TZ------ 1	0.00	.13				ZA------ 1	0.00	.13
SU------ 11	1.04	.56				XT------ 7	0.85	.48	ZI------ 1	0.00	.13
SP------ 10	1.00	.55	UR------ 31	1.49	.75	XA------ 2	0.30	.25	5,000		
SD------ 5	0.70	.42	UN------ 21	1.32	.68	XC------ 2	0.30	.25			

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6–A.

TABLE 9–A.—*The 428 different digraphs of Table 6–A, arranged first alphabetically according to their final letters and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

F	$L_{10}(F)$	$L_{224}(2F)$	F	$L_{10}(F)$	$L_{224}(2F)$	F	$L_{10}(F)$	$L_{224}(2F)$	F	$L_{10}(F)$	$L_{224}(2F)$
RA...... 39	1.59	.80	EC...... 32	1.51	.76	RE...... 98	1.99	.96	GF...... 2	0.30	.25
MA...... 36	1.56	.78	IC...... 22	1.34	.69	TE...... 71	1.85	.91	PF...... 2	0.30	.25
EA...... 35	1.54	.78	NC...... 19	1.28	.67	NE...... 57	1.76	.87	CF...... 1	0.00	.13
DA...... 32	1.51	.76	AC...... 14	1.15	.61	VE...... 57	1.76	.87	MF...... 1	0.00	.13
LA...... 28	1.45	.74	SC...... 13	1.11	.59	SE...... 49	1.69	.84	UF...... 1	0.00	.13
TA...... 28	1.45	.74	RC...... 9	0.95	.53	EE...... 42	1.62	.81	XF...... 1	0.00	.13
NA...... 26	1.41	.72	OC...... 8	0.90	.51	LE...... 37	1.57	.79			
SA...... 24	1.38	.71	TC...... 6	0.78	.45	DE...... 33	1.52	.77			
CA...... 20	1.30	.67	DC...... 4	0.60	.38	CE...... 32	1.51	.76	NG...... 27	1.43	.73
HA...... 20	1.30	.67	YC...... 4	0.60	.38	ME...... 26	1.41	.72	IG...... 19	1.28	.67
PA...... 14	1.15	.61	CC...... 3	0.48	.33	PE...... 23	1.36	.70	UG...... 8	0.90	.51
WA...... 12	1.08	.58	HC...... 3	0.48	.33	WE...... 22	1.34	.69	RG...... 7	0.85	.48
IA...... 8	0.90	.51	LC...... 3	0.48	.33	HE...... 20	1.30	.67	AG...... 6	0.78	.45
GA...... 7	0.85	.48	MC...... 3	0.48	.33	BE...... 18	1.26	.66	EG...... 4	0.60	.38
OA...... 7	0.85	.48	UC...... 3	0.48	.33	GE...... 14	1.15	.61	DG...... 2	0.30	.25
VA...... 6	0.78	.45	FC...... 2	0.30	.25	IE...... 13	1.11	.59	OG...... 2	0.30	.25
YA...... 6	0.78	.45	GC...... 2	0.30	.25	UE...... 11	1.04	.56	SG...... 2	0.30	.25
FA...... 5	0.70	.42	XC...... 2	0.30	.25	FE...... 10	1.00	.55	FG...... 1	0.00	.13
UA...... 5	0.70	.42	KC...... 1	0.00	.13	YE...... 9	0.95	.53	GG...... 1	0.00	.13
BA...... 4	0.60	.38	PC...... 1	0.00	.13	KE...... 6	0.78	.45	LG...... 1	0.00	.13
AA...... 3	0.48	.33				OE...... 3	0.48	.33	TG...... 1	0.00	.13
XA...... 2	0.30	.25				JE...... 2	0.30	.25	YG...... 1	0.00	.13
JA...... 1	0.00	.13	ED...... 60	1.78	.88	ZE...... 2	0.30	.25			
KA...... 1	0.00	.13	ND...... 52	1.72	.85	AE...... 1	0.00	.13			
ZA...... 1	0.00	.13	AD...... 27	1.43	.73	XE...... 1	0.00	.13			
			RD...... 17	1.23	.64				TH...... 78	1.89	.92
AB...... 6	0.78	.45	OD...... 12	1.08	.58				SH...... 26	1.41	.72
MB...... 6	0.78	.45	LD...... 9	0.95	.53				GH...... 20	1.30	.67
DB...... 4	0.60	.38	DD...... 8	0.90	.51	OF...... 25	1.40	.72	CH...... 14	1.15	.61
EB...... 4	0.60	.38	ID...... 6	0.78	.45	EF...... 18	1.26	.66	EH...... 7	0.85	.48
OB...... 4	0.60	.38	TD...... 6	0.78	.45	SF...... 12	1.08	.58	NH...... 4	0.60	.38
LB...... 3	0.48	.33	SD...... 5	0.70	.42	FF...... 11	1.04	.56	WH...... 4	0.60	.38
SB...... 3	0.48	.33	YD...... 4	0.60	.38	YF...... 11	1.04	.56	OH...... 3	0.48	.33
TB...... 3	0.48	.33	UD...... 3	0.48	.33	IF...... 10	1.00	.55	PH...... 3	0.48	.33
UB...... 3	0.48	.33	HD...... 2	0.30	.25	NF...... 9	0.95	.53	RH...... 3	0.48	.33
IB...... 2	0.30	.25	CD...... 1	0.00	.13	DF...... 8	0.90	.51	AH...... 2	0.30	.25
NB...... 2	0.30	.25	FD...... 1	0.00	.13	TF...... 7	0.85	.48	DH...... 2	0.30	.25
RB...... 2	0.30	.25	GD...... 1	0.00	.13	RF...... 6	0.78	.45	LH...... 1	0.00	.13
YB...... 2	0.30	.25	MD...... 1	0.00	.13	HF...... 5	0.70	.42	MH...... 1	0.00	.13
HB...... 1	0.00	.13	PD...... 1	0.00	.13	AF...... 4	0.60	.38	XH...... 1	0.00	.13
PB...... 1	0.00	.13	XD...... 1	0.00	.13	LF...... 3	0.48	.33	YH...... 1	0.00	.13

TABLE 9-A, Continued.—*The 428 different digraphs of Table 6-A, arranged first alphabetically according to their final letters and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

F		$L_{10}(F)$	$L_{224}(2F)$
TI	45	1.65	.82
FI	39	1.59	.80
SI	34	1.53	.77
HI	33	1.52	.77
NI	30	1.48	.75
RI	30	1.48	.75
DI	27	1.43	.73
EI	27	1.43	.73
LI	20	1.30	.67
AI	17	1.23	.64
WI	13	1.11	.59
VI	12	1.08	.58
MI	9	0.95	.53
CI	7	0.85	.48
PI	6	0.78	.45
GI	5	0.70	.42
OI	5	0.70	.42
UI	5	0.70	.42
YI	3	0.48	.33
BI	2	0.30	.25
KI	2	0.30	.25
XI	2	0.30	.25
ZI	1	0.00	.13
AJ	1	0.00	.13
BJ	1	0.00	.13
DJ	1	0.00	.13
EJ	1	0.00	.13
GJ	1	0.00	.13
NJ	1	0.00	.13
OJ	1	0.00	.13
RJ	1	0.00	.13
CK	4	0.60	.38
AK	2	0.30	.25
IK	2	0.30	.25
NK	2	0.30	.25
OK	2	0.30	.25
RK	1	0.00	.13
SK	1	0.00	.13
AL	32	1.51	.76
EL	29	1.46	.74

F		$L_{10}(F)$	$L_{224}(2F)$
LL	27	1.43	.73
IL	23	1.36	.70
OL	19	1.28	.67
PL	13	1.11	.59
BL	6	0.78	.45
UL	6	0.78	.45
CL	5	0.70	.42
NL	5	0.70	.42
RL	5	0.70	.42
TL	5	0.70	.42
DL	3	0.48	.33
FL	2	0.30	.25
GL	2	0.30	.25
SL	2	0.30	.25
YL	2	0.30	.25
HL	1	0.00	.13
KL	1	0.00	.13
WL	1	0.00	.13
OM	25	1.40	.72
AM	14	1.15	.61
EM	14	1.15	.61
MM	13	1.11	.59
IM	9	0.95	.53
RM	9	0.95	.53
TM	6	0.78	.45
DM	5	0.70	.42
NM	5	0.70	.42
UM	5	0.70	.42
PM	4	0.60	.38
SM	3	0.48	.33
HM	2	0.30	.25
LM	2	0.30	.25
YM	2	0.30	.25
BM	1	0.00	.13
CM	1	0.00	.13
FM	1	0.00	.13
GM	1	0.00	.13
QM	1	0.00	.13
EN	111	2.05	.99
ON	77	1.89	.92
IN	75	1.88	.92

F		$L_{10}(F)$	$L_{224}(2F)$
AN	64	1.81	.89
UN	21	1.32	.68
NN	8	0.90	.51
RN	7	0.85	.48
TN	7	0.85	.48
YN	6	0.78	.45
DN	4	0.60	.38
SN	4	0.60	.38
GN	3	0.48	.33
HN	3	0.48	.33
WN	2	0.30	.25
CN	1	0.00	.13
KN	1	0.00	.13
LN	1	0.00	.13
PN	1	0.00	.13
XN	1	0.00	.13
TO	50	1.70	.84
CO	41	1.61	.80
IO	41	1.61	.80
FO	40	1.60	.80
RO	28	1.45	.74
HO	20	1.30	.67
WO	19	1.28	.67
NO	18	1.26	.66
PO	17	1.23	.64
DO	16	1.20	.63
SO	15	1.18	.62
LO	13	1.11	.59
EO	12	1.08	.58
MO	10	1.00	.55
YO	10	1.00	.55
GO	6	0.78	.45
OO	6	0.78	.45
BO	4	0.60	.38
AO	2	0.30	.25
JO	2	0.30	.25
UO	1	0.00	.13
VO	1	0.00	.13
XO	1	0.00	.13
OP	25	1.40	.72
EP	20	1.30	.67

F		$L_{10}(F)$	$L_{224}(2F)$
RP	13	1.11	.59
AP	12	1.08	.58
PP	11	1.04	.56
SP	10	1.00	.55
MP	8	0.90	.51
IP	7	0.85	.48
DP	5	0.70	.42
LP	3	0.48	.33
NP	3	0.48	.33
YP	3	0.48	.33
GP	2	0.30	.25
TP	2	0.30	.25
UP	2	0.30	.25
XP	2	0.30	.25
FP	1	0.00	.13
HP	1	0.00	.13
EQ	12	1.08	.58
DQ	2	0.30	.25
HQ	1	0.00	.13
NQ	1	0.00	.13
TQ	1	0.00	.13
ER	87	1.94	.94
OR	64	1.81	.89
AR	44	1.64	.82
UR	31	1.49	.75
IR	27	1.43	.73
PR	18	1.26	.66
HR	17	1.23	.64
TR	17	1.23	.64
DR	12	1.08	.58
RR	11	1.04	.56
FR	9	0.95	.53
GR	5	0.70	.42
SR	5	0.70	.42
CR	4	0.60	.38
NR	4	0.60	.38
YR	4	0.60	.38
BR	2	0.30	.25
LR	2	0.30	.25
MR	2	0.30	.25
QR	1	0.00	.13
WR	1	0.00	.13
XR	1	0.00	.13

TABLE 9–A, Concluded.—*The 428 different digraphs of Table 6-A, arranged first alphabetically according to their final letters and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	L_{224} (2F)		F	$L_{10}(F)$	L_{224} (2F)		F	$L_{10}(F)$	L_{224} (2F)		F	$L_{10}(F)$	L_{224} (2F)
ES	54	1.73	.86	OT	19	1.28	.67	JU	2	0.30	.25	PW	1	0.00	.13
AS	41	1.61	.80	TT	19	1.28	.67	LU	2	0.30	.25	YW	1	0.00	.13
IS	35	1.54	.78	DT	15	1.18	.62	MU	2	0.30	.25				
RS	31	1.49	.75	YT	15	1.18	.62	YU	1	0.00	.13	IX	15	1.18	.62
NS	24	1.38	.71	CT	14	1.15	.61					EX	7	0.85	.48
SS	19	1.28	.67	UT	12	1.08	.58	IV	25	1.40	.72	OX	1	0.00	.13
TS	19	1.28	.67	FT	11	1.04	.56	EV	20	1.30	.67				
OS	14	1.15	.61	LT	8	0.90	.51	AV	7	0.85	.48	TY	41	1.61	.80
DS	13	1.11	.59	PT	8	0.90	.51	OV	7	0.85	.48	AY	12	1.08	.58
US	12	1.08	.58	XT	7	0.85	.48	RV	5	0.70	.42	LY	10	1.00	.55
YS	11	1.04	.56	GT	4	0.60	.38	DV	3	0.48	.33	RY	9	0.95	.53
LS	6	0.78	.45	MT	2	0.30	.25	NV	3	0.48	.33	BY	7	0.85	.48
PS	6	0.78	.45	BT	1	0.00	.13	LV	2	0.30	.25	NY	5	0.70	.42
HS	4	0.60	.38	VT	1	0.00	.13	PV	1	0.00	.13	EY	4	0.60	.38
MS	4	0.60	.38					SV	1	0.00	.13	MY	2	0.30	.25
FS	3	0.48	.33	OU	37	1.57	.79	UV	1	0.00	.13	OY	2	0.30	.25
GS	3	0.48	.33	QU	15	1.18	.62					CY	1	0.00	.13
BS	1	0.00	.13	AU	13	1.11	.59	TW	36	1.56	.78	DY	1	0.00	.13
CS	1	0.00	.13	SU	11	1.04	.56	OW	8	0.90	.51	FY	1	0.00	.13
KS	1	0.00	.13	HU	8	0.90	.51	EW	7	0.85	.48	HY	1	0.00	.13
WS	1	0.00	.13	NU	7	0.85	.48	DW	4	0.60	.38	PY	1	0.00	.13
XS	1	0.00	.13	DU	5	0.70	.42	RW	4	0.60	.38	SY	1	0.00	.13
				RU	5	0.70	.42	SW	4	0.60	.38	WY	1	0.00	.13
NT	82	1.91	.93	TU	5	0.70	.42	AW	3	0.48	.33				
ST	63	1.80	.88	CU	4	0.60	.38	NW	3	0.48	.33	IZ	2	0.30	.25
AT	47	1.67	.83	EU	3	0.48	.33	LW	2	0.30	.25	EZ	1	0.00	.13
RT	42	1.62	.81	FU	3	0.48	.33	CW	1	0.00	.13	TZ	1	0.00	.13
ET	37	1.57	.79	PU	3	0.48	.33	FW	1	0.00	.13		5,000		
HT	28	1.45	.74	BU	2	0.30	.25	GW	1	0.00	.13				
IT	27	1.43	.73	GU	2	0.30	.25	HW	1	0.00	.13				

TABLE 9–B.—*The 18 digraphs composing 25% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$
ED	60	1.78	.88	IN	75	1.88	.92
ND	52	1.72	.85	ON	77	1.89	.92
NE	57	1.76	.87	TO	50	1.70	.84
RE	98	1.99	.96				
SE	49	1.69	.84	ER	87	1.94	.94
TE	71	1.85	.91	OR	64	1.81	.89
VE	57	1.76	.87				
				ES	54	1.73	.86
TH	78	1.89	.92				
				NT	82	1.91	.93
AN	64	1.81	.89	ST	63	1.80	.88
EN	111	2.05	.99		1,249		

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$
ED	60	1.78	.88	IN	75	1.88	.92
ND	52	1.72	.85	AN	64	1.81	.89
RE	98	1.99	.96	TO	50	1.70	.84
TE	71	1.85	.91				
NE	57	1.76	.87	ER	87	1.94	.94
VE	57	1.76	.87	OR	64	1.81	.89
SE	49	1.69	.84				
				ES	54	1.73	.86
TH	78	1.89	.92				
				NT	82	1.91	.93
EN	111	2.05	.99	ST	63	1.80	.88
ON	77	1.89	.92		1,249		

TABLE 9–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$	F		$L_{10}(F)$	$L_{224}(2F)$
DA	32	1.51	.76	NE	57	1.76	.87	AN	64	1.81	.89	AS	41	1.61	.80
EA	35	1.54	.78	RE	98	1.99	.96	EN	111	2.05	.99	ES	54	1.73	.86
LA	28	1.45	.74	SE	49	1.69	.84	IN	75	1.88	.92	IS	35	1.54	.78
MA	36	1.56	.78	TE	71	1.85	.91	ON	77	1.89	.92	RS	31	1.49	.75
RA	39	1.59	.80	VE	57	1.76	.87								
TA	28	1.45	.74									AT	47	1.67	.83
				TH	78	1.89	.92	CO	41	1.61	.80	ET	37	1.57	.79
EC	32	1.51	.76					FO	40	1.60	.80	HT	28	1.45	.74
				FI	39	1.59	.80	IO	41	1.61	.80	NT	82	1.91	.93
				HI	33	1.52	.77	RO	28	1.45	.74	RT	42	1.62	.81
ED	60	1.78	.88	NI	30	1.48	.75	TO	50	1.70	.84	ST	63	1.80	.88
ND	52	1.72	.85	RI	30	1.48	.75								
				SI	34	1.53	.77					OU	37	1.57	.79
				TI	45	1.65	.82								
CE	32	1.51	.76					AR	44	1.64	.82	TW	36	1.56	.78
DE	33	1.52	.77					ER	87	1.94	.94				
EE	42	1.62	.81	AL	32	1.51	.76	OR	64	1.81	.89	TY	41	1.61	.80
LE	37	1.57	.79	EL	29	1.46	.74	UR	31	1.49	.75		2,495		

TABLE 9–C, Concluded.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

	F	$L_{10}(F)$	$L_{234}(2F)$
RA	39	1.59	.80
MA	36	1.56	.78
EA	35	1.54	.78
DA	32	1.51	.76
LA	28	1.45	.74
TA	28	1.45	.74
EC	32	1.51	.76
ED	60	1.78	.88
ND	52	1.72	.85
RE	98	1.99	.96
TE	71	1.85	.91
NE	57	1.76	.87
VE	57	1.76	.87
SE	49	1.69	.84

	F	$L_{10}(F)$	$L_{234}(2F)$
EE	42	1.62	.81
LE	37	1.57	.79
DE	33	1.52	.77
CE	32	1.51	.76
TH	78	1.89	.92
TI	45	1.65	.82
FI	39	1.59	.80
SI	34	1.53	.77
HI	33	1.52	.77
NI	30	1.48	.75
RI	30	1.48	.75
AL	32	1.51	.76
EL	29	1.46	.74

	F	$L_{10}(F)$	$L_{234}(2F)$
EN	111	2.05	.99
ON	77	1.89	.92
IN	75	1.88	.92
AN	64	1.81	.89
TO	50	1.70	.84
CO	41	1.61	.80
IO	41	1.61	.80
FO	40	1.60	.80
RO	28	1.45	.74
ER	87	1.94	.94
OR	64	1.81	.89
AR	44	1.64	.82
UR	31	1.49	.75

	F	$L_{10}(F)$	$L_{234}(2F)$
ES	54	1.73	.86
AS	41	1.61	.80
IS	35	1.54	.78
RS	31	1.49	.75
NT	82	1.91	.93
ST	63	1.80	.88
AT	47	1.67	.83
RT	42	1.62	.81
ET	37	1.57	.79
HT	28	1.45	.74
OU	37	1.57	.79
TW	36	1.56	.78
TY	41	1.61	.80
	2,495		

TABLE 9–D.—*The 122 digraphs composing 75% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

	F	$L_{10}(F)$	$L_{234}(2F)$
CA	20	1.30	.67
DA	32	1.51	.76
EA	35	1.54	.78
HA	20	1.30	.67
LA	28	1.45	.74
MA	36	1.56	.78
NA	26	1.41	.72
PA	14	1.15	.61
RA	39	1.59	.80
SA	24	1.38	.71
TA	28	1.45	.74
AC	14	1.15	.61
EC	32	1.51	.76
IC	22	1.34	.69
NC	19	1.28	.67
AD	27	1.43	.73
ED	60	1.78	.88

	F	$L_{10}(F)$	$L_{234}(2F)$
ND	52	1.72	.85
RD	17	1.23	.64
BE	18	1.26	.66
CE	32	1.51	.76
DE	33	1.52	.77
EE	42	1.62	.81
GE	14	1.15	.61
HE	20	1.30	.67
IE	13	1.11	.59
LE	37	1.57	.79
ME	26	1.41	.72
NE	57	1.76	.87
PE	23	1.36	.70
RE	98	1.99	.96
SE	49	1.69	.84
TE	71	1.85	.91
VE	57	1.76	.87
WE	22	1.34	.69

	F	$L_{10}(F)$	$L_{234}(2F)$
EF	18	1.26	.66
OF	25	1.40	.72
IG	19	1.28	.67
NG	27	1.43	.73
CH	14	1.15	.61
GH	20	1.30	.67
SH	26	1.41	.72
TH	78	1.89	.92
AI	17	1.23	.64
DI	27	1.43	.73
EI	27	1.43	.73
FI	39	1.59	.80
HI	33	1.52	.77
LI	20	1.30	.67
NI	30	1.48	.75
RI	30	1.48	.75

	F	$L_{10}(F)$	$L_{234}(2F)$
SI	34	1.53	.77
TI	45	1.65	.82
AL	32	1.51	.76
EL	29	1.46	.74
IL	23	1.36	.70
LL	27	1.43	.73
OL	19	1.28	.67
AM	14	1.15	.61
EM	14	1.15	.61
OM	25	1.40	.72
AN	64	1.81	.89
EN	111	2.05	.99
IN	75	1.88	.92
ON	77	1.89	.92
UN	21	1.32	.68

TABLE 9–D, Continued.—*The 122 digraphs composing 75% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(1) AND ACCORDING TO THEIR INITIAL LETTERS—Concluded

	F	$L_{10}(F)$	$L_{224}(2F)$
CO	41	1.61	.80
DO	16	1.20	.63
FO	40	1.60	.80
HO	20	1.30	.67
IO	41	1.61	.80
LO	13	1.11	.59
NO	18	1.26	.66
PO	17	1.23	.64
RO	28	1.45	.74
SO	15	1.18	.62
TO	50	1.70	.84
WO	19	1.28	.67
EP	20	1.30	.67
OP	25	1.40	.72

	F	$L_{10}(F)$	$L_{224}(2F)$
AR	44	1.64	.82
ER	87	1.94	.94
HR	17	1.23	.64
IR	27	1.43	.73
OR	64	1.81	.89
PR	18	1.26	.66
TR	17	1.23	.64
UR	31	1.49	.75
AS	41	1.61	.80
DS	13	1.11	.59
ES	54	1.73	.86
IS	35	1.54	.78
NS	24	1.38	.71
OS	14	1.15	.61

	F	$L_{10}(F)$	$L_{224}(2F)$
RS	31	1.49	.75
SS	19	1.28	.67
TS	19	1.28	.67
AT	47	1.67	.83
CT	14	1.15	.61
DT	15	1.18	.62
ET	37	1.57	.79
HT	28	1.45	.74
IT	27	1.43	.73
NT	82	1.91	.93
OT	19	1.28	.67
RT	42	1.62	.81
ST	63	1.80	.88

	F	$L_{10}(F)$	$L_{224}(2F)$
TT	19	1.28	.67
YT	15	1.18	.62
AU	13	1.11	.59
OU	37	1.57	.79
QU	15	1.18	.62
EV	20	1.30	.67
IV	25	1.40	.72
TW	36	1.56	.78
IX	15	1.18	.62
TY	41	1.61	.80
	3,745		

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

	F	$L_{10}(F)$	$L_{224}(2F)$
RA	39	1.59	.80
MA	36	1.56	.78
EA	35	1.54	.78
DA	32	1.51	.76
LA	28	1.45	.74
TA	28	1.45	.74
NA	26	1.41	.72
SA	24	1.38	.71
CA	20	1.30	.67
HA	20	1.30	.67
PA	14	1.15	.61
EC	32	1.51	.76
IC	22	1.34	.69
NC	19	1.28	.67
AC	14	1.15	.61
ED	60	1.78	.88
ND	52	1.72	.85
AD	27	1.43	.73
RD	17	1.23	.64

	F	$L_{10}(F)$	$L_{224}(2F)$
RE	98	1.99	.96
TE	71	1.85	.91
NE	57	1.76	.87
VE	57	1.76	.87
SE	49	1.69	.84
EE	42	1.62	.81
LE	37	1.57	.79
DE	33	1.52	.77
CE	32	1.51	.76
ME	26	1.41	.72
PE	23	1.36	.70
WE	22	1.34	.69
HE	20	1.30	.67
BE	18	1.26	.66
GE	14	1.15	.61
IE	13	1.11	.59
OF	25	1.40	.72
EF	18	1.26	.66
NG	27	1.43	.73
IG	19	1.28	.67

	F	$L_{10}(F)$	$L_{224}(2F)$
TH	78	1.89	.92
SH	26	1.41	.72
GH	20	1.30	.67
CH	14	1.15	.61
TI	45	1.65	.82
FI	39	1.59	.80
SI	34	1.53	.77
HI	33	1.52	.77
NI	30	1.48	.75
RI	30	1.48	.75
DI	27	1.43	.73
EI	27	1.43	.73
LI	20	1.30	.67
AI	17	1.23	.64
AL	32	1.51	.76
EL	29	1.46	.74
LL	27	1.43	.73
IL	23	1.36	.70
OL	19	1.28	.67

	F	$L_{10}(F)$	$L_{224}(2F)$
OM	25	1.40	.72
AM	14	1.15	.61
EM	14	1.15	.61
EN	111	2.05	.99
ON	77	1.89	.92
IN	75	1.88	.92
AN	64	1.81	.89
UN	21	1.32	.68
TO	50	1.70	.84
CO	41	1.61	.80
IO	41	1.61	.80
FO	40	1.60	.80
RO	28	1.45	.74
HO	20	1.30	.67
WO	19	1.28	.67
NO	18	1.26	.66
PO	17	1.23	.64
DO	16	1.20	.63
SO	15	1.18	.62
LO	13	1.11	.59

~~CONFIDENTIAL~~

TABLE 9–D, Concluded.—*The 122 digraphs composing 75% of the 5,000 digraphs of Table 6–A, accompanied by the logarithms of their assigned probabilities, arranged alphabetically according to their final letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES—Concluded

	F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$		F	$L_{10}(F)$	$L_{224}(2F)$
OP	25	1.40	.72	ES	54	1.73	.86	NT	82	1.91	.93	OU	37	1.57	.79
EP	20	1.30	.67	AS	41	1.61	.80	ST	63	1.80	.88	QU	15	1.18	.62
				IS	35	1.54	.78	AT	47	1.67	.83	AU	13	1.11	.59
				RS	31	1.49	.75	RT	42	1.62	.81				
ER	87	1.94	.94	NS	24	1.38	.71	ET	37	1.57	.79	IV	25	1.40	.72
OR	64	1.81	.89	SS	19	1.28	.67	HT	28	1.45	.74	EV	20	1.30	.67
AR	44	1.64	.82	TS	19	1.28	.67	IT	27	1.43	.73	TW	36	1.56	.78
UR	31	1.49	.75	OS	14	1.15	.61	OT	19	1.28	.67				
IR	27	1.43	.73	DS	13	1.11	.59	TT	19	1.28	.67	IX	15	1.18	.62
PR	18	1.26	.66					DT	15	1.18	.62	TY	41	1.61	.80
HR	17	1.23	.64					YT	15	1.18	.62		3,745		
TR	17	1.23	.64					CT	14	1.15	.61				

TABLE 9–E.—*All the 428 different digraphs of Table 6–A, arranged alphabetically first according to their final letters and then according to their initial letters*

(SEE TABLE 6–A.—READ DOWN THE COLUMNS)

TABLE 10–A.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{586}(F)$		F	$L_{10}(F)$	$L_{586}(F)$		F	$L_{10}(F)$	$L_{586}(F)$
ENT	569	2.76	.99	TOP	174	2.24	.82	EIG	135	2.13	.79
ION	260	2.41	.88	NTH	171	2.23	.82	FIV	135	2.13	.79
AND	228	2.36	.86	TWE	170	2.23	.82	MEN	131	2.12	.78
ING	226	2.35	.86	TWO	163	2.21	.81	SEV	131	2.12	.78
IVE	225	2.35	.86	ATI	160	2.20	.81	ERS	126	2.10	.78
TIO	221	2.34	.85	THR	158	2.20	.81	UND	125	2.10	.78
FOR	218	2.34	.85	NTY	157	2.20	.81	NET	118	2.07	.77
OUR	211	2.32	.85	HRE	153	2.18	.80	PER	115	2.06	.76
THI	211	2.32	.85	WEN	153	2.18	.80	STA	115	2.06	.76
ONE	210	2.32	.85	FOU	152	2.18	.80	TER	115	2.06	.76
NIN	207	2.32	.85	ORT	146	2.16	.80	EQU	114	2.06	.76
STO	202	2.31	.84	REE	146	2.16	.80	RED	113	2.05	.76
EEN	196	2.29	.84	SIX	146	2.16	.80	TED	112	2.05	.76
GHT	196	2.29	.84	ASH	143	2.16	.80	ERI	109	2.04	.76
INE	192	2.28	.83	DAS	140	2.15	.79	HIR	106	2.03	.75
VEN	190	2.28	.83	IGH	140	2.15	.79	IRT	105	2.02	.75
EVE	177	2.25	.82	ERE	138	2.14	.79	DER	101	2.00	.74
EST	176	2.25	.82	COM	136	2.13	.79	DRE	100	2.00	.74
TEE	174	2.24	.82	ATE	135	2.13	.79				

~~CONFIDENTIAL~~

TABLE 10–B.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	L_{10} (F)	L_{556} (F)		F	L_{10} (F)	L_{556} (F)		F	L_{10} (F)	L_{556} (F)
AND	228	2.36	.86	GHT	196	2.29	.84	REE	146	2.16	.80
ATI	160	2.20	.81					RED	113	2.05	.76
ASH	143	2.16	.80	HRE	153	2.18	.80				
ATE	135	2.13	.79	HIR	106	2.03	.75	STO	202	2.31	.84
								SIX	146	2.16	.80
COM	136	2.13	.79	ION	260	2.41	.88	SEV	131	2.12	.78
				ING	226	2.35	.86	STA	115	2.06	.76
DAS	140	2.15	.79	IVE	225	2.35	.86				
DER	101	2.00	.74	INE	192	2.28	.83	TIO	221	2.34	.85
DRE	100	2.00	.74	IGH	140	2.15	.79	THI	211	2.32	.85
				IRT	105	2.02	.75	TEE	174	2.24	.82
ENT	569	2.76	.99					TOP	174	2.24	.82
EEN	196	2.29	.84	MEN	131	2.12	.78	TWE	170	2.23	.82
EVE	177	2.25	.82					TWO	163	2.21	.81
EST	176	2.25	.82	NIN	207	2.32	.85	THR	158	2.20	.81
ERE	138	2.14	.79	NTH	171	2.23	.82	TER	115	2.06	.76
EIG	135	2.13	.79	NTY	157	2.20	.81	TED	112	2.05	.76
ERS	126	2.10	.78	NET	118	2.07	.77				
EQU	114	2.06	.76					UND	125	2.10	.78
ERI	109	2.04	.76	OUR	211	2.32	.85				
				ONE	210	2.32	.85				
FOR	218	2.34	.85	ORT	146	2.16	.80	VEN	190	2.28	.83
FOU	152	2.18	.80								
FIV	135	2.13	.79	PER	115	2.06	.76	WEN	153	2.18	.80

TABLE 10–C.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their central letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	L_{10} (F)	L_{686} (2F)
DAS	140	2.15	.79
EEN	196	2.29	.84
VEN	190	2.28	.83
TEE	174	2.24	.82
WEN	153	2.18	.80
REE	146	2.16	.80
MEN	131	2.12	.78
SEV	131	2.12	.78
NET	118	2.07	.77
PER	115	2.06	.76
TER	115	2.06	.76
RED	113	2.05	.76
TED	112	2.05	.76
DER	101	2.00	.74
IGH	140	2.15	.79
THI	211	2.32	.85
GHT	196	2.29	.84
THR	158	2.20	.81

	F	L_{10} (F)	L_{686} (2F)
TIO	221	2.34	.85
NIN	207	2.32	.85
SIX	146	2.16	.80
EIG	135	2.13	.79
FIV	135	2.13	.79
HIR	106	2.03	.75
ENT	569	2.76	.99
AND	228	2.36	.86
ING	226	2.35	.86
ONE	210	2.32	.85
INE	192	2.28	.83
UND	125	2.10	.78
ION	260	2.41	.88
FOR	218	2.34	.85
TOP	174	2.24	.82
FOU	152	2.18	.80
COM	136	2.13	.79
EQU	114	2.06	.76

	F	L_{10} (F)	L_{686} (2F)
HRE	153	2.18	.80
ORT	146	2.16	.80
ERE	138	2.14	.79
ERS	126	2.10	.78
ERI	109	2.04	.76
IRT	105	2.02	.75
DRE	100	2.00	.74
EST	176	2.25	.82
ASH	143	2.16	.80
STO	202	2.31	.84
NTH	171	2.23	.82
ATI	160	2.20	.81
NTY	157	2.20	.81
ATE	135	2.13	.79
STA	115	2.06	.76
OUR	211	2.32	.85
IVE	225	2.35	.86
EVE	177	2.25	.82
TWE	170	2.23	.82
TWO	163	2.21	.81

TABLE 10–D.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their final letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	L_{10} (F)	L_{586} (F)
STA	115	2.06	.76
AND	228	2.36	.86
UND	125	2.10	.78
RED	113	2.05	.76
TED	112	2.05	.76
IVE	225	2.35	.86
ONE	210	2.32	.85
INE	192	2.28	.83
EVE	177	2.25	.82
TEE	174	2.24	.82
TWE	170	2.23	.82
HRE	153	2.18	.80
REE	146	2.16	.80
ERE	138	2.14	.79
ATE	135	2.13	.79
DRE	100	2.00	.74
ING	226	2.35	.86
EIG	135	2.13	.79
NTH	171	2.23	.82
ASH	143	2.16	.80
IGH	140	2.15	.79
THI	211	2.32	.85
ATI	160	2.20	.81
ERI	109	2.04	.76
COM	136	2.13	.79
ION	260	2.41	.88
NIN	207	2.32	.85
EEN	196	2.29	.84
VEN	190	2.28	.83
WEN	153	2.18	.80
MEN	131	2.12	.78
TIO	221	2.34	.85
STO	202	2.31	.84
TWO	163	2.21	.81
TOP	174	2.24	.82
FOR	218	2.34	.85
OUR	211	2.32	.85
THR	158	2.20	.81
PER	115	2.06	.76
TER	115	2.06	.76
HIR	106	2.03	.75
DER	101	2.00	.74
DAS	140	2.15	.79
ERS	126	2.10	.78
ENT	569	2.76	.99
GHT	196	2.29	.84
EST	176	2.25	.82
ORT	146	2.16	.80
NET	118	2.07	.77
IRT	105	2.02	.75
FOU	152	2.18	.80
EQU	114	2.06	.76
FIV	135	2.13	.79
SEV	131	2.12	.78
SIX	146	2.16	.80
NTY	157	2.20	.81

TABLE 11–A.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$
TION	218	2.34	.99	THIR	104	2.02	.87	ASHT	64	1.81	.79
EVEN	168	2.23	.95	EENT	102	2.01	.87	HUND	64	1.81	.79
TEEN	163	2.21	.94	REQU	98	1.99	.86	DRED	63	1.80	.79
ENTY	161	2.21	.94	HIRT	97	1.99	.86	RIOD	63	1.80	.79
STOP	154	2.19	.93	COMM	93	1.97	.85	IVED	62	1.79	.78
WENT	153	2.18	.93	QUES	87	1.94	.84	ENTS	62	1.79	.78
NINE	153	2.18	.93	UEST	87	1.94	.84	FFIC	62	1.79	.78
TWEN	152	2.18	.93	EQUE	86	1.93	.84	FROM	59	1.77	.78
THRE	149	2.17	.93	NDRE	77	1.89	.82	IRTY	59	1.77	.78
FOUR	144	2.16	.92	OMMA	71	1.85	.81	RTEE	59	1.77	.78
IGHT	140	2.15	.92	LLAR	71	1.85	.81	UNDR	59	1.77	.78
FIVE	135	2.13	.91	OLLA	70	1.85	.81	NAUG	56	1.75	.77
HREE	134	2.13	.91	VENT	70	1.85	.81	OURT	56	1.75	.77
DASH	132	2.12	.91	DOLL	68	1.83	.80	UGHT	56	1.75	.77
EIGH	132	2.12	.91	LARS	68	1.83	.80	STAT	54	1.73	.76
SEVE	121	2.08	.89	THIS	68	1.83	.80	AUGH	52	1.72	.76
ENTH	114	2.06	.89	PERI	67	1.83	.80	CENT	52	1.72	.76
MENT	111	2.05	.88	ERIO	66	1.82	.80	FICE	50	1.70	.75

TABLE 11–B.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$
ASHT	64	1.81	.79	HREE	134	2.13	.91	REQU	98	1.99	.86
AUGH	52	1.72	.76	HIRT	97	1.99	.86	RIOD	63	1.80	.79
COMM	93	1.97	.85	HUND	64	1.81	.79	RTEE	59	1.77	.78
CENT	52	1.72	.76	IGHT	140	2.15	.92	STOP	154	2.19	.93
DASH	132	2.12	.91	IVED	62	1.79	.78	SEVE	121	2.08	.89
DOLL	68	1.83	.80	IRTY	59	1.77	.78	STAT	54	1.73	.76
DRED	63	1.80	.79	LLAR	71	1.85	.81	TION	218	2.34	.99
EVEN	168	2.23	.95	LARS	68	1.83	.80	TEEN	163	2.21	.94
ENTY	161	2.21	.94	MENT	111	2.05	.88	TWEN	152	2.18	.93
EIGH	132	2.12	.91	NINE	153	2.18	.93	THRE	149	2.17	.93
ENTH	114	2.06	.89	NDRE	77	1.89	.82	THIR	104	2.02	.87
EENT	102	2.01	.87	NAUG	56	1.75	.77	THIS	68	1.83	.80
EQUE	86	1.93	.84	OMMA	71	1.85	.81	UEST	87	1.94	.84
ERIO	66	1.82	.80	OLLA	70	1.85	.81	UNDR	59	1.77	.78
ENTS	62	1.79	.78	OURT	56	1.75	.77	UGHT	56	1.75	.77
FOUR	144	2.16	.92	PERI	67	1.83	.80	VENT	70	1.85	.81
FIVE	135	2.13	.91	QUES	87	1.94	.84	WENT	153	2.18	.93
FFIC	62	1.79	.78								
FROM	59	1.77	.78								
FICE	50	1.70	.75								

TABLE 11–C.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their second letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities.*

	F	$L_{10}(F)$	$L_{244}(F)$
DASH	132	2.12	.91
LARS	68	1.83	.80
NAUG	56	1.75	.77
NDRE	77	1.89	.82
TEEN	163	2.21	.94
WENT	153	2.18	.93
SEVE	121	2.08	.89
MENT	111	2.05	.88
EENT	102	2.01	.87
REQU	98	1.99	.86
UEST	87	1.94	.84
VENT	70	1.85	.81
PERI	67	1.83	.80
CENT	52	1.72	.76
FFIC	62	1.79	.78
IGHT	140	2.15	.92
UGHT	56	1.75	.77
THRE	149	2.17	.93
THIR	104	2.02	.87
THIS	68	1.83	.80

	F	$L_{10}(F)$	$L_{244}(F)$
TION	218	2.34	.99
NINE	153	2.18	.93
FIVE	135	2.13	.91
EIGH	132	2.12	.91
HIRT	97	1.99	.86
RIOD	63	1.80	.79
FICE	50	1.70	.75
LLAR	71	1.85	.81
OLLA	70	1.85	.81
OMMA	71	1.85	.81
ENTY	161	2.21	.94
ENTH	114	2.06	.89
ENTS	62	1.79	.78
UNDR	59	1.77	.78
FOUR	144	2.16	.92
COMM	93	1.97	.85
DOLL	68	1.83	.80
EQUE	86	1.93	.84

	F	$L_{10}(F)$	$L_{244}(F)$
HREE	134	2.13	.91
ERIO	66	1.82	.80
DRED	63	1.80	.79
FROM	59	1.77	.78
IRTY	59	1.77	.78
ASHT	64	1.81	.79
STOP	154	2.19	.93
RTEE	59	1.77	.78
STAT	54	1.73	.76
QUES	87	1.94	.84
HUND	64	1.81	.79
OURT	56	1.75	.77
AUGH	52	1.72	.76
EVEN	168	2.23	.95
IVED	62	1.79	.78
TWEN	152	2.18	.93

TABLE 11–D.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their third letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities.*

	F	$L_{10}(F)$	$L_{244}(F)$
LLAR	71	1.85	.81
STAT	54	1.73	.76
FICE	50	1.70	.75
UNDR	59	1.77	.78
EVEN	168	2.23	.95
TEEN	163	2.21	.94
TWEN	152	2.18	.93
HREE	134	2.13	.91
QUES	87	1.94	.84
DRED	63	1.80	.79
IVED	62	1.79	.78
RTEE	59	1.77	.78

	F	$L_{10}(F)$	$L_{244}(F)$
EIGH	132	2.12	.91
AUGH	52	1.72	.76
IGHT	140	2.15	.92
ASHT	64	1.81	.79
UGHT	56	1.75	.77
THIR	104	2.02	.87
THIS	68	1.83	.80
ERIO	66	1.82	.80
FFIC	62	1.79	.78
OLLA	70	1.85	.81
DOLL	68	1.83	.80

	F	$L_{10}(F)$	$L_{244}(F)$
COMM	93	1.97	.85
OMMA	71	1.85	.81
WENT	153	2.18	.93
NINE	153	2.18	.93
MENT	111	2.05	.88
EENT	102	2.01	.87
VENT	70	1.85	.81
HUNT	64	1.81	.79
CENT	52	1.72	.76
TION	218	2.34	.99
STOP	154	2.19	.93
RIOD	63	1.80	.79
FROM	59	1.77	.78

TABLE 11–D, Concluded.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their third letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$
REQU	98	1.99	.86	DASH	132	2.12	.91	FOUR	144	2.16	.92
				UEST	87	1.94	.84	EQUE	86	1.93	.84
THRE	149	2.17	.93					NAUG	56	1.75	.77
HIRT	97	1.99	.86								
NDRE	77	1.89	.82	ENTY	161	2.21	.94				
LARS	68	1.83	.80	ENTH	114	2.06	.89				
PERI	67	1.83	.80	ENTS	62	1.79	.78	FIVE	135	2.13	.91
OURT	56	1.75	.77	IRTY	59	1.77	.78	SEVE	121	2.08	.89

TABLE 11–E.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Governmental plaintext telegrams, arranged first alphabetically according to their final letters, and then according to their absolute frequencies, accompanied by the logarithms of their assigned probabilities*

	F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$		F	$L_{10}(F)$	$L_{244}(F)$
OMMA	71	1.85	.81	DASH	132	2.12	.91	QUES	87	1.94	.84
OLLA	70	1.85	.81	EIGH	132	2.12	.91	THIS	68	1.83	.80
				ENTH	114	2.06	.89	LARS	68	1.83	.80
				AUGH	52	1.72	.76	ENTS	62	1.79	.78
FFIC	62	1.79	.78								
				PERI	67	1.83	.80				
								WENT	153	2.18	.93
HUND	64	1.81	.79	DOLL	68	1.83	.80	IGHT	140	2.15	.92
DRED	63	1.80	.79					MENT	111	2.05	.88
RIOD	63	1.80	.79	COMM	93	1.97	.85	EENT	102	2.01	.87
IVED	62	1.79	.78	FROM	59	1.77	.78	HIRT	97	1.99	.86
								UEST	87	1.94	.84
				TION	218	2.34	.99	VENT	70	1.85	.81
NINE	153	2.18	.93	EVEN	168	2.23	.95	ASHT	64	1.81	.79
THRE	149	2.17	.93	TEEN	163	2.21	.94	OURT	56	1.75	.77
FIVE	135	2.13	.91	TWEN	152	2.18	.93	UGHT	56	1.75	.77
HREE	134	2.13	.91					STAT	54	1.73	.76
SEVE	121	2.08	.89	ERIO	66	1.82	.80	CENT	52	1.72	.76
EQUE	86	1.93	.84								
NDRE	77	1.89	.82	STOP	154	2.19	.93				
RTEE	59	1.77	.78					REQU	98	1.99	.86
FICE	50	1.70	.75	FOUR	144	2.16	.92				
				THIR	104	2.02	.87				
				LLAR	71	1.85	.81	ENTY	161	2.21	.94
NAUG	56	1.75	.77	UNDR	59	1.77	.78	IRTY	59	1.77	.78

TABLE 12.—*Average length of words and messages*

Number of letters in word x	Number of times x-letter word appears	Number of letters
1	378	378
2	973	1,946
3	1,307	3,921
4	1,635	6,540
5	1,410	7,050
6	1,143	6,858
7	1,009	7,063
8	717	5,736
9	476	4,284
10	274	2,740
11	161	1,771
12	86	1,032
13	23	299
14	23	322
15	4	60
	9,619	50,000

(1) Average length of words .... 5.2 letters.
(2) Average length of messages .... 217 letters.
(3) Modal (most frequent) length of messages .... 105–114 letters.
(4) It is extremely unusual to find five consecutive letters without at least one vowel.
(5) The average number of letters between vowels is two.

TABLE 13.—*Four-square individual frequencies* [1]

[Based on count of 5,000 digraphs]

P_1					C_1				
A	B	C	D	E	244	225	375	394	197
F	G	H	I J	K	125	98	193	271	95
L	M	N	O	P	229	199	188	350	251
Q	R	S	T	U	148	162	258	427	295
V	W	X	Y	Z	42	12	34	91	97
212	317	358	308	249	A	B	C	D	E
120	108	216	256	85	F	G	H	I J	K
216	140	152	435	269	L	M	N	O	P
206	121	306	364	284	Q	R	S	T	U
38	29	21	147	43	V	W	X	Y	Z
C_2					P_2				

[1] The numbers in the C_1 C_2 squares represent the frequency of the individual components of the cipher digraph used to replace a given P_1 P_2 digraph in accordance with a digraphic four-square system where P_1 and P_2 are the plaintext squares.

TABLE 14.—*Relative logarithmic values of frequencies of English digraphs*

[Based on a count of 5,000 digraphs. To obtain logarithm to base 10 (Log 10) divide by 100]

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	48	78	115	143	00	60	78	30	123	00	30	151	115	181	30	108	*	164	161	167	111	85	48	*	108	*
B	60	*	*	*	126	*	*	*	30	00	*	78	00	*	60	*	*	30	00	00	30	*	*	*	85	*
C	130	*	48	00	151	00	*	115	85	*	60	70	00	00	161	*	*	60	00	115	60	*	00	*	00	*
D	151	60	60	90	152	90	30	30	143	00	*	48	70	60	120	70	30	108	111	118	70	48	60	*	00	*
E	154	60	151	178	162	126	60	85	143	00	*	146	115	205	108	130	108	194	173	157	48	130	85	85	60	00
F	70	*	30	00	100	104	00	*	159	*	*	30	00	*	160	00	*	95	48	104	48	*	00	*	00	*
G	85	*	30	00	115	30	00	130	70	00	*	30	00	48	78	30	*	70	48	60	30	*	00	*	*	*
H	130	00	48	30	130	70	*	*	152	*	*	00	30	48	130	00	00	123	60	145	90	*	00	*	00	*
I	90	30	135	78	111	100	128	*	*	*	30	136	95	188	161	85	*	143	154	143	*	140	*	118	*	30
J	00	*	*	*	30	*	*	*	*	*	*	*	*	*	30	*	*	*	*	*	30	*	*	*	*	*
K	00	*	00	*	78	*	*	*	30	*	*	00	*	00	*	*	*	*	00	*	*	*	*	*	*	*
L	145	48	48	95	157	48	00	00	130	*	*	143	30	00	111	48	*	30	78	90	30	30	30	*	100	*
M	156	78	48	00	141	00	*	00	95	*	*	*	111	*	100	90	*	30	60	30	30	*	*	*	30	*
N	141	30	128	172	176	95	43	60	148	00	30	70	70	90	126	48	00	60	138	191	85	48	48	*	70	*
O	85	60	90	108	48	140	30	48	70	00	30	128	140	189	78	140	*	181	115	128	157	85	90	00	30	*
P	115	00	00	00	136	30	*	48	78	*	*	111	60	00	123	104	*	126	78	90	48	00	00	*	00	*
Q	*	*	*	*	*	*	*	*	*	*	*	*	00	*	*	*	*	00	*	*	118	*	*	*	*	*
R	159	30	95	123	199	78	85	48	148	00	00	70	95	85	145	111	*	104	149	162	70	70	60	*	95	*
S	138	48	111	70	169	108	30	142	153	*	00	30	48	60	118	100	*	70	128	180	104	00	60	*	00	*
T	145	48	78	78	185	85	00	189	165	*	*	70	78	85	170	30	00	123	128	128	70	*	156	*	161	00
U	70	48	48	48	104	00	90	*	70	*	*	78	70	132	00	30	*	149	108	108	*	00	*	*	*	*
V	78	*	*	*	176	*	*	*	108	*	*	*	*	*	00	*	*	*	*	00	*	*	*	*	*	*
W	108	*	*	*	134	*	*	60	111	*	*	00	*	30	128	*	*	00	00	*	*	*	*	*	00	*
X	30	*	30	00	00	00	*	00	30	*	*	*	*	00	00	30	*	00	00	85	*	*	*	*	*	*
Y	78	30	60	60	95	104	00	00	48	*	*	30	30	78	100	48	*	60	104	118	00	*	00	*	*	*
Z	00	*	*	*	30	*	*	*	00	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

*In computations, assign a value of −100 as the log for these digraphs. These combinations do not usually occur in 5,000 digraphs. Do not assign "0" to these combinations as that is the logarithmic value for a frequency of one, and these combinations have a frequency of less than one.

TABLE 15.—*Relative logarithmic values* ($Log_e\ 2F$) *of frequencies of English digraphs*[1]

[Based on a count of 5,000 digraphs]

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	.33	.45	.61	.73	.13	.38	.45	.25	.64	.13	.25	.76	.61	.89	.25	.58	0	.82	.80	.83	.59	.48	.33	0	.58	0
B	.38	0	0	0	.66	0	0	0	.25	.13	0	.45	.13	0	.38	0	0	.25	.13	.13	.25	0	0	0	.48	0
C	.67	0	.33	.13	.76	.13	0	.61	.48	0	.38	.42	.13	.13	.80	0	0	.38	.13	.61	.38	0	.13	0	.13	0
D	.76	.38	.38	.51	.77	.51	.25	.25	.73	.13	0	.33	.42	.38	.63	.42	.25	.58	.59	.62	.42	.33	.38	0	.13	0
E	.78	.38	.76	.88	.81	.66	.38	.48	.73	.13	0	.74	.61	.99	.58	.67	.58	.94	.86	.79	.33	.67	.48	.48	.38	.13
F	.42	0	.25	.13	.55	.56	.13	0	.80	0	0	.25	.13	0	.80	.13	0	.53	.33	.56	.33	0	.13	0	.13	0
G	.48	0	.25	.13	.61	.25	.13	.67	.42	.13	0	.25	.13	.33	.45	.25	0	.42	.33	.38	.25	0	.13	0	0	0
H	.67	.13	.33	.25	.67	.42	0	0	.77	0	0	.13	.25	.33	.67	.13	.13	.64	.38	.74	.51	0	.13	0	.13	0
I	.51	.25	.69	.45	.59	.55	.67	0	0	0	.25	.70	.53	.92	.80	.48	0	.73	.78	.73	0	.72	0	.62	0	.25
J	.18	0	0	0	.25	0	0	0	0	0	0	0	0	0	.25	0	0	0	0	0	.25	0	0	0	0	0
K	.13	0	.13	0	.45	0	0	0	.25	0	0	.13	0	.13	0	0	0	0	.13	0	0	0	0	0	0	0
L	.74	.33	.33	.53	.79	.33	.13	.13	.67	0	0	.73	.25	.13	.59	.33	0	.25	.45	.51	.25	.25	.25	0	.55	0
M	.78	.45	.33	.13	.72	.13	0	.13	.53	0	0	0	.59	0	.55	.51	0	.25	.38	.25	.25	0	0	0	.25	0
N	.72	.25	.67	.85	.87	.53	.73	.38	.75	.13	.25	.42	.42	.51	.66	.33	.13	.38	.71	.93	.48	.33	.33	0	.42	0
O	.48	.38	.51	.58	.33	.72	.25	.33	.42	.13	.25	.67	.72	.92	.45	.72	0	.89	.61	.67	.79	.48	.51	.13	.25	0
P	.61	.13	.13	.13	.70	.25	0	.33	.45	0	0	.59	.38	.13	.64	.56	0	.66	.45	.51	.33	.13	.13	0	.13	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	.13	0	0	0	0	.13	0	0	.62	0	0	0	0	0
R	.80	.25	.53	.64	.96	.45	.48	.33	.75	.13	.13	.42	.53	.48	.74	.59	0	.56	.75	.81	.42	.42	.38	0	.53	0
S	.71	.33	.59	.42	.84	.58	.25	.72	.77	0	.13	.25	.33	.38	.62	.55	0	.42	.67	.88	.56	.13	.38	0	.13	0
T	.74	.33	.45	.45	.91	.48	.13	.92	.82	0	0	.42	.45	.48	.84	.25	.13	.64	.67	.67	.42	0	.78	0	.80	.13
U	.42	.33	.33	.33	.56	.13	.51	0	.42	0	0	.45	.42	.68	.13	.25	0	.75	.58	.58	0	.13	0	0	0	0
V	.45	0	0	0	.87	0	0	0	.58	0	0	0	0	0	.13	0	0	0	0	.13	0	0	0	0	0	0
W	.58	0	0	0	.69	0	0	.38	.59	0	0	.13	0	.25	.67	0	0	.13	.13	0	0	0	0	0	.13	0
X	.25	0	.25	.13	.13	.13	0	.13	.25	0	0	0	0	.13	.13	.25	0	.13	.13	.48	0	0	0	0	0	0
Y	.45	.25	.38	.38	.53	.56	.13	.13	.33	0	0	.25	.25	.45	.55	.33	0	.38	.56	.62	.13	0	.13	0	0	0
Z	.13	0	0	0	.25	0	0	0	.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[1] See pp. 259–260 for details.

TABLE 16–A.—*Frequency distribution of digraphs, based on 64,365 letters of decrypted U. S. Government messages in which* Z *was used as a word separator and* X *was used for both* X_p *and* Z_p

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	28	154	142	137	17	90	99	13	118	16	43	220	157	427	18	112	2	625	586	347	56	52	20	3	66	546
B	63	14	7	1	193	1		1	43	33		148	6	18	61	2		59	17	8	15	1	1	3	60	19
C	123	1	19	8	260	22	28	183	115		48	95	390	5	414	3	1	63	66	161	47	1	5	3	27	122
D	360	12	33	30	270	4	16		141	2	1	7	4	6	102	11	11	33	32	34	38	38	17	1	11	1026
E	180	34	226	383	620	131	35	13	275	3	6	185	144	758	75	118	91	857	329	187	40	210	28	76	29	1715
F	44	16	10	3	100	122	4	1	365	2		28	23	4	536	68		114	8	32	34	1	1	2	3	343
G	78	29	7	18	258	5	31	260	25		1	11	5	31	20	18		73	29	17	25	2			1	275
H	194	1	6	12	193	14	1	24	213	3	9	7	2	24	93	3	24	229	26	257	17	2	6	1	3	428
I	85	10	209	30	152	53	330	5	5	1	46	181	40	704	200	92	1	128	303	217	2	272	2	193	1	56
J	26		3	2	31	3		1	18	20		3	1	4	35	1		5	2	18	7	2	1		2	19
K	28		2	6	108	2			54	3	20	11	3	10	9		1	1	9	2	1	1	2	1	10	59
L	159	6	6	48	328	14		4	194	2	1	237	20	65	120	5		5	41	25	41	5		1	71	296
M	581	68	36	12	198	1	58	1	92	4	1	2	62	4	43	101		10	53	20	17	1	3	6	86	231
N	112	13	157	286	733	77	244	4	234		14	15	9	76	169	16	16	13	135	267	64	10	7	7	14	910
O	25	67	46	100	56	317	66	26	23	6	23	161	230	873	59	57	2	418	129	143	413	49	59	92	13	916
P	304	5	8	363	169		2	37	27	3		75	46	9	145	104	3	153	26	351	44	2	2	1	4	122
Q	2	1	1		7			4	1			1	5	11	1	1	9	5	7		117		1			46
R	261	5	44	86	967	26	59	5	191	5	30	61	122	45	570	310	4	72	208	179	60	19	14	13	74	733
S	143	14	66	6	389	85	52	426	334	1	16	16	34	6	99	47	13	5	143	305	138	13	12	1	43	788
T	171	1	67	22	357	32	6	572	275	2	10	27	18	49	372	9	2	119	99	156	37	1	313	10	48	1106
U	45	48	26	60	87	4	61	2	35	1	3	56	61	96	32	38		453	140	48	5	5	5	1	1	44
V	39		10	2	496	1	1		91		1	3	1	8	19	4	1	3	4	7	1	9	1	1	7	34
W	111	1	3	7	34	1	11	33	107	2	1	10		12	367	7	2	3	11	5			13	13	2	30
X	9	1	8	7	350	9		2	10	1	2		2	2	10	20	3	12	9	32	1	1		32	3	203
Y	8	3	6	3	14	6	3	2	5			4	9	10	49	27		3	18	8	4	1	1		8	432
Z	902	264	1058	613	364	844	120	171	328	98	69	135	294	349	750	823	36	700	768	1046	130	46	278	271	42	
	4081	768	2206	2245	6751	1864	1227	1790	3319	208	345	1699	1638	3606	4368	1997	222	4161	3198	3872	1354	744	792	732	620	10499

In the text from which this and the following two tables are derived, the frequently-used punctuation signs "comma" and "period" were abbreviated as CMA and PD, respectively, and the procedure term "repeat" was abbreviated as RPT; thus the digraphs CM, PD, PT, and RP, which usually do not occur frequently (see Table 6–A), are of relatively high frequency here. Furthermore, the high doublet rate was caused by a convention which required at a certain point in each message a string of five identical plaintext letters (excepting Z).

TABLE 16–B.—*Frequency distribution of digraphs, based on the text used for Table 16–A, from which the word separator Z has been omitted (total: 53,866 letters)*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	78	175	190	164	40	136	111	26	139	19	52	227	166	439	58	147	3	657	619	395	65	58	40	23	67	
B	63	14	9	2	193	5		1	43	32		149	6	18	62	2		62	17	13	15	2	3	3	60	
C	133	1	31	20	263	32	29	184	119		48	98	393	11	416	8	2	78	. 79	180	47	1	6	4	27	
D	443	66	102	74	307	86	26	13	183	7	5	23	32	22	151	97	16	142	118	153	59	40	55	2	18	
E	299	70	384	481	690	283	48	37	326	21	12	201	190	855	181	278	93	931	476	367	53	215	87	136	34	
F	60	19	42	25	109	137	7	2	380	3	1	39	25	10	582	80	1	148	56	67	49	3	9	3	7	
G	102	39	20	59	266	19	32	262	37	5	2	12	10	41	45	38	4	91	53	38	31	2	3	7	1	
H	270	8	34	28	215	54	13	31	220	14	11	8	13	34	139	14	23	239	64	315	18	3	16	5	3	
I	86	10	213	41	156	55	330	8	5	1	46	182	40	705	202	96	1	148	303	218	2	270	3	196	1	
J	28		7	2	31	7		1	21	20		3	1	5	36	2		6	2	19	7	2	2		2	
K	35	4	7	10	108	10	2		56	3	20	11	4	13	12	7	1	6	11	5	2	1	4	1	10	
L	197	21	38	61	338	47	2	13	207	7	4	243	26	68	134	19		21	59	50	44	8	14	1	72	
M	595	72	66	18	206	22	64	4	96	6	1	6	67	17	63	123	3	26	61	40	22	2	10	15	86	
N	213	27	280	336	748	139	254	12	263	6	19	31	47	86	234	92	24	66	202	352	75	23	28	28	17	
O	63	82	191	155	93	426	72	47	37	13	27	172	252	910	99	112	2	473	204	214	419	51	68	170	17	
P	311	7	16	388	170	5	3	40	29	4		76	46	11	150	111	3	179	37	365	44	2	2	1	5	
Q	14	4	3		7	2		4	5		1	2	5	11	8	2	9	10	10	2	117		3	1		
R	298	12	131	146	1011	84	66	14	207	17	40	69	142	59	639	369	8	103	266	263	67	19	29	30	74	
S	237	37	143	31	396	149	55	453	369	5	19	25	60	36	173	129	16	62	178	385	144	14	34	2	43	
T	277	30	167	70	400	97	21	592	308	14	16	43	67	100	463	95	5	195	150	282	52	12	338	30	57	
U	48	48	33	61	88	7	61	2	36	4	4	56	61	97	35	40		454	148	50	6	5	6	6	1	
V	41		13	5	499	7	1		92		2	4	3	8	21	6	2	4	9	8	1	9	1	1	7	
W	113	6	6	9	37	2	12	35	107	3	1	10	1	14	367	10	2	3	11	6	1		13	13	4	
X	18	2	23	22	361	20		4	12	3	10	2	9	11	24	41	3	26	29	47	4	1		54	3	
Y	59	14	57	37	19	33	18	5	22	1	4	7	22	25	74	79	1	31	36	38	10	1	18		13	
Z																										
	4081	768	2206	2245	6751	1864	1227	1790	3319	208	345	1690	1688	3606	4368	1997	222	4161	3198	3872	1354	744	792	732	629	

TABLE 16–C.—*The 53 digraphs from Table 6–A which comprise 50% of the total, arranged according to frequencies reduced to a base of 5,000 digraphs, shown with the corresponding frequencies of the same digraphs from Table 16–B (also reduced to a base of 5,000).*[1]

Digraph	*6–A*	*16–B*	*Digraph*	*6–A*	*16–B*
EN	111	79	FO	40	54
RE	98	94	FI	39	35
ER	87	86	RA	39	28
NT	82	33	ET	37	34
TH	78	55	LE	37	31
ON	77	84	OU	37	39
IN	75	65	MA	36	55
TE	71	36	TW	36	31
AN	64	41	EA	35	28
OR	64	44	IS	35	28
ST	63	36	SI	34	34
ED	60	45	DE	33	29
NE	57	69	HI	33	[1] 20
VE	57	46	AL	32	[1] 21
ES	54	44	CE	32	24
ND	52	31	DA	32	41
TO	50	43	EC	32	36
SE	49	37	RS	31	25
AT	47	37	UR	31	42
TI	45	29	NI	30	24
AR	44	61	RI	30	[1] 19
EE	42	64	EL	29	[1] 19
RT	42	24	HT	28	29
AS	41	57	LA	28	[1] 18
CO	41	39	RO	28	59
IO	41	[1] 19	TA	28	26
TY	41	[1] 15			

[1] With the exception of AL, EL, HI, IO, LA, RI, TY, the digraphs of this table are all from among the 65 digraphs from Table 16–B which comprise 50% of the total.

APPENDIX 3

WORD AND PATTERN LISTS—ENGLISH

WORD AND PATTERN LISTS—ENGLISH

A. LIST OF WORDS USED IN MILITARY TEXT ARRANGED ALPHABETICALLY ACCORDING TO WORD LENGTH

TWO LETTER WORDS

AM	BY	EM	IN	MM	OK	TO
AN	CO	GO	IS	MP	ON	US
AS	CP	HE	IT	MY	OR	WD
AT	CQ	HQ	MC	NO	QM	WE
BE	DO	IF	ME	OF	SO	WO
BN						

THREE LETTER WORDS

ACT	BIG	EAT	HER	MIX	PVT	TEN
ADD	BOX	END	HIM	NAN	QMC	THE
ADJ	BUT	EYE	HIS	NET	RED	TIN
AGE	BUY	FAR	HOW	NEW	RID	TON
AGO	CAM	FEW	ILL	NOT	ROB	TOO
AID	CAN	FIT	ITS	NOW	RUN	TOP
AIM	CAR	FIX	JIG	OFF	SAW	TRY
AIR	CAV	FOR	JOB	OLD	SAY	TUB
ALL	COL	FOX	KEG	ONE	SEA	TWO
AND	CPL	GAL	LAW	OUR	SEE	USE
ANY	CUT	GAS	LAY	OUT	SET	VAT
APT	CWT	GEN	LET	OWE	SGT	WAR
ARC	DAY	GET	LOT	OWN	SHE	WAS
ARE	DID	GHQ	LOW	PAR	SIX	WAY
ARM	DIE	GOT	MAJ	PAY	SPY	WET
ASK	DOG	GUN	MAN	PEN	SUM	WGT
BAD	DRY	HAD	MAT	PER	SUN	WON
BAG	DUE	HAM	MAY	PIN	TAN	YET
BAR	DUN	HAS	MEN	PUT	TAX	YOU
BID						

FOUR LETTER WORDS

ABLE	BOTH	EACH	FLEE	HIGH	LATE	MAIN
AIDE	BULB	EAST	FORM	HILL	LEAD	MANY
ALLY	BULK	EASY	FOUR	HITS	LEAK	MASK
ALSO	CALL	EDGE	FROM	HOLD	LEFT	MASS
AREA	CELL	EYES	FULL	HOOK	LESS	MEAT
ARMY	CITY	FALL	FUSE	INTO	LIEU	MEET
ASIA	CODE	FARM	FUZE	ITEM	LINE	MESS
AWAY	COOK	FAST	GUNS	JOIN	LIST	MIKE
AXIS	DARK	FEEL	HALF	JULY	LOAD	MILE
BACK	DASH	FEET	HALT	JUNE	LONG	MINE
BASE	DATE	FELL	HAND	JUST	LOOK	MORE
BEEN	DAYS	FILE	HARD	KEEP	LOSS	MOVE
BLUE	DIRT	FIRE	HAVE	KIND	LOST	MTCL
BODY	DOWN	FIRM	HEAD	KING	LOVE	MULE
BOMB	DRAW	FIVE	HERD	LAND	MADE	NAVY
BOOK	DUMP	FLAG	HERE	LAST	MAIM	NEAR

FOUR LETTER WORDS—Continued

NEXT	PARK	REAR	SHOT	TEAM	TOOK	WEST
NINE	PASS	RIOT	SIDE	TENT	TOOL	WHAT
NOON	PIPE	ROAD	SOME	TEXT	TOWN	WHEN
NOTE	PLAN	ROUT	SOON	THAN	TYPE	WILL
OBOE	POST	RULE	STOP	THAT	UNIT	WIRE
OMIT	PUMP	RUSH	SUNK	THEM	VARY	WITH
ONCE	PUSH	SAID	TAKE	THEN	VERY	XRAY
ONLY	RAID	SAME	TALK	THEY	WEAK	YOKE
OPEN	RAIL	SANK	TANK	THIS	WEEK	YOUR
ORAL	RAIN	SEEN	TARE	TIME	WELL	ZERO
OVER	RANK	SHIP	TASK	TONS	WERE	ZONE

FIVE LETTER WORDS

ABOUT	BOATS	DECKS	FIGHT	LATER	PRIOR	SHIPS	TITLE
AFTER	BOMBS	DEFER	FIRES	LEAST	PROOF	SHORE	TODAY
AGAIN	BOOTH	DELAY	FIRST	LEAVE	PROVE	SIEGE	TOTAL
AGENT	BREAK	DEPOT	FLANK	LEVEL	QUEEN	SIGHT	TRACT
ALARM	BRIBE	DEPTH	FLARE	LIGHT	QUICK	SIXTH	TRAIN
ALERT	BROKE	DOCKS	FLATS	LIMIT	QUIET	SIXTY	TROOP
ALIGN	BURST	DRAWN	FLEET	LOCAL	RADIO	SLOPE	TRUCE
ALINE	CANAL	DRESS	FOGGY	MAJOR	RAFTS	SMALL	TRUCK
ALLOW	CASES	DRILL	FORCE	MARCH	RAIDS	SMOKE	UNDER
ALONG	CAUSE	DRIVE	FORTY	METER	RALLY	SOUTH	UNION
AMONG	CEASE	EAGER	FRESH	MILES	RANGE	SPEED	UNITS
ANNEX	CHECK	EARLY	FRONT	MOTOR	RAPID	SPELL	USUAL
APPLY	CHIEF	EIGHT	GATES	NAVAL	REACH	SPLIT	VALOR
APRIL	CLEAR	ENEMY	GAUGE	NIGHT	READY	SQUAD	VISIT
AREAS	CLERK	ENTER	GIVEN	NINTH	REFER	STAFF	VITAL
ARMOR	CLOSE	EQUAL	GOING	NORTH	REPEL	STAKE	VOCAL
ASSET	COAST	EQUIP	GROUP	ORDER	RIDGE	START	VOICE
AWAIT	COLON	ERASE	GUARD	OTHER	RIGHT	STEEL	WAGON
AWARD	COMMA	ERROR	GUEST	PACKS	RIGID	SUGAR	WEIGH
BAKER	CORPS	ETHER	HEAVY	PAIRS	RIVER	TAKEN	WHEEL
BANKS	COUNT	EVERY	HONOR	PARTY	ROGER	TANKS	WHERE
BARGE	COVER	FATAL	HORSE	PETER	ROUTE	TENTH	WHICH
BEACH	CREEK	FEARS	HOURS	PLACE	SCALE	THEIR	WIDTH
BEGIN	CREST	FERRY	HOUSE	PLAIN	SEIZE	THERE	WIPED
BEING	CROSS	FIELD	ISSUE	PLANS	SEVEN	THESE	WOODS
BLACK	CURVE	FIFTH	JAPAN	POINT	SHELL	THIRD	YARDS
BLIND	DAILY	FIFTY	LARGE	PRESS	SHIFT	THREE	ZEBRA

SIX LETTER WORDS

ACCEPT	BOMBED	DEGREE	FIERCE	LESSON	OTHERS	RESUME	SUFFER
ACCESS	BOMBER	DEPART	FILING	LETTER	OUTPUT	RETIRE	SUMMER
ACROSS	BOTTOM	DEPEND	FINISH	LINING	PANAMA	RETURN	SUMMIT
ACTION	BRANCH	DEPLOY	FIRING	LIQUID	PARADE	REVIEW	SUMMON
ACTIVE	BREACH	DESERT	FLIGHT	LITTER	PARLEY	RIDING	SUNDAY
ADJUST	BREEZE	DETACH	FLYING	LITTLE	PASSED	ROCKET	SUNKEN
ADVICE	BRIDGE	DETAIL	FOLLOW	LOCATE	PASSES	ROUTED	SUNSET
ADVISE	BROKEN	DEVICE	FORCES	LOSSES	PATROL	ROUTES	SUPPLY
AFFAIR	BUREAU	DEVISE	FORMAL	MANAGE	PERIOD	RUBBER	SURVEY
ALASKA	CANADA	DIRECT	FORMED	MANNER	PICKET	RUNNER	SWITCH
ALLEGE	CANCEL	DIVERT	FOUGHT	MANUAL	PINCER	SALARY	SYSTEM
ALLIED	CANNOT	DIVIDE	FOURTH	MEAGER	PISTOL	SCHEME	TABLES
ALLIES	CANVAS	DOCTOR	FRIDAY	MEDIUM	PLACES	SCHOOL	TANKER
ALWAYS	CASUAL	DOLLAR	FUTURE	MEMBER	PLANES	SCORED	TARGET
ANIMAL	CAUSED	DOWNED	GARAGE	METHOD	POINTS	SCREEN	TATTOO
ANNUAL	CENTER	DRYRUN	GEORGE	METRIC	POISON	SEAMAN	TERROR
ANYWAY	CHANGE	DUGOUT	GREASE	MINING	POLICE	SEAMEN	THIRTY
APPEAR	CHARGE	DURING	GROUND	MINUTE	PONTON	SEARCH	THOUGH
ARABIA	CHEESE	EFFECT	GUNNER	MIRROR	POSTAL	SECOND	THREAT
ARMIES	CHURCH	EFFORT	HALTED	MOBILE	PREFER	SECTOR	TRAINS
ARMORY	CIPHER	EIGHTH	HAMMER	MONDAY	PROMPT	SECURE	TRENCH
ARREST	CIRCLE	EIGHTY	HAPPEN	MORALE	PROPER	SELECT	TROOPS
ARRIVE	COFFEE	EITHER	HARBOR	MORTAR	PURSUE	SERIAL	TURRET
ASSETS	COLORS	ELEVEN	HELPER	MOVING	RADIAL	SETTLE	TWELVE
ASSIST	COLUMN	EMBARK	HIGHER	MURDER	RAIDED	SEVERE	TWENTY
ASSURE	COMBAT	EMPLOY	HOURLY	MUZZLE	RATION	SHELLS	UNABLE
ATTACH	COMMIT	ENCODE	INDEED	NAUGHT	RAVINE	SIGCOM	UNITED
ATTACK	COMMON	ENGAGE	INFORM	NEARER	RECORD	SIGNAL	UNLESS
ATTAIN	CONVEY	ENGINE	INLAND	NINETY	REDUCE	SINGLE	VALLEY
AUGUST	CONVOY	ENROLL	INTEND	NORMAL	REFILL	SLIGHT	VERBAL
BANNER	COURSE	ENTIRE	INTENT	NOTING	REFUGE	SPHERE	VERIFY
BARBED	CREDIT	ERASER	INVENT	NOUGHT	REFUSE	SPOOLS	VESSEL
BARGES	CRISIS	ESCORT	ISLAND	NOVICE	REJECT	SPOONS	VICTIM
BATTEN	CRITIC	EUROPE	ISSUES	NOZZLE	RELIEF	STATES	VICTOR
BATTLE	DAMAGE	EXCEPT	KEEPER	NUMBER	REMAIN	STATUS	VISITS
BEETLE	DEBARK	EXCESS	KILLED	OCCUPY	REMEDY	STRAFE	VISUAL
BEFORE	DECIDE	EXCITE	LADDER	OFFEND	REPAIR	STREET	WEIGHT
BETTER	DECODE	EXPECT	LANDED	OFFICE	REPORT	STRESS	WIRING
BEYOND	DECREE	EXPELS	LAUNCH	OPPOSE	RESCUE	STRIPS	WITHIN
BILLET	DEFEAT	EXPEND	LEADER	ORDERS	RESIST	SUBMIT	WOODED
BITTER	DEFECT	EXTEND	LEAGUE	ORIENT	RESULT	SUDDEN	ZIGZAG
BODIES	DEFEND	EXTENT					

SEVEN LETTER WORDS

ABANDON	ALMANAC	APPOINT	ASIATIC	AVIATOR	BATTERY	BETWEEN
ABSENCE	AMMETER	APPROVE	ASSAULT	AWKWARD	BATTLES	BICYCLE
ADDRESS	ANALYZE	ARMORED	ATTACKS	BAGGAGE	BEARING	BINDING
ADVANCE	ANOTHER	ARRANGE	ATTEMPT	BALLOON	BECAUSE	BIVOUAC
AGAINST	ANTENNA	ARRIVAL	AVERAGE	BARRAGE	BEDDING	BOMBARD

~~CONFIDENTIAL~~

SEVEN LETTER WORDS—Continued

BOMBERS	DEBOUCH	FISHING	LANDING	PACKAGE	REQUEST	SUPPOSE
BOMBING	DECIDED	FITTING	LEADING	PASSAGE	REQUIRE	SURPLUS
BOYCOTT	DECLARE	FOGHORN	LECTURE	PASSIVE	RESERVE	SUSPEND
BRIBERY	DECODED	FORCING	LIAISON	PATROLS	RESPECT	TACTICS
BRIGADE	DEFENSE	FORGING	LIBRARY	PAYROLL	RESPOND	TALKING
CALIBER	DELAYED	FORWARD	LICENSE	PLACING	RETIRED	TARGETS
CALIBRE	DELIVER	FOXHOLE	LIFTING	PLATOON	RETREAT	TERRAIN
CAPTAIN	DERRICK	FUELOIL	LOADING	POUNDER	REVENUE	THATTHE
CAPTIVE	DESTROY	FURNISH	LOGICAL	PRAIRIE	REVERSE	THROUGH
CARRIER	DETRAIN	FURTHER	LOOKOUT	PRECEDE	REVOLVE	TOBACCO
CAVALRY	DETRUCK	GASSING	MACHINE	PREPARE	ROUTINE	TONIGHT
CENTRAL	DEVELOP	GENERAL	MANDATE	PRESENT	RUNNING	TONNAGE
CHANGES	DIAGRAM	GETTING	MANNING	PRESSED	SAILORS	TORPEDO
CHANNEL	DISCUSS	GLASSES	MAPPING	PRIMARY	SATISFY	TRACTOR
CHARLIE	DISEASE	GRADUAL	MARCHED	PROCEED	SECRECY	TRAFFIC
CHASSIS	DISMISS	GRENADE	MARSHAL	PROGRAM	SECTION	TRAWLER
CIRCUIT	DISTILL	GUARDED	MARTIAL	PROMOTE	SECTORS	TRIGGER
CLIPPER	DROPPED	HALTING	MAXIMUM	PROPOSE	SERVICE	TUESDAY
COASTAL	EASTERN	HASBEEN	MEDICAL	PROTECT	SESSION	TWELFTH
COLLECT	ECHELON	HEADING	MESSAGE	PROTEST	SETBACK	UNKNOWN
COLLEGE	ELEMENT	HEAVIER	MESSING	PROVOST	SEVENTH	UNUSUAL
COLONEL	ELEVATE	HIGHEST	MILITIA	PURPOSE	SEVENTY	USELESS
COMMAND	EMBASSY	HOLDING	MINIMUM	PURSUIT	SEVERAL	UTILITY
COMMEND	ENCODED	HORIZON	MISFIRE	PUSHING	SHELLED	VACANCY
COMMENT	ENEMIES	HOSTILE	MISSING	QUARTER	SHORTLY	VARYING
COMMUTE	ENFORCE	HUNDRED	MISSION	QUICKLY	SIGNIFY	VESSELS
COMPANY	ENGAGED	ICEBERG	MORNING	RADIATE	SIMILAR	VICTORY
COMPASS	ENTENTE	ILLEGAL	NATURAL	RAIDING	SIMPLEX	VILLAGE
CONCEAL	ENTRAIN	ILLNESS	NEAREST	RAILWAY	SINKING	VISIBLE
CONDEMN	ENTRUCK	INCLUDE	NIGHTLY	RAINING	SIXTEEN	VISITOR
CONDUCT	ENVELOP	INFLICT	NOTHING	RAPIDLY	SLOPING	WARFARE
CONFINE	EVENING	INITIAL	NUMBERS	REACHED	SMOKING	WARSHIP
CONTACT	EXCLUDE	INQUIRE	OSBERVE	RECEIPT	SOLDIER	WEATHER
CONTAIN	EXPLAIN	INQUIRY	OCTOBER	RECEIVE	STARTER	WESTERN
CONTROL	EXPRESS	INSPIRE	OFFENSE	RECOVER	STATION	WHETHER
CORRECT	EXTRACT	INSTALL	OFFICER	RECRUIT	STEAMER	WILLIAM
COUNCIL	EXTREME	INSTANT	OMITTED	REDUCED	STOPPED	WINDAGE
COURIER	FALLING	INVADED	OPERATE	REFUGEE	STORAGE	WITHOUT
COVERED	FARTHER	ISLANDS	OPINION	REGULAR	SUCCESS	WITHTHE
CROSSED	FEDERAL	ISSUING	ORDERED	RELEASE	SUGGEST	WITNESS
CRUISER	FIFTEEN	JANUARY	OUTPOST	RELIEVE	SUMMARY	WOUNDED
CURRENT	FIGHTER	JUMPOFF	OUTSIDE	REPAIRS	SUNRISE	WRECKED
CYCLONE	FILLING	KITCHEN	PACIFIC	REPLACE	SUPPORT	WRITTEN
DAMAGED	FINDING	KILLING				

EIGHT LETTER WORDS

ACTIVITY	ADVANCED	AIRBORNE	AIRPLANE	ANNOUNCE	APPROACH	ASSEMBLE
ACTUALLY	ADVANCES	AIRCRAFT	ALTITUDE	ANTITANK	APPROVAL	ASSEMBLY
ADJACENT	ADVISING	AIRDROME	AMERICAN	APPARENT	ARMAMENT	ASSIGNED
ADJUTANT	ADVISORY	AIRFIELD	ANALYSIS	APPEARED	ARRESTED	ASSOONAS

EIGHT LETTER WORDS—Continued

ATLANTIC
ATTACKED
ATTEMPTS
AVIATION
BARRACKS
BARRAGES
BATTERED
BATTLING
BESIEGED
BILLETED
BOUNDARY
BREAKING
BUILDING
BULLETIN
BUSINESS
CALAMITY
CAMPAIGN
CANISTER
CAPACITY
CAPTURED
CARELESS
CARRIAGE
CARRIERS
CARRYING
CASUALTY
CAUSEWAY
CEMETERY
CENTERED
CHAPLAIN
CHEMICAL
CIRCULAR
CITATION
CIVILIAN
CLERICAL
CODEBOOK
COMMANDS
COMMENCE
COMMERCE
COMPLETE
COMPOSED
CONCLUDE
CONCRETE
CONFLICT
CONGRESS
CONTINUE
CONTRACT
CORPORAL
CORRIDOR
COVERING
CRITICAL
CRITIQUE
CROSSING
CRUISERS
DAMAGING
DARKNESS
DAYLIGHT
DECEMBER
DECIPHER
DECISION
DECISIVE
DECLARED
DECREASE
DEDICATE
DEFEATED
DEFENDED
DEFENDER
DEFENSES
DEFERRED
DEFINITE
DELAYING
DEMANDED
DEPARTED
DEPLOYED
DEPORTED
DESCRIBE
DESERTED
DESERTER
DESPATCH
DETACHED
DETECTOR
DETONATE
DEVELOPE
DICTATED
DICTATOR
DIMINISH
DIRECTOR
DISARMED
DISASTER
DISLODGE
DISPATCH
DISPERSE
DISTANCE
DISTRESS
DISTRICT
DIVIDING
DIVISION
DOCTRINE
DOMINANT
DRESSING
DRIFTING
EASTERLY
EASTWARD
ECONOMIC
EFFECTED
EFFICACY
EIGHTEEN
ELEMENTS
ELEVENTH
ELIGIBLE
EMPLOYEE
EMPLOYER
ENCIPHER
ENCIRCLE
ENFILADE
ENGAGING
ENGINEER
ENLISTED
ENORMOUS
ENROLLED
ENTERING
ENTRENCH
ENVELOPE
EQUALIZE
EQUIPAGE
ESCORTED
ESTIMATE
EUROPEAN
EVACUATE
EXCAVATE
EXCHANGE
EXERCISE
EXPANDED
EXPEDITE
EXPELLED
EXPENDED
EXPENSES
EXTENDED
EXTERIOR
FACTIONS
FATALITY
FEBRUARY
FERRYING
FIGHTERS
FIGHTING
FINISHED
FLANKING
FLEXIBLE
FOOTHOLD
FORENOON
FORTRESS
FOURTEEN
FRONTAGE
FUSELAGE
GARRISON
GROUNDED
GROUPING
GUARDING
HAVEBEEN
HINDERED
HOSPITAL
HOWITZER
IDENTIFY
IGNITION
IMPROPER
IMPROVED
INCIDENT
INDICATE
INDIRECT
INFANTRY
INFECTED
INITIATE
INSECURE
INSIGNIA
INSTRUCT
INTEREST
INTERIOR
INTERNAL
INTRENCH
INVADING
INVASION
INVENTED
JETPLANE
JUNCTION
LANGUAGE
LATITUDE
LETTERED
LIMITING
LOCATION
LUMINOUS
MAINTAIN
MANDATED
MANEUVER
MARCHING
MARITIME
MATERIAL
MATERIEL
MECHANIC
MEDICINE
MEMORIAL
MERCIFUL
MESSAGES
MIDNIGHT
MILITARY
MISFIRES
MISSIONS
MOBILIZE
MONOPOLY
MOUNTAIN
MOVEMENT
NATIONAL
NAUTICAL
NINETEEN
NORTHERN
NOVEMBER
OBSERVED
OBSERVER
OBSOLETE
OBSTACLE
OCCUPIED
OFFENDED
OFFICERS
OFFICIAL
OPERATOR
OPPOSING
OPPOSITE
ORDINATE
ORDNANCE
OUTBOARD
OUTGUARD
OUTPOSTS
PAINTING
PARALLAX
PARALLEL
PASSPORT
PLANNING
POLITICS
PONTOONS
POSITION
POSITIVE
POSSIBLE
POSTPONE
PREPARED
PRESERVE
PRESSING
PRESSURE
PRINTING
PRIORITY
PRISONER
PROBABLE
PROBABLY
PROGRESS
PROHIBIT
PROTESTS
PROTOCOL
PURPOSES
QUARTERS
RAILHEAD
RAILROAD
RALLYING
RECEIVER
RECORDER
REDCROSS
REENLIST
REGIMENT
REGISTER
REJECTED
REJECTOR
REMEDIES
REMEMBER
REPAIRED
REPEATER
REPELLED
REPLACED
REPORTED
REPULSED
REQUIRED
RESEARCH
RESERVES
RESPECTS
RESTORED
RETIRING
RETURNED
REVIEWED
REVOLVER
RIGOROUS
SABOTAGE
SANITARY
SATURDAY
SCHEDULE
SEABORNE
SEALEVEL
SELECTED
SENTENCE
SENTINEL
SEPARATE
SERGEANT
SHELLING
SHIPPING
SIGHTING
SKIRMISH
SOLDIERS
SOUTHERN
SPECIFIC
SPOTTING
SQUADRON
STANDARD
STATIONS
STRATEGY
SUFFERED
SUITABLE
SUPERIOR
SUPPLIES
SURPRISE
SURROUND
SURVIVED
SUSPENSE
SWEEPING
SWIMMING
TACTICAL
TAXATION
TELEGRAM
TERRIBLE
TERRIFIC
THATHAVE
THIRTEEN
THOUSAND
THURSDAY
TOMORROW
TOTALING
TRAILERS
TRAINING
TRANSFER
TRAVERSE
TRAWLERS
VEHICLES
VICINITY
VIGOROUS
WARSHIPS
WESTERLY
WESTWARD
WINDWARD
WIRELESS
WITHDRAW
WITHDREW

CONFIDENTIAL

NINE LETTER WORDS

ACCESSORY	CENTERING	DEVELOPED	FORMATION	MOVEMENTS	PROTECTOR
ACCOMPANY	CHALLENGE	DIETITIAN	FORTIFIED	MUNITIONS	PROTESTED
ACCORDING	CHARACTER	DIFFERENT	FRONTLINE	NAVALBASE	PROVISION
ADDRESSED	CHAUFFEUR	DIFFICULT	GROUPMENT	NECESSARY	PROXIMITY
ADDRESSES	CHRONICAL	DIMENSION	GYROMETER	NECESSITY	RADIATION
ADMISSION	CIGARETTE	DIRECTION	HOSTILITY	NEGLIGENT	RADIOGRAM
ADVANCING	CIRCULATE	DIRIGIBLE	HURRICANE	NEWSPAPER	READINESS
ADVANTAGE	CIVILIANS	DISAPPEAR	IDENTICAL	NORTHEAST	REARGUARD
AERODROME	CLEARANCE	DISCUSSED	IMMEDIATE	NORTHERLY	REBELLION
AEROPLANE	COALITION	DISINFECT	IMPORTANT	NORTHWARD	RECEIVING
AFTERNOON	COLLAPSED	DISMISSAL	IMPRESSED	NORTHWEST	RECOGNIZE
AGREEMENT	COLLISION	DISPERSED	INCENTIVE	NUMBERING	RECOMMEND
AIRDROMES	COMBATANT	DISTRICTS	INCIDENCE	OBJECTION	REENFORCE
AIRPLANES	COMMANDED	DIVISIONS	INCIDENTS	OBJECTIVE	REFERENCE
ALLOTMENT	COMMANDER	DOMINANCE	INCLINING	OBTAINING	REFILLING
ALLOWANCE	COMMITTEE	DOMINATED	INCLUDING	OCCUPYING	REGARDING
ALTERNATE	COMPANIES	ECHELONED	INCLUSIVE	OFFENSIVE	REINFORCE
AMBULANCE	COMPELLED	EFFECTIVE	INCREASED	OFFICIALS	REINSTATE
AMUSEMENT	COMPLETED	EFFICIENT	INDEMNITY	OPERATING	REMAINDER
ANNOUNCED	CONDEMNED	ELABORATE	INDICATED	OPERATION	REMAINING
ANONYMOUS	CONDENSED	ELEVATION	INFLATION	OSCILLATE	REPRESENT
APPARATUS	CONDITION	ELSEWHERE	INFLICTED	OUTSKIRTS	REPRISALS
APPOINTED	CONFERRED	EMBASSIES	INFLUENCE	PARACHUTE	REQUESTED
ARBITRARY	CONFIDENT	EMERGENCY	INHABITED	PARAGRAPH	REQUIRING
ARTILLERY	CONFLICTS	EMPLOYING	INSTANTLY	PARTITION	RESOURCES
ASCENSION	CONQUERED	ENDURANCE	INTEGRITY	PASSENGER	RESTRAINT
ASSAULTED	CONTINUAL	ENGINEERS	INTENSIVE	PATRIOTIC	RETENTION
ASSISTANT	CONTINUED	ENLISTING	INTENTION	PENETRATE	RETURNING
ASSOCIATE	CONTINUES	ENTRAINED	INTERCEPT	PERMANENT	REVIEWING
ASSURANCE	COOPERATE	EQUIPMENT	INTERDICT	PERSONNEL	SCREENING
ATTACKING	CORRECTED	ESTABLISH	INTERFERE	PLACEMENT	SEAPLANES
ATTEMPTED	CRITICISE	ESTIMATED	INTERMENT	POLITICAL	SECRETARY
ATTENTION	CRITICISM	ESTIMATES	INTERPOSE	POPULATED	SEMICOLON
AUTOMATIC	DEBARKING	EXCESSIVE	INTERRUPT	POSITIONS	SEMIRIGID
AVAILABLE	DECREASED	EXCLUSION	INTERVENE	PRACTICAL	SEPTEMBER
BALLISTIC	DEFECTIVE	EXCLUSIVE	INTERVIEW	PRECEDING	SERIOUSLY
BAROMETER	DEFENSIVE	EXECUTIVE	INVENTION	PREFERRED	SERVICING
BATTALION	DEFICIENT	EXERCISES	IRREGULAR	PREMATURE	SEVENTEEN
BATTERIES	DEPARTURE	EXHIBITED	KILOMETER	PREPARING	SHELLFIRE
BEACHHEAD	DEPENDENT	EXPANSION	LAUNCHING	PRESIDENT	SITUATION
BEGINNING	DESCRIBED	EXPANSIVE	LIABILITY	PRINCIPAL	SIXTEENTH
BLOCKADED	DESIGNATE	EXPENSIVE	LOGISTICS	PRINCIPLE	SOUTHEAST
BOMBARDED	DESTITUTE	EXPLOSION	LONGITUDE	PRISONERS	SOUTHWARD
BRIGADIER	DESTROYED	EXPLOSIVE	MAINTAINS	PROCEDURE	SOUTHWEST
BUILDINGS	DESTROYER	EXTENDING	MANGANESE	PROCEEDED	SPEARHEAD
CABLEGRAM	DETENTION	EXTENSION	MECHANISM	PROJECTOR	STANDARDS
CAMPAIGNS	DETERMINE	EXTENSIVE	MEMORANDA	PROMOTION	STATEMENT
CANCELLED	DETONATED	FIFTEENTH	MESSENGER	PROPOSALS	STRAGGLER
CARTRIDGE	DETRAINED	FIREALARM	MOTORIZED	PROTECTED	STRATEGIC

NINE LETTER WORDS—Continued

SUBMITTED	SUSPENDED	TELEPHONE	THEREFORE	UNTENABLE	WEDNESDAY
SUCCEEDED	SUSPICION	TENTATIVE	TRANSPORT	VARIATION	WITNESSES
SURRENDER	TECHNICAL	TERRITORY	TWENTIETH	WATERTANK	YESTERDAY
SUSPECTED	TECHNIQUE				

TEN LETTER WORDS

ACCEPTABLE	COLLISIONS	DESPATCHES	EXPENDABLE	MAINTAINED
ACCEPTANCE	COMMANDANT	DESTROYERS	EXPERIENCE	MANAGEMENT
ACCIDENTAL	COMMANDEER	DETACHMENT	EXPERIMENT	MECHANIZED
ACCORDANCE	COMMANDING	DETERMINED	EXPLOSIONS	MEMORANDUM
ACTIVITIES	COMMISSARY	DETONATION	EXTINGUISH	MILLIMETER
ADDITIONAL	COMMISSION	DETRAINING	FACILITIES	MOTORCYCLE
AIRCONTROL	COMMITMENT	DETRUCKING	FLASHLIGHT	NATURALIZE
AIRSUPPORT	COMMUNIQUE	DIFFERENCE	FORMATIONS	NAVIGATION
ALLEGIANCE	COMPENSATE	DIPLOMATIC	FOUNDATION	NEGLIGENCE
ALLOCATION	COMPLETELY	DIRECTIONS	FOURTEENTH	NEWSPAPERS
AMBASSADOR	COMPRESSED	DISCIPLINE	FRONTLINES	NINETEENTH
AMMUNITION	CONCERNING	DISCUSSION	GEOGRAPHIC	OBJECTIVES
ANTEDATING	CONCESSION	DISPATCHED	GONIOMETER	OCCUPATION
ANTICIPATE	CONCLUSION	DISPATCHER	GOVERNMENT	ONEHUNDRED
APPARENTLY	CONDITIONS	DISPATCHES	GYROSCOPIC	OPERATIONS
APPEARANCE	CONFERENCE	DISPERSION	HYDROMETER	OPPOSITION
APPROACHED	CONFESSION	DISTRESSED	HYGROMETER	OVERCOMING
ARMOREDCAR	CONFIDENCE	DISTRIBUTE	ILLITERATE	PATROLLING
ARTIFICIAL	CONNECTING	DIVEBOMBER	ILLUMINATE	PERMISSION
ASPOSSIBLE	CONNECTION	DOMINATION	ILLUSTRATE	PERSISTENT
ASSEMBLIES	CONSPIRACY	EFFICIENCY	IMPASSABLE	PHOSPHORUS
ASSESSMENT	CONSTITUTE	EIGHTEENTH	IMPOSSIBLE	POPULATION
ASSIGNMENT	CONTINGENT	ELEMENTARY	IMPRESSION	POSSESSION
ASSISTANCE	CONTINUOUS	EMPLOYMENT	IMPRESSIVE	POSTOFFICE
ATOMICBOMB	CONTRABAND	ENCIPHERED	INCENDIARY	PRECEDENCE
ATTACHMENT	CONVENIENT	ENCIRCLING	INDICATING	PREFERENCE
ATTAINMENT	COORDINATE	ENEMYTANKS	INDICATION	PRESCRIBED
ATTEMPTING	CORRECTION	ENGAGEMENT	INDIVIDUAL	PROHIBITED
AUDIBILITY	CREDENTIAL	ENLISTMENT	INFLICTING	PROPORTION
AUTOMOBILE	CROSSROADS	ENROLLMENT	INSECURITY	PROTECTION
BALLISTICS	DEBOUCHING	ENTERPRISE	INSPECTION	PROVISIONS
BATTLESHIP	DECIPHERED	ENTRENCHED	INSTRUCTED	QUARANTINE
BEENNEEDED	DECORATION	ENTRUCKING	INSTRUCTOR	RECEPTACLE
BRIDGEHEAD	DEDICATION	EQUIVALENT	INSTRUMENT	RECREATION
CAMOUFLAGE	DEFICIENCY	ESTIMATION	INTERNMENT	RECRUITING
CAPABILITY	DEFINITION	EVACUATING	INVITATION	REENFORCED
CASUALTIES	DEMOBILIZE	EVACUATION	IRRIGATION	REENLISTED
CENSORSHIP	DEPARTMENT	EVALUATION	KILOMETERS	REGIMENTAL
CENTRALIZE	DEPENDABLE	EXCAVATION	LABORATORY	REGULATION
CIRCUITOUS	DEPLOYMENT	EXCITEMENT	LIEUTENANT	REINFORCED
COASTGUARD	DEPRESSION	EXHIBITION	LIMITATION	RESISTANCE
COLLECTING	DESIGNATED	EXPEDITING	LOCOMOTIVE	RESPECTFUL
COLLECTION	DESPATCHED	EXPEDITION	MACHINEGUN	RESTRICTED

TEN LETTER WORDS—Continued

REVOLUTION	SUBMISSION	SUSPENSION	TRANSPORTS	UNEXPENDED
SANITATION	SUBSTITUTE	SUSPICIONS	TRANSVERSE	UNSUITABLE
SEPARATION	SUCCESSFUL	SUSPICIOUS	TROOPSHIPS	VICTORIOUS
SIGNALLING	SUCCESSIVE	THIRTEENTH	TWENTYFIVE	VISIBILITY
SIMILARITY	SUFFICIENT	THREATENED	UNDERSTAND	WILLATTACK
STATISTICS	SUPPORTING	TRAJECTORY	UNDERSTOOD	WITHDRAWAL
SUBMARINES				

ELEVEN LETTER WORDS

ACCESSORIES	CONCEALMENT	EMBARKATION	INTERCEPTED	REAPPOINTED
AERONAUTICS	CONCENTRATE	EMPLACEMENT	INTERESTING	RECOGNITION
ACKNOWLEDGE	CONFINEMENT	ENCOUNTERED	INTERFERING	RECOMMENDED
ALTERNATING	CONSTITUTED	ENEMYPLANES	INTERPRETER	RECONNOITER
APPLICATION	CONSUMPTION	ENFORCEMENT	INTERRUPTED	REPLACEMENT
APPOINTMENT	CONTINENTAL	ENGAGEMENTS	INTERVENING	REQUIREMENT
APPROACHING	CONTROVERSY	ENGINEERING	INVESTIGATE	REQUISITION
APPROPRIATE	COOPERATION	ESTABLISHED	LEGISLATION	RESERVATION
APPROXIMATE	CORPORATION	ESTIMATEDAT	LIGHTBOMBER	RESIGNATION
ARBITRATION	CORRECTNESS	EXAMINATION	MAINTENANCE	RESPONSIBLE
ARMOREDCARS	CREDENTIALS	EXPLANATION	MANUFACTURE	RESTRICTION
ARRANGEMENT	CUSTOMHOUSE	EXTENSIVELY	MEASUREMENT	RETALIATION
ASSESSMENTS	DEBARKATION	EXTERMINATE	NATIONALISM	RETROACTIVE
ASSIGNMENTS	DEMONSTRATE	FINGERPRINT	NATIONALITY	SCHOOLHOUSE
ASSOCIATION	DESCRIPTION	FIRECONTROL	NAVALATTACK	SEVENTEENTH
BATTLEFIELD	DESCRIPTIVE	HEAVYBOMBER	NAVALBATTLE	SEVENTYFIVE
BATTLESHIPS	DESIGNATION	HEAVYLOSSES	NAVALFORCES	SIGNIFICANT
BELLIGERENT	DESTRUCTION	HOSTILITIES	NECESSITATE	SMOKESCREEN
BLOCKBUSTER	DETERIORATE	IMMEDIATELY	OBSERVATION	STRATEGICAL
BOMBARDMENT	DEVELOPMENT	IMMIGRATION	OVERWHELMED	SUBSISTENCE
CATASTROPHE	DISAPPEARED	IMPEDIMENTA	PARENTHESES	SUITABILITY
CERTIFICATE	DISCONTINUE	IMPROVEMENT	PARENTHESIS	SUPERIORITY
CIRCULATION	DISCREPANCY	INCOMPETENT	PENETRATION	SURRENDERED
COEFFICIENT	DISINFECTED	INDEPENDENT	PERFORMANCE	SYNCHRONIZE
COINCIDENCE	DISPOSITION	INFLAMMABLE	PHILIPPINES	TEMPERATURE
COMMUNICATE	DISTINCTION	INFORMATION	PHOTOGRAPHY	THERMOMETER
COMMUNIQUES	DISTINGUISH	INSPIRATION	PREARRANGED	TOPOGRAPHIC
COMPARTMENT	DYNAMOMETER	INSTITUTION	PREPARATION	TRADITIONAL
COMPETITION	ECHELONMENT	INSTRUCTION	PRELIMINARY	TRANSFERRED
COMPOSITION	EFFECTIVELY	INSTRUMENTS	PROGRESSIVE	WITHDRAWING
COMPUTATION	ELECTRICITY	INTELLIGENT	RANGEFINDER	

TWELVE LETTER WORDS

ADVANTAGEOUS	CARELESSNESS	CONCENTRATED	CONSIDERABLE	COORDINATION
AGRICULTURAL	COMMENCEMENT	CONCILIATION	CONSTITUTING	DECENTRALIZE
ANNOUNCEMENT	COMMENDATION	CONFIDENTIAL	CONSTITUTION	DECIPHERMENT
ANTIAIRCRAFT	COMMISSIONED	CONFIRMATION	CONSTRUCTION	DEMONSTRATED
ANTICIPATION	COMMISSIONER	CONFISCATION	CONTINUATION	DEPARTMENTAL
BREAKTHROUGH	COMPENSATION	CONFORMATION	CONVALESCENT	DIFFICULTIES
CANCELLATION	COMPLETENESS	CONSCRIPTION	CONVERSATION	DISORGANIZED

TWELVE LETTER WORDS—Continued

DISPLACEMENT	HYDROGRAPHIC	INTERVENTION	PREPAREDNESS	SHARPSHOOTER
DISSEMINATED	ILLUMINATING	INTRODUCTION	PRESERVATION	SIGNIFICANCE
DISTRIBUTING	ILLUMINATION	INTRODUCTORY	PRESIDENTIAL	SIMULTANEOUS
DISTRIBUTION	ILLUSTRATION	IRREGULARITY	PROCLAMATION	SOUTHWESTERN
EMPLACEMENTS	INAUGURATION	LIGHTBOMBERS	PSYCHROMETER	SUBSTITUTION
ENCIPHERMENT	INCOMPETENCE	MARKSMANSHIP	RADIOSTATION	SUCCESSFULLY
ENTANGLEMENT	INEFFICIENCY	MEASUREMENTS	RECREATIONAL	TRANSFERRING
ENTERPRISING	INSTRUCTIONS	MEDIUMBOMBER	REENLISTMENT	TRANSMISSION
FIGHTERPLANE	INTELLIGENCE	MOBILIZATION	REGISTRATION	TRANSPACIFIC
GENERALALARM	INTERCEPTION	NONCOMBATANT	REPLACEMENTS	UNIDENTIFIED
GENERALSTAFF	INTERDICTION	NORTHWESTERN	RESPECTFULLY	UNITEDSTATES
GEOGRAPHICAL	INTERFERENCE	OBSTRUCTIONS	ROADJUNCTION	UNSUCCESSFUL
HEADQUARTERS	INTERMEDIATE	ORGANIZATION	SATISFACTORY	VERIFICATION
HEAVYBOMBERS	INTERRUPTION	PREPARATIONS	SEARCHLIGHTS	VETERINARIAN

THIRTEEN LETTER WORDS

ACCOMMODATION	CORRESPONDING	DISTINGUISHED	INSTANTANEOUS	REAPPOINTMENT
APPROXIMATELY	COUNTERATTACK	ENTERTAINMENT	INTERNATIONAL	REENFORCEMENT
CHRONOLOGICAL	DECENTRALIZED	ESTABLISHMENT	INVESTIGATION	REIMBURSEMENT
CIRCUMSTANCES	DEMONSTRATION	EXTERMINATION	MEDIUMBOMBERS	REINFORCEMENT
COMMUNICATION	DEPENDABILITY	EXTRAORDINARY	MISCELLANEOUS	REINSTATEMENT
CONCENTRATING	DETERMINATION	FIGHTERPLANES	PRELIMINARIES	REVOLUTIONARY
CONCENTRATION	DISAPPEARANCE	IMPRACTICABLE	QUALIFICATION	SPECIFICATION
CONGRESSIONAL	DISCREPANCIES	INDETERMINATE	QUARTERMASTER	TRANSATLANTIC
CONSIDERATION	DISSEMINATION	INSTALLATIONS		

FOURTEEN LETTER WORDS

ADMINISTRATION	DEMOBILIZATION	IRREGULARITIES	RECONSTRUCTION
ADMINISTRATIVE	DISCONTINUANCE	METEOROLOGICAL	REORGANIZATION
CENTRALIZATION	DISTINGUISHING	NATURALIZATION	REPRESENTATIVE
CHARACTERISTIC	IDENTIFICATION	RECOMMENDATION	RESPONSIBILITY
CIRCUMSTANTIAL	INTERPRETATION	RECONNAISSANCE	SATISFACTORILY
CLASSIFICATION	INVESTIGATIONS	RECONNOITERING	TRANSPORTATION
CORRESPONDENCE			

B. LIST OF WORDS USED IN MILITARY TEXT ARRANGED IN RHYMING ORDER ACCORDING TO WORD LENGTH

THREE LETTER WORDS

SEA	SEE	MAJ	TAN	TOP	EAT	BUT	FIX
JOB	AGE	ADJ	GEN	GHQ	MAT	CUT	MIX
ROB	SHE	ASK	MEN	BAR	VAT	OUT	SIX
TUB	THE	GAL	PEN	CAR	ACT	PUT	BOX
QMC	DIE	ALL	TEN	FAR	GET	PVT	FOX
ARC	ONE	ILL	PIN	PAR	LET	CWT	DAY
BAD	ARE	COL	TIN	WAR	NET	YOU	LAY
HAD	USE	CPL	TON	HER	SET	CAV	MAY
ADD	DUE	CAM	WON	PER	WET	LAW	PAY
RED	OWE	HAM	DUN	AIR	YET	SAW	SAY
AID	EYE	AIM	GUN	FOR	SGT	FEW	WAY
BID	OFF	HIM	RUN	OUR	WGT	NEW	ANY
DID	BAG	ARM	SUN	GAS	FIT	HOW	SPY
RID	KEG	SUM	OWN	HAS	GOT	LOW	DRY
OLD	BIG	CAN	AGO	WAS	LOT	NOW	TRY
AND	JIG	MAN	TOO	HIS	NOT	TAX	BUY
END	DOG	NAN	TWO	ITS	APT		

FOUR LETTER WORDS

AREA	MIKE	BASE	WEEK	WELL	NOON	PASS	LIST
ASIA	YOKE	FUSE	TALK	HILL	SOON	LESS	LOST
BULB	ABLE	DATE	BULK	WILL	DOWN	MESS	POST
BOMB	FILE	LATE	RANK	FULL	TOWN	LOSS	JUST
HEAD	MILE	NOTE	SANK	TOOL	ZERO	HITS	ROUT
LEAD	MULE	BLUE	TANK	TEAM	ALSO	DAYS	NEXT
LOAD	RULE	HAVE	SUNK	THEM	INTO	MEAT	TEXT
ROAD	SAME	FIVE	BOOK	ITEM	KEEP	THAT	LIEU
RAID	TIME	LOVE	COOK	MAIM	SHIP	WHAT	DRAW
SAID	SOME	MOVE	HOOK	FROM	DUMP	FEET	XRAY
HOLD	LINE	FUZE	LOOK	FARM	PUMP	MEET	AWAY
HAND	MINE	HALF	TOOK	FIRM	STOP	LEFT	BODY
LAND	NINE	FLAG	DARK	FORM	NEAR	OMIT	THEY
KIND	ZONE	KING	PARK	THAN	REAR	UNIT	ALLY
HARD	JUNE	LONG	MASK	PLAN	OVER	HALT	ONLY
HERD	OBOE	EACH	TASK	BEEN	FOUR	TENT	JULY
ONCE	PIPE	HIGH	ORAL	SEEN	YOUR	SHOT	ARMY
MADE	TYPE	DASH	MTCL	THEN	EYES	RIOT	MANY
AIDE	TARE	PUSH	FEEL	WHEN	THIS	DIRT	VARY
SIDE	HERE	RUSH	RAIL	OPEN	AXIS	EAST	VERY
CODE	WERE	WITH	CALL	MAIN	TONS	FAST	EASY
FLEE	FIRE	BOTH	FALL	RAIN	GUNS	LAST	CITY
EDGE	WIRE	LEAK	CELL	JOIN	MASS	WEST	NAVY
TAKE	MORE	BACK	FELL				

FIVE LETTER WORDS

COMMA	SCALE	ALONG	CANAL	WAGON	PRIOR	DRESS	START
ZEBRA	TITLE	AMONG	FATAL	UNION	MAJOR	PRESS	ALERT
SQUAD	ALINE	BEACH	VITAL	COLON	VALOR	CROSS	LEAST
SPEED	SLOPE	REACH	TOTAL	DRAWN	ARMOR	FLATS	COAST
WIPED	FLARE	WHICH	EQUAL	RADIO	HONOR	BOATS	CREST
RIGID	THERE	MARCH	USUAL	EQUIP	ERROR	RAFTS	GUEST
RAPID	WHERE	WEIGH	NAVAL	TROOP	MOTOR	UNITS	FIRST
FIELD	SHORE	FRESH	WHEEL	GROUP	AREAS	TRACT	BURST
BLIND	CEASE	WIDTH	STEEL	CLEAR	BOMBS	FLEET	ABOUT
GUARD	ERASE	FIFTH	REPEL	SUGAR	RAIDS	QUIET	ALLOW
AWARD	THESE	TENTH	LEVEL	UNDER	WOODS	ASSET	ANNEX
THIRD	CLOSE	NINTH	APRIL	ORDER	YARDS	SHIFT	TODAY
BRIBE	HORSE	BOOTH	SMALL	DEFER	MILES	EIGHT	DELAY
PLACE	CAUSE	DEPTH	SHELL	REFER	FIRES	FIGHT	READY
VOICE	HOUSE	NORTH	SPELL	EAGER	CASES	LIGHT	FOGGY
FORCE	ROUTE	SOUTH	DRILL	ROGER	GATES	NIGHT	DAILY
TRUCE	ISSUE	SIXTH	ALARM	ETHER	PACKS	RIGHT	RALLY
THREE	LEAVE	BREAK	JAPAN	OTHER	DECKS	SIGHT	APPLY
RIDGE	DRIVE	BLACK	QUEEN	BAKER	DOCKS	AWAIT	EARLY
SIEGE	PROVE	CHECK	TAKEN	LATER	BANKS	SPLIT	ENEMY
RANGE	CURVE	QUICK	SEVEN	METER	TANKS	LIMIT	EVERY
BARGE	SEIZE	TRUCK	GIVEN	PETER	PLANS	VISIT	FERRY
LARGE	CHIEF	CREEK	ALIGN	AFTER	SHIPS	AGENT	FIFTY
GAUGE	STAFF	FLANK	AGAIN	ENTER	CORPS	POINT	PARTY
STAKE	PROOF	CLERK	PLAIN	RIVER	FEARS	FRONT	FORTY
SMOKE	BEING	LOCAL	TRAIN	COVER	PAIRS	COUNT	SIXTY
BROKE	GOING	VOCAL	BEGIN	THEIR	HOURS	DEPOT	HEAVY

SIX LETTER WORDS

CANADA	HALTED	DEVICE	CHARGE	SEVERE	ARRIVE	TRENCH	MANUAL
ARABIA	ROUTED	NOVICE	GEORGE	RETIRE	ACTIVE	LAUNCH	ANNUAL
ALASKA	LIQUID	FIERCE	REFUGE	ENTIRE	TWELVE	SEARCH	CASUAL
PANAMA	INLAND	REDUCE	MORALE	BEFORE	BREEZE	CHURCH	VISUAL
METRIC	ISLAND	PARADE	UNABLE	SECURE	RELIEF	SWITCH	CANCEL
CRITIC	DEFEND	DECIDE	CIRCLE	ASSURE	ZIGZAG	THOUGH	VESSEL
BOMBED	OFFEND	DIVIDE	SINGLE	FUTURE	RIDING	FINISH	DETAIL
BARBED	DEPEND	DECODE	MOBILE	GREASE	FILING	EIGHTH	REFILL
RAIDED	EXPEND	ENCODE	BEETLE	CHEESE	LINING	FOURTH	ENROLL
LANDED	INTEND	COFFEE	BATTLE	ADVISE	MINING	ATTACK	SCHOOL
WOODED	EXTEND	DECREE	SETTLE	DEVISE	FIRING	DEBARK	PATROL
INDEED	SECOND	DEGREE	LITTLE	OPPOSE	WIRING	EMBARK	PISTOL
ALLIED	BEYOND	STRAFE	NOZZLE	COURSE	DURING	VERBAL	SYSTEM
KILLED	GROUND	ENGAGE	MUZZLE	REFUSE	NOTING	RADIAL	VICTIM
FORMED	METHOD	DAMAGE	SCHEME	LOCATE	MOVING	SERIAL	SIGCOM
DOWNED	PERIOD	MANAGE	RESUME	EXCITE	FLYING	ANIMAL	BOTTOM
SCORED	RECORD	GARAGE	ENGINE	MINUTE	BREACH	FORMAL	INFORM
PASSED	OFFICE	BRIDGE	RAVINE	RESCUE	DETACH	NORMAL	MEDIUM
CAUSED	POLICE	ALLEGE	EUROPE	LEAGUE	ATTACH	SIGNAL	SUDDEN
UNITED	ADVICE	CHANGE	SPHERE	PURSUE	BRANCH	POSTAL	SCREEN

SIX LETTER WORDS—Continued

SUNKEN	MORTAR	RUNNER	FORCES	COLORS	TARGET	CANNOT	MONDAY
BROKEN	RUBBER	KEEPER	BARGES	ACCESS	PICKET	ACCEPT	SUNDAY
SEAMEN	MEMBER	HELPER	BODIES	EXCESS	ROCKET	EXCEPT	ANYWAY
HAPPEN	BOMBER	PROPER	ALLIES	UNLESS	BILLET	PROMPT	REMEDY
BATTEN	NUMBER	NEARER	ARMIES	STRESS	TURRET	DEPART	VALLEY
ELEVEN	PINCER	ERASER	TABLES	ACROSS	SUNSET	DESERT	PARLEY
REMAIN	LEADER	CENTER	PLANES	ASSETS	WEIGHT	DIVERT	CONVEY
ATTAIN	LADDER	BETTER	PASSES	VISITS	FLIGHT	ESCORT	SURVEY
WITHIN	MURDER	LETTER	LOSSES	POINTS	SLIGHT	EFFORT	VERIFY
COLUMN	PREFER	BITTER	STATES	STATUS	NAUGHT	REPORT	SUPPLY
RATION	SUFFER	LITTER	ROUTES	ALWAYS	FOUGHT	ARREST	HOURLY
ACTION	MEAGER	AFFAIR	ISSUES	COMBAT	NOUGHT	RESIST	DEPLOY
COMMON	HIGHER	REPAIR	CRISIS	DEFEAT	CREDIT	ASSIST	EMPLOY
SUMMON	CIPHER	HARBOR	SHELLS	THREAT	SUBMIT	AUGUST	CONVOY
POISON	EITHER	TERROR	SPOOLS	DEFECT	COMMIT	ADJUST	OCCUPY
LESSON	TANKER	MIRROR	TRAINS	EFFECT	SUMMIT	DUGOUT	SALARY
PONTON	HAMMER	SECTOR	SPOONS	REJECT	RESULT	OUTPUT	ARMORY
RETURN	SUMMER	VICTOR	STRIPS	SELECT	ORIENT	BUREAU	NINETY
DRYRUN	BANNER	DOCTOR	TROOPS	EXPECT	INTENT	REVIEW	EIGHTY
TATTOO	MANNER	CANVAS	ORDERS	DIRECT	EXTENT	FOLLOW	TWENTY
APPEAR	GUNNER	PLACES	OTHERS	STREET	INVENT	FRIDAY	THIRTY
DOLLAR							

SEVEN LETTER WORDS

MILITIA	RETIRED	WINDAGE	DECLARE	COMMUTE	FISHING	VARYING
ANTENNA	ARMORED	BAGGAGE	PREPARE	REVENUE	PUSHING	ICEBERG
ALMANAC	PRESSED	PACKAGE	CALIBRE	RELIEVE	NOTHING	DEBOUCH
BIVOUAC	CROSSED	VILLAGE	MISFIRE	RECEIVE	TALKING	THROUGH
TRAFFIC	OMITTED	TONNAGE	INSPIRE	PASSIVE	SINKING	FURNISH
PACIFIC	DELAYED	AVERAGE	REQUIRE	CAPTIVE	SMOKING	TWELFTH
ASIATIC	COMMAND	STORAGE	INQUIRE	REVOLVE	FALLING	SEVENTH
REDUCED	COMMEND	BARRAGE	LECTURE	APPROVE	FILLING	SETBACK
INVADED	SUSPEND	PASSAGE	RELEASE	OBSERVE	KILLING	DERRICK
DECIDED	RESPOND	MESSAGE	DISEASE	RESERVE	EVENING	DETRUCK
DECODED	BOMBARD	COLLEGE	SUNRISE	ANALYZE	RAINING	ENTRUCK
ENCODED	AWKWARD	ARRANGE	LICENSE	JUMPOFF	MANNING	MEDICAL
WOUNDED	FORWARD	WITHTHE	DEFENSE	BOMBING	RUNNING	LOGICAL
GUARDED	REPLACE	THATTHE	OFFENSE	PLACING	MORNING	CONCEAL
PROCEED	SERVICE	CHARLIE	PROPOSE	FORCING	SLOPING	ILLEGAL
ENGAGED	ADVANCE	PRAIRIE	SUPPOSE	HEADING	MAPPING	MARSHAL
DAMAGED	ABSENCE	VISIBLE	PURPOSE	LEADING	BEARING	INITIAL
REACHED	ENFORCE	BICYCLE	REVERSE	LOADING	GASSING	MARTIAL
MARCHED	BRIGADE	HOSTILE	BECAUSE	BEDDING	MESSING	FEDERAL
WRECKED	GRENADE	EXTREME	MANDATE	RAIDING	MISSING	GENERAL
SHELLED	PRECEDE	CONFINE	RADIATE	HOLDING	LIFTING	SEVERAL
DROPPED	OUTSIDE	MACHINE	OPERATE	LANDING	HALTING	CENTRAL
STOPPED	INCLUDE	ROUTINE	ELEVATE	BINDING	GETTING	NATURAL
HUNDRED	EXCLUDE	CYCLONE	ENTENTE	FINDING	FITTING	COASTAL
ORDERED	REFUGEE	WARFARE	PROMOTE	FORGING	ISSUING	GRADUAL
COVERED						

SEVEN LETTER WORDS—Continued

UNUSUAL	ENTRAIN	ENVELOP	STARTER	SUCCESS	ASSAULT	RAILWAY
ARRIVAL	CONTAIN	SIMILAR	QUARTER	USELESS	INSTANT	SECRECY
CHANNEL	CAPTAIN	REGULAR	DELIVER	ILLNESS	ELEMENT	VACANCY
COLONEL	CONDEMN	CALIBER	RECOVER	WITNESS	COMMENT	SIGNIFY
COUNCIL	ABANDON	OCTOBER	AVIATOR	ADDRESS	CURRENT	SATISFY
FUELOIL	OPINION	OFFICER	TRACTOR	EXPRESS	PRESENT	RAPIDLY
INSTALL	SESSION	POUNDER	VISITOR	DISMISS	APPOINT	QUICKLY
DISTILL	MISSION	TRIGGER	TACTICS	DISCUSS	RECEIPT	NIGHTLY
PAYROLL	STATION	WEATHER	ISLANDS	TARGETS	ATTEMPT	SHORTLY
CONTROL	SECTION	WHETHER	CHANGES	SURPLUS	SUPPORT	COMPANY
WILLIAM	ECHELON	ANOTHER	ENEMIES	RETREAT	SUGGEST	DESTROY
DIAGRAM	BALLOON	FARTHER	BATTLES	EXTRACT	HIGHEST	PRIMARY
PROGRAM	PLATOON	FURTHER	GLASSES	CONTACT	NEAREST	SUMMARY
MINIMUM	LIAISON	SOLDIER	CHASSIS	COLLECT	PROTEST	LIBRARY
MAXIMUM	HORIZON	CARRIER	ATTACKS	RESPECT	REQUEST	JANUARY
HASBEEN	EASTERN	COURIER	VESSELS	CORRECT	AGAINST	BRIBERY
FIFTEEN	WESTERN	HEAVIER	PATROLS	PROTECT	OUTPOST	BATTERY
SIXTEEN	FOGHORN	TRAWLER	BOMBERS	INFLICT	PROVOST	INQUIRY
BETWEEN	UNKNOWN	STEAMER	NUMBERS	CONDUCT	BOYCOTT	CAVALRY
KITCHEN	TOBACCO	CLIPPER	REPAIRS	TONIGHT	WITHOUT	VICTORY
WRITTEN	TORPEDO	CRUISER	SAILORS	CIRCUIT	LOOKOUT	EMBASSY
EXPLAIN	WARSHIP	AMMETER	SECTORS	RECRUIT	SIMPLEX	UTILITY
TERRAIN	DEVELOP	FIGHTER	COMPASS	PURSUIT	TUESDAY	SEVENTY
DETRAIN						

EIGHT LETTER WORDS

INSIGNIA	EXPELLED	DICTATED	STANDARD	LANGUAGE	ENVELOPE	OPPOSITE
SPECIFIC	ENROLLED	EFFECTED	OUTBOARD	DISLODGE	INSECURE	CONTINUE
TERRIFIC	DISARMED	INFECTED	OUTGUARD	EXCHANGE	PRESSURE	CRITIQUE
ECONOMIC	ASSIGNED	REJECTED	WINDWARD	PROBABLE	DECREASE	THATHAVE
MECHANIC	RETURNED	SELECTED	EASTWARD	SUITABLE	EXERCISE	DECISIVE
ATLANTIC	APPEARED	BILLETED	WESTWARD	ELIGIBLE	SURPRISE	POSITIVE
RAILHEAD	DECLARED	INVENTED	DESCRIBE	TERRIBLE	SUSPENSE	PRESERVE
RAILROAD	PREPARED	DEPARTED	ORDNANCE	POSSIBLE	DISPERSE	EQUALIZE
REPLACED	HINDERED	DESERTED	DISTANCE	FLEXIBLE	TRAVERSE	MOBILIZE
ADVANCED	SUFFERED	ESCORTED	COMMENCE	ASSEMBLE	DEDICATE	INVADING
DEMANDED	CENTERED	DEPORTED	SENTENCE	OBSTACLE	INDICATE	DIVIDING
EXPANDED	BATTERED	REPORTED	ANNOUNCE	ENCIRCLE	INITIATE	BUILDING
DEFENDED	LETTERED	ARRESTED	COMMERCE	SCHEDULE	ESTIMATE	GUARDING
OFFENDED	REPAIRED	ENLISTED	ENFILADE	MARITIME	ORDINATE	ENGAGING
EXPENDED	REQUIRED	SURVIVED	CONCLUDE	AIRDROME	DETONATE	DAMAGING
EXTENDED	RESTORED	IMPROVED	LATITUDE	AIRPLANE	SEPARATE	MARCHING
GROUNDED	DEFERRED	OBSERVED	ALTITUDE	JETPLANE	EVACUATE	BREAKING
BESIEGED	CAPTURED	REVIEWED	EMPLOYEE	MEDICINE	EXCAVATE	FLANKING
DETACHED	REPULSED	DEPLOYED	CARRIAGE	DOCTRINE	OBSOLETE	TOTALING
FINISHED	COMPOSED	AIRFIELD	FUSELAGE	POSTPONE	COMPLETE	SHELLING
OCCUPIED	MANDATED	FOOTHOLD	EQUIPAGE	SEABORNE	CONCRETE	BATTLING
ATTACKED	DEFEATED	THOUSAND	FRONTAGE	AIRBORNE	EXPEDITE	SWIMMING
REPELLED	REPEATED	SURROUND	SABOTAGE	DEVELOPE	DEFINITE	TRAINING

EIGHT LETTER WORDS—Continued

PLANNING	ELEVENTH	CAMPAIGN	PRISONER	VEHICLES	RESPECTS	WITHDRAW
SWEEPING	ANTITANK	CHAPLAIN	IMPROPER	MISFIRES	ELEMENTS	WITHDREW
SHIPPING	CODEBOOK	MAINTAIN	REPEATER	DEFENSES	ATTEMPTS	TOMORROW
GROUPING	CHEMICAL	MOUNTAIN	DESERTER	EXPENSES	PROTESTS	PARALLAX
ENTERING	CLERICAL	BULLETIN	DISASTER	PURPOSES	OUTPOSTS	SATURDAY
COVERING	TACTICAL	INVASION	REGISTER	RESERVES	ENORMOUS	THURSDAY
RETIRING	CRITICAL	DECISION	CANISTER	ANALYSIS	LUMINOUS	CAUSEWAY
ADVISING	NAUTICAL	DIVISION	RECEIVER	BARRACKS	RIGOROUS	EFFICACY
OPPOSING	OFFICIAL	LOCATION	REVOLVER	MISSIONS	VIGOROUS	IDENTIFY
DRESSING	MATERIAL	AVIATION	OBSERVER	STATIONS	CONTRACT	STRATEGY
PRESSING	MEMORIAL	CITATION	MANEUVER	FACTIONS	INDIRECT	PROBABLY
CROSSING	NATIONAL	TAXATION	EMPLOYER	PONTOONS	CONFLICT	ASSEMBLY
DRIFTING	INTERNAL	JUNCTION	HOWITZER	WARSHIPS	DISTRICT	ACTUALLY
FIGHTING	CORPORAL	IGNITION	CORRIDOR	OFFICERS	INSTRUCT	MONOPOLY
SIGHTING	HOSPITAL	POSITION	SUPERIOR	SOLDIERS	AIRCRAFT	EASTERLY
LIMITING	APPROVAL	FORENOON	INTERIOR	CARRIERS	DAYLIGHT	WESTERLY
PAINTING	MATERIEL	SQUADRON	EXTERIOR	TRAILERS	MIDNIGHT	BOUNDARY
PRINTING	PARALLEL	GARRISON	OPERATOR	TRAWLERS	PROHIBIT	MILITARY
SPOTTING	SENTINEL	NORTHERN	DICTATOR	CRUISERS	SERGEANT	SANITARY
DELAYING	SEALEVEL	SOUTHERN	REJECTOR	FIGHTERS	DOMINANT	FEBRUARY
RALLYING	PROTOCOL	CIRCULAR	DIRECTOR	QUARTERS	ADJUTANT	CEMETERY
CARRYING	MERCIFUL	DECEMBER	DETECTOR	CARELESS	ADJACENT	ADVISORY
FERRYING	TELEGRAM	REMEMBER	ASSOONAS	WIRELESS	INCIDENT	INFANTRY
APPROACH	AMERICAN	NOVEMBER	POLITICS	BUSINESS	ARMAMENT	CAPACITY
ENTRENCH	EUROPEAN	DEFENDER	COMMANDS	DARKNESS	MOVEMENT	FATALITY
INTRENCH	CIVILIAN	RECORDER	ADVANCES	CONGRESS	REGIMENT	CALAMITY
RESEARCH	HAVEBEEN	ENGINEER	BARRAGES	PROGRESS	APPARENT	VICINITY
DESPATCH	NINETEEN	TRANSFER	MESSAGES	FORTRESS	PASSPORT	PRIORITY
DISPATCH	EIGHTEEN	DECIPHER	REMEDIES	DISTRESS	INTEREST	ACTIVITY
SKIRMISH	THIRTEEN	ENCIPHER	SUPPLIES	REDCROSS	REENLIST	CASUALTY
DIMINISH	FOURTEEN					

NINE LETTER WORDS

MEMORANDA	CANCELLED	IMPRESSED	ATTEMPTED	ASSURANCE	AERODROME
STRATEGIC	COMPELLED	DISCUSSED	PROTESTED	ALLOWANCE	HURRICANE
AUTOMATIC	DETRAINED	INDICATED	REQUESTED	INCIDENCE	AEROPLANE
PATRIOTIC	ENTRAINED	POPULATED	SUBMITTED	REFERENCE	INTERVENE
BALLISTIC	CONDEMNED	ESTIMATED	CONTINUED	INFLUENCE	FRONTLINE
BEACHHEAD	ECHELONED	DOMINATED	DESTROYED	REENFORCE	DETERMINE
SPEARHEAD	DEVELOPED	DETONATED	MOTORIZED	REINFORCE	TELEPHONE
DESCRIBED	CONQUERED	SUSPECTED	SEMIRIGID	LONGITUDE	INTERFERE
ANNOUNCED	PREFERRED	CORRECTED	RECOMMEND	COMMITTEE	ELSEWHERE
BLOCKADED	CONFERRED	PROTECTED	REARGUARD	ADVANTAGE	SHELLFIRE
SUCCEEDED	DECREASED	INFLICTED	NORTHWARD	CARTRIDGE	THEREFORE
PROCEEDED	INCREASED	COMPLETED	SOUTHWARD	CHALLENGE	PROCEDURE
COMMANDED	CONDENSED	INHABITED	AMBULANCE	AVAILABLE	PREMATURE
SUSPENDED	COLLAPSED	EXHIBITED	DOMINANCE	UNTENABLE	DEPARTURE
BOMBARDED	DISPERSED	ASSAULTED	CLEARANCE	DIRIGIBLE	NAVALBASE
FORTIFIED	ADDRESSED	APPOINTED	ENDURANCE	PRINCIPLE	MANGANESE

NINE LETTER WORDS—Continued

CRITICISE	REGARDING	PERSONNEL	INVENTION	CONTINUES	STATEMENT
INTERPOSE	ACCORDING	CABLEGRAM	PROMOTION	BUILDINGS	EQUIPMENT
ASSOCIATE	INCLUDING	RADIOGRAM	SEMICOLON	OFFICIALS	GROUPMENT
IMMEDIATE	LAUNCHING	FIREALARM	AFTERNOON	REPRISALS	INTERMENT
OSCILLATE	ATTACKING	CRITICISM	DISAPPEAR	PROPOSALS	ALLOTMENT
CIRCULATE	DEBARKING	MECHANISM	IRREGULAR	CIVILIANS	PERMANENT
DESIGNATE	REFILLING	DIETITIAN	SEPTEMBER	CAMPAIGNS	DIFFERENT
ALTERNATE	SCREENING	SEVENTEEN	COMMANDER	MAINTAINS	REPRESENT
COOPERATE	REMAINING	SUSPICION	SURRENDER	DIVISIONS	RESTRAINT
ELABORATE	OBTAINING	BATTALION	REMAINDER	MUNITIONS	INTERCEPT
PENETRATE	INCLINING	REBELLION	PASSENGER	POSITIONS	INTERRUPT
REINSTATE	BEGINNING	COLLISION	MESSENGER	ENGINEERS	TRANSPORT
CIGARETTE	RETURNING	PROVISION	BRIGADIER	PRISONERS	NORTHEAST
PARACHUTE	PREPARING	EXPANSION	STRAGGLER	READINESS	SOUTHEAST
DESTITUTE	NUMBERING	ASCENSION	NEWSPAPER	CONFLICTS	NORTHWEST
TECHNIQUE	CENTERING	DIMENSION	CHARACTER	DISTRICTS	SOUTHWEST
EXPANSIVE	REQUIRING	EXTENSION	KILOMETER	INCIDENTS	INTERVIEW
DEFENSIVE	OPERATING	EXPLOSION	BAROMETER	MOVEMENTS	YESTERDAY
OFFENSIVE	ENLISTING	ADMISSION	GYROMETER	OUTSKIRTS	WEDNESDAY
EXPENSIVE	RECEIVING	EXCLUSION	DESTROYER	ANONYMOUS	EMERGENCY
INTENSIVE	REVIEWING	RADIATION	PROJECTOR	APPARATUS	NORTHERLY
EXTENSIVE	EMPLOYING	VARIATION	PROTECTOR	DISINFECT	SERIOUSLY
EXPLOSIVE	OCCUPYING	INFLATION	CHAUFFEUR	INTERDICT	INSTANTLY
EXCESSIVE	PARAGRAPH	FORMATION	LOGISTICS	DIFFICULT	ACCOMPANY
INCLUSIVE	ESTABLISH	OPERATION	STANDARDS	COMBATANT	ARBITRARY
EXCLUSIVE	TWENTIETH	SITUATION	RESOURCES	IMPORTANT	NECESSARY
TENTATIVE	FIFTEENTH	ELEVATION	COMPANIES	ASSISTANT	SECRETARY
DEFECTIVE	SIXTEENTH	OBJECTION	BATTERIES	CONFIDENT	ARTILLERY
EFFECTIVE	WATERTANK	DIRECTION	EMBASSIES	PRESIDENT	ACCESSORY
OBJECTIVE	TECHNICAL	CONDITION	AIRDROMES	DEPENDENT	TERRITORY
INCENTIVE	CHRONICAL	COALITION	SEAPLANES	NEGLIGENT	LIABILITY
EXECUTIVE	PRACTICAL	PARTITION	AIRPLANES	DEFICIENT	HOSTILITY
RECOGNIZE	POLITICAL	DETENTION	EXERCISES	EFFICIENT	PROXIMITY
SERVICING	IDENTICAL	RETENTION	WITNESSES	PLACEMENT	INDEMNITY
ADVANCING	PRINCIPAL	INTENTION	ADDRESSES	AGREEMENT	INTEGRITY
PRECEDING	DISMISSAL	ATTENTION	ESTIMATES	AMUSEMENT	NECESSITY
EXTENDING	CONTINUAL				

TEN LETTER WORDS

ATOMICBOMB	APPROACHED	COMPRESSED	UNDERSTOOD	CONFIDENCE
GEOGRAPHIC	ENTRENCHED	DISTRESSED	COASTGUARD	NEGLIGENCE
GYROSCOPIC	DESPATCHED	DESIGNATED	POSTOFFICE	EXPERIENCE
DIPLOMATIC	DISPATCHED	RESTRICTED	ACCORDANCE	PREFERENCE
BRIDGEHEAD	THREATENED	INSTRUCTED	ALLEGIANCE	DIFFERENCE
PRESCRIBED	MAINTAINED	PROHIBITED	APPEARANCE	CONFERENCE
REENFORCED	DETERMINED	REENLISTED	ACCEPTANCE	CAMOUFLAGE
REINFORCED	ONEHUNDRED	MECHANIZED	RESISTANCE	DEPENDABLE
BEENNEEDED	DECIPHERED	CONTRABAND	ASSISTANCE	EXPENDABLE
UNEXPENDED	ENCIPHERED	UNDERSTAND	PRECEDENCE	IMPASSABLE

TEN LETTER WORDS—Continued

UNSUITABLE	EVACUATING	ALLOCATION	GONIOMETER	CONTINGENT
ACCEPTABLE	COLLECTING	FOUNDATION	HYDROMETER	SUFFICIENT
IMPOSSIBLE	CONNECTING	RECREATION	HYGROMETER	CONVENIENT
ASPOSSIBLE	INFLICTING	IRRIGATION	AMBASSADOR	EQUIVALENT
RECEPTACLE	EXPEDITING	NAVIGATION	INSTRUCTOR	ENGAGEMENT
MOTORCYCLE	RECRUITING	REGULATION	BALLISTICS	MANAGEMENT
AUTOMOBILE	ATTEMPTING	POPULATION	STATISTICS	EXCITEMENT
DISCIPLINE	SUPPORTING	ESTIMATION	CROSSROADS	DETACHMENT
QUARANTINE	EXTINGUISH	DOMINATION	DESPATCHES	ATTACHMENT
ENTERPRISE	NINETEENTH	DETONATION	DISPATCHES	EXPERIMENT
TRANSVERSE	EIGHTEENTH	OCCUPATION	ASSEMBLIES	ENROLLMENT
COORDINATE	THIRTEENTH	SEPARATION	FACILITIES	ASSIGNMENT
ILLUMINATE	FOURTEENTH	DECORATION	ACTIVITIES	ATTAINMENT
ANTICIPATE	WILLATTACK	LIMITATION	CASUALTIES	INTERNMENT
ILLITERATE	ARTIFICIAL	SANITATION	FRONTLINES	GOVERNMENT
ILLUSTRATE	CREDENTIAL	INVITATION	SUBMARINES	ASSESSMENT
COMPENSATE	ADDITIONAL	EVACUATION	OBJECTIVES	COMMITMENT
DISTRIBUTE	ACCIDENTAL	EVALUATION	ENEMYTANKS	DEPARTMENT
SUBSTITUTE	REGIMENTAL	EXCAVATION	SUSPICIONS	ENLISTMENT
CONSTITUTE	INDIVIDUAL	COLLECTION	COLLISIONS	INSTRUMENT
COMMUNIQUE	WITHDRAWAL	CONNECTION	PROVISIONS	DEPLOYMENT
TWENTYFIVE	AIRCONTROL	INSPECTION	EXPLOSIONS	EMPLOYMENT
SUCCESSIVE	SUCCESSFUL	CORRECTION	FORMATIONS	PERSISTENT
IMPRESSIVE	RESPECTFUL	PROTECTION	OPERATIONS	AIRSUPPORT
LOCOMOTIVE	MEMORANDUM	EXHIBITION	DIRECTIONS	CONSPIRACY
CENTRALIZE	SUSPENSION	EXPEDITION	CONDITIONS	DEFICIENCY
NATURALIZE	DISPERSION	DEFINITION	TROOPSHIPS	EFFICIENCY
DEMOBILIZE	CONCESSION	AMMUNITION	NEWSPAPERS	COMPLETELY
COMMANDING	CONFESSION	OPPOSITION	KILOMETERS	APPARENTLY
DEBOUCHING	DEPRESSION	PROPORTION	DESTROYERS	INCENDIARY
DETRUCKING	IMPRESSION	REVOLUTION	TRANSPORTS	COMMISSARY
ENTRUCKING	POSSESSION	MACHINEGUN	SUSPICIOUS	ELEMENTARY
ENCIRCLING	SUBMISSION	BATTLESHIP	VICTORIOUS	LABORATORY
SIGNALLING	COMMISSION	CENSORSHIP	CIRCUITOUS	TRAJECTORY
PATROLLING	PERMISSION	ARMOREDCAR	CONTINUOUS	CAPABILITY
OVERCOMING	DISCUSSION	DIVEBOMBER	PHOSPHORUS	AUDIBILITY
DETRAINING	CONCLUSION	COMMANDEER	FLASHLIGHT	VISIBILITY
CONCERNING	DEDICATION	DISPATCHER	COMMANDANT	SIMILARITY
INDICATING	INDICATION	MILLIMETER	LIEUTENANT	INSECURITY
ANTEDATING				

ELEVEN LETTER WORDS

IMPEDIMENTA	SURRENDERED	CONSTITUTED	CATASTROPHE	CUSTOMHOUSE
TOPOGRAPHIC	ENCOUNTERED	BATTLEFIELD	INFLAMMABLE	CERTIFICATE
RECOMMENDED	TRANSFERRED	PERFORMANCE	RESPONSIBLE	COMMUNICATE
PREARRANGED	DISINFECTED	MAINTENANCE	NAVALBATTLE	INVESTIGATE
ESTABLISHED	REAPPOINTED	COINCIDENCE	TEMPERATURE	APPROPRIATE
OVERWHELMED	INTERCEPTED	SUBSISTENCE	MANUFACTURE	APPROXIMATE
DISAPPEARED	INTERRUPTED	ACKNOWLEDGE	SCHOOLHOUSE	EXTERMINATE

ELEVEN LETTER WORDS—Continued

DETERIORATE
CONCENTRATE
DEMONSTRATE
NECESSITATE
DISCONTINUE
SEVENTYFIVE
PROGRESSIVE
RETROACTIVE
DESCRIPTIVE
SYNCHRONIZE
APPROACHING
INTERVENING
ENGINEERING
INTERFERING
ALTERNATING
INTERESTING
WITHDRAWING
DISTINGUISH
SEVENTEENTH
NAVALATTACK
STRATEGICAL
TRADITIONAL
CONTINENTAL
FIRECONTROL

NATIONALISM
SMOKESCREEN
APPLICATION
ASSOCIATION
RETALIATION
DEBARKATION
EMBARKATION
LEGISLATION
CIRCULATION
INFORMATION
EXPLANATION
DESIGNATION
RESIGNATION
EXAMINATION
PREPARATION
COOPERATION
IMMIGRATION
INSPIRATION
CORPORATION
PENETRATION
ARBITRATION
COMPUTATION
OBSERVATION
RESERVATION

RESTRICTION
DISTINCTION
DESTRUCTION
INSTRUCTION
RECOGNITION
REQUISITION
COMPOSITION
DISPOSITION
COMPETITION
DESCRIPTION
CONSUMPTION
INSTITUTION
LIGHTBOMBER
HEAVYBOMBER
RANGEFINDER
DYNAMOMETER
THERMOMETER
INTERPRETER
RECONNOITER
BLOCKBUSTER
AERONAUTICS
NAVALFORCES
ACCESSORIES
HOSTILITIES

ENEMYPLANES
PHILIPPINES
PARENTHESES
HEAVYLOSSES
COMMUNIQUES
PARENTHESIS
CREDENTIALS
BATTLESHIPS
ARMOREDCARS
CORRECTNESS
ENGAGEMENTS
ASSIGNMENTS
ASSESSMENTS
INSTRUMENTS
ESTIMATEDAT
SIGNIFICANT
INDEPENDENT
INTELLIGENT
COEFFICIENT
BOMBARDMENT
REPLACEMENT
EMPLACEMENT
ENFORCEMENT
ARRANGEMENT

CONFINEMENT
REQUIREMENT
MEASUREMENT
IMPROVEMENT
CONCEALMENT
ECHELONMENT
DEVELOPMENT
APPOINTMENT
COMPARTMENT
BELLIGERENT
INCOMPETENT
FINGERPRINT
DISCREPANCY
PHOTOGRAPHY
IMMEDIATELY
EXTENSIVELY
EFFECTIVELY
PRELIMINARY
CONTROVERSY
ELECTRICITY
NATIONALITY
SUITABILITY
SUPERIORITY

TWELVE LETTER WORDS

TRANSPACIFIC
HYDROGRAPHIC
UNIDENTIFIED
COMMISSIONED
DISSEMINATED
CONCENTRATED
DEMONSTRATED
DISORGANIZED
SIGNIFICANCE
INTELLIGENCE
INTERFERENCE
INCOMPETENCE
CONSIDERABLE
FIGHTERPLANE
INTERMEDIATE
DECENTRALIZE
GENERALSTAFF
TRANSFERRING
ENTERPRISING
ILLUMINATING
DISTRIBUTING

CONSTITUTING
BREAKTHROUGH
GEOGRAPHICAL
CONFIDENTIAL
PRESIDENTIAL
RECREATIONAL
AGRICULTURAL
DEPARTMENTAL
UNSUCCESSFUL
GENERALALARM
VETERINARIAN
TRANSMISSION
VERIFICATION
CONFISCATION
COMMENDATION
CONCILIATION
CANCELLATION
PROCLAMATION
CONFIRMATION
CONFORMATION
COORDINATION

ILLUMINATION
ANTICIPATION
REGISTRATION
ILLUSTRATION
INAUGURATION
COMPENSATION
CONVERSATION
RADIOSTATION
CONTINUATION
PRESERVATION
MOBILIZATION
ORGANIZATION
INTERDICTION
ROADJUNCTION
INTRODUCTION
CONSTRUCTION
INTERVENTION
INTERCEPTION
CONSCRIPTION
INTERRUPTION
DISTRIBUTION

SUBSTITUTION
CONSTITUTION
NORTHWESTERN
SOUTHWESTERN
MARKSMANSHIP
MEDIUMBOMBER
COMMISSIONER
PSYCHROMETER
SHARPSHOOTER
DIFFICULTIES
UNITEDSTATES
PREPARATIONS
OBSTRUCTIONS
INSTRUCTIONS
LIGHTBOMBERS
HEAVYBOMBERS
HEADQUARTERS
PREPAREDNESS
COMPLETENESS
CARELESSNESS
SEARCHLIGHTS

REPLACEMENTS
EMPLACEMENTS
MEASUREMENTS
ADVANTAGEOUS
SIMULTANEOUS
ANTIAIRCRAFT
NONCOMBATANT
CONVALESCENT
DISPLACEMENT
COMMENCEMENT
ANNOUNCEMENT
ENTANGLEMENT
DECIPHERMENT
ENCIPHERMENT
REENLISTMENT
INEFFICIENCY
SUCCESSFULLY
RESPECTFULLY
SATISFACTORY
INTRODUCTORY
IRREGULARITY

THIRTEEN LETTER WORDS

TRANSATLANTIC	CHRONOLOGICAL	DETERMINATION	FIGHTERPLANES	REINSTATEMENT
DISTINGUISHED	CONGRESSIONAL	EXTERMINATION	INSTALLATIONS	ESTABLISHMENT
DECENTRALIZED	INTERNATIONAL	CONSIDERATION	MEDIUMBOMBERS	ENTERTAINMENT
DISAPPEARANCE	SPECIFICATION	CONCENTRATION	MISCELLANEOUS	REAPPOINTMENT
IMPRACTICABLE	QUALIFICATION	DEMONSTRATION	INSTANTANEOUS	APPROXIMATELY
INDETERMINATE	COMMUNICATION	QUARTERMASTER	REENFORCEMENT	EXTRAORDINARY
CORRESPONDING	ACCOMMODATION	CIRCUMSTANCES	REINFORCEMENT	REVOLUTIONARY
CONCENTRATING	INVESTIGATION	DISCREPANCIES	REIMBURSEMENT	DEPENDABILITY
COUNTERATTACK	DISSEMINATION	PRELIMINARIES		

FOURTEEN LETTER WORDS

CHARACTERISTIC	RECONNOITERING	ADMINISTRATION	REORGANIZATION
RECONNAISSANCE	METEOROLOGICAL	INTERPRETATION	RECONSTRUCTION
DISCONTINUANCE	CIRCUMSTANTIAL	TRANSPORTATION	IRREGULARITIES
CORRESPONDENCE	CLASSIFICATION	CENTRALIZATION	INVESTIGATIONS
ADMINISTRATIVE	IDENTIFICATION	NATURALIZATION	SATISFACTORILY
REPRESENTATIVE	RECOMMENDATION	DEMOBILIZATION	RESPONSIBILITY
DISTINGUISHING			

C. LIST OF WORDS USED IN MILITARY TEXT ARRANGED ALPHABETICALLY ACCORDING TO WORD PATTERN

PATTERN AA

A CC EPT
A CC ORDING
O CC UPY
A DD
SU DD EN
LA DD ER
BE DD ING
FL EE
S EE
THR EE
PROC EE D
SP EE D
CR EE K
W EE K
F EE L
ST EE L
WH EE L
B EE N
FOURT EE N
HASB EE N
QU EE N
SCR EE N
S EE N
SIXT EE N
R EE NLIST
K EE P
SW EE PING
F EE T
FL EE T
M EE T
JUMPO FF
O FF
STA FF
O FF END
SU FF ER
TRA FF IC
O FF ICE
O FF ICER
E FF ORT
FO GG Y
A LL
CA LL
CE LL
DRI LL
ENRO LL
FA LL
FE LL
FU LL
HI LL
I LL
INSTA LL
PAYRO LL
REFI LL
SHE LL
SMA LL
SPE LL
WE LL
WI LL
VI LL AGE
CO LL APSED
DO LL AR
OSCI LL ATE
KI LL ED
BI LL ET
BU LL ETIN
VA LL EY
A LL IED
A LL IES
FA LL ING
PATRO LL ING
SHE LL ING
A LL OW
A LL Y
RA LL Y
CO MM A
CO MM AND
CO MM ANDER
SU MM ARY
CO MM END
CO MM ENT
HA MM ER
SU MM ER
CO MM IT
SU MM IT
SU MM ON
CO MM UTE
TO NN AGE
CHA NN EL
BA NN ER
GU NN ER
MA NN ER
A NN EX
CA NN OT
T OO
W OO DS
PR OO F
B OO K
C OO K
H OO K
L OO K
T OO K
SCH OO L
T OO L
PLAT OO N
S OO N
TR OO PS
C OO RDINATE
B OO TH
STO PP ED
HA PP EN
CLI PP ER
MA PP ING
A PP LY
SU PP LY
A PP OINT
A PP OINTED
SU PP ORT
SU PP ORTING
A PP ROVE
TE RR AIN
CU RR ENT
A RR EST
HU RR ICANE
DE RR ICK
GA RR ISON
A RR IVE
CA RR Y
FE RR Y
ACRO SS
COMPA SS
CONGRE SS
CRO SS
DARKNE SS
DRE SS
LE SS

~~CONFIDENTIAL~~

PATTERN AA—Continued

LO SS
MA SS
ME SS
PA SS
PRE SS
UNLE SS
WITNE SS
PA SS ED
A SS EMBLY
A SS ET
PO SS IBLE
A SS IGNED
CRO SS ING
DRE SS ING
ME SS ING
PA SS IVE
LE SS ON
I SS UE
A SS URE
EMBA SS Y
OMI TT ED
SUBMI TT ED
BA TT EN
WRI TT EN
BI TT ER
LI TT ER
BA TT ERY
SPO TT ING
BA TT LE
BA TT LESHIP
MU ZZ LE
NO ZZ LE

MISCELLANEOUS PATTERNS

AABA	AGR EEME NT	AABCB	SU FFICI ENT
AABA	K EEPE R	AABCB	A LLEGE
AABA	CH EESE	AABCB	CO LLEGE
AABA	BR EEZE	AABCB	BI LLETE D
AABA	MA NNIN G	AABCB	A MMETE R
AABA	PLA NNIN G	AABCB	W OODED
AABA	RU NNIN G	AABCB	TE RRIFI C
AABA	L OOKO UT	AABCB	BA TTERE D
AABA	E RROR	AABCBDEB	DI FFERENCE
AABA	MI RROR	AABCC	A CCESS
AABA	TE RROR	AABCC	A CCESS ORY
AABA	GLA SSES	AABCC	CO MMISS ARY
AABA	LO SSES	AABCCB	WI LLATTA CK
AABA	PA SSES	AABCCDD	CO MMITTEE
AABA	CHA SSIS	AABCCDEFBC	A CCESSORIES
AABA	A SSIS T	AABCDA	I LLEGAL
AABAACB	A SSESSME NT	AABCDA	A TTEMPT
AABAACBDEA	A SSESSMENTS	AABCDAB	A TTEMPTE D
AABAB	PROC EEDED	AABCDB	O FFENSE
AABB	CO FFEE	AABCDB	CHA LLENGE
AABB	BA LLOO N	AABCDB	BA LLISTI C
AABBAACAC	B EENNEEDED	AABCDB	A RRESTE D
AABBCBC	SU CCEEDED	AABCDB	PA SSENGE R
AABCA	B EETLE	AABCDB	BA TTERIE S
AABCA	A NNOUN CE	AABCDBA	SU RRENDER
AABCA	F OOTHO LD	AABCDBABD	SU RRENDERED
AABCA	CA RRIER	AABCDBC	CO MMANDAN T
AABCA	A SSETS	AABCDBD	O FFENDED
AABCA	I SSUES	AABCDBEC	BA LLISTICS
AABCADEC	CO MMITMENT	AABCDC	E FFICAC Y
AABCADEC	A TTENTION	AABCDD	A DDRESS
AABCADEFEA	A NNOUNCEMEN T	AABCDD	I LLNESS
AABCB	SCR EENIN G	AABCDDCA	A DDRESSED
AABCB	SU FFERE D	AABCDDCD	A DDRESSES
AABCB	DI FFERE NT	AABCDEB	CO MMUNIQU E
AABCB	O FFICI AL	AABCDEB	TR OOPSHIP

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

AABCDEB	A SSEMBLE
AABCDEBC	TR OOPSHIPS
AABCDEC	CO MMANDIN G
AABCDECB	BA TTLEFIEL D
AABCDED	CO MMANDED
AABCDEDFC	A MMUNITION
AABCDEE	CO MMANDEE R
AABCDEFA	R EENLISTE D
AABCDEFA	I RREGULAR
AABCDEFB	O FFENSIVE
AABCDEFBA	A SSEMBLIES
AABCDEFC	A LLOTMENT
AABCDEFC	C OOPERATE
AABCDEFD	I LLUSTRAT E
AABCDEFD	A SSIGNMEN T
AABCDEFDGA	A SSIGNMENTS
AABCDEFGA	C OOPERATIO N
AABCDEFGABF	R EENLISTMENT
AABCDEFGD	BA TTLESHIPS
AABCDEFGDAE	C OORDINATION
AABCDEFGDE	A PPOINTMENT
ABA	AGA IN
ABA	AGA INST
ABA	C ALA MITY
ABA	ALA RM
ABA	S ALA RY
ABA	D AMA GE
ABA	M ANA GE
ABA	C ANA L
ABA	ANA LYZE
ABA	J APA N
ABA	P ARA CHUTE
ABA	P ARA DE
ABA	SEP ARA TION
ABA	F ATA L
ABA	N AVA L
ABA	N AVA LFORCES
ABA	C AVA LRY
ABA	EXC AVA TION
ABA	AWA IT
ABA	AWA RD
ABA	AWA Y
ABA	PRO BAB LE
ABA	PRO BAB LY
ABA	BI CYC LE
ABA	CYC LONE
ABA	BLOCKA DED

ABA	GROUN DED
ABA	GUAR DED
ABA	INVA DED
ABA	LAN DED
ABA	RAI DED
ABA	WOUN DED
ABA	DID
ABA	IC EBE RG
ABA	PR ECE DING
ABA	R ECE IPT
ABA	CR EDE NTIAL
ABA	F EDE RAL
ABA	D EFE AT
ABA	D EFE CT
ABA	D EFE R
ABA	SI EGE
ABA	R EJE CT
ABA	S ELE CT
ABA	T ELE GRAM
ABA	ELE VATION
ABA	SCH EME
ABA	R EME DY
ABA	DISPLAC EME NT
ABA	PLAC EME NT
ABA	ENE MY
ABA	G ENE RAL
ABA	R EPE L
ABA	H ERE
ABA	SPH ERE
ABA	TH ERE
ABA	W ERE
ABA	WH ERE
ABA	CONQU ERE D
ABA	COV ERE D
ABA	TH ESE
ABA	PR ESE NT
ABA	D ESE RT
ABA	COMPL ETE
ABA	KILOM ETE R
ABA	M ETE R
ABA	P ETE R
ABA	D EVE LOP
ABA	S EVE N
ABA	S EVE NTH
ABA	S EVE NTY
ABA	S EVE RAL
ABA	EVE RY

MISCELLANEOUS PATTERNS—Continued

ABA		EYE	
ABA		FIF	TH
ABA		FIF	TY
ABA	EIG	HTH	
ABA	L	IAI	SON
ABA	PROH	IBI	T
ABA	SERV	ICI	NG
ABA	RA	IDI	NG
ABA	R	IDI	NG
ABA	R	IGI	D
ABA	F	ILI	NG
ABA	M	ILI	TARY
ABA	MOB	ILI	ZE
ABA	S	IMI	LAR
ABA	L	IMI	T
ABA	PROX	IMI	TY
ABA	F	INI	SH
ABA	F	IRI	NG
ABA	RET	IRI	NG
ABA	W	IRI	NG
ABA	V	ISI	BLE
ABA	D	ISI	NFECT
ABA	ADV	ISI	NG
ABA	DEC	ISI	ON
ABA	V	ISI	T
ABA	V	ISI	TOR
ABA	POL	ITI	CS
ABA	CR	ITI	QUE
ABA	POS	ITI	VE
ABA		MEM	ORIAL
ABA		NAN	
ABA	DOMI	NAN	CE
ABA	ORD	NAN	CE
ABA	DOMI	NAN	T
ABA		NIN	E
ABA		NIN	ETY
ABA	MOR	NIN	G
ABA		NIN	TH
ABA		OBO	E
ABA	C	OLO	N
ABA	SEMIC	OLO	N
ABA	C	OLO	RS
ABA	AUT	OMO	BILE
ABA	PR	OMO	TE
ABA	H	ONO	R
ABA	VIG	ORO	US
ABA	M	OTO	R

ABA	M	OTO	RIZED
ABA	PR	OVO	ST
ABA		PIP	E
ABA		POP	ULATED
ABA	LIB	RAR	Y
ABA	AI	RDR	OME
ABA	CA	RTR	IDGE
ABA	D	RYR	UN
ABA	DI	SAS	TER
ABA	CA	SES	
ABA	RE	SIS	T
ABA		SUS	PEND
ABA		SYS	TEM
ABA	S	TAT	ION
ABA	DIC	TAT	OR
ABA		TIT	LE
ABA	AL	TIT	UDE
ABA	LA	TIT	UDE
ABA		TOT	AL
ABA		TOT	ALING
ABA	A	UGU	ST
ABA		USU	AL
ABA	F	UTU	RE
ABA	SUR	VIV	ED
ABAA	HAV	EBEE	N
ABAA		SESS	ION
ABAACC		TATTOO	
ABAB	DETRA	ININ	G
ABAB	L	ININ	G
ABAB	M	ININ	G
ABAB	OBTA	ININ	G
ABAB	RA	ININ	G
ABAB	REMA	ININ	G
ABAB	TRA	ININ	G
ABAB	CR	ISIS	
ABAB	WI	THTH	E
ABAB	PAR	TITI	ON
ABACA	C	ANADA	
ABACA	P	ANAMA	
ABACA	PR	ECEDE	
ABACA		ELEME	NT
ABACA		ELEME	NTARY
ABACA		ELEVE	N
ABACA	C	EMETE	RY
ABACA	S	EVERE	
ABACA	AUD	IBILI	TY
ABACA	EXH	IBITI	ON

ABACA	V	ICINI TY
ABACA	M	ILITI A
ABACA	FAC	ILITI ES
ABACA	D	IMINI SH
ABACA	L	IMITI NG
ABACA		INITI AL
ABACA	DEF	INITI ON
ABACA	D	IRIGI BLE
ABACA	SEM	IRIGI D
ABACA	REQU	ISITI ON
ABACA	C	IVILI AN
ABACA	D	IVISI ON
ABACA	L	OCOMO TIVE
ABACA	M	ONOPO LY
ABACA	PR	OTOCO L
ABACA	CONS	TITUT E
ABACA		UNUSU AL
ABACADA	V	ISIBILI TY
ABACADB	DEF	INITION
ABACADBA	PR	ECEDENCE
ABACADC		INITIAT E
ABACADD	COMPL	ETENESS
ABACADDA	N	AVALATTA CK
ABACADEC	D	IVISIONS
ABACB	V	ACANC Y
ABACB	COMB	ATANT
ABACB	C	ATAST ROPHE
ABACB	D	ETECT OR
ABACB	V	ISITS
ABACB		MEMBE R
ABACBDEC	D	ETENTION
ABACBDEC	R	ETENTION
ABACBDEFGFAG		NONCOMBATANT
ABACC	R	EBELL ION
ABACC	N	ECESS ARY
ABACC	N	ECESS ITY
ABACC	CAR	ELESS
ABACC	WIR	ELESS
ABACCA	P	ARALLA X
ABACCA	R	EPELLE D
ABACCA	T	OMORRO W
ABACCDACC	CAR	ELESSNESS
ABACCDC	P	ARALLEL
ABACCDEFEA	N	ECESSITATE
ABACDA		ALASKA
ABACDA		ARABIA
ABACDA	N	AVALBA SE
ABACDA	R	ECEIVE
ABACDA	D	ECEMBE R
ABACDA	D	EFENSE
ABACDA	R	EJECTE D
ABACDA	R	ELEASE
ABACDA	S	ELECTE D
ABACDA	R	EMEDIE S
ABACDA		EMERGE NCY
ABACDA		ENEMIE S
ABACDA	R	EPEATE D
ABACDA	R	EVENUE
ABACDA	U	NKNOWN
ABACDA	PR	OMOTIO N
ABACDAAC	S	EVENTEEN
ABACDAACD	S	EVENTEENT H
ABACDAC	D	ESERTER
ABACDAD	D	EFENSES
ABACDAED		AVAILABL E
ABACDAEEC	N	AVALBATTL E
ABACDB	F	ATALIT Y
ABACDB	A	NONYMO US
ABACDB	C	OLONEL
ABACDBA	TH	EREFORE
ABACDC	R	ECEIVI NG
ABACDC		EVENIN G
ABACDC	DYNA	MOMETE R
ABACDCA	L	IMITATI ON
ABACDCCA		NINETEEN
ABACDCCAD		NINETEENT H
ABACDCEA	S	TATEMENT
ABACDCECFGHIE	M	ETEOROLOGICAL
ABACDD		FIFTEE N
ABACDD	FO	RTRESS
ABACDDEC		FIFTEENT H
ABACDEA		ELEVATE
ABACDEA	D	EVELOPE
ABACDEA	VER	IFICATI ON
ABACDEA	S	IMILARI TY
ABACDEAD		SUSPENSE
ABACDEAFGE		SUSPENSION
ABACDEB	EXPL	ANATION
ABACDEB	T	OPOGRAP HIC
ABACDEBFA	R	ECEPTACLE
ABACDEC		ABANDON
ABACDEC	D	AMAGING
ABACDEC	QU	ARANTIN E
ABACDECA	P	ENETRATE

MISCELLANEOUS PATTERNS—Continued

ABACDECFBA	D	ETERIORATE
ABACDECFGB	P	ENETRATION
ABACDED	C	APABILI TY
ABACDED	M	OTORCYC LE
ABACDED		SUSPICI ON
ABACDEDEDC	G	ENERALALAR M
ABACDEDFBA		SUSPICIOUS
ABACDEDFGA		SUSPICIONS
ABACDEFA	D	EFECTIVE
ABACDEFA	D	EFENSIVE
ABACDEFA	T	ELEPHONE
ABACDEFA	D	ETERMINE
ABACDEFA	D	EVELOPME NT
ABACDEFA		EXERCISE
ABACDEFAF		EXERCISES
ABACDEFB		DEDICATE
ABACDEFB		ENEMYTAN KS
ABACDEFC		DEDICATI ON
ABACDEFCDFE	V	ETERINARIAN
ABACDEFCFD		ELECTRICIT Y
ABACDEFD		SUSPECTE D
ABACDEFDF		SUSPENDED
ABACDEFE		ANALYSIS
ABACDEFGA		EXECUTIVE
ABACDEFGB		POPULATIO N
ABACDEFGBA		ENEMYPLANE S
ABACDEFGBA	S	EVENTYFIVE
ABACDEFGBEHF	D	ETERMINATION
ABACDEFGDHH	G	ENERALSTAFF
ABACDEFGE		MEMORANDA
ABACDEFGHA		MEMORANDUM
ABACDEFGHIA	D	ECENTRALIZE
ABBA		AFFA IR
ABBA		APPA RENT
ABBA		APPA RENTLY
ABBA	B	ARRA CKS
ABBA	B	ARRA GE
ABBA		ARRA NGE
ABBA		ASSA ULT
ABBA	P	ASSA GE
ABBA	IMP	ASSA BLE
ABBA		ATTA CH
ABBA		ATTA CK
ABBA		ATTA IN
ABBA	B	ATTA LION
ABBA	IN	DEED
ABBA		EFFE CT
ABBA	COMP	ELLE D
ABBA	SH	ELLE D
ABBA	CONF	ERRE D
ABBA	COMPR	ESSE D
ABBA	IMPR	ESSE D
ABBA	PR	ESSE D
ABBA	V	ESSE L
ABBA	CIGAR	ETTE
ABBA	B	ETTE R
ABBA	L	ETTE R
ABBA	D	IFFI CULT
ABBA	W	ILLI AM
ABBA	F	ILLI NG
ABBA	K	ILLI NG
ABBA	REF	ILLI NG
ABBA	SW	IMMI NG
ABBA	SH	IPPI NG
ABBA	M	ISSI NG
ABBA	ADM	ISSI ON
ABBA	M	ISSI ON
ABBA	PERM	ISSI ON
ABBA	F	ITTI NG
ABBA	AFTER	NOON
ABBA		NOON
ABBA	F	OLLO W
ABBA	C	OMMO N
ABBA		OPPO SE
ABBA		OPPO SITE
ABBA	B	OTTO M
ABBAB	B	AGGAG E
ABBAB	WITN	ESSES
ABBACA		APPARA TUS
ABBACA	L	ETTERE D
ABBACB	V	ESSELS
ABBACDA	M	ESSENGE R
ABBACDA		EFFECTE D
ABBACDB	M	ISSIONS
ABBACDEA		IRRIGATI ON
ABBACDEDA		OPPOSITIO N
ABBACDEFA		EFFECTIVE
ABBACDEFA	D	IFFICULTI ES
ABBACDEFA		IMMIGRATI ON
ABBACDEFCD		ILLITERATE
ABBACDEFDB		ATTAINMENT
ABBACDEFEC		ARRANGEMEN T
ABBACDEFGB		ATTACHMENT
ABBCA		ANNUA L

MISCELLANEOUS PATTERNS—Continued

ABBCA		APPEA R
ABBCA	DIS	APPEA R
ABBCA	C	ARRIA GE
ABBCA	S	ETTLE
ABBCA		ISSUI NG
ABBCA	FOUR	TEENT H
ABBCA	SIX	TEENT H
ABBCA	CHA	UFFEU R
ABBCA	S	URROU ND
ABBCADAEFC		APPEARANCE
ABBCADAEFC	DIS	APPEARANCE
ABBCADC		APPEARE D
ABBCBBDA	P	OSSESSIO N
ABBCBDA		ASSISTA NCE
ABBCBDAED		ASSISTANT
ABBCCDAB		ASSOONAS
ABBCDA		ALLOWA NCE
ABBCDA		APPROA CH
ABBCDA		ARRIVA L
ABBCDA		ASSURA NCE
ABBCDA	M	ESSAGE
ABBCDA		ILLUMI NATE
ABBCDAB	M	ESSAGES
ABBCDAB	C	ORRIDOR
ABBCDAEA	B	ELLIGERE NT
ABBCDAEFC		ALLOCATIO N
ABBCDAEFC		IMMEDIATE
ABBCDAEFGAE		ILLUMINATIN G
ABBCDAEFGAHE		ILLUMINATION
ABBCDAEFGAHE	D	ISSEMINATION
ABBCDBCEA		APPROPRIA TE
ABBCDCA		EFFICIE NT
ABBCDCA	C	OLLISIO N
ABBCDCAED		EFFICIENC Y
ABBCDCAED	C	OLLISIONS
ABBCDCEFA		ADDITIONA L
ABBCDDCA	C	OMMISSIO N
ABBCDDCA	C	OMMISSIO NER
ABBCDDCEAFGC		ACCOMMODATIO N
ABBCDEA		ACCOMPA NY
ABBCDEA		APPROVA L
ABBCDEA		ASSOCIA TE
ABBCDEA	SH	ELLFIRE
ABBCDEA	T	ERRIBLE
ABBCDEAFB		ACCORDANC E
ABBCDEAFB		REENFORCE
ABBCDEAFBC		ACCEPTANCE
ABBCDEAFBGBC		REENFORCEMEN T
ABBCDEAFD		APPLICATI ON
ABBCDEAFEC		ASSOCIATIO N
ABBCDEAFGC		ACCEPTABLE
ABBCDEAFGC		ALLEGIANCE
ABBCDEAFGHF	C	ORRESPONDIN G
ABBCDEFGA		ACCIDENTA L
ABBCDEFGA		APPROXIMA TE
ABBCDEFGA		OCCUPATIO N
ABBCDEFGBA		IRREGULARI TY
ABBCDEFGBAHAC		IRREGULARITIE S
ABBCDEFGEA		ILLUSTRATI ON
ABBCDEFGHAD	C	OMMENDATION
ABCA	P	ACKA GE
ABCA	EV	ACUA TING
ABCA	EV	ACUA TION
ABCA	R	ADIA L
ABCA	R	ADIA TE
ABCA		ADJA CENT
ABCA	GR	ADUA L
ABCA		ADVA NCE
ABCA	DI	AGRA M
ABCA	EV	ALUA TION
ABCA		ALWA YS
ABCA	C	AMPA IGN
ABCA	M	ANDA TE
ABCA	M	ANUA L
ABCA	J	ANUA RY
ABCA	C	ANVA S
ABCA	CH	APLA IN
ABCA	C	APTA IN
ABCA		AREA
ABCA	DEB	ARKA TION
ABCA	EMB	ARKA TION
ABCA		ASIA
ABCA	CO	ASTA L
ABCA	C	ASUA L
ABCA	C	ASUA LTY
ABCA		AVIA TOR
ABCA		BARB ED
ABCA		BOMB
ABCA		BOMB ARD
ABCA		BOMB ER
ABCA	LIGHT	BOMB ER
ABCA		BRIB E
ABCA		BULB
ABCA		CANC EL

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

ABCA		CHEC	K
ABCA		CIRC	LE
ABCA		CIRC	ULATE
ABCA		CONC	EAL
ABCA		CONC	LUDE
ABCA	HUN	DRED	
ABCA	L	EADE	R
ABCA		EAGE	R
ABCA	M	EAGE	R
ABCA	S	EAME	N
ABCA	ST	EAME	R
ABCA	N	EARE	ST
ABCA	C	EASE	
ABCA	GR	EASE	
ABCA	INCR	EASE	D
ABCA	L	EAVE	
ABCA		ECHE	LON
ABCA	WR	ECKE	D
ABCA	INF	ECTE	D
ABCA		EDGE	
ABCA	S	EIZE	
ABCA	R	ELIE	F
ABCA	H	ELPE	R
ABCA	TW	ELVE	
ABCA	NOV	EMBE	R
ABCA	ABS	ENCE	
ABCA	LIC	ENSE	
ABCA	C	ENTE	R
ABCA		ENTE	R
ABCA		ENVE	LOP
ABCA	R	EQUE	ST
ABCA	FI	ERCE	
ABCA	S	ERGE	ANT
ABCA	MAT	ERIE	L
ABCA	REV	ERSE	
ABCA	OBS	ERVE	
ABCA	R	ESPE	CT
ABCA	W	ESTE	RLY
ABCA	W	ESTE	RN
ABCA		ETHE	R
ABCA	MAN	EUVE	R
ABCA	R	EVIE	W
ABCA		EXCE	PT
ABCA		EXPE	CT
ABCA		EXPE	ND
ABCA		EXTE	ND
ABCA		GAUG	E
ABCA		GEOG	RAPHIC
ABCA	FOR	GING	
ABCA	W	HICH	
ABCA		HIGH	
ABCA		HIGH	ER
ABCA		HIGH	EST
ABCA	V	ICTI	M
ABCA	M	IDNI	GHT
ABCA	DR	IFTI	NG
ABCA	L	IFTI	NG
ABCA	S	IGNI	FY
ABCA	BU	ILDI	NG
ABCA		INDI	CATE
ABCA		INDI	RECT
ABCA	DESCR	IPTI	ON
ABCA	L	IQUI	D
ABCA	A	IRFI	ELD
ABCA		REPR	ISAL
ABCA	M	ISFI	RE
ABCA	F	ISHI	NG
ABCA	W	ITHI	N
ABCA	FUE	LOIL	
ABCA		MAIM	
ABCA	LA	NDIN	G
ABCA	I	NFAN	TRY
ABCA	CO	NFIN	E
ABCA	U	NION	
ABCA	SU	NKEN	
ABCA	FLA	NKIN	G
ABCA	I	NLAN	D
ABCA	I	NTEN	D
ABCA	CO	NTIN	UAL
ABCA	CO	NTIN	UE
ABCA	I	NVEN	T
ABCA		OCTO	BER
ABCA	D	OCTO	R
ABCA	F	OGHO	RN
ABCA	P	OISO	N
ABCA	C	OMPO	SED
ABCA	C	ONVO	Y
ABCA	EN	ORMO	US
ABCA	EXPL	OSIO	N
ABCA		PUMP	
ABCA		PURP	OSE
ABCA	HA	RBOR	
ABCA	AI	RBOR	NE
ABCA	MU	RDER	

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

ABCA	O	RDER
ABCA	O	RDER S
ABCA		REAR
ABCA		RECR UIT
ABCA	COU	RIER
ABCA	P	RIOR
ABCA	SUPE	RIOR
ABCA	A	RMOR
ABCA	A	RMOR Y
ABCA	P	ROGR AM
ABCA	MO	RTAR
ABCA	QUA	RTER
ABCA	QUA	RTER S
ABCA	FEB	RUAR Y
ABCA	FO	RWAR D
ABCA	CEN	SORS HIP
ABCA		SUNS ET
ABCA	IMPOR	TANT
ABCA	S	TART
ABCA	PRO	TECT
ABCA		TENT
ABCA		TENT H
ABCA	PRO	TEST
ABCA		TEXT
ABCA		THAT
ABCA	S	TRAT EGIC
ABCA	S	TRAT EGY
ABCA	D	UGOU T
ABCA		UNSU ITABLE
ABCA	P	URSU E
ABCA	P	URSU IT
ABCA	O	UTGU ARD
ABCAA	D	ECREE
ABCAA	D	EGREE
ABCAA	B	ETWEE N
ABCAA	DI	SCUSS
ABCAA	A	SPOSS IBLE
ABCAAB	P	ONTOON
ABCAAB		THATTH E
ABCAACDEB	P	REARRANGE D
ABCAB	W	ARFAR E
ABCAB	S	ECREC Y
ABCAB	OBS	ERVER
ABCAB	W	HETHE R
ABCAB	B	INDIN G
ABCAB	F	INDIN G
ABCAB	S	INKIN G
ABCAB	PA	INTIN G
ABCAB	PR	INTIN G
ABCAB	I	NTENT
ABCAB	P	ONTON
ABCAB	C	ORPOR AL
ABCAB		RECRE ATION
ABCAB	P	RIORI TY
ABCAB	SUPE	RIORI TY
ABCAB	DI	SEASE
ABCAB	PRO	TECTE D
ABCAB	PRO	TESTE D
ABCAB	O	UTPUT
ABCABA	INT	ERFERE
ABCABB	D	ISMISS
ABCABB	D	ISMISS AL
ABCABC		THATHA VE
ABCABCA		ENTENTE
ABCABDA	S	ENTENCE
ABCABDB		REPRESE NT
ABCABDBEFGFHIB		REPRESENTATIVE
ABCABDBEFGFHIED		REPRESENTATIONS
ABCABDC		RETREAT
ABCABDED	M	ANGANESE
ABCABDEFA	C	ORPORATIO N
ABCABDEFGHD		RECREATIONA L
ABCAC		ARMAM ENT
ABCAC	N	EARER
ABCAC		PROPO SE
ABCAC	P	RAIRI E
ABCAC	PRO	TESTS
ABCACA	D	IETITI AN
ABCACB	O	RDERED
ABCACBDEC		PROPORTIO N
ABCACDEFD		PROPOSALS
ABCADA		ALMANA C
ABCADA	R	ELIEVE
ABCADA	C	ENTERE D
ABCADA	B	ESIEGE D
ABCADA	R	EVIEWE D
ABCADAB	CO	NTINENT AL
ABCADAC	S	EALEVEL
ABCADAC		INDIVID UAL
ABCADAEC		IGNITION
ABCADAEFB		TENTATIVE
ABCADAEFC	S	IGNIFICAN T
ABCADAEFCE	S	IGNIFICANC E
ABCADAEFGHF		SUBSISTENCE

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

ABCADB	ATLANT IC	ABCADEAB	CO NTINGENT
ABCADB	BRIBER Y	ABCADEAE	EXPENDED
ABCADB	CIRCUI T	ABCADEAE	EXPENSES
ABCADB	W EDNESD AY	ABCADEAE	EXTENDED
ABCADB	LOG ISTICS	ABCADEAFA	ELSEWHERE
ABCADB	EXPL OSIONS	ABCADEAFGA	EXPERIENCE
ABCADB	PREPAR ING	ABCADEB	C ENTERIN G
ABCADB	IM PROPER	ABCADEB	ENTERIN G
ABCADB	PROPER	ABCADEB	R ESPECTS
ABCADBA	INSIGNI A	ABCADEB	INCIDEN T
ABCADBC	PREPARE	ABCADEB	M ISFIRES
ABCADBCEFCGG	PREPAREDNESS	ABCADEBCE	INCIDENCE
ABCADBD	PREPARA TION	ABCADEC	M ANDATED
ABCADBEFD	CIRCUITOU S	ABCADEC	S ECRETAR Y
ABCADC	R ADIATI ON	ABCADEC	GYR OSCOPIC
ABCADC	ST ANDARD	ABCADECA	REARGUAR D
ABCADC	V ARIATI ON	ABCADECAFD	D ISTINCTION
ABCADC	ASIATI C	ABCADECFC	CONCERNIN G
ABCADC	AVIATI ON	ABCADEDA	CO NFINEMEN T
ABCADC	R EVIEWI NG	ABCADEDAFB	INVITATION
ABCADC	EXTENT	ABCADEDBD	SUBSTITUT E
ABCADC	I NVENTE D	ABCADEDBDE	SUBSTITUTI ON
ABCADC	TACTIC S	ABCADEDC	LI EUTENANT
ABCADC	S TARTER	ABCADEDFGA	ENTERPRISE
ABCADC	ZIGZAG	ABCADEDFGDBC	CONCILIATION
ABCADCA	CO NVENIEN T	ABCADEDFGFB	ENTERPRISIN G
ABCADCB	CO NDENSED	ABCADEE	P ROGRESS
ABCADCB	TACTICA L	ABCADEEBFGHC	CANCELLATION
ABCADCEFBGABC	ENTERTAINMENT	ABCADEED	CANCELLE D
ABCADCEFGED	CONCENTRATE	ABCADEEFBC	CONCESSION
ABCADCEFGEHC	CONCENTRATIN G	ABCADEEFGD	P ROGRESSIVE
ABCADCEFGEHBC	CONCENTRATION	ABCADEFA	ECHELONE D
ABCADD	D EPRESS ION	ABCADEFA	ENVELOPE
ABCADD	EXCESS	ABCADEFA	EXPEDITE
ABCADD	D ISTILL	ABCADEFA	EXPERIME NT
ABCADD	P OSTOFF ICE	ABCADEFAB	INDICATIN G
ABCADD	B OYCOTT	ABCADEFAB	D ISTINGUIS H
ABCADDA	AMBASSA DOR	ABCADEFABGADE	D ISTINGUISHING
ABCADDA	EXPELLE D	ABCADEFAGB	INDICATION
ABCADDECCFA	UNSUCCESSFU L	ABCADEFB	ADVANCED
ABCADDEFA	EXCESSIVE	ABCADEFBA	EXT RAORDINAR Y
ABCADEA	ADVANTA GE	ABCADEFC	BOMBARDM ENT
ABCADEA	ADVANTA GEOUS	ABCADEFC	CIRCULAR
ABCADEA	D ECREASE	ABCADEFC	U NTENABLE
ABCADEA	S EPTEMBE R	ABCADEFCGHB	RETROACTIVE
ABCADEA	R EQUESTE D	ABCADEFD	ADVANCIN G
ABCADEA	D ISCIPLI NE	ABCADEFD	EXTENDIN G

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

ABCADEFD		EXTERIOR
ABCADEFE		CONCRETE
ABCADEFE		EXPEDITI NG
ABCADEFE		EXPEDITI ON
ABCADEFE		OBSOLETE
ABCADEFE	G	ONIOMETE R
ABCADEFE		PURPOSES
ABCADEFE		RECRUITI NG
ABCADEFEA	C	OMPOSITIO N
ABCADEFGA		EXPENSIVE
ABCADEFGA		EXTENSIVE
ABCADEFGAF		ECHELONMEN T
ABCADEFGB	C	ASUALTIES
ABCADEFGB		CIRCULATI ON
ABCADEFGBC		CONCLUSION
ABCADEFGC		INDICATED
ABCADEFGC	S	TRATEGICA L
ABCADEFGD		EXTENSION
ABCADEFGDC		CONCEALMEN T
ABCADEFGE		REPRISALS
ABCADEFGF		BOMBARDED
ABCADEFGHAB	C	ONFORMATION
ABCADEFGHCA		EXTERMINATE
ABCADEFGHCFIG		EXTERMINATION
ABCADEFGHEIGCF		REORGANIZATION
ABCADEFGHH	R	ESPECTFULL Y
ABCADEFGHIAJF		CIRCUMSTANCES
ABCADEFGHIB		RETROACTIVE
ABCADEFGHIE		GEOGRAPHICA L
ABCADEFGHIGBH		CIRCUMSTANTIA L
ABCBA	COMP	LETEL Y
ABCBA		AWKWA RD
ABCBA		CAPAC ITY
ABCBA	PA	CIFIC
ABCBA	SPE	CIFIC
ABCBA	HIN	DERED
ABCBA		DIVID E
ABCBA		GARAG E
ABCBA	C	ITATI ON
ABCBA		LEVEL
ABCBA	P	REFER
ABCBA		REFER
ABCBA	P	RESER VATION
ABCBA		RESER VATION
ABCBA		TAXAT ION
ABCBA	HOS	TILIT Y
ABCBA	U	TILIT Y
ABCBA	AC	TIVIT Y
ABCBAA	U	SELESS
ABCBAAB	P	REFERRE D
ABCBAB		DIVIDI NG
ABCBAB	AC	TIVITI ES
ABCBABDEB	P	REFERENCE
ABCBABDEB		REFERENCE
ABCBADA		MINIMUM
ABCBADB	P	RESERVE
ABCBADB		RESERVE
ABCBADB		REVERSE
ABCBADBC		RESERVES
ABCBADEB	SPE	CIFICATI ON
ABCBCDBA		REMEMBER
ABCBDA		DEFEND
ABCBDA		DEPEND
ABCBDA	MU	NITION S
ABCBDA		RESEAR CH
ABCBDA		STATES
ABCBDA		STATUS
ABCBDA	IN	TEREST
ABCBDAB		DEFENDE R
ABCBDAB	E	NGAGING
ABCBDABA		DEFENDED
ABCBDABD		DEPENDEN T
ABCBDABDEA		STATISTICS
ABCBDAEFGB		DEPENDABLE
ABCBDAEFGHG		DEPENDABILI TY
ABCBDCBA		PARAGRAP H
ABCBDDBA		DEFERRED
ABCBDEA	E	CONOMIC
ABCBDEA		DAMAGED
ABCBDEA	PO	LITICAL
ABCBDEAEC		MANAGEMEN T
ABCBDEBA		DEFEATED
ABCBDEBA		DESERTED
ABCBDEBA		RECEIVER
ABCBDEBA		REPEATER
ABCBDEFA		REJECTOR
ABCBDEFA		STATIONS
ABCBDEFBA		DEVELOPED
ABCBDEFGA	R	ESISTANCE
ABCBDEFGBA		DETERMINED
ABCBDEFGHFA		DISINFECTED
ABCBDEFGHIJBA		DECENTRALIZED
ABCCA		LITTL E
ABCCA		PASSP ORT

~~CONFIDENTIAL~~

MISCELLANEOUS PATTERNS—Continued

ABCCA	S TREET	ABCDA	M ARTIA L
ABCCABDEC	C ROSSROADS	ABCDA	E ASTWA RD
ABCCBADED	MILLIMETE R	ABCDA	N ATURA L
ABCCBCA	BE GINNING	ABCDA	N ATURA LIZE
ABCCBDA	INF LAMMABL E	ABCDA	TE CHNIC AL
ABCCDA	COLLEC T	ABCDA	COUNC IL
ABCCDA	CORREC T	ABCDA	R EACHE D
ABCCDA	T RIGGER	ABCDA	L EAGUE
ABCCDA	RUBBER	ABCDA	EASTE RLY
ABCCDA	RUNNER	ABCDA	EASTE RN
ABCCDA	SPOOLS	ABCDA	W EATHE R
ABCCDA	SPOONS	ABCDA	H EAVIE R
ABCCDA	SUGGES T	ABCDA	INS ECURE
ABCCDA	SUPPOS E	ABCDA	S ECURE
ABCCDA	TURRET	ABCDA	R EDUCE
ABCCDAA	SUCCESS	ABCDA	SCH EDULE
ABCCDAAEB	SUCCESSFU L	ABCDA	B EFORE
ABCCDAAEBFF	SUCCESSFULL Y	ABCDA	R EFUGE
ABCCDAAEFD	SUCCESSIVE	ABCDA	R EFUSE
ABCCDAB	P RESSURE	ABCDA	R EGIME NT
ABCCDAEC	TERRITOR Y	ABCDA	R EGIME NTAL
ABCCDAED	CORRECTE D	ABCDA	EITHE R
ABCCDAEFB	COLLECTIO N	ABCDA	FUS ELAGE
ABCCDAEFB	CORRECTIO N	ABCDA	D ELIVE R
ABCCDAEFBC	CONNECTION	ABCDA	GR ENADE
ABCCDAEFC	CONNECTIN G	ABCDA	ERASE
ABCCDAEFDGG	CORRECTNESS	ABCDA	OP ERATE
ABCCDEA	GASSING	ABCDA	R ESCUE
ABCCDEA	GETTING	ABCDA	PR ESIDE NT
ABCCDEA	ST RAGGLER	ABCDA	R ESUME
ABCCDEA	IN TERRUPT	ABCDA	D EVICE
ABCCDEAB	IN TERRUPTE D	ABCDA	D EVISE
ABCCDEAD	COMMENCE	ABCDA	GOING
ABCCDEAD	COMMERCE	ABCDA	T HOUGH
ABCCDEADCDE	COMMENCEMEN T	ABCDA	C HURCH
ABCCDEBFGHDA	DISSEMINATED	ABCDA	F IGHTI NG
ABCCDEFA	COMMUNIC ATE	ABCDA	INFLI CT
ABCCDEFA	SUPPLIES	ABCDA	EXT INGUI SH
ABCCDEFAGHFBE	COMMUNICATION	ABCDA	INQUI RE
ABCCDEFBGHDGAD	CORRESPONDENCE	ABCDA	INQUI RY
ABCCDEFGA	R EAPPOINTE D	ABCDA	INSPI RE
ABCCDEFGHAFG	R EAPPOINTMENT	ABCDA	LOCAL
ABCDA	S ABOTA GE	ABCDA	LAU NCHIN G
ABCDA	R AILWA Y	ABCDA	CO NDEMN
ABCDA	ANIMA L	ABCDA	MACHI NEGUN
ABCDA	S ANITA RY	ABCDA	NOTIN G
ABCDA	M ARSHA L	ABCDA	EXPA NSION

MISCELLANEOUS PATTERNS—Continued

ABCDA	CO	NTAIN
ABCDA	MOU	NTAIN
ABCDA	I	NTERN AL
ABCDA	FRO	NTLIN E
ABCDA	I	NTREN CH
ABCDA	C	ONTRO L
ABCDA	H	ORIZO N
ABCDA		OUTBO ARD
ABCDA		PROMP T
ABCDA		RECOR D
ABCDA		REPOR T
ABCDA		RETUR N
ABCDA	P	RIMAR Y
ABCDA		RIVER
ABCDA		ROGER
ABCDA	FA	RTHER
ABCDA	FU	RTHER
ABCDA	NO	RTHER LY
ABCDA		SATIS FY
ABCDA		SHIPS
ABCDA	WAR	SHIPS
ABCDA		THIRT Y
ABCDA	WI	THOUT
ABCDA	EX	TRACT
ABCDA		TRACT
ABCDA	INS	TRUCT
ABCDA	DES	TRUCT ION
ABCDA		TWENT Y
ABCDA	B	UREAU
ABCDA		WESTW ARD
ABCDAA	R	EFUGEE
ABCDAA	C	ODEBOO K
ABCDAA	BU	SINESS
ABCDAA	DI	STRESS
ABCDAA		STRESS
ABCDAAD	F	ORENOON
ABCDAB		DECIDE
ABCDAB		DECODE
ABCDAB	SP	EARHEA D
ABCDAB	R	EDUCED
ABCDAB		ENTREN CH
ABCDAB		ERASER
ABCDAB		GEORGE
ABCDAB		POSTPO NE
ABCDAB		RETIRE
ABCDAB	ES	TIMATI ON
ABCDABA		DECIDED
ABCDABAB		INCLININ G
ABCDABC	M	AINTAIN
ABCDABC	M	AINTAIN ED
ABCDABCEFD		PHOSPHORUS
ABCDABEFA		ENTRENCHE D
ABCDAC	L	ANGUAG E
ABCDAC		ANYWAY
ABCDAC	GOV	ERNMEN T
ABCDAC	I	NSTANT
ABCDAC	I	NSTANT LY
ABCDAC	DI	SPERSE
ABCDAC	RES	TRICTI ON
ABCDAC	PA	TRIOTI C
ABCDACB	CO	NDEMNED
ABCDACDAEFGB	I	NSTANTANEOUS
ABCDACEFDAF		COINCIDENCE
ABCDAD		MOVEME NT
ABCDAD	A	MUSEME NT
ABCDAD		RIGORO US
ABCDADC	S	ANITATI ON
ABCDADEDAFB		INSTITUTION
ABCDADEFEAGC		ANTIAIRCRAFT
ABCDAEA		EXTREME
ABCDAEA		MAXIMUM
ABCDAEAB	SU	ITABILIT Y
ABCDAEABD	UNI	TEDSTATES
ABCDAEAE	PAR	ENTHESES
ABCDAEB	F	IGHTING
ABCDAEB	S	IGHTING
ABCDAEB		RAILROA D
ABCDAEB		REPORTE D
ABCDAEB		RETURNE D
ABCDAEB		TRACTOR
ABCDAEB	INS	TRUCTOR
ABCDAEBA		RECORDER
ABCDAEBC	DE	TONATION
ABCDAEBFBDC	U	NIDENTIFIED
ABCDAEBFC		SATISFACT ORY
ABCDAEC		AVERAGE
ABCDAEC	D	ISTRICT
ABCDAEC		OUTPOST
ABCDAECA		TWENTIET H
ABCDAECAB	I	NTERNMENT
ABCDAECB	D	ISTRICTS
ABCDAECD	L	ABORATOR Y
ABCDAECE		OUTPOSTS
ABCDAECFD	EX	AMINATION

MISCELLANEOUS PATTERNS—Continued

ABCDAED	T	RAVERSE
ABCDAEE		ACTUALL Y
ABCDAEE		EXPRESS
ABCDAEE		THIRTEE N
ABCDAEEFAB		THIRTEENTH
ABCDAEFA	OV	ERWHELME D
ABCDAEFAB		INFLICTIN G
ABCDAEFB	P	RESCRIBE D
ABCDAEFBE	O	NEHUNDRED
ABCDAEFC	M	ANUFACTU RE
ABCDAEFC	PR	ESIDENTI AL
ABCDAEFC	D	ISTRIBUT E
ABCDAEFCA	D	ISTRIBUTI NG
ABCDAEFCA	D	ISTRIBUTI ON
ABCDAEFD	F	LASHLIGH T
ABCDAEFD	C	ONTROVER SY
ABCDAEFD	A	SCENSION
ABCDAEFD		WINDWARD
ABCDAEFDB		RESTRICTE D
ABCDAEFDE		RESTRICTI ON
ABCDAEFE	PAR	ENTHESIS
ABCDAEFE		RETURNIN G
ABCDAEFEGE	RE	SPONSIBILI TY
ABCDAEFF		REDCROSS
ABCDAEFGAHB		INSPIRATION
ABCDAEFGC		REGARDING
ABCDAEFGD		RESTRAINT
ABCDAEFGFE	TR	ANSPACIFIC
ABCDAEFGHC		TWENTYFIVE
ABCDAEFGHFBC		CONSCRIPTION
ABCDBA	PR	ACTICA L
ABCDBA	W	ATERTA NK
ABCDBA	DIV	EBOMBE R
ABCDBA		ENGINE
ABCDBA	S	ENTINE L
ABCDBA	R	EVOLVE
ABCDBA	S	ITUATI ON
ABCDBAA		ENGINEE R
ABCDBAAEDBC		ENGINEERING
ABCDBAB		LIABILI TY
ABCDBAD	RE	TALIATI ON
ABCDBAEAD	D	ISPOSITIO N
ABCDBAEBE	U	NEXPENDED
ABCDBBA		ANTENNA
ABCDBBA	D	ISCUSSI ON
ABCDBBDEA	TRA	NSMISSION
ABCDBCAEB	INTENTION	

ABCDBCEA	A	ERODROME
ABCDBEA		INCENDI ARY
ABCDBEA	PR	OTECTIO N
ABCDBEA	IN	TERCEPT
ABCDBEAB	IN	TERCEPTE D
ABCDBEAE	C	ONTINUOU S
ABCDBEAFB		INVENTION
ABCDBEAFCDB	QU	ARTERMASTER
ABCDBEAFD		INCENTIVE
ABCDBEAFD		INTENSIVE
ABCDBECA	E	NCIRCLIN G
ABCDBEFAGABC		ENTANGLEMENT
ABCDBEFAGEB		TEMPERATURE
ABCDBEBA		DECREASED
ABCDBEFCDAB	C	ONTINUATION
ABCDBEFGA		YESTERDAY
ABCDBEFGAB		ARMOREDCAR
ABCDBEFGBCHIA		DISTINGUISHED
ABCDBEFGHA	P	ERFORMANCE
ABCDCA		AIRCRA FT
ABCDCA		CRITIC
ABCDCA		CRITIC AL
ABCDCA	D	EFICIE NT
ABCDCA		ENGAGE
ABCDCA	P	OSITIO N
ABCDCA	PR	OVISIO N
ABCDCA	FI	REALAR M
ABCDCAAC		PHILIPPI NES
ABCDCAB		ANTITAN K
ABCDCABCA	I	NDEPENDEN T
ABCDCAC		CRITICI SE
ABCDCAC		CRITICI SM
ABCDCAD		OPINION
ABCDCAEAB		ENGAGEMEN T
ABCDCAEB	P	OSITIONS
ABCDCAED	D	EFICIENC Y
ABCDCAED	PR	OVISIONS
ABCDCAEFD		CHARACTER
ABCDCAEFDGHEGA		CHARACTERISTIC
ABCDCBABC	IN	TERPRETER
ABCDCBCEA	HO	STILITIES
ABCDCEA	BRI	DGEHEAD
ABCDCEA	M	EDICINE
ABCDCEA	D	EFINITE
ABCDCEA	S	EPARATE
ABCDCEA		SURPRIS E
ABCDCEAFC	QU	ALIFICATI ON

MISCELLANEOUS PATTERNS—Continued

ABCDCEAFE	P	ERSISTENT
ABCDCEBA		ELIGIBLE
ABCDCECA	D	ESTITUTE
ABCDCECDA	CO	NSTITUTIN G
ABCDCEFGAB		PHOTOGRAPH Y
ABCDCEFGCA	DEM	OBILIZATIO N
ABCDCEFGCA	M	OBILIZATIO N
ABCDDA	R	ECOMME ND
ABCDDA	T	OBACCO
ABCDDA		SHELLS
ABCDDAB	B	EACHHEA D
ABCDDAEACBE		INEFFICIENC Y
ABCDDAEFAF	R	ECOMENDED
ABCDDAEFGHICE	R	ECOMMENDATION
ABCDDEA		DROPPED
ABCDDEA	AI	RSUPPOR T
ABCDDEA	A	RTILLER Y
ABCDDEAEC		COEFFICIE NT
ABCDDECDFA		SCHOOLHOUS E
ABCDDEFCGHA	MI	SCELLANEOUS
ABCDDEFEACGE		CLASSIFICATI ON
ABCDDEFGGEDBA	R	ECONNAISSANCE
ABCDEA		AERONA UTICS
ABCDEA	R	AILHEA D
ABCDEA		AIRPLA NE
ABCDEA		AMBULA NCE
ABCDEA	CO	ASTGUA RD
ABCDEA	M	ATERIA L
ABCDEA	S	ATURDA Y
ABCDEA	C	AUSEWA Y
ABCDEA	N	AUTICA L
ABCDEA		BLOCKB USTER
ABCDEA	ME	CHANIC
ABCDEA		CHEMIC AL
ABCDEA		CONDUC T
ABCDEA		DISLOD GE
ABCDEA		DOWNED
ABCDEA	B	ECAUSE
ABCDEA	D	ECIPHE R
ABCDEA	D	ECLARE
ABCDEA	OBJ	ECTIVE
ABCDEA	L	ECTURE
ABCDEA	V	EHICLE S
ABCDEA		ENCODE
ABCDEA	COMP	ENSATE
ABCDEA		ENTIRE
ABCDEA	R	EPLACE
ABCDEA	R	EPULSE D
ABCDEA	CONSID	ERABLE
ABCDEA	INT	ERPOSE
ABCDEA	S	ERVICE
ABCDEA		EUROPE
ABCDEA		EUROPE AN
ABCDEA		EXCITE
ABCDEA	T	HROUGH
ABCDEA		IDENTI CAL
ABCDEA		IDENTI FY
ABCDEA		INHABI TED
ABCDEA	D	IRECTI ON
ABCDEA		MEDIUM
ABCDEA	SY	NCHRON IZE
ABCDEA	JU	NCTION
ABCDEA	CO	NFIDEN T
ABCDEA		NOTHIN G
ABCDEA	E	NTRAIN
ABCDEA	L	OCATIO N
ABCDEA	REV	OLUTIO N
ABCDEA	DEC	ORATIO N
ABCDEA	T	ORPEDO
ABCDEA		OVERCO MING
ABCDEA	T	RAILER S
ABCDEA	T	RAWLER
ABCDEA	DI	RECTOR
ABCDEA		REPAIR
ABCDEA	NO	RTHWAR D
ABCDEA	C	RUISER
ABCDEA	I	SLANDS
ABCDEA		STRIPS
ABCDEA		SUNRIS E
ABCDEA		TARGET
ABCDEA	NOR	THEAST
ABCDEA		THREAT
ABCDEA	NOR	THWEST
ABCDEA		TWELFT H
ABCDEA	L	UMINOU S
ABCDEAA		EIGHTEE N
ABCDEAAE		SUBMISSI ON
ABCDEAAFED		EIGHTEENTH
ABCDEAB		INVADIN G
ABCDEAB	F	LEXIBLE
ABCDEAB		NATIONA L
ABCDEAB		REQUIRE
ABCDEAB		RESTORE D
ABCDEAB	OU	TSKIRTS

MISCELLANEOUS PATTERNS—Continued

ABCDEABA		DEMANDED	ABCDEBAED		ARBITRATI ON
ABCDEABD		IMPEDIME NTA	ABCDEBFA	B	RIGADIER
ABCDEABE	AT	OMICBOMB	ABCDEBFAGA		ENCOUNTERE D
ABCDEABF		REPAIRED	ABCDEBFCAGBF		INTERNATIONA L
ABCDEABFB		REQUIREME NT	ABCDEBFDGA		NAVIGATION
ABCDEABFD		NATIONALI SM	ABCDEBFGAF	H	EADQUARTER S
ABCDEABFDC		NATIONALIT Y	ABCDEBFGHA	R	ESPONSIBLE
ABCDEABFE		MARKSMANS HIP	ABCDEBFGHBCGIA		NATURALIZATION
ABCDEABFFGHD		SHARPSHOOTER	ABCDECA	E	NLISTIN G
ABCDEAC		AUTOMAT IC	ABCDECA		PRINCIP AL
ABCDEAC	AI	RCONTRO L	ABCDECA		PRINCIP LE
ABCDEACFB		ANTEDATIN G	ABCDECA		SKIRMIS H
ABCDEAD		CONTACT	ABCDECAB	I	NTERMENT
ABCDEAD	V	ICTORIO US	ABCDECAC	I	NTERVENE
ABCDEAD	C	RUISERS	ABCDECACFE	M	AINTENANCE
ABCDEADFD		THREATENE D	ABCDECAFCDA		TRANSATLANT IC
ABCDEAE		ENCODED	ABCDECBA		NEGLIGEN T
ABCDEAE	P	ERMANEN T	ABCDECBA		REVOLVER
ABCDEAE		FORTIFI ED	ABCDECBA	P	ROTECTOR
ABCDEAE		REQUIRI NG	ABCDECBAFB		NEGLIGENCE
ABCDEAEFGC		TRADITIONA L	ABCDECCFA		DISCUSSED
ABCDEAFA	R	EPLACEME NT	ABCDECDCAFC	I	NTERFERENCE
ABCDEAFAGE		EXCITEMENT	ABCDECFA		ENCIRCLE
ABCDEAFAGHEAID		IDENTIFICATION	ABCDECFA		EVACUATE
ABCDEAFB		CLERICAL	ABCDECFBA		SEAPLANES
ABCDEAFB		INVASION	ABCDECFEA		STANDARDS
ABCDEAFBC		RESOURCES	ABCDEDA	N	EWSPAPE R
ABCDEAFC	DES	IGNATION	ABCDEDA		MARITIM E
ABCDEAFC	RES	IGNATION	ABCDEDA	CO	NTRABAN D
ABCDEAFC	CO	NFIDENTI AL	ABCDEDA	C	OALITIO N
ABCDEAFD	D	IMENSION	ABCDEDA	BA	ROMETER
ABCDEAFE		ADJUTANT	ABCDEDA	GY	ROMETER
ABCDEAFE		INTERIOR	ABCDEDA	HYD	ROMETER
ABCDEAFE	I	NFLUENCE	ABCDEDA	HYG	ROMETER
ABCDEAFF	R	EADINESS	ABCDEDA	PSYCH	ROMETER
ABCDEAFGA	D	ECIPHERME NT	ABCDEDAB	C	ONDITION
ABCDEAFGAFB		MEDIUMBOMBE R	ABCDEDAC	REC	OGNITION
ABCDEAFGD		LEGISLATI ON	ABCDEDAFC	N	EWSPAPERS
ABCDEAFGE	CO	MPARTMENT	ABCDEDFA		DICTATED
ABCDEAFGEE		SMOKESCREE N	ABCDEDFA		EXCAVATE
ABCDEBA		DELAYED	ABCDEDFA		EXHIBITE D
ABCDEBA	D	ETONATE	ABCDEDFAC		ANTICIPAT E
ABCDEBA		INDEMNI TY	ABCDEDFAC		CLEARANCE
ABCDEBA	D	ISPERSI ON	ABCDEDFACDGB		ANTICIPATION
ABCDEBA		RECOVER	ABCDEDFCAB		INTERESTIN G
ABCDEBA		SURPLUS	ABCDEDFCGAHB		INAUGURATION
ABCDEBAB		ARBITRAR Y	ABCDEDFDA		ARTIFICIA L

MISCELLANEOUS PATTERNS—Continued

ABCDEDFDEAB	C	ONSTITUTION
ABCDEDFDGHAIF		CHRONOLOGICAL
ABCDEDFGA	PR	OCLAMATIO N
ABCDEDFGA	P	RELIMINAR Y
ABCDEDFGABHED		INDETERMINATE
ABCDEDFGADB	P	RELIMINARIE S
ABCDEDFGHAGD		ADMINISTRATI VE
ABCDEDFGHAGDIE		ADMINISTRATION
ABCDEEA		ENROLLE D
ABCDEEA	P	ERSONNE L
ABCDEEA		IMPOSSI BLE
ABCDEEACB	S	IGNALLING
ABCDEEAFDBC		INTELLIGENT
ABCDEEAFDBGD		INTELLIGENCE
ABCDEEDFGBA		RECONNOITER
ABCDEEDFGBAFE		RECONNOITERIN G
ABCDEEFAB		ENROLLMEN T
ABCDEEFAB	C	ONFESSION
ABCDEEFAE		EMBASSIES
ABCDEEFDGFA		DISAPPEARED
ABCDEEFGCAHB		INTERRUPTION
ABCDEFA	C	ABLEGRA M
ABCDEFA		AMERICA N
ABCDEFA	C	AMOUFLA GE
ABCDEFA		CHRONIC AL
ABCDEFA		CONFLIC T
ABCDEFA	DIS	CREPANC Y
ABCDEFA	S	EABORNE
ABCDEFA		EMPLOYE R
ABCDEFA		ENCIPHE R
ABCDEFA		ENFORCE
ABCDEFA		ENLISTE D
ABCDEFA	D	EPLOYME NT
ABCDEFA		EQUIPME NT
ABCDEFA	FIGHT	ERPLANE
ABCDEFA		ESCORTE D
ABCDEFA	D	ESCRIBE
ABCDEFA	J	ETPLANE
ABCDEFA		EXCLUDE
ABCDEFA		INCLUSI VE
ABCDEFA		LOGICAL
ABCDEFA	F	ORMATIO N
ABCDEFA	T	RANSFER
ABCDEFA		REGULAR
ABCDEFA	P	RISONER
ABCDEFA		SAILORS
ABCDEFA		SECTORS

ABCDEFA		SERIOUS LY
ABCDEFA	E	STABLIS H
ABCDEFA		TONIGHT
ABCDEFAA		EMPLOYEE
ABCDEFAAF	T	RANSFERRE D
ABCDEFAAGC	T	RANSFERRIN G
ABCDEFAB		INCLUDIN G
ABCDEFAB		RADIOGRA M
ABCDEFAB	P	REMATURE
ABCDEFABA		EMPLACEME NT
ABCDEFAC		INTEGRIT Y
ABCDEFAC	P	RISONERS
ABCDEFACB	IN	TRODUCTOR Y
ABCDEFACD		ALTERNATE
ABCDEFACGF		ALTERNATIN G
ABCDEFAD		CONTRACT
ABCDEFAD	D	ESTROYER
ABCDEFAD		INTERVIE W
ABCDEFAD		OPERATOR
ABCDEFAD	FI	RECONTRO L
ABCDEFAD	P	ROCEDURE
ABCDEFADB	D	ESTROYERS
ABCDEFADF	T	RANSVERSE
ABCDEFAE	D	ISCONTIN UE
ABCDEFAEGHEC	D	ISCONTINUANC E
ABCDEFAF		EXPANDED
ABCDEFAF	I	MPROVEME NT
ABCDEFAFCD	R	ADIOSTATIO N
ABCDEFAGA		ENCIPHERE D
ABCDEFAGAB		ENFORCEMEN T
ABCDEFAGB		AEROPLANE
ABCDEFAGB	D	ETACHMENT
ABCDEFAGB		INFLATION
ABCDEFAGB		REINFORCE
ABCDEFAGB		TRAJECTOR Y
ABCDEFAGBDB		REIMBURSEME NT
ABCDEFAGBHBD		REINFORCEMEN T
ABCDEFAGC		INTERDICT
ABCDEFAGCAHB		INTERDICTION
ABCDEFAGE	D	EPARTMENT
ABCDEFAGEC	D	EPARTMENTA L
ABCDEFAGFD		REGISTRATI ON
ABCDEFAGHAB		ENCIPHERMEN T
ABCDEFAGHEBC		CONFISCATION
ABCDEFAGHFD		INVESTIGATE
ABCDEFAGHFAIB		INVESTIGATION
ABCDEFAGHFAIBE		INVESTIGATIONS

MISCELLANEOUS PATTERNS—Continued

ABCDEFAGHIF	B	REAKTHROUGH
ABCDEFBA		DECLARED
ABCDEFBA		DEPARTED
ABCDEFBA		DEPLOYED
ABCDEFBA		DEPORTED
ABCDEFBA		DETACHED
ABCDEFBA		EMPLOYME NT
ABCDEFBA		ENTRAINE D
ABCDEFBA		REGISTER
ABCDEFBA	P	ROJECTOR
ABCDEFBAB		MEASUREME NT
ABCDEFBABGHD		MEASUREMENTS
ABCDEFBGA		ENDURANCE
ABCDEFBGBA		DECIPHERED
ABCDEFCA		ESTIMATE
ABCDEFCA		NORTHERN
ABCDEFCAB		ESTIMATES
ABCDEFCAD	D	OMINATION
ABCDEFCAGFC		ESTIMATEDAT
ABCDEFCBA		DETONATED
ABCDEFCCFA		DISTRESSED
ABCDEFCEA		DISPERSED
ABCDEFCGA		ELABORATE
ABCDEFDA	D	EPARTURE
ABCDEFDAB	C	USTOMHOUS E
ABCDEFDBAB		INTERVENIN G
ABCDEFDBCAGB		INTERVENTION
ABCDEFDEAB		INTERFERIN G
ABCDEFDGAB	DEM	ONSTRATION
ABCDEFDGAHCD		INTERMEDIATE
ABCDEFDGHA		HYDROGRAPH IC
ABCDEFEA	R	EINSTATE
ABCDEFEAB	F	INGERPRIN T
ABCDEFEAGACE	R	EINSTATEMENT
ABCDEFEAGDB		CERTIFICATE
ABCDEFECACD		THERMOMETER
ABCDEFECAE		CONFERENCE
ABCDEFEDCGCAHB		INTERPRETATION
ABCDEFEFA	C	OMPETITIO N
ABCDEFEGA	D	EMOBILIZE
ABCDEFEGA	C	OMPUTATIO N
ABCDEFFA	UN	DERSTOOD
ABCDEFFA		IMPRESSI ON
ABCDEFFAGE		IMPRESSIVE
ABCDEFFEDAGBC		INSTALLATIONS
ABCDEFFGAB	C	ONGRESSION AL
ABCDEFGA		DISARMED

ABCDEFGA	M	ECHANIZE D
ABCDEFGA	T	ECHNIQUE
ABCDEFGA	R	ECOGNIZE
ABCDEFGA		ENFILADE
ABCDEFGA		EQUALIZE
ABCDEFGA		EQUIPAGE
ABCDEFGA		EQUIVALE NT
ABCDEFGA	D	ESIGNATE
ABCDEFGA		EXCHANGE
ABCDEFGA		GROUPING
ABCDEFGA		GUARDING
ABCDEFGA		INSECURI TY
ABCDEFGA	D	IPLOMATI C
ABCDEFGA	E	NTRUCKIN G
ABCDEFGA		NUMBERIN G
ABCDEFGA		OBJECTIO N
ABCDEFGA		OPERATIO N
ABCDEFGA		SOLDIERS
ABCDEFGA	DI	SPATCHES
ABCDEFGA		WITHDRAW
ABCDEFGA		WITHDREW
ABCDEFGAB	D	ESPATCHES
ABCDEFGAB	U	NDERSTAND
ABCDEFGAB		WITHDRAWI NG
ABCDEFGABF		ENLISTMENT
ABCDEFGAC	I	NSTRUMENT
ABCDEFGAC	F	OUNDATION
ABCDEFGACB	I	NSTRUMENTS
ABCDEFGAD		SOUTHEAST
ABCDEFGAD		SOUTHWEST
ABCDEFGADG		SOUTHWESTE RN
ABCDEFGAEHBC		CONSTRUCTION
ABCDEFGAFE		IMPRACTICA BLE
ABCDEFGAG		WITHDRAWA L
ABCDEFGAHB		INSPECTION
ABCDEFGAHCGIDE		RECONSTRUCTION
ABCDEFGBA		DESCRIBED
ABCDEFGBA		DESTROYED
ABCDEFGBA		DETRAINED
ABCDEFGBA		REMAINDER
ABCDEFGBA		TRANSPORT
ABCDEFGBACAHGD		TRANSPORTATION
ABCDEFGBAE		TRANSPORTS
ABCDEFGBHA		ESTABLISHE D
ABCDEFGBHIAKC		ESTABLISHMENT
ABCDEFGCAG		CONFIDENCE
ABCDEFGCHEA		RANGEFINDER

MISCELLANEOUS PATTERNS—Continued

ABCDEFGDAHB	INSTRUCTION
ABCDEFGDAHBC	INSTRUCTIONS
ABCDEFGDBFHA	CE NTRALIZATION
ABCDEFGDHAIC	OBSTRUCTIONS
ABCDEFGDHFAE	ORGANIZATION
ABCDEFGEA	H EAVYBOMBE R
ABCDEFGEHA	D ESCRIPTIVE
ABCDEFGFABF	I NCOMPETENCE
ABCDEFGFAG	I NCOMPETENT
ABCDEFGGAG	H EAVYLOSSES
ABCDEFGHA	CONSPIRAC Y
ABCDEFGHA	DOMINATED
ABCDEFGHA	C ENTRALIZE
ABCDEFGHA	EXCLUSIVE
ABCDEFGHA	EXPANSIVE
ABCDEFGHA	EXPLOSIVE
ABCDEFGHA	MECHANISM
ABCDEFGHAB	C ONSUMPTION
ABCDEFGHADB	INFORMATION
ABCDEFGHAGC	CONVALESCEN T
ABCDEFGHBA	DESIGNATED
ABCDEFGHBA	DESPATCHED
ABCDEFGHBIKA	DISORGANIZED
ABCDEFGHCAEB	INTRODUCTION
ABCDEFGHCAEB	D ISCREPANCIES
ABCDEFGHDAB	C ONFIRMATION
ABCDEFGHDGCA	NORTHWESTERN
ABCDEFGHDIKA	REVOLUTIONAR Y
ABCDEFGHEEHA	COUNTERATTAC K
ABCDEFGHFA	D EMONSTRATE
ABCDEFGHFCAG	AGRICULTURAL
ABCDEFGHIA	DISPATCHED
ABCDEFGHIA	OBSERVATIO N
ABCDEFGHIA	SUBMARINES
ABCDEFGHIAB	C ONVERSATION
ABCDEFGHIAE	C OMPENSATION
ABCDEFGHIAF	R OADJUNCTION
ABCDEFGHIDAB	C ONSIDERATION
ABCDEFGHIFKA	SEARCHLIGHTS
ABCDEFGHIGBA	DEMONSTRATED
ABCDEFGHIJDA	SIMULTANEOUS

D. DIGRAPHIC IDIOMORPHS: GENERAL [1]

			AB	AB				
–G	EN	ER	AL	AL	AR	M–		
		NE	ED	ED				
–P	RO	CE	ED	ED				
–S	UC	CE	ED	ED				
–D	ET	RA	IN	IN	G–			
		–L	IN	IN	G–			
		–M	IN	IN	G–			
	OB	TA	IN	IN	G–			
		QU	IN	IN	E–			
		RA	IN	IN	G–			
	RE	MA	IN	IN	G–			
		SH	IN	IN	G–			
	–T	RA	IN	IN	G–			
		CR	IS	IS				
PO	SI	TI	ON	ON				
		–A	RE	RE	EN	FO	RC	ED
		–A	SU	SU	AL			
		BO	TH	TH	E–			
		WI	TH	TH	E–			
	–P	AR	TI	TI	ON			
	RE	PE	TI	TI	ON			
			VI	VI	D–			

		AB	—	AB		
	–M	AI	NT	AI	N–	
	RE	AR	GU	AR	D–	
		CH	UR	CH		
		DE	CI	DE		
		DE	CO	DE		
		DI	VI	DI	NG	
	SP	EA	RH	EA	D–	
	–R	ED	UC	ED		
–S	CH	ED	UL	ED		
	–B	EE	NN	EE	DE	D–
		EM	BL	EM		
	AM	EN	DM	EN	T–	
CO	NT	EN	TM	EN	T–	
–S	EV	EN	TE	EN		
–S	EV	EN	TE	EN	TH	
		EN	TR	EN	CH	
		ER	AS	ER		

			AB	—	AB		
		TH	ER	EF	ER	EN	CE
		TH	ER	ES	ER	VE	
		WH	ER	EV	ER		
–C	AR	EL	ES	SN	ES	S–	
			GE	OR	GE		
		SC	HO	OL	HO	US	E–
–I	LL	UM	IN	AT	IN	G–	
			IN	CL	IN	E–	
	–F	IR	IN	GL	IN	E–	
		MA	IN	TA	IN		
–I	NF	AL	LI	BI	LI	TY	
		–A	ME	ND	ME	NT	
		SO	ME	TI	ME		
		–O	NE	NI	NE		
			NO	TK	NO	WN	
			NO	WK	NO	WN	
–A	PP	OI	NT	ME	NT		
–C	ON	TE	NT	ME	NT		
		–C	OM	PR	OM	IS	E–
		–P	ON	TO	ON		
	–T	HR	OU	GH	OU	T–	
		–N	OW	KN	OW	N–	
			PH	OS	PH	OR	US
			PO	ST	PO	NE	
	TR	OO	PS	HI	PS		
		PA	RA	PH	RA	SE	
		–P	RE	FE	RE	NC	E–
			RE	FE	RE	NC	E–
	–T	HE	RE	FO	RE		
		–P	RE	PA	RE		
			RE	TI	RE		
			RE	VE	RE	NT	
		–C	RO	SS	RO	AD	S–
CA	RE	LE	SS	NE	SS		
		AT	TE	MP	TE	D–	
			TH	AT	EH	E–	
	–F	OR	TH	WI	TH		
–I	NV	ES	TI	GA	TI	ON	
		ES	TI	MA	TI	ON	
	–D	ES	TI	NA	TI	ON	
		AC	TI	VI	TI	ES	
		–H	UM	DR	UM		

[1] See subpar. 68*e*, Chapter IX.

	AB	—	—	AB	
-P	AN	AM	AC	AN	AL
	AR	BI	TR	AR	Y-
	AS	SO	ON	AS	
AC	CE	PT	AN	CE	
	EM	PL	AC	EM	EN T-
-Q UA RT	ER	MA	ST	ER	
-I NT	ER	PR	ET	ER	
-A CC	ES	SO	RI	ES	
	IN	CL	UD	IN	G-
-D	IR	EC	TF	IR	E-
TO	MO	RR	OW	MO	RN IN G-
PA	NA	MA	CA	NA	L-
-I	NT	ER	ME	NT	
-I	NT	ER	VE	NT	IO N-
CO	NT	IN	GE	NT	
-C	ON	DI	TI	ON	
-T OM	OR	RO	WM	OR	NI NG
	RA	DI	OG	RA	M-
	RE	AS	SU	RE	
-P	RE	MA	TU	RE	
-D EF EN	SI	VE	PO	SI	TI ON
IN	TE	RD	IC	TE	D-
QU AR	TE	RM	AS	TE	R-
IN	TE	RP	RE	TE	R-
IN	TE	RR	UP	TE	D-
-F OR	TI	FI	CA	TI	ON

	AB	—	—	—	AB	
	AR	MO	RE	DC	AR	
	EN	FO	RC	EM	EN	T-
RE	EN	FO	RC	EM	EN	TS
	IN	DE	TE	RM	IN	AT E-
	IN	TE	RE	ST	IN	G-
	IN	TE	RF	ER	IN	G-
	IN	TE	RV	EN	IN	G-
-I	NC	OM	PE	TE	NC	E-
-C	ON	GR	ES	SI	ON	AL
-D EM	ON	ST	RA	TI	ON	
-C	ON	SU	MP	TI	ON	
	PH	OT	OG	RA	PH	
	TH	IR	TE	EN	TH	

	AB	—	—	—	—	AB
-I	NS	TA	LL	AT	IO	NS
-C	ON	CE	NT	RA	TI	ON
-C	ON	FL	AG	RA	TI	ON
-C	ON	SI	DE	RA	TI	ON

	AB	—	AB	AB	
	IN	CL	IN	IN	G-
MA	IN	TA	IN	IN	G-

E. DIGRAPHIC IDIOMORPHS: PLAYFAIR[1]

	AB	BA	
SC	AB	BA	RD
	AF	FA	BL E–
	AF	FA	IR
–B	AG	GA	GE
–H AW	AI	IA	N–
	AL	LA	RE AS
–B	AL	LA	ST
–F	AL	LA	CY
IN ST	AL	LA	TI ON S–
–P AR	AL	LA	X–
	AP	PA	RA TU S–
	AP	PA	RE L–
	AP	PA	RE NT
	AP	PA	RE NT LY
	AR	RA	NG E–
	AR	RA	Y–
–B	AR	RA	CK S–
–B	AR	RA	GE
–E MB	AR	RA	SX SE D–
–N	AR	RA	TI ON
	AS	SA	IL AN T
	AS	SA	UL T–
–A MB	AS	SA	DO R–
–I MP	AS	SA	BL E–
–M	AS	SA	CR E–
–P	AS	SA	GE
	AT	TA	CH
	AT	TA	CK
	AT	TA	IN
–B	AT	TA	LI ON
–R	AT	TA	N–
	BO	OB	YT RA P–
IN	DE	ED	
–W	EB	BE	D–
	EF	FE	CT
	EF	FE	CT IV E–
CO MP	EL	LE	D–
–EX XC	EL	LE	NC E–
–E XC	EL	LE	NT
–E XP	EL	LE	D–
–I MP	EL	LE	D–
–P	EL	LE	T–
PR OP	EL	LE	D–
–R EP	EL	LE	D–

	AB	BA	
SH	EL	LE	D–
–H	EM	ME	DI N–
ST	EM	ME	D–
ST	EP	PE	D–
AV	ER	RE	D–
CO NF	ER	RE	D–
–I NT	ER	RE	D–
–R EF	ER	RE	D–
	ES	SE	NC E–
	ES	SE	NT IA L–
AD DR	ES	SE	S–
–C OM PR	ES	SE	D–
CO NF	ES	SE	D–
IM PR	ES	SE	D–
–L	ES	SE	N–
–M	ES	SE	NG ER
PR	ES	SE	D–
PR OF	ES	SE	D–
–P RO GR	ES	SE	D–
–S TR	ES	SE	D–
–S TR	ES	SE	S–
–V	ES	SE	L–
WI TN	ES	SE	S–
AB	ET	TE	D–
–C IG AR	ET	TE	S–
–B	ET	TE	R–
–L	ET	TE	R–
–E IG	HT	TH	RE E–
–R	IB	BI	NG
FO RB	ID	DI	NG
–D	IF	FI	CU LT
–B	IL	LI	ON
–F	IL	LI	NG
–K	IL	LI	NG
–M	IL	LI	ME TE R–
–M	IL	LI	NG
–M	IL	LI	ON
SH	IL	LI	NG
SP	IL	LI	NG
–T	IL	LI	NG
–W	IL	LI	AM
–W	IL	LI	NG
	IM	MI	GR AN T–
	IM	MI	GR AT IO N–

[1] See footnote 35 on p. 168.

			AB	BA			
			IM	MI	NE	NT	
		SW	IM	MI	NG		
	-B	EG	IN	NI	NG		
		SP	IN	NI	NG		
		-W	IN	NI	NG		
		CL	IP	PI	NG		
		SH	IP	PI	NG		
	-S	TR	IP	PI	NG		
			IR	RI	GA	TI	ON
		-M	IS	SI	NG		
		-M	IS	SI	ON		
	-A	DM	IS	SI	ON		
		EM	IS	SI	ON		
		-H	IS	SI	NG		
	PE	RM	IS	SI	ON		
TR	AN	SM	IS	SI	ON		
		EM	IT	TI	NG		
		-F	IT	TI	NG		
	-S	PL	IT	TI	NG		
	PE	RM	IT	TI	NG		
-A	FT	ER	NO	ON			
	FO	RE	NO	ON			
			NO	ON	TI	ME	
		-F	OL	LO	W-		
		-H	OL	LO	W-		
		-C	OM	MO	N-		
		-C	OM	MO	TI	ON	
PO	SI	TI	ON	NO	RT	HO	F-
	-R	EC	ON	NO	IT	ER	
			OP	PO	RT	UN	E-
			OP	PO	RT	UN	IT Y-
			OP	PO	SE		
			OP	PO	SI	TE	
			OP	PO	SI	TI	ON
		-C	OR	RO	BO	RA	TE
		-C	OR	RO	DE		
	-T	OM	OR	RO	W-		
		-B	OT	TO	M-		
		-C	OT	TO	N-		
		CA	RE	ER			
		-S	UC	CU	MB	ED	

			AB	—	BA		
		PR	AC	TI	CA	BL	E-
		PR	AC	TI	CA	L-	
		-T	AC	TI	CA	L-	
	-D	IV	EB	OM	BE	R-	
			EN	GI	NE	ER	
		-G	EN	UI	NE		
	-I	NT	ER	FE	RE		
	-I	NT	ER	FE	RE	NC	E-
	-P	EN	ET	RA	TE		
		-R	EV	OL	VE	R-	
			IN	FI	NI	TE	
		-D	IS	PO	SI	TI	ON
		-S	IT	UA	TI	ON	
		CA	NA	DI	AN		
VE	TE	RI	NA	RI	AN		
		NI	NE	TE	EN		
		NI	NE	TE	EN	TH	
			PE	RC	EP	TI	ON
		-P	RE	MI	ER		
	-S	UR	RE	ND	ER		
	-O	UR	SE	LV	ES		
	TH	EM	SE	LV	ES		
		DE	SE	RV	ES		
		RE	SE	RV	ES		
			SE	RV	ES		

	AB	—	—	BA		
	DE	BA	RK	ED		
	DE	CL	AR	ED		
	DE	FE	ND	ED		
	DE	MA	ND	ED		
	DE	PA	RT	ED		
	DE	PL	OY	ED		
	DE	PO	RT	ED		
	DE	SE	RT	ED		
	DE	TA	CH	ED		
PR	EC	ED	EN	CE		
	EM	PL	OY	ME	NT	
	EN	TR	AI	NE	D-	
	ME	AS	UR	EM	EN	T-
	NE	GL	IG	EN	CE	
	NO	TA	TI	ON		
	PA	RA	GR	AP	H-	
	RE	CE	IV	ER		
	RE	CO	RD	ER		
	RE	GI	ST	ER		
	RE	PE	AT	ER		
	RE	PO	RT	ER		
	RE	VO	LV	ER		
-P	RO	JE	CT	OR		
AS	SE	MB	LI	ES		

	AB	—	—	—	BA	
	DE	SE	CR	AT	ED	
	DE	SI	GN	AT	ED	
	DE	SP	AT	CH	ED	
	EN	EM	YP	LA	NE	S-
-D	ET	ER	IO	RA	TE	
-S	EV	EN	TY	FI	VE	
	IR	RE	GU	LA	RI	TY
	NO	MI	NA	TI	ON	
	SU	SP	IC	IO	US	

AB	—	—	—	—	BA
DE	MO	NS	TR	AT	ED
NO	TI	FI	CA	TI	ON

F. DIGRAPHIC IDIOMORPHS: FOUR-SQUARE [1]

(Grouped by number of significant letters in the idiomorphic pattern)

Two letters

A– A–

B LO CK	AD ED	
I NV	AD ED	
D	AM AG	E
CO MM	AN DS	
I SL	AN DS	
A IR PL	AN ES	
E NE MY PL	AN ES	
DE SI GN	AT ED	
E ST IM	AT ED	
I ND IC	AT ED	
C	AV AL	RY
M	AV AL	
P RO	CE DU	RE
ME	CH AN	IZ ED
IM ME	DI AT	EL Y
WI TH	DR AW	
WI TH	DR EW	
	EM ER	GE NC Y
L IE UT	EN AN	T
	FI FT	EE N
	FI FT	H
	FI FT	Y
BR ID	GE HE	AD
V	IC IN	IT Y
W	IT HD	RA W
A DD	IT IO	NA L
A MM UN	IT IO	N
CO ND	IT IO	N
RE CO GN	IT IO	N
E	LE ME	NT
	MI LI	TA RY
	MI NI	MU M
	NI NT	H
P	OI NT	
T	OM OR	RO W
	PO NT	ON
	RE QU	ES T
	RE QU	IR E
P	RI SO	NE R
RE	SI ST	AN CE
D IS PO	SI TI	ON
PO	SI TI	ON
	SO UT	H

A–_A–

	SQ UA	DR ON
FI GH	TE RP	LA NE
MO	TO RI	ZE D
D EP AR	TU RE	
	UN US	UA L

A– — A–

S	AB OT AG	E
D ET	AC HM EN	T
H	AS BE EN	
	BA TT AL	IO N
	BO MB ED	
	CA SU AL	TI ES
	CA SU AL	TY
	CO MB AT	
	CO OR DI	NA TE S
	DI RE CT	IO N
	DI SP AT	CH
ME	DI UM BO	MB ER
	DI VE BO	MB ER
R OA	DJ UN CT	IO N
R	EP LA CE	ME NT
R	ET RE AT	
S	EV ER AL	
JU	NC TI ON	
CO	NF IR MA	TI ON
I	NF OR MA	TI ON
I	NT EL LI	GE NC E
	PA TR OL	
	SA BO TA	GE
	SE VE RE	
AC	TI VI TY	
A	TT EN TI	ON
S	UC CE SS	FU LL Y

A– — — A–

	AR TI LL ER	Y
	AT TA CK ED	
R	EE NF OR CE	
R	EE NF OR CE	ME NT
	ID EN TI FY	
	IM PO SS IB	LE

A– — — A–

	MO VE ME NT	
E	MP LA CE ME	NT
	PE RS ON NE	L
A	RT IL LE RY	

A– — — — A–

	CO MM UN IC AT	IO NS
	CO NC EN TR AT	E
R	EO RG AN IZ AT	IO N
	LI EU TE NA NT	
CO	NS TR UC TI ON	

A– — — — — A–

CO MM IS SI ON ED

–B –B

UN	AB LE	
OB ST	AC LE	
	AD VA	NC E
	AG AI	NS T
R	AI LH	EA D
PR EP	AR AT	IO N
A SS	AU LT	
B OM	BA RD	
A IR	BO RN	E
S EA	BO RN	E
A DV AN	CI NG	
VI	CI NI	TY
	DE TA	CH
	DE TA	CH ME NT
H AV	EB EE	N
M OV	EM EN	T
	EN EM	Y
R	ES ER	VE
R	ET UR	N
	FL AN	K
	FO LL	OW
B AG	GA GE	
	HA SB	EE N
A PP RO AC	HI NG	
DE BO UC	HI NG	
L AU NC	HI NG	
I MM ED	IA TE	LY

[1] See subpar. 69d Chapter IX.

Two letters (cont.)

```
         -B -B
   IN IT|IA TE|
       F|IF TH|
   TE RR|IT OR|Y
       S|IX TY|
 M IS CE|LL AN|EO US
       E|LE VA|TI ON
       E|LE VE|N
        |LI AI|SO N
      DA|MA GE|
        |MO RN|IN G
       U|NU SU|AL
        |OB JE|CT IV E
       C|OL ON|
       C|OL ON|EL
SU PE RI|OR IT|Y
       M|OT OR|IZ ED
        |OU TS|KI RT S
   EQ UI|PM EN|T
    A VE|RA GE|
    B AR|RA GE|
      AI|RC RA|FT
AN TI AI|RC RA|FT
        |RE MA|IN
R EQ UI|RE ME|NT
    M IS|SI NG|
```

```
         -B -B
    P ER|SO NN|EL
ES TI MA|TE DA|T
    P LA|TO ON|
       S|UP PL|Y
       S|UP PO|RT
      NA|VA LB|AS E
    F OR|WA RD|
   WI ND|WA RD|

         -B — -B
       C|AS UA LT|Y
       P|AT RO LS|
 B AT TL|ES HI PS|
        |GE NE RA|L
    W IL|LA TT AC|K
    T RA|NS MI SS|IO N
    R EC|OG NI TI|ON
    T RO|OP SH IP|
        |RE GI ME|NT
      CA|RR IE RS|
      MI|SS IO NS|
        |TW EN TY|
  R EQ|UE ST ED|
```

```
  -B — — -B
I|DE NT IF IC|AT IO N
M|EC HA NI ZE|D
D|EP LO YM EN|T
M|ES SE NG ER|
D|ES TR OY ER|
A|IR SU PP OR|T
V|IS IB IL IT|Y
 |ME SS EN GE|R
I|MP AS SA BL|E
I|MP OS SI BL|E
A|NT IA IR CR|AF T
C|OM MA ND IN|G
 |OP ER AT IO|N
 |PR IS ON ER|
 |PR OC ED UR|E
 |RE EN FO RC|E
 |TR AN SP OR|TA TI ON
 |YE ST ER DA|Y

        -B — — — -B
     R|EC OM ME ND ED|
      |HE AV YL OS SE|S
  R EC|OM ME ND AT IO|N
     C|OM MU NI CA TI|ON
  R EC|ON NO IT ER IN|G
```

Three letters

```
     A- A- A-
    N|AV AL BA|SE
 R EQ|UI SI TI|ON
```

```
A- A- — A-
RE QU ES TE D
```

```
        -B -B -B
  B OM|BA RD ME|NT
      |EL EM EN|TS
    EN|GA GE ME|NT
```

Four letters

```
   AB A- -B
  H|EA DQ UA|RT ER S
   |EL EV EN|

   AB -B A-
   |CA NC EL|
 RE|CO NN AI|SS AN CE

   AB -B — A-
   |AD VA NC ED|
   |EN EM YT AN|KS

   AB — A- -B
   SI GH TI NG
```

```
 A- AB -B
 AD DI TI ON AL

 A- AB — -B
 SO UT HW ES T

   A- A- -B -B
 W IT HD RA WA L

 A- A- — A- A-
 CO MM AN DI NG

 A- A- — -B -B
 RE QU IR EM EN T
```

```
   A- -B AB
 M|OR NI NG|
 P|OS TP ON|E

   A- -B -B — A-
   RE CO NN OI TE R

   A- -B — AB
   IN TE RD IC T

   A- -B — A- -B
 S AT IS FA CT OR Y

   A- — A- C- C-
   DI SP AT CH ES
```

Four letters (cont.)

A- — — C- A- C-
RO AD JU NC TI ON

-B AB A-
DI SP|OS IT IO|N
P|OS IT IO|N
|PR ES EN|T
RE|PR ES EN|T

-B A- AB
RE PE AT ED

-B A- A- -B
DE ST RO YE R

-B A- -B — A-
UN ID EN TI FI ED

-B A- — AB
U NS UC CE SS FU L

-B A- — A- -B
ME DI UM BO MB ER

-B A- — -B A-
VI SI BI LI TY

-B A- — — AB
IN FO RM AT IO N

-B A- — — A- -B
IN ST AL LA TI ON

-B -D -B — -D
CR OS SR OA DS

-B -D -D -B
AI RS UP PO RT

-B -D — -D -B
|IN ST RU CT IO|N
C|ON ST RU CT IO|N

-B — A- AB
F IG HT ER PL AN ES

-B — A- — — AB
E ST AB LI SH ME NT

-B — -B A- A-
EN CO UN TE RE D

-B — — -B -D -D
RE IN FO RC EM EN T

Five letters

A- -B AB — -B
NA VA LA TT AC K

A- -B — -B AB
R EC ON NA IS SA NC E

-B A- A- — AB
DI ST RI BU TI ON

-B A- -B AB
RE PL AC EM EN T

-B -D — -D -B -D
IN ST RU CT IO NS

Six letters

AB CB C- A-
P OS IT IO NS

AB -D -D AB
C|ON DI TI ON|
|RA DI OG RA|M

A- A- -B AB A-
RE QU IS IT IO N

A- CB — A- CB
Q UA RT ER MA ST ER

A- CB — CB A-
SC HO OL HO US E

A- — CB A- — CB
ID EN TI FI CA TI ON

-B AB AD -D
A DM IN IS TR AT IV E

Seven letters

-B AD — -B -D AD
RE EN FO RC EM EN T

Eight letters

AB -B AD — -B AD
QU AR TE RM AS TE R

AB -B C- AB CB
EM PL AC EM EN T

AB -D C- AD C- -B
IN TE RD IC TI ON

APPENDIX 4

SERVICE TERMINOLOGY AND STEREOTYPES

Familiarity with the style and peculiar phraseology which exist in military messages greatly facilitates the cryptanalytic recovery of the plain text of any such messages which have been encrypted. Thus, this appendix has been compiled to comprise notes on those idiosyncrasies present in military messages which are of particular interest and aid to the cryptanalyst. The notes which are applicable to the messages of all Services are grouped together in Section *A*; those which are applicable to messages of ground, naval, and air origin, respectively, constitute Sections *B*, *C*, and *D*; those which apply to special types of messages, such as weather messages, are contained in Section *E*; and remarks on stereotypic beginnings and endings of messages comprise Section *F*.

Although the notes contained herein derive primarily from U. S. military communications, many apply as well to the military communications of other countries. At the very least, this appendix indicates the types of information on message style and phraseology which, when known concerning the messages of any source, can be quite helpful in the cryptanalysis of such messages.

SERVICE TERMINOLOGY AND STEREOTYPES

A. Remarks Applicable to Messages of All Services

1. When mention is made of time in military messages, it is conventionally specified in terms of the 24-hour clock system (ending at midnight), in which time is expressed as a group of four numerals. The first two numerals of the group denote the hour and the last two numerals, the minute after the hour; for example, 6:00 AM appears as 0600 and 6:00 PM appears as 1800. For any current month, the day may be indicated by prefixing the four-digit time group with a two-digit date group, indicating the day of the month; for example, 080600 denotes 6:00 AM on the 8th day of the month. In some instances, a four-digit time group or six-digit date-time group occurring in a message may be found with a literal suffix, giving rise to such groups as 1800Z, 080600Q, etc; this suffix may be any one of the letters A to I or K to Z and is a type of designator used in communications practices to refer to a particular one of the 24 time zones of the earth.

2. Administrative messages in general often contain many sequences of numbers, brought about by numerous references to previous messages and to various Service regulations (among other items), reference generally being made on the basis of identifying serial numbers and dates which such items usually bear; specific illustrations of this fact appear in several of the succeeding paragraphs in this appendix. Furthermore, administrative messages contain references to items having literal designations; to minimize errors in this connection such designations are often spelled out phonetically, by means of a phonetic alphabet, such as one of the following:

ABLE	JIG	SUGAR	ALFA	JULIETT	SIERRA
BAKER	KING	TARE	BRAVO	KILO	TANGO
CHARLIE	LOVE	UNCLE	CHARLIE	LIMA	UNIFORM
DOG	MIKE	VICTOR	DELTA	MIKE	VICTOR
EASY	NAN	WILLIAM	ECHO	NOVEMBER	WHISKEY
FOX	OBOE	XRAY	FOXTROT	OSCAR	XRAY
GEORGE	PETER	YOKE	GOLF	PAPA	YANKEE
HOW	QUEEN	ZEBRA	HOTEL	QUEBEC	ZULU
ITEM	ROGER		INDIA	ROMEO	

3. The messages of all Services exhibit a high content of abbreviations; for this reason, the following list of frequently-encountered abbreviations is included:

Navy officer ranks

FADM_______fleet admiral
ADM________admiral
VADM_______vice admiral
RADM_______rear admiral
COMO_______commodore
CAPT_______captain
CDR________commander
LCDR_______lieut. commander
LT_________lieutenant
LTJG_______lieut. jr. grade
ENS________ensign

Army, air officer ranks

GEN________general
LTGEN______lieut. general
MAJGEN_____major general
BRIGGEN___brigadier general
COL________colonel
LTCOL______lieut. colonel
MAJ________major
CAPT_______captain
1ST LT_____first lieutenant
2D LT______second lieutenant

Punctuation		*Miscellaneous*			
CLN	colon	CG	commanding general	HQ	headquarters
CMA	comma	CO	commanding officer	LAT	latitude
PARA	paragraph	COM	commander	LONG	longitude
PAREN	parenthesis	COMDT	commandant	LTR	letter
PD	period	DET	detachment	MSG	message
RPT	repeat	ETA	estimated time of arrival	NR	number
		ETD	estimated time of departure	RE	reference
		GMT	Greenwich mean time	UR	your

4. The identity of the person originating a military message may appear as a signature at the end of a message and the addressee'si dentity may appear at the beginning. Either, or both, of these may be "buried" in the middle of the message, set off by parentheses. If the signature is at the end of the message, it may be preceded by STOP (or PERIOD, or SIGNED, or both. The identification of the originator or addressee may consist merely of his command designation (e. g., COMMANDING GENERAL, FIRST ARMY) or it may consist of his name and rank, followed by COMMANDING or some other appropriate amplifying data (e. g., in the Army, his branch of service).

Examples:

JONES, COLONEL, ARTILLERY
COMMANDING OFFICER, THIRD REGIMENT
COMMANDER, DESTROYER SQUADRON SIX
SMITH, FLIGHT LEADER, SECOND SQUADRON

5. In military communications, long messages are often broken into parts, each part subsequently being treated as a separate message. Thus, messages arise which begin "PART (number) OF (number) PARTS", or "(number) PART MESSAGE PART (number)", often separated from the following message text by STOP or simply by an "X".

B. Remarks on Items Appearing in Ground (Army) Messages

1. When mention of an army unit appears in a military message, its size (echelon) is indicated, generally preceded by a numerical or literal designation and, as further information concerning the unit, its branch of service may be included. The several echelons of the U. S. Army, listed in descending order of size, are: army, corps, division (DIV), brigade, regiment (REGT), battalion (BN), company[1] (CO), platoon. Some of the branches of service which may appear, as mentioned above, are: Infantry (INF), Artillery (ARTY), Signal Corps (SIG C), Armor, Ordnance (ORD), Engineers (ENG), Quartermaster (QM).

Examples of unit designations:

(a) "A" Company, 39th Infantry Regiment, 9th Infantry Division
(b) 1st Armored Division
(c) 57th Signal Battalion

2. In connection with 1, above, an army is the normal command of a general (four stars); a corps being the command of a lieutenant general; a division, the command of a major general; and a brigade, the command of a brigadier general. A regiment is normally commanded by a

[1] An artillery unit at this echelon is termed a *attery.*

colonel; a battalion may be commanded by a lieutenant colonel or a major; a company, by a captain; and a platoon, by a lieutenant.

3. For reference purposes, when giving locations of units, readily-recognizable landmarks such as hills, crossroads and road junctions may be referred to in terms of their altitude above sea level (in number of feet), if such landmarks do not bear proper names which are suitable for the purpose. Thus, a reference, in a military message, to CROSSROADS SIX FIVE ZERO would apply to that particular crossroads within a preselected area which is located at an altitude of 650 feet. If, within any preselected area of reference, there are two or more landmarks of any given type which are both at the same altitude, it is necessary to affix a distinctive letter or number to the altitude designation of each, in order to identify a particular one clearly; thus, such a reference as CROSSROADS SIX FIVE ZERO DASH [hyphen] B may be encountered. In this connection, CROSSROADS may be found abbreviated as "CR", and ROADJUNCTION as "RJ".

4. The location of any particular unit may be specified in terms of its location with respect to a particular place or locality, or to a particular landmark. However, its location may also be specified by stating how it is located on a specific map or portion of a map. This gives rise in military messages to phrases such as COORDINATES ONE FIVE POINT TWO FOUR DASH ONE NINE FOUR POINT SEVEN, wherein the numbers before the "dash" indicate the unit's location with respect to the horizontal grid lines of a preselected map and the numbers after the "dash" indicate its location with respect to the vertical grid lines.

5. Specific highways, turnpikes, and other roadways are often identified in military messages by stating the place names of their terminal points; thus the highway running between Grizurbeto and Bolzano could be called the GRIZURBETO DASH [hyphen] BOLZANO HIGHWAY. Similarly, when reference is made to an imaginary straight line across the terrain in a particular area, such a line may be identified by specifying any recognizable landmarks between which the line runs; for example, LINE CROSSROADS THREE ONE FIVE DASH ROADJUNCTION TWO NINE EIGHT.

6. Reference to military weapons may be made in terms of the diameter of the weapons' bore (expressed in millimeters, inches or by caliber). Included below is a brief list of U. S. Military weapons with their sizes as they are generally expressed.

Rifles
30 Cal.

Machine Guns
30 Cal.
50 Cal.

Antitank Guns
37 MM
75 MM (recoilless rifle)
105 MM (recoilless rifle)

Field Artillery Howitzers
105 MM
155 MM
240 MM

Antiaircraft Guns
40 MM
90 MM
120 MM

Mortars
60 MM
81 MM
4.2" (chemical)

7. Included below is a brief list of frequent words appearing in low-echelon ground traffic; the abbreviation for certain ones are appended in parentheses after them. In addition to the words listed below, numbers and ranks/titles will be found to have a high frequency of occurrence.

ACROSS
ACTIVITY
ADDITIONAL
ADVANCE
ADVISE
AFTERNOON
AIRPLANE
AMMUNITION (AMMO)
AREA
ARMORED
ARMY
ARRIVE
ARTILLERY (ARTY)
ATTACK
BARRAGE
BATTALION (BN)
BRIDGE
CAPTURE
CASUALTIES
COMMA
COMMAND POST (CP)
COMMUNICATION (COMM)
COMPANY (CO)
CONCENTRATION
COUNTERATTACK
CROSSROADS (CR)
DAILY
DASH
DEFEND
DEFENSIVE
DISPOSITION
DIVISION (DIV)
EAST
EMPLACEMENTS
ENEMY
ENLISTED PERSONNEL
FIRE
FLANK
FORCE
FROM (FM)
HEADQUARTERS (HQ)
HEAVY
HILL
HOSTILE
IDENTIFICATION (IDENT)
IMMEDIATELY
INFANTRY (INF)
INFORMATION (INFO)
LARGE
LEFT
LIGHT
LINE
LOCATION
MACHINEGUN (MG)
MESSAGE (MSG)
MORNING
MORTAR
MOUNTAIN
MOVE
MOVEMENT
NEAR
NEUTRALIZE
NIGHT
NORTH
NOTHING
OBJECTIVE
OFFENSIVE
OFFICER
ORDER
PATROL
PLANE
POSITION
PREPARE
PRISONER
PROCEED
RADIO
RAILROAD
READY
RECEIVE
RECONNAISSANCE (RCN)
REFERENCE (RE) (REF)
REGIMENT (REGT)
REINFORCEMENTS
REPORTS
REQUEST
REQUIRE
REQUISITION
RESERVES
RIGHT
RIVER
ROAD
ROADJUNCTION (RJ)
ROCKET
SEND
SMALL
SOUTH
STOP
SUPPLY
SUPPORT
TANKS
TODAY
TOMORROW
TONIGHT
TROOPS
VICINITY
WEST
WOODS
YESTERDAY

C. Remarks on Items Appearing in Naval Messages

1. Mention of various sized groupings of vessels are found in messages of naval origin, among which those mentioned below are quite frequently encountered. A major naval force is called a *fleet*, and the levels of echelonment (or subdivision) within a fleet are the *task force*, *task group*, and *task unit* (in descending order of size). The basic unit of all fleet vessels is termed a *division*, and is comprised of two or more vessels of the same type; in this connection, when mention is made of a division in a naval message, the particular type of vessel of which the division is made up is often specified; for example, CRUISER DIVISION. A *squadron* is made up of two or more divisions of submarines, destroyers, landing ships or other light vessels; and a *flotilla* comprises two or more such squadrons.

2. In connection with 1, above, a fleet is normally commanded by an admiral (four stars); a task force being the command of a vice admiral; and a task group, the command of a rear admiral. Furthermore, in time of war the officer in command of a convoy or flotilla often holds the rank of commodore; the officer commanding an individual ship may range in rank from captain on down, depending on the type of ship.

3. A list of the main combat vessels is included below; the approximate maximum speed of each, which is expressed in *knots* (i. e., nautical miles per hour), is shown in parentheses.

BATTLESHIP	(35)
CARRIER	(35)
CARRIER ESCORT	(15)
CRUISER	(30)
DESTROYER	(35)
DESTROYER ESCORT	(25)
SUBMARINE	(20, on surface; 10, submerged)

When a particular vessel is mentioned in a naval message, it may be identified by a numerical designation, by a group of letters, or by some proper name.

4. In naval messages, the direction of an object from a ship, or the course of a particular naval vessel or unit at sea is given as a compass bearing expressed in degrees (from 0 to 359); for example, BEARING ZERO EIGHT FIVE. In some instances the statement of a bearing will be followed by the word TRUE or MAGNETIC, indicating that the bearing is measured from the geographical pole (true north) or the magnetic pole (not corrected for variation), respectively.

5. The position of a particular naval vessel or unit at sea may be specified in a naval message by stating its latitude and longitude in degrees and minutes. The latitude may be from 0 to 90 degrees and the longitude from 0 to 180 degrees; a specified latitude is generally followed by NORTH or SOUTH (as appropriate) and, similarly, longitude is followed by EAST or WEST. For example, LATITUDE THREE ZERO DEGREES TWO ONE MINUTES NORTH LONGITUDE ONE FOUR TWO DEGREES ONE SIX MINUTES WEST, or more briefly LATITUDE THREE ZERO DASH TWO ONE NORTH LONGITUDE ONE FOUR TWO DASH ONE SIX WEST. If position is stated in conjunction with a bearing, it is not necessary to state both latitude and longitude; and the location, NORTH or SOUTH, with respect to the equator or EAST or WEST with respect to Greenwich Meridian may be found omitted where no ambiguity arises. Positions are also sometimes given as a bearing and distance in miles from a specific point.

6. The following words may be expected to appear frequently in a selection of naval messages of various types:

AIRCRAFT	EXECUTE	RADAR
ALTITUDE	FLEET	REJOIN
BEACH	FLIGHT	RENDEZVOUS
BLOCKADE	FORMATION	SAIL
BOMBED	FUEL	SEA
CARGO	GUARDING	SHIFT
CHANNEL	HARBOR	SHIP
COASTAL	KNOTS	SORTIE
COMMAND	LATITUDE	SQUADRON
CONTACT	LONGITUDE	STARBOARD
CONVOY	MILES	STRAFED
CORRECTED	MINE (FIELD)	STRAIT
COURSE	MISSION	TARGET
CRAFT	NAVAL	VESSELS
DEPART	NAVY	VIA
DEPLOY	OPERATIONS	VOYAGE
EMBARK	PILOT	WEATHER
ESCORT	PORT	

D. Remarks on Items Appearing in Air Messages

1. The various elements of which an air force is composed and which may be mentioned in air messages are given below. The smallest grouping of aircraft, composed of one or more aircraft of a particular type, is called a *flight*. A *squadron* is two or more *flights* of the same type; a *group* is made up of two or more squadrons; a *wing* comprises two or more groups; an *air division* is composed of two or more wings; and two or more divisions constitute an *air force*.

2. In connection with 1, above, a flight is usually commanded by a major; a squadron being the command of a lieut. colonel; a group being the command of a colonel; a wing, the command of a brigadier general; and an air force, the command of a major general.

3. Some of the types of aircraft which may be mentioned frequently in air messages are listed below; an indication of the range of speed of each type, expressed in knots, is shown in parentheses.

BOMBER	(250–400)
CARGO PLANE	(150–350)
FIGHTER	(250–500)
JET BOMBER	(350–600)
JET FIGHTER	(250–500)
LIAISON PLANE	(65–150)

4. The position of a particular aircraft may be specified in an air message by stating its latitude and longitude in degrees and minutes, sometimes in conjunction with its altitude in feet. (Latitude may be from 0 to 90 degrees and longitude from 0 to 180 degrees.)

E. Remarks on Special Types of Messages

1. *Weather messages.* Any generalization on the specific elements which a weather message will contain would perforce be rather tenuous, the contents of a particular weather message generally being dependent on its source and purpose. However, there are certain elements which may be expected to appear in most weather messages; these are listed below with an indication of the terms in which each is generally expressed:

a. Identification of the originating station (by code number, or location).

b. Wind speed (knots) and direction (tens of degrees, from 00–36).

c. Amount of low clouds (tenths of sky covered) and their height (hundreds of feet).

d. Types of low, medium, and high clouds (e. g., cumulus, stratus, cirrus, etc.).

e. Temperature of the air and temperature of the dew point (both in degrees Fahrenheit).

f. Present and past weather (e. g., clear, partly cloudy, cloudy or overcast, fog, drizzle, rain, snow, showers, thunderstorm, etc.).

g. Horizontal visibility (miles).

h. Atmospheric pressure (tens, units, and tenths digits in millibars) and barometric tendency (e. g., falling, steady, rising, etc.).

2. *Air-to-ground position reports.* Position reports made by aircraft in flight may be expected to contain the majority of the following elements of information:

a. Position of the aircraft (in latitude and longitude or with respect to some locality on the ground).

b. Time.

c. Speed.

d. Altitude.

e. Weather conditions.

f. Estimated time of arrival at next reporting point or at destination.

F. Stereotypic Beginnings and Endings

Within the confines of the comparatively limited scope of military messages, stereotypy of phraseology is inevitable. Particularly in the beginnings of messages is this limitation apparent; thus these positions lend themselves most readily to attack by the cryptanalyst. The following is a list of stereotypes having a high frequency of positional occurrence, and therefore providing a fruitful source for cribs. It is to be noted that a stereotypic initial word often may suggest a whole opening phrase. For example, if a message of low-echelon ground origin begins with the word HEAVY, then it is not too unlikely that the opening phrase is "HEAVY

ARTILLERY (FIRE, BARRAGE) (FALLING, INTERDICTING) ....," which might be expanded into "HEAVY ARTILLERY FIRE FALLING ON OUR POSITIONS (NORTH, EAST, SOUTH, WEST) OF...."

BEGINNINGS

ACKNOWLEDGE
ADVANCE
ADVISE (THIS COMMAND)
(THIS HEADQUARTERS)
ARRIVE
ATTACK
ATTENTION
CANCEL
CITE
COMMANDING (GENERAL) (OFFICER)
COMMUNICATION (OFFICER)
CONCENTRATE (-ION OF)
CONFIRM
CONTINUE
DEPART (-URE)
DISCONTINUE
EFFECTIVE
ENEMY
EQUIPMENT
EXPEDITE (SHIPMENT)
FOLLOWING (ARE) (IS)
FOR
FROM
HEAVY
HOSTILE
INFORM (-ATION)
IN REPLY (TO YOUR) (MESSAGE)
LOCATION (OF)
NUMBERS (1, 1st, 2, 2d, etc.)
ORDERS
OUR
PARAPHRASE
PLEASE
POSITION
PREPARE (TO) (-ATIONS FOR)
PROCEED
RECEIPT
RECEIVE
RECOMMEND (-ATION) (-ED)
REFER (-RING) (TO) (YOUR)
REFERENCE (YOUR, MY) (MESSAGE, RADIOGRAM, TELEGRAM) (NUMBER) (DATED, OF)
REPEAT
REPORT
REQUEST
REQUIRE
RERAD
REURAD
SEND
SITUATION REPORT
STATUS REPORT
SUPPLY
VERIFY
YOUR (COMMAND) (ORGANIZATION)

ENDINGS

ACKNOWLEDGE
ADVISE (IMMEDIATELY)
CONFIRM
END
END OF MESSAGE
IMMEDIATELY
NUMBERS (1, 1st, 2, 2d, etc.)
PERIOD
REPLY
REFERENCE
REQUESTED
SIGNED (name)
STOP
TITLES (MAJ, COL, etc.)

Appendix 5

LETTER FREQUENCY DATA—FOREIGN LANGUAGES

The letter frequency data contained herein have been complied from selected newspaper and magazine articles comprising war communiqués and other military-type text. In the material which was processed there were place names and words foreign to each particular language; these words account for the presence of certain non-characteristic letters in the data given herein for those languages which make use of the Roman alphabet.

LETTER FREQUENCY DATA—FOREIGN LANGUAGES

A. German Letter Frequency Data

1–a.—*Absolute frequencies of single letters of German plain text, arranged alphabetically, based on 60,046 letters of text.* (The letters X and Y are derived from foreign words contained in German plain text.)

A------	3, 601	G-----	1, 921	L-----	1, 988	Q-----	6	V----	523
B------	1, 023	H-----	2, 477	M-----	1, 360	R-----	4, 339	W----	899
C------	1, 620	I-----	4, 879	N-----	6, 336	S-----	4, 127	X----	12
D------	3, 248	J-----	192	O-----	1, 635	T-----	3, 447	Y----	24
E------	10, 778	K-----	747	P-----	499	U-----	2, 753	Z----	654
F------	958								60, 046

1–b.—*Monographic kappa plain, German language*=.0787 (I. C.=2.05)

1–c.—*Frequency distribution of single letters based on 60,046 letters of German plain text, reduced to 1,000 letters, arranged according to their relative frequencies.*

E---------	180	T--------	57	G--------	32	F--------	16	P-------	8
N---------	106	D--------	54	O--------	27	W--------	15	J-------	3
I---------	81	U--------	46	C--------	27	K--------	13	Y-------	----
R---------	72	H--------	41	M--------	23	Z--------	11	X-------	----
S---------	69	L--------	33	B--------	17	V--------	9	Q-------	----
A---------	60								1, 000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants in 60,046 letters of German plain text. Percentage of 8 most frequent letters in German plain text.*

Vowels A, E, I, O, U, and Y=39.4%
High-Frequency Consonants D, N, R, S, and T=35.8%
Medium-Frequency Consonants B, C, F, G, H, L, M, and W=20.4%
Low-Frequency Consonants J, K, P, Q, V, X, and Z=4.4%

8 most frequent letters (in descending order of frequency) E, N, I, R, S, A, T, and D=67.9%

1–e.—*Absolute frequencies of single letters as initial letters of 9,568 words in German plain text, arranged according to their frequencies.*

D-------	1, 716	U------	550	Z------	343	K------	263	O-----	135
A-------	762	W------	544	M------	339	P------	181	T-----	106
S-------	698	G------	461	N------	306	R------	167	C-----	22
E-------	686	B------	460	F------	280	L------	158	Q-----	2
I-------	581	V------	408	H------	265	J------	135		9, 568

~~CONFIDENTIAL~~

2–a.—*Frequency distribution of digraphs based on 60,046 letters of German plain text, reduced to 5,000 digraphs.*

First Letter \ Second Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	4	14	10	4	33	7	9	7	1	1	2	33	13	48		2		22	27	23	36	1	1			1
B	6				48		1	1	5			3			3			11	2	1	3		1			1
C								130			5															
D	29	2		8	127	1	2	2	60		1	3	2	2	4	1		5	6	2	9	2	2			2
E	13	22	10	31	13	12	32	24	90	2	6	28	25	235	3	6		195	68	28	24	9	15			7
F	7	1		3	15	7	2		2			2	1	1	3			10	2	10	12					
G	10	1		8	78	1	2	2	8		2	7	1	3	1			11	8	5	8	2	1			1
H	29	1		8	64	1	2	1	14		2	8	3	6	6	1		20	4	23	7	2	3			1
I	3	1	39	7	91	2	18	7	2		7	12	11	84	13	1		7	53	44	1	2	1			1
J	4				8																3					
K	12	1		1	11		1	1	1			5			9			10	1	5	4					
L	26	3	1	6	27	1	2		37		3	20	1	2	4				10	12	6	1				1
M	16	3		4	26	2	2	1	14	1	2	1	11	1	8	5		1	3	3	9	1	1			1
N	39	12	1	118	58	9	57	8	35	4	10	6	10	18	8	5		4	36	27	20	10	17			14
O	1	3	5	3	11	3	3	3			1	18	6	33	1	5		18	12	4	1	1	5			1
P	10				5	4		1	2			1			7	2		7		1	1					
Q																					1					
R	34	11	5	35	60	9	12	9	37	2	11	6	8	12	19	3		6	22	18	26	6	8			5
S	14	6	55	13	46	3	7	3	30	1	5	4	7	3	16	6		2	40	57	9	5	5		1	5
T	25	3		17	88	2	4	6	40	1	3	7	3	4	4			14	20	7	16	2	10			13
U	1	2	8	2	37	15	5	1			2	2	11	76		2		18	28	14	1	1	2			1
V	1				19				3						21											
W	16				24				20	3					6						6					
X																										
Y																										
Z	1			1	8				5			1			2					4	27		4			

~~CONFIDENTIAL~~

2-b.—*Digraphic kappa plain, German language*=.0111 (I. C.=7.50).

2-c.—*The 95 digraphs comprising 75% of German plain text, based on the table of 5,000 digraphs (Item 2-a), arranged according to their relative frequencies.*

EN...	235	RE...	60	NA...	39	ED...	31	TA...	25	HR...	20	TU....	16
ER...	195	DI...	60	LI...	37	SI...	30	EM...	25	LL...	20	WA....	16
CH...	130	NE...	58	UE...	37	HA...	29	EH...	24	VE...	19	UF....	15
DE...	127	NG...	57	RI...	37	DA...	29	EU...	24	RO...	19	FE....	15
ND...	118	ST...	57	AU...	36	EL...	28	WE...	24	OR...	18	EW....	15
IE...	91	SC...	55	NS...	36	US...	28	HT...	23	UR...	18	AB....	14
EI...	90	IS...	53	NI...	35	ET...	28	AT...	23	NN...	18	HI....	14
TE...	88	BE...	48	RD...	35	AS...	27	AR...	22	RT...	18	TR....	14
IN...	84	AN...	48	RA...	34	LE...	27	RS...	22	OL...	18	SA....	14
GE...	78	SE...	46	AE...	33	NT...	27	EB...	22	IG...	18	MI....	14
	[1] 1,236	IT...	44		[2] 2,508	ZU...	27	VO...	21	NW...	17	NZ....	14
UN...	76	SS...	40	ON...	33	LA...	26	NU...	20	TD...	17	UT....	14
ES...	68	TI...	40	AL...	33	ME...	26	WI...	20	MA...	16	SD....	13
HE...	64	IC...	39	EG...	32	RU...	26	TS...	20	SO...	16		3,750

2-d.—*Frequent digraphs in German plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

EN..	235	NE..	58	IE..	91	EI..	90	ES..	68	SE..	46	AN..	48	NA...	39
ER..	195	RE..	60	IN..	84	NI..	35	IS..	53	SI..	30	IT..	44	TI...	40
DE..	127	ED..	31	GE..	78	EG..	32								

2-e.—*Frequent digraphs in German plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

CH..	130	HC..	0	ND..	113	DN..	2	NG..	57	GN...	3	SC..	55	CS...	0

2-f.—*Doublets occurring in German plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2-a).*

SS....	40	EE...	13	FF...	7	RR...	6	GG...	2	PP...	2	OO....	1
LL....	20	MM...	11	TT...	7	AA...	4	II...	2	HH...	1	UU....	1
NN....	18	DD...	8										

2-g.—*The 22 digraphs appearing 100 or more times as beginnings of words in 9,568 words in German plain text, arranged according to their absolute frequencies.*

DE...	805	EI..	300	DA..	244	WE..	192	ER..	153	ZU..	124	ST..	112
DI...	567	GE..	299	VO..	214	VE..	172	HA..	140	MI..	117	IN..	111
UN...	428	BE..	252	SI..	197	WI..	155	AL..	134	SN..	112	SE..	111
AU...	318												

[1] The 10 digraphs above this line compose 25% of German plain text.
[2] The 37 digraphs above this line compose 50% of German plain text.

3–a.—*The 102 trigraphs appearing 100 or more times in 60,046 letters of German plain text, arranged according to their absolute frequencies.*

SCH	666	ERE	313	NEN	198	AUS	162	IST	142	HRE	124	NAU	108
DER	602	ENS	270	SSE	191	TIS	159	STA	141	HER	122	TSC	107
CHE	599	CHT	264	REI	190	BER	157	DES	140	ACH	119	ENN	106
DIE	564	NGE	263	TER	188	ENI	157	FUE	139	GES	118	ERG	106
NDE	541	NDI	259	REN	185	ENG	155	NTE	139	ABE	117	RIT	106
EIN	519	IND	254	EIT	184	ION	154	UER	138	ERA	117	EHR	105
END	481	ERD	248	EBE	178	SEN	152	ERU	137	BEN	116	CHA	104
DEN	457	INE	247	ENE	175	ITI	151	TUN	136	MEN	115	VON	104
ICH	453	AND	246	LIC	175	AUF	149	SEI	133	RIE	112	SIC	103
TEN	425	RDE	239	EGE	173	IES	149	ESE	132	VER	110	IGE	102
UNG	377	ENA	214	DAS	172	ASS	148	ERT	128	LAN	109	ITE	101
HEN	332	ERS	212	ENU	171	ENW	148	NDA	127	ENB	108	ENZ	100
UND	331	EDE	209	NUN	169	ENT	146	IED	126	ESS	108	ERB	100
GEN	321	STE	205	NER	166	ERI	143	ERN	125	LLE	108	EUT	100
ISC	317	VER	204	RUN	163	EST	142						

3–b.—*The 25 trigraphs appearing 50 or more times as beginnings of words in 9,568 words in German plain text, arranged according to their absolute frequencies.*

EIN	242	DAS	79	SCH	73	AUF	64	DEU	61	UNT	57	UEB	53
VER	170	BRI	79	AUS	69	NER	63	GES	60	GRO	56	POL	52
FUE	89	DIE	76	SEI	68	IND	62	GEG	59	AUC	55	WIR	51
SIC	86	NIC	73	STA	65	ALL	61						

4.—*The 121 tetragraphs appearing 50 or more times in 60,046 letters of German plain text, arranged according to their absolute frequencies.*

SCHE	398	NUND	106	NICH	80	ATIO	65	RSCH	60	ENZU	54
ISCH	317	ITIS	104	UNGD	80	GEND	65	EDEN	59	ITEN	54
CHEN	296	SICH	103	EITE	79	TEND	65	ERGE	59	KRIE	54
NDER	243	RUNG	101	DEUT	78	EBER	64	ESSE	59	RIEG	54
EINE	218	ANDE	100	FUER	78	GEGE	64	UNTE	59	SDIE	54
ENDE	216	UNGE	100	CHTE	77	POLI	64	EICH	58	URCH	53
NDIE	176	EREI	94	EGEN	76	SIND	64	TLIC	58	ALLE	52
LICH	168	TION	93	NEIN	76	TUNG	64	INER	57	DERS	52
ICHT	151	SEIN	92	IESE	75	ENSI	62	EBEN	56	ENWE	52
TISC	146	IEDE	91	ERST	74	FUTS	62	ENDA	56	HABE	52
ERDE	144	LAND	91	RDIE	74	LITI	62	ENST	56	ONEN	52
ENDI	141	SSEN	90	ERDI	72	UEBE	62	IGEN	56	SCHI	52
NDEN	136	BRIT	89	STEN	72	UTSC	62	ONDE	56	DEND	51
RDEN	133	DASS	86	CHER	71	AUCH	61	TENS	56	DISC	51
ENUN	120	NTER	86	INDI	71	DENS	61	EDIE	55	ENEN	51
ICHE	120	EDER	83	REIN	71	EIND	61	ERTE	55	NACH	51
INDE	111	EREN	83	DERE	70	OLIT	61	HREN	55	NDAS	51
NGEN	110	ENGE	81	NGDE	70	SCHA	61	TDIE	55	UNGS	51
ERUN	109	ENAU	80	ENBE	68	SCHL	61	ATEN	54	ABEN	50
DIES	108	ENIN	80	RITI	66	WERD	61	DIEB	54	NBER	50
TSCH	107										

5.—*Average length of words in German plain text*=6.3 letters.

B. *French Letter Frequency Data*

1–a.—*Absolute frequencies of single letters of French plain text, arranged alphabetically, based on 55,758 letters of text.*

A	4,480	G	624	L	2,737	Q	616	V	801
B	406	H	276	M	1,617	R	4,117	W	6
C	1,944	I	4,230	N	4,406	S	4,564	X	317
D	2,198	J	184	O	3,255	T	4,057	Y	100
E	9,334	K	25	P	1,689	U	3,045	Z	84
F	646								55,758

1–b.—*Monographic kappa plain, French language*=.0777 (I. C.=2.02).

1–c.—*Frequency distribution of single letters based on 55,758 letters in French plain text reduced to 1,000 letters, and arranged according to their relative frequencies.*

E	167	T	73	C	35	G	11	J	3
S	82	O	58	P	30	Q	11	Y	2
A	80	U	55	M	29	B	7	Z	2
N	79	L	49	V	14	X	6	K	1
I	76	D	39	F	12	H	5	W	----
R	74								1,000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants in 55,758 letters of French plain text. Percentage of 8 most frequent letters in French plain text.*

Vowels A, E, I, O, U, and Y=43.8%
High-Frequency Consonants N, R, S, and T=30.7%
Medium-Frequency Consonants C, D, L, M, and P=18.3%
Low-Frequency Consonants B, F, G, H, J, K, Q, V, W, X, and Z=7.2%

8 most frequent letters (in descending order of frequency) E, S, A, N, I, R, T, and O=68.9%

1–e.—*Absolute frequencies of single letters as initial letters of 10,748 words in French plain text, arranged according to their frequencies.* (*One-letter words have been omitted.*)

D	1,445	L	784	I	315	U	240	H	67
P	929	S	664	F	313	O	177	Z	7
E	894	Q	394	T	305	G	146	K	5
A	866	R	389	N	278	B	115	W	3
C	816	M	337	V	263	J	98	Y	3
									9,853

~~CONFIDENTIAL~~

2–a.—*Frequency distribution of digraphs based on 55,758 letters of French plain text, reduced to 5,000 digraphs.*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	2	6	20	12	4	6	11		50	1		36	12	68	1	21	3	41	17	46	29	13			2	1
B	4				4				4			12			4			5	2	1	2					
C	15		6		47			11	20			5			48			4	1	8	8					
D	18			1	109			1	20	1			1		10	1		6	2		26					
E	30	4	49	48	30	15	14	3	13	5		56	58	105	4	38	12	89	154	58	27	17		8		3
F	10		2	1	9	6			8			1			8	1		10	1		1					
G	6				16		1		2			3	1	7	6			8		4	2					
H	6				6				4						3			1			4					
I	9	3	12	10	41	4	4			1		27	8	49	51	5	12	27	52	47		9		7		1
J	4				6										5						2					
K															1											
L	57		1	5	95	1		1	23			26		3	10	1			5	4	12				1	
M	22	9	1	1	52				23				13		8	9			1		4					
N	19	1	29	40	54	9	11	1	20	1		3	2	10	19	6	4	3	53	99	4	7				1
O		5	7	3	1	1	2	1	21	1		10	21	109		7		27	13	8	52	2			2	
P	30		1	1	13			2	3			11			35	9		34	1	6	4					
Q			1																		54					
R	62	2	10	13	127	2	6		24	1		16	11	8	27	5	3	7	14	19	6	7				1
S	42	2	16	32	75	5	2	1	36	2		15	8	6	22	24	11	8	41	33	24	4			1	
T	40	1	7	22	78	4	1	2	67	1		12	4	4	14	11	7	44	23	10	11	2				
U	12	3	10	5	39	4	3	1	24	3		13	6	26	1	8	1	48	26	19	1	8		13	1	
V	9				24				16						16			5			2					
W																										
X	4		3	3	3			1	1				1	1		4	1	1	2	3		1				
Y	2				2										1				2							
Z					3				1						1											

~~CONFIDENTIAL~~

2-b.—*Digraphic kappa plain, French language*=. 0093(I. C.=6.29).

2-c.—*The 87 digraphs comprising 75% of French plain text, based on the table of 5,000 digraphs (Item 2-a), arranged according to their relative frequencies.*

ES... 154	RA... 62	AI... 50	SS... 41	EA... 30	UI... 24	OM.... 21
RE... 127	[1] 1,237	EC... 49	ND... 40	EE... 30	SP... 24	NI.... 20
ON... 109	ET... 58	IN... 49	[2] 2,515	NC... 29	SU... 24	DI.... 20
DE... 109	EM... 58	ED... 48	TA... 40	AU... 29	RI... 24	CI.... 20
EN... 105	LA... 57	CO... 48	UE... 39	IR... 27	VE... 24	AC.... 20
NT... 99	EL... 56	UR... 48	EP... 38	EU... 27	TS... 23	UT.... 19
LE... 95	QU... 54	CE... 47	AL... 36	IL... 27	MI... 23	NO.... 19
ER... 89	NE... 54	IT... 47	SI... 36	RO... 27	LI... 23	RT.... 19
TE... 78	NS... 53	AT... 46	PO... 35	OR... 27	SO... 22	NA.... 19
SE... 75	ME... 52	TR... 44	PR... 34	DU... 26	MA... 22	DA.... 18
AN... 68	IS... 52	SA... 42	ST... 33	LL... 26	TD... 22	AS.... 17
TI... 67	OU... 52	IE... 41	SD... 32	US... 26	AP... 21	EV.... 17
	IO... 51	AR... 41	PA... 30	UN... 26	OI... 21	3,751

2-d.—*Frequent digraphs in French plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

ES.. 154	SE.. 75	LE.. 95	EL.. 56	RA.. 62	AR.. 41	IS.. 52	SI... 36
RE.. 127	ER.. 89	TE.. 78	ET.. 58	EM.. 58	ME.. 52	EC.. 49	CE... 47
DE.. 109	ED.. 48	TI.. 67	IT.. 47	LA.. 57	AL.. 36	AT.. 46	TA... 40
EN.. 105	NE.. 54						

2-e.—*Frequent digraphs in French plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

NT... 99	TN.. 4	QU.. 54	UQ.. 1	NS.. 53	SN.. 6	OU.. 52	UO..... 1

2-f.—*Doublets occurring in French plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2-a).*

SS.... 41	LL... 26	NN... 10	PP... 9	CC... 6	AA... 2	GG... 1
EE.... 30	MM... 13	TT... 10	RR... 7	FF... 6	DD... 1	UU... 1

2-g.—*The 22 digraphs appearing 100 or more times as beginnings of words in 10,748 words in French plain text, arranged according to their absolute frequencies.*

DE... 501	RE.. 283	PO... 222	SU.. 168	AU.. 150	DI.. 124	SO... 117
CO... 394	PA.. 268	IN.. 178	CE.. 163	NO.. 133	AL.. 122	VO... 112
QU... 347	LE.. 240	SE.. 178	ET.. 153	TR.. 127	UN.. 122	FR... 101
PR... 291						

[1] The 13 digraphs above this line compose 25% of French plain text.
[2] The 39 digraphs above this line compose 50% of French plain text.

3–a.—*The 97 trigraphs appearing 100 or more times in 55,758 letters of French plain text, arranged according to their absolute frequencies.*

ENT	588	CON	271	EST	188	ESS	151	NSE	130	EUR	115	TRA	105
ION	555	ERE	267	ERA	185	AIT	147	REN	127	NTA	115	ISS	104
TIO	433	ANT	238	ECO	184	POU	146	SQU	124	SER	115	INT	103
ONS	373	ESE	230	ESD	179	TER	146	AIR	123	ESO	112	TEN	103
RES	367	ELA	227	OND	175	COM	143	EPA	120	DEC	110	UEL	102
QUE	338	LLE	216	LEM	173	ESP	139	QUI	120	EPR	110	ANS	101
DES	313	PAR	213	NCE	173	OUS	139	SET	120	ALL	109	BLE	101
EDE	305	NDE	211	ELE	172	AIS	137	REC	119	ECE	109	QUA	101
EME	288	SDE	210	ESA	163	EMA	137	AND	118	UNE	108	CES	100
ATI	287	DEL	209	TDE	163	IER	136	ETA	118	RAI	106	ETE	100
LES	284	PRE	206	ITE	162	NTS	135	SEN	118	RLE	106	ETR	100
NTE	281	OUR	205	SSE	160	TES	135	PRO	117	SSI	106	OMM	100
TRE	280	RAN	196	ONT	157	EQU	133	ISE	116	ENE	105	TAT	100
MEN	272	IRE	191	ANC	153	IQU	131	REP	116	SUR	105		

3–b.—*The 20 trigraphs appearing 50 or more times as beginnings of words in 10,748 words in French plain text, arranged according to their absolute frequencies.*

CON	213	COM	129	FRA	93	INT	75	ETA	69	SER	61	VOU	56
POU	144	PRO	105	PAR	87	CEN	72	DAN	68	TRA	57	FAI	50
PRE	135	ALL	104	QUA	80	NOU	69	RED	65	RES	56		

4.—*The 82 tetragraphs appearing 50 or more times in 55,758 letters of French plain text, arranged according to their absolute frequencies.*

TION	431	CONS	98	LEME	83	ERAL	71	EREN	58	RESS	55
MENT	251	EPAR	98	QUEL	83	ERES	70	ESSE	58	IERE	53
ATIO	220	RESE	96	LEMA	80	DANS	67	NOUS	58	IRES	53
IONS	208	ENTE	95	PORT	80	OUVE	67	TRES	58	TEDE	53
EMEN	200	LLEM	93	ENTS	78	EMAN	66	ENER	57	EQUE	52
POUR	136	FRAN	91	EPRE	77	SENT	66	NDES	57	NDEL	52
IQUE	128	PRES	91	EDES	76	ANDE	63	NSEI	57	ECOM	51
IOND	124	ENTA	90	ESET	76	PART	62	NTDE	57	GENE	51
DELA	120	RANC	90	INTE	76	SDES	62	CAIS	56	SEIL	51
AIRE	117	ANCE	89	ALLE	75	ESEN	61	ESTI	56	ELES	50
ONDE	107	SION	89	ANTE	75	RAIT	61	ITIO	55	ETAT	50
ECON	102	COMM	88	MAND	75	ENTD	60	NEME	55	ILLE	50
ESDE	102	ELLE	84	CENT	74	SSIO	60	NERA	55	SQUE	50
ONSE	101	NTER	84	QUES	72	ENCE	59				

5.—*Average length of words in French plain text*=5.2 letters.

C. Italian Letter Frequency Data

(*In all calculations, accented letters have been combined with the corresponding unaccented letter.*)

1–a.—*Absolute frequencies of single letters of Italian plain text, arranged alphabetically, based on 57,906 letters of text.*

A	6,771	G	1,168	L	3,592	Q	227	V	1,024
B	527	H	493	M	1,441	R	4,037	W	13
C	2,367	I	6,568	N	4,094	S	2,967	X	9
D	2,258	J	18	O	5,022	T	4,139	Y	14
E	6,784	K	28	P	1,616	U	1,547	Z	527
F	655								57,906

1–b.—*Monographic kappa plain, Italian language*=.0745 (I. C.=1.94).

1–c.—*Frequency distribution of single letters based on 57,906 letters in Italian plain text, reduced to 1,000 letters, arranged according to their relative frequencies.*

E	117	R	70	P	28	F	11	K	-----
A	117	L	62	U	27	B	9	J	-----
I	113	S	51	M	25	Z	9	Y	-----
O	87	C	41	G	20	H	9	W	-----
T	72	D	39	V	18	Q	4	X	-----
N	71								1,000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants in 57,906 letters of Italian plain text. Percentage of 8 most frequent letters in Italian plain text.*

Vowels A,E,I,O,U, and Y=46.1%
High-Frequency Consonants L, N, R, and T=27.4%
Medium-Frequency Consonants C, D, G, M, P, S, and V=22.2%
Low-Frequency Consonants B, F, H, J, K, Q, W, X, and Z=4.3%
8 most frequent letters (in descending order of frequency) E, A, I, O, T, N, R, and L=70.8%

1–e.—*Absolute frequencies of single letters as initial letters of 10,481 words in Italian plain text, arranged according to their frequencies.* (*One letter words have been omitted.*)

D	1,381	L	500	T	337	U	217	J	13
C	1,041	R	403	G	333	Q	172	W	9
S	885	N	396	F	298	B	153	K	6
P	830	E	374	V	263	H	69	Y	3
A	822	M	371	O	235	Z	29	X	2
I	685								10,481

2–a.—*Frequency distribution of digraphs based on 57,847 letters of Italian plain text, reduced to 5,000 digraphs.*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	18	9	39	41	14	12	22	1	19			76	24	78	5	24	4	57	36	63	6	24				12
B	10	7			7				10			1			4			4			2					
C	32		10		20			33	33			2			64		1	5			6					
D	31			1	65				64						23			2			9					
E	23	7	31	53	15	8	22	2	25			66	18	73	6	22	4	96	62	27	6	17				4
F	9				11	7			11			1			10			6			3					
G	9				11		8	2	20			17		8	9			11			6					
H	6				27				9																	
I	66	8	52	30	31	11	11	2	11			35	31	62	44	20	3	20	48	45	15	16				7
J																					1					
K																										
L	62	3	8	6	49	2	7		56			52	4	2	21	5	1	3	6	15	7	3				
M	31	5			35				17				4		18	13					2					
N	32	1	15	26	51	6	11	1	37			3	1	10	50	4	5	2	11	66	8	4				11
O	17	4	22	27	10	5	10	1	20			45	24	86	4	25	2	55	40	14	3	18				2
P	23				30				14			2			28	11		23			7					
Q																					20					
R	64	1	8	8	71	1	7		63			4	13	9	45	2		12	9	16	10	3				3
S	20		15	1	32	2			45			2	3		25	9			31	58	12	1				
T	83		1		65	1			59			1		1	56			43	1	37	10					
U	12	2	4	3	15	1	3		10			6	3	24	8	6		9	11	15						1
V	26				23				23						10			2			2	2				
W																										
X																										
Y																										
Z	13				4				20						3											5

2–b.—*Digraphic kappa plain, Italian language*=.0081 (I. C.=5.48)

2–c.—*The 89 digraphs comprising 75% of Italian plain text, based on the table of 5,000 digraphs (Item 2–a), arranged according to their relative frequencies.*

ER.... 96	RI... 63	LL... 52	AC... 38	MA... 31	HE... 25	VE.... 23
ON.... 86	IA... 63	IC... 51	TT... 37	SS... 31	OP... 25	OC.... 22
TA.... 78	LA... 62	NE... 50	[2] 2,495	DA... 31	AM... 24	AG.... 22
AN.... 78	IN... 62	NO... 50	NI... 37	EC... 30	UN... 24	EG.... 22
AL.... 76	[1] 1,260	LE... 49	ME... 35	PE... 30	EI... 24	EP.... 22
EN.... 73	RA... 62	IS... 48	AS... 35	ID... 30	AV... 24	LO.... 21
RE.... 71	ES... 61	IT... 45	IL... 35	IE... 30	OM... 24	IP.... 20
NT.... 66	TI... 59	OL... 45	CH... 33	PO... 28	PA... 23	ZI.... 20
DE.... 65	ST... 58	RO... 45	CI... 33	OD... 27	DO... 23	SA.... 20
TE.... 65	AR... 57	SI... 44	RA... 32	ET... 27	VI... 23	CE.... 20
EL.... 65	TO... 56	IO... 43	SE... 32	VA... 26	AP... 23	QU.... 20
DI.... 64	LI... 56	TR... 43	CA... 32	ND... 26	PR... 23	GI.... 20
CO.... 64	OR... 55	OS... 40	IM... 31	SO... 25	EA... 23	3.762
AT.... 63	ED... 52	AD... 39				

2–d.—*Frequent digraphs in Italian plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

ER...... 96	RE..... 71	EL..... 66	LE..... 49	LI..... 56	IL...... 35
ON...... 86	NO..... 50	DE..... 65	ED..... 53	OR..... 55	RO...... 45
TA...... 83	AT..... 63	RA..... 64	AR..... 57	IC..... 52	CI...... 33
AN...... 78	NA..... 32	IN..... 62	NI..... 37	IS..... 48	SI...... 45
AL...... 76	LA..... 62	ES..... 62	SE..... 32	AD..... 41	DA...... 31
EN...... 73	NE..... 51	TI..... 59	IT..... 45	AC..... 39	CA...... 32

2–e.—*Frequent digraphs in Italian plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

NT...... 66	TN..... 1	ST..... 58	TS..... 1	CH..... 33	HC...... 0

2–f.—*Doublets occurring in Italian plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2–a).*

LL.... 52	AA... 18	II... 11	NN... 10	FF... 7	MM... 4	VV.... 2
TT.... 37	EE... 15	PP... 11	GG... 8	ZZ... 5	OO... 4	DD.... 1
SS.... 31	RR... 12	CC... 10	BB... 7			

2–g.—*The 26 digraphs appearing 100 or more times as beginnings of words in 10,481 words in Italian plain text, arranged according to their absolute frequencies.*

CO... 543	PE.. 210	PR.. 184	NO.. 154	SE.. 121	MA.. 112	RE... 108
DE... 505	CH.. 197	QU.. 172	PA.. 153	SO.. 121	UN.. 111	ES... 107
ST... 222	AL.. 186	NE.. 169	PO.. 141	TR.. 121	SU.. 109	TE... 103
DI... 215	IN.. 185	RI.. 162	CA.. 132	DA.. 120		

[1] The 18 digraphs above this line comprise 25% of Italian plain text.
The 43 digraphs above this line comprise 50% of Italian plain text.

3–a.—*The 90 trigraphs appearing 100 or more times in 57,906 letters of Italian plain text, arranged according to their absolute frequencies.*

DEL.. 348	STA. 215	ERE. 169	ICA. 145	SSI. 130	ODI. 114	ESI.. 107
ENT.. 348	ALI. 213	ZIO. 166	RAN. 145	NEL. 127	ORI. 114	COR.. 106
ELL.. 314	EDI. 212	ATO. 165	STR. 145	ACO. 125	RMA. 114	IAN.. 106
CON.. 306	ALL. 201	NTI. 165	ALE. 144	ATI. 125	AME. 113	TAN.. 105
CHE.. 276	ITA. 198	ANT. 163	IDI. 143	IDE. 123	ETT. 113	ATE.. 104
LLA.. 274	ANO. 197	ERA. 163	COM. 139	ADI. 121	ODE. 113	NON.. 103
ION.. 265	OST. 196	TRA. 160	ECO. 137	AND. 121	PRE. 112	VER.. 103
ONE.. 247	ERI. 187	ESS. 158	LLE. 137	TEN. 120	NDO. 110	ICA.. 101
PER.. 238	ARE. 186	ATT. 157	ONT. 136	ONO. 119	ONI. 110	OLA.. 101
EDE.. 228	TAL. 184	NTO. 156	TER. 136	ARI. 117	AZI. 109	STI.. 101
NTE.. 227	LIA. 180	ADE. 155	TAT. 134	NTR. 117	ENE. 109	OCO.. 100
ICO.. 216	IST. 174	EST. 151	TTA. 132	PAR. 116	ELA. 107	RIA.. 100
MEN.. 216	GLI. 171	RES. 146	ATA. 130	TRO. 116	ERO. 107	

3–b.—*The 19 trigraphs appearing 50 or more times as beginnings of words in 10,481 words in Italian plain text, arranged according to their absolute frequencies.*

DEL.. 217	STA. 106	QUA.. 81	PRE.. 62	DAL.. 57	PER.. 55	GRA... 53
CON.. 195	ALL. 100	PRO.. 75	NEL.. 57	ANC.. 56	RUS.. 55	STO... 51
COM.. 137	ITA.. 94	QUE.. 74				

4.—*The 57 tetragraphs appearing 50 or more times in 57,906 letters of Italian plain text, arranged according to their absolute frequencies.*

DELL... 209	ALIA... 99	ICON... 74	AGLI... 66	LIAN... 59	OPER.... 56
MENT... 188	CONT... 93	VANO... 74	ICHE... 66	TORI... 59	RUSS.... 56
IONE... 160	ADEL... 92	ECON... 73	IDEL... 64	ALLE... 58	TATO.... 55
ELLA... 150	OSTR... 88	IONI... 71	ELLE... 63	ANDO... 58	TEDE.... 55
ZION... 147	ENTO... 87	STAT... 70	NELL... 63	DALL... 58	OCON.... 54
TALI... 125	AMEN... 83	STRA... 70	IMEN... 61	NTRO... 58	SION.... 53
AZIO... 106	ALLA... 81	GLIA... 69	ANTI... 60	OCHE... 58	TANT.... 53
EDEL... 106	ENZA... 75	ISTA... 68	ATTA... 60	ANTE... 57	STOP.... 52
ITAL... 106	ONTR... 75	ODEL... 68	PART... 60	EPER... 57	NOST.... 51
ENTE... 105	ENTI... 74	ACON... 66			

5.—*Average length of words in Italian plain text*=5.5 letters.

D. *Spanish Letter Frequency Data*

1–a.—*Absolute frequencies of single letters of Spanish plain text, arranged alphabetically, based on 60,115 letters of text.*

A	6,681	G	823	L	2,174	Q	346	V	602
B	799	H	367	M	1,740	R	4,628	W [2]	36
C	3,137	I	4,920	N [1]	4,823	S	4,140	X	127
D	2,687	J	190	O	5,859	T	3,180	Y	413
E	7,801	K	22	P	1,785	U	2,172	Z	182
F	481								60,115

1–b.—*Monographic kappa plain, Spanish language*=.0747 (I. C. =1.94)

1–c.—*Frequency distribution of single letters based on 60,115 letters in Spanish plain text, reduced to 1,000 letters, arranged according to their relative frequencies.*

E	130	S	69	U	36	V	10	J	3
A	111	T	53	P	30	F	8	Z	3
O	97	C	52	M	29	Y	7	X	2
I	82	D	45	G	14	H	6	W	1
N	80	L	36	B	13	Q	6	K	----
R	77								1,000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants in 60,115 letters of Spanish plain text. Percentage of 7 most frequent letters in Spanish plain text.*

Vowels A, E, I, O, U, and Y=46.3%
High-Frequency Consonants N, R, and S=22.6%
Medium-Frequency Consonants C, D, L, M, P, and T=24.5%
Low-Frequency Consonants B, F, G, H, J, K, Q, V, W, X, and Z=6.6%

7 most frequent letters (in descending order of frequency) E, A, O, I, N, R, and S=64.6%

1–e.—*Absolute frequencies of single letters as initial letters of 10,129 words in Spanish plain text, arranged according to their frequencies.* (*One-letter words have been omitted.*)

P	1,128	L	435	Q	286	V	183	Y	27
C	1,081	R	425	I	281	F	177	W	19
D	1,012	M	403	H	230	O	169	Z	2
E	989	N	346	U	219	B	124	K	1
S	789	T	298	G	206	J	47	X	----
A	761								10,129

[1] Includes Ñ throughout all tables.
[2] From foreign words appearing in Spanish plain text.

2–a.—*Frequency distribution of digraphs based on 60,115 letters of Spanish plain text, reduced to 5,000 digraphs.*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	12	14	54	64	15	5	8	4	10	8		41	30	64	4	24	5	81	62	18	9	9			11	4
B	11				5				14	1		12			5			12	2	1	3					
C	39		5		17			8	80			3			69			6		13	18					
D	32		1	2	84			1	30					1	59	2	1	3	1		6				1	
E	20	5	47	26	17	8	21	6	9	3		44	26	126	5	23	4	94	119	17	5	10	1	8	2	3
F	2				9				12			1			7			4			5					
G	12				12				5			1		2	15			11		1	11					
H	15				3				5						6						1					
I	43	8	42	29	40	5	8			1		14	16	50	67	4	1	16	27	24	1	8				5
J	4				5										3						3					
K					1																					
L	44		5	5	35	1	3		28			9	5	1	17	5	1	2	4	5	5	3			1	
M	32	10			42				30						15	10					6					
N	41	2	33	37	41	10	6	2	28	1		5	4	3	43	10	2	4	21	91	12	6			1	1
O	19	17	28	26	16	6	5	5	4	1		22	33	104	4	29	7	58	73	12	3	5		2	9	1
P	30		1		16				5			8			31			34	1	3	19					
Q																					29					
R	74	1	12	10	94	1	12		45	1	1	6	15	11	43	7	3	10	10	15	9	6			1	1
S	32	2	18	15	57	3	2	4	41	1		5	7	5	22	26	4	6	10	57	23	2			4	
T	60		1		67				35						56			34			11					
U	13	6	11	5	52	1	3		9			9	6	34	1	3		9	10	4		1			2	
V	12			1	15				15						7											
W	1				1																				1	
X			1						4							3				2						
Y	5	1	3	2	5	1	1					1	1	1	5	2	1	1	3	1	1					
Z	6		1	1											3						2					

2–b.—*Digraphic kappa plain, Spanish language*=.0091 (I. C.=6.15)

2–c.—*The 87 digraphs comprising 75% of Spanish plain text, based on the table of 5,000 digraphs (Item 2–a), arranged according to their relative frequencies.*

EN--- 126	TE--- 67	IN--- 50	NA--- 41	MA--- 32	IS--- 27	EA----- 20
ES--- 119	AN--- 64	EC--- 47	IE--- 40	SA--- 32	EM--- 26	OA---- 19
ON--- 104	[1] 1,287	RI--- 45	[2] 2,513	PO--- 31	SP--- 26	PU---- 19
ER--- 94	AD--- 64	EL--- 44	CA--- 39	MI--- 30	ED--- 26	SC---- 18
RE--- 94	AS--- 62	LA--- 44	ND--- 37	PA--- 30	OD--- 26	AT---- 18
NT--- 91	TA--- 60	RO--- 43	TI--- 35	AD--- 30	AP--- 24	CU---- 18
DE--- 84	DO--- 59	NO--- 43	LE--- 35	DI--- 30	IT--- 24	EE---- 17
AR--- 81	OR--- 58	IA--- 43	TR--- 34	ID--- 29	EP--- 23	OB---- 17
CI--- 80	SE--- 57	IC--- 42	UN--- 34	QU--- 29	SU--- 23	CE---- 17
RA--- 74	ST--- 57	ME--- 42	PR--- 34	OP--- 29	SO--- 22	ET---- 17
OS--- 73	TO--- 56	AL--- 41	OM--- 33	LI--- 28	OL--- 22	LO---- 17
CO--- 69	AC--- 54	SI--- 41	NC--- 33	NI--- 28	NS--- 21	3,753
IO--- 67	UE--- 52	NE--- 41	DA--- 32	OC--- 28	EG--- 21	

2–d.—*Frequent digraphs in Spanish plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

EN-- 126	NE-- 41	AR-- 81	RA-- 74	AS-- 62	SA-- 32	LA-- 44	AL--- 41
ES-- 119	SE-- 57	CI-- 80	IC-- 42	OR-- 58	RO-- 43	EL-- 44	LE-- 35
ON-- 104	NO-- 43	AN-- 64	NA-- 41	AC-- 54	CA-- 39	MA-- 32	AM---- 30
ER-- 94	RE-- 94	AD-- 64	DA-- 32				

2–e.—*Frequent digraphs in Spanish plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

NT--- 91	TN--- 0	ST-- 57	TS--- 0	ND-- 37	DN-- 1	NC-- 33	CN--- 0
IO--- 67	OI-- 4						

2–f.—*Doublets occurring in Spanish plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2–a).*

EE-- 17	AA- 12	RR- 10	SS- 10	LL- 9	CC- 5	OO- 4	NN- 3	DD-- 2

2–g.—*The 21 digraphs appearing 100 or more times as beginnings of words in 10,129 words in Spanish plain text, arranged according to their absolute frequencies.*

CO--- 684	PR-- 307	PA-- 263	SE-- 189	CA-- 151	PE-- 111	MA--- 101
RE--- 335	ES-- 286	PO-- 247	DI-- 175	SI-- 137	UN-- 109	CU--- 100
DE--- 323	QU-- 286	IN-- 235	PU-- 157	MI-- 117	HA-- 108	SO--- 100

[1] The 15 digraphs above this line comprise 25% of Spanish plain text.
[2] The 40 digraphs above this line comprise 50% of Spanish plain text.

3-a.—*The 105 trigraphs appearing 100 or more times in 60,115 letters of Spanish plain text, arranged according to their absolute frequencies.*

ENT.. 596	ARA. 229	POR. 176	OSE. 147	ERO. 131	NDE. 121	PER.. 111
ION.. 564	ONE. 227	TER. 174	ONS. 144	ONT. 131	RAN. 121	ASE.. 109
CIO.. 502	ESE. 217	ODE. 168	REC. 144	ANA. 130	STE. 119	CAN.. 109
NTE.. 429	ADE. 202	ERE. 166	ORE. 143	ARE. 130	REN. 118	UNI.. 108
CON.. 415	PAR. 193	ERA. 165	OCO. 142	UNT. 129	ARI. 117	OSI.. 107
EST.. 355	CIA. 190	TRA. 165	EDE. 141	ANO. 127	TEN. 116	GEN.. 105
RES.. 335	ENC. 190	AME. 163	ICI. 140	TAR. 127	OND. 115	NCO.. 105
ADO.. 307	NCI. 188	ERI. 162	END. 139	ANT. 126	RIA. 115	RIO.. 105
QUE.. 294	PRE. 184	MER. 159	SEN. 139	ESA. 126	ECI. 114	ERN.. 104
ACI.. 277	DEL. 183	ELA. 158	TAD. 138	IER. 126	IST. 113	OMI.. 104
NTO.. 270	NDO. 183	PRO. 155	ECO. 135	ADA. 125	ONA. 113	SCO.. 104
IEN.. 267	NES. 183	ACO. 153	STR. 134	DEN. 124	DAD. 112	TES.. 103
COM.. 246	DOS. 182	ENE. 151	TOS. 133	AND. 123	INT. 112	BIE.. 101
ICA.. 242	MEN. 181	UES. 150	IDA. 132	DES. 121	NTR. 112	NTI.. 100
STA.. 240	NTA. 176	ESP. 149	SDE. 132	IDO. 121	ESI. 111	TOR.. 100

3-b.—*The 19 trigraphs appearing 50 or more times as beginnings of words in 10,129 words in Spanish plain text, arranged according to their absolute frequencies.*

CON.. 298	PAR. 154	PUN.. 93	INT.. 72	UNI.. 55	CUA.. 52	REP... 51
COM.. 218	PRO. 139	PER.. 80	RES.. 72	DES.. 53	TRA.. 52	ARG... 50
EST.. 194	PRE. 114	GOB.. 77	NUE.. 66	INF.. 53		

4.—*The 86 tetragraphs appearing 50 or more times in 60,115 letters of Spanish plain text, arranged according to their absolute frequencies.*

CION... 444	CONS.. 104	ERNO... 79	AMER... 72	FORM... 62	EEST.... 55
ACIO... 252	CONT... 99	IERN... 78	IEND... 72	SENT... 62	SCON.... 55
ENTE... 233	PUNT... 95	OQUE... 78	IDAD... 71	ICIO... 61	SIDE.... 55
ESTA... 174	ANDO... 91	IONA... 77	ENDO... 70	ONTR... 60	CIEN.... 54
IONE... 159	TADO... 91	UEST... 77	ERIC... 70	SION... 60	NFOR.... 54
MENT... 150	ACON... 90	BIER... 76	NTOS... 70	CCIO... 59	OPOR.... 54
ONES... 146	ANTE... 89	ICAN... 76	MIEN... 69	GENT... 58	RESP.... 54
IENT... 141	NTER... 85	RESE... 76	IOND... 67	COMA... 57	ARIO.... 53
ENTO... 137	INTE... 84	GOBI... 75	MERI... 67	ESDE... 57	ESTR.... 53
ENCI... 128	NTES... 82	OBIE... 75	NTRA... 67	ORES... 57	ARGE.... 51
PARA... 117	ADOS... 81	ECON... 74	DELA... 65	RECI... 57	ECTO.... 51
ENTA... 115	AMEN... 81	RGEN... 73	ENTI... 64	AQUE... 56	PART.... 51
NCIA... 115	OCON... 81	RICA... 73	NTIN... 64	IONP... 56	POSI.... 51
PRES... 111	ESEN... 80	STAD... 73	COMI... 63	QUES... 56	EPRE.... 50
UNTO... 111	ONDE... 80				

5.—*Average length of words in Spanish plain text*=5.9 letters.

E. Portuguese Letter Frequency Data

1–a.—*Absolute frequencies of single letters of Portuguese plain text, arranged alphabetically, based on 45,106 letters of text.*

A	5,362	G	724	L	1,245	Q	348	V	737
B	470	H	304	M	1,699	R	3,292	W	24
C	2,285	I	3,314	N	2,912	S	3,409	X	166
D	1,900	J	160	O	5,001	T	2,679	Y	22
E	5,441	K	17	P	1,377	U	1,491	Z	207
F	520								45,106

1–b. *Monographic kappa plain, Portuguese language*=.0746 (I. C.=1.94)

1–c.—*Frequency distribution of single letters based on 45,106 letters of Portuguese plain text, reduced to 1,000 letters, arranged according to their relative frequencies.*

E	121	N	65	U	33	F	11	X	4
A	119	T	59	P	30	B	10	J	3
O	111	C	51	L	28	Q	8	W	1
S	76	D	42	V	16	H	7	Y	–
I	73	M	38	G	16	Z	5	K	–
R	73								1,000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonants in 45,106 letters of Portuguese plain text. Percentage of 8 most frequent letters in Portuguese plain text.*

Vowels A, E, I, O, U, and Y=45.8%
High-Frequency Consonants N, R, and S=21.3%
Medium-Frequency Consonants C, D, L, M, P, and T=24.8%
Low-Frequency Consonants B, F, G, H, J, K, Q, V, W, X, Y, and Z=8.1%

8 most frequent letters (in descending order of frequency) E, A, O, S, I, R, N, and T=69.7%

1–e.—*Absolute frequencies of single letters as initial letters of 7,058 words in Portuguese plain text, arranged according to their frequencies.*

P	847	M	405	I	264	B	113	Z	14
C	731	T	348	F	222	G	111	W	11
E	608	R	316	Q	222	J	92	K	7
S	601	N	299	O	187	U	77	Y	4
A	597	V	271	L	143	H	60	X	2
D	506								7,058

2–a.—*Frequency distribution of digraphs based on 45,106 letters of Portuguese plain text, reduced to 5,000 digraphs.*

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	11	11	52	60	15	9	14	2	18	2		38	36	56	49	23	8	68	72	22	8	16	1			5
B	11			1	10				5			2	1		9			9	2	1	2					
C	60		2		30			4	39			5		1	85			7		8	12					
D	45				61				33				1		61			2	1	1	5					
E	15	5	48	22	11	11	23	1	27	6	1	31	44	97	6	18	6	76	95	20	7	12	1	15		5
F	9				14				13			1			15			2			3					
G	15				14				4			1		1	14			14			15					
H	10				8				3						11						1					
I	42	3	34	31	6	7	9		1			16	22	53	26	5	2	25	39	27	2	10		2		7
J	7				2										2						7					
K																										
L	24	1	4	4	24	1	5	9	21			2	4	2	14	4	2	1	4	7	6	2				
M	41	10	3	4	51	1			26	1		1	2	1	16	15	1	3	5	2	6	2				
N	31		29	35	14	7	8	12	18						25	1			19	114	4	4				1
O	21	9	32	25	27	10	7	3	20	4		20	36	79	5	35	8	71	85	18	12	22	1	1	1	1
P	26		2		25				2			4		1	60	1	1	28	1	1	3					
Q					1																37					
R	75	2	14	9	86	3	7	1	46	1		2	18	8	34	7	3	11	8	18	4	6				1
S	41	6	22	10	62	6	3	2	23	2		3	12	5	23	35	7	4	40	47	18	5				
T	65		1	1	69	1			26					1	88			33		1	13					
U	22	5	5	7	26	1	4		18	1		14	11	17	2	4		9	6	11		1				2
V	11				37				23						9			1								
W	1																									
X	10		3		1				2							3				1						
Y																										
Z	7		1		9				2				1		1		1	1								

2b.—*Digraphic kappa plain, Portuguese language*=.0084 (I. C.=5.68).

2-c.—*The 91 digraphs comprising 75% of Portuguese plain text, based on the table of 5,000 digraphs (Item 2-a), arranged according to their relative frequencies.*

NT--- 114	TA--- 65	ST--- 47	AM--- 36	CE--- 30	OD--- 25	AT---- 22
EN--- 97	[1] 1,224	RI--- 46	[2] 2,505	NC--- 29	NO--- 25	UA---- 22
ES--- 95	SE--- 62	DA--- 45	ND--- 35	PR--- 28	LA--- 24	OA---- 21
TO--- 88	DO--- 61	EM--- 44	OP--- 35	IT--- 27	LE--- 24	LI---- 21
RE--- 86	DE--- 61	IA--- 42	SP--- 35	OE--- 27	AP--- 23	OL---- 20
CO--- 85	AD--- 60	MA--- 41	RO--- 34	EI--- 27	EG--- 23	ET---- 20
OS--- 85	PO--- 60	SA--- 41	IC--- 34	UE--- 26	VI--- 23	OI---- 20
ON--- 79	CA--- 60	SS--- 40	TR--- 33	MI--- 26	SO--- 23	NS---- 19
ER--- 76	AN--- 56	CI--- 39	DI--- 33	IO--- 26	SI--- 23	SU---- 18
RA--- 75	IN--- 53	IS--- 39	OC--- 32	PA--- 26	OV--- 22	RT---- 18
AS--- 72	AC--- 52	AL--- 38	EL--- 31	TI--- 26	SC--- 22	EP---- 18
OR--- 71	ME--- 51	VE--- 37	ID--- 31	PE--- 25	IM--- 22	UI---- 18
TE--- 69	AO--- 49	QU--- 37	NA--- 31	IR--- 25	ED--- 22	3,755
AR--- 68	EC--- 48	OM--- 36				

2-d.—*Frequent digraphs in Portuguese plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

ES------ 95	SE----- 62	OR----- 71	RO----- 34	ME----- 51	EM------ 44
RE------ 86	ER----- 76	CA----- 60	AC----- 52	EC----- 48	CE------ 30
CO------ 85	OC----- 32	AD----- 60	DA----- 45	MA----- 41	AM------ 36
RA------ 75	AR----- 68	PO----- 60	OP----- 35	CI----- 39	IC------ 34
AS------ 72	SA----- 41	AN----- 56	NA----- 31	DI----- 33	ID------ 31

2-e.—*Frequent digraphs in Portuguese plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2-a).*

NT----- 114	TN----- 1	ST----- 47	TS----- 0	ND----- 35	DN------ 0

2-f.—*Doublets occurring in Portuguese plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2-a).*

SS---- 40	EE--- 11	OO--- 5	LL--- 2	II--- 1	PP--- 1	TT---- 1
AA---- 11	RR--- 11	CC--- 2	MM--- 2			

2-g.—*The 20 digraphs appearing 100 or more times as beginnings of words in 6,803 words in Portuguese plain text, arranged according to their absolute frequencies.*

CO--- 464	RE-- 276	IN-- 188	PA-- 143	MA-- 130	ME-- 111	TR--- 103
PO--- 386	DE-- 259	ES-- 173	NA-- 133	PE-- 122	MI-- 105	DI--- 102
SE--- 333	QU-- 220	PR-- 169	TE-- 132	VE-- 115	NO-- 104	

[1] The 15 digraphs above this line comprise 25% of Portuguese plain text.
[2] The 42 digraphs above this line comprise 50% of Portuguese plain text.

CONFIDENTIAL

3-a.—*The 59 trigraphs appearing 100 or more times in 45,106 letters of Portuguese plain text, arranged according to their absolute frequencies.*

ENT	474	TOS	191	ERE	150	IDA	133	OSE	126	ECE	115	ASE	105
NTO	457	EST	186	CIA	145	TER	132	ARE	125	NCI	114	ITO	104
ONT	303	ACA	182	ADE	143	OPO	130	ESE	124	REC	113	ELE	103
NTE	284	RES	181	STA	143	SPO	130	OVE	124	PAR	112	ERI	103
CON	255	QUE	172	ICA	142	ADA	129	SSA	124	ESS	110	PRO	102
PON	236	NTA	167	OCO	140	TRA	129	DES	123	DAD	109	AME	101
CAO	227	POR	159	ARA	136	NDO	127	ECO	121	ORE	108	OSS	101
ADO	211	ACO	158	DOS	134	ENC	126	ODE	118	EDI	107	IME	100
MEN	205	COM	154	OES	134								

3-b.—*The 19 trigraphs appearing 50 or more times as beginnings of words in 6,803 words in Portuguese plain text, arranged according to their absolute frequencies.*

CON	224	QUE	109	PRO	93	QUA	83	TRA	66	VEX	53	RES	52
PON	213	EST	105	POR	88	DES	71	MIL	61	IND	52	REC	51
COM	136	PAR	93	NAO	86	SER	70	REF	56				

4.—*The 38 tetragraphs appearing 50 or more times in 45,106 letters of Portuguese plain text, arranged according to their absolute frequencies.*

ONTO	233	ENTA	97	AMEN	81	CONT	68	CONS	58	RENT	52
PONT	221	NCIA	95	PARA	81	FORM	67	NTES	58	TELE	52
MENT	183	PORT	87	COES	73	OCON	66	ANDO	57	EGRA	51
ENTO	173	DADE	86	IDAD	71	ELEG	61	ANTE	57	NFOR	51
ENTE	147	ESTA	85	CENT	70	ADOS	60	ORMA	54	OPON	51
ACAO	142	ENCI	83	INTE	70	IMEN	60	VEXA	54	LEGR	50
NTOS	141	SPON	83								

5.—*Average length of words in Portuguese plain text*=6.48 letters.

CONFIDENTIAL

F. Russian Letter Frequency Data

1–a.—*Absolute frequencies of single letters of Russian plain text, arranged alphabetically, based on 67,850 letters of text.*

А	5,122	З	1,280	Н	4,463	У	1,578	Щ	257
Б	1,095	И	4,923	О	8,078	Ф	127	Ы	1,421
В	3,543	Й	961	П	1,815	Х	941	Ь	960
Г	1,141	К	2,324	Р	3,427	Ц	369	Э	173
Д	2,076	Л	2,747	С	3,917	Ч	902	Ю	455
Е	5,537	М	1,936	Т	4,041	Ш	554	Я	1,185
Ж	502								67,850

1–b.—*Monographic kappa plain, Russian language*=.0568 (I. C.=1.76)

1–c.—*Frequency distribution of single letters based on 67,850 letters of Russian plain text, reduced to 1,000 letters, arranged according to their relative frequencies.*

О	119	В	52	П	27	Б	16	Ж	7
Е	82	Р	50	У	23	Й	14	Ю	7
А	75	Л	40	Ы	21	Ь	14	Ц	5
И	73	К	34	З	19	Х	14	Щ	4
Н	66	Д	31	Я	17	Ч	13	Э	3
Т	60	М	29	Г	17	Ш	8	Ф	2
С	58								1,000

1–d.—*Percentage of vowels, high-frequency consonants, medium-frequency consonants, and low-frequency consonents in 67,850 letters of Russian plain text. Percentage of 10 most frequent letters in Russian plain text.*

Vowels А, Е, И, Й, О, У, Ы, Э, Ю, and Я=43.4%.

High-Frequency Consonants, В, Н, Р, С, and Т=28.6%

Medium-Frequency Consonants Б, Г, Д, З, К, Л, М, П, Х, Ч, and Ь=25.4%

Low-Frequency Consonants Ж, Ф, Ц, Ш, and Щ=2.6%

10 most frequent letters (in descending order of frequency) О, Е, А, И, Н, Т, С, В, Р, and Л=67.5%

1–e.—*Absolute frequencies of single letters as initial letters of 10,601 words in Russian plain text, arranged according to their frequencies.* (*One-letter words have been omitted.*)

П	1,210	Д	496	И	321	Х	120	Ф	58
С	983	М	446	Г	292	А	116	Ц	47
Н	800	Р	429	У	229	Е	92	Я	41
В	731	Т	418	Ч	182	Ж	72	Ю	34
О	650	Э	404	Э	147	Ш	63	Щ	2
К	555	Б	344	Л	146				10,601

2–a.—*Frequency distribution of digraphs based on 67,850 letters of Russian plain text, reduced to 5,000 digraphs.*

SECOND LETTER

FIRST LETTER	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ы	Ь	Э	Ю	Я
А	2	12	35	8	14	7	6	15	7	7	19	27	19	45	5	11	26	31	27	3	1	10	6	7	10	1			2	6	9
Б	5					9	1		6			6		2	21		8	1		6						1	11				2
В	35	1	5	3	3	32		2	17		7	10	3	9	58	6	6	19	6	7		1	1	2	4	1	18	1	2		3
Г	7				3	3			5		1	5		1	50		7			2											
Д	25		3	1	1	29	1	1	13		1	5	1	13	22	3	6	8	1	10			1	1	1		5	1			1
Е	2	9	18	11	27	7	5	10	6	15	13	35	24	63	7	16	39	37	33	3	1	8	3	7	3	3			1	1	2
Ж	5	1			6	12			5					6				1													
З	35	1	7	1	5	3			4		2	1	2	9	9	1	3	1		2							4				4
И	4	6	22	5	10	21	2	23	19	11	19	21	20	32	8	13	11	29	29	3	1	17	3	11	1	1			1	3	17
Й	1	1	4	1	3		1	2	4		5	1	2	7	9	7	3	10	2				1	3	2						
К	24	1	4	1		4	1	1	26		1	4	1	2	66	2	10	3	7	10			1								
Л	25	1	1	1	1	33	2	1	36		1	2	1	8	30	2		3	1	6		4		1			2	30		4	9
М	18	2	4	1	1	21	1	2	23		3	1	3	7	19	5	2	5	3	9	1			2			5	1	1		3
Н	54	1	2	3	3	34			58		3		1	24	67	2	1	9	9	7	1		5	2			36	3			5
О	1	28	84	32	47	15	7	18	12	29	19	41	38	30	9	18	43	50	39	3	2	5	2	12	4	3			2	3	2
П	7					15			4			9		1	46		41	1		6							2				2
Р	55	1	4	4	3	37	3	1	24		3	1	3	7	56	2	1	5	9	16		1	1	1	2		8	3			5
С	8	1	7	1	2	25			6		40	13	3	9	27	11	4	11	82	6		1	1	2	2		1	8			17
Т	35	1	27	1	3	31		1	28		5	1	1	11	56	4	26	18	2	10				1			11	21			4
У	1	4	4	4	11	2	6	3	2		8	5	5	5	1	5	7	14	7			1		8	3	2				9	1
Ф	2					2			2						1		1	1													
Х	4	1	4	1	3	1		2	3		4	3	3	4	18	5	3	4	2	2	1			1							
Ц	3					7			10		2				1					1							1				
Ч	12					23			13		2			6					7	1					1			1			
Ш	5					11			14		1	2		2	2					1								1			
Щ	3					8			6					1						1											
Ы		1	9	1	3	12		2	4	7	3	6	6	3	2	10	3	9	4	1		16		1	2						
Ь		2	4	1	1	2		2	2		6		3	13	2	4	1	11	3					1	4				1	3	1
Э											1			1				1	9												
Ю		2	1	2	1			3	1		1		1	1	1	3	1	1	7				1	1		4					
Я	1	3	9	1	3	3	1	5	3	2	3	3	4	6	3	6	3	6	10			2	1	4	1	1			1	1	1

2–b.—*Digraphic kappa plain, Russian language*=.0052 (I. C.=5.00)

2–c.—*The 159 digraphs comprising 75% of Russian plain text, based on the table of 5,000 digraphs (Item 2–a), arranged according to their relative frequencies.*

ОВ 84	ЕР 39	ЛО 30	ЕМ 24	АМ 19	АД 14	ИР 11	ВЛ 10
СТ 82	ОМ 38	ЛЬ 30	РИ 24	ОК 19	ШИ 14	СС 11	АШ 10
НО 67	РЕ 37	ДЕ 29	НН 24	ТС 18	УС 14	ШЕ 11	ЛЯ 9
КО 66	[1] 1,258	ИТ 29	ИЗ 23	ВЫ 18	ЬН 13	АП 11	РТ 9
ЕН 63	ЕС 37	ОЙ 29	ЧЕ 23	ОЗ 18	СЛ 13	ИЧ 11	ЯВ 9
НИ 58	ЛИ 36	ИС 29	[2] 2,492	МА 18	ДН 13	ТН 11	ВН 9
ВО 58	НЫ 36	ТИ 28	МИ 23	ХО 18	ДИ 13	ИЙ 11	НС 9
РО 56	ВА 35	ОБ 28	ДО 22	ОП 18	ЕК 13	УД 11	БЕ 9
ТО 56	ЗА 35	АТ 27	ИВ 22	ЕВ 18	ИП 13	ЕГ 11	ЗН 9
РА 55	ЕЛ 35	ТВ 27	ИЛ 21	СЯ 17	ЧИ 13	ИД 10	ЕБ 9
НА 54	АВ 35	ЕД 27	ТЬ 21	ИХ 17	ОИ 12	ЕЗ 10	МУ 9
ГО 50	ТА 35	АЛ 27	МЕ 21	ИЯ 17	ЖЕ 12	ЙС 10	ЫВ 9
ОС 50	НЕ 34	СО 27	ИЕ 21	ВИ 17	АБ 12	ЫП 10	НТ 9
АН 48	ЕТ 33	КИ 26	БО 21	РУ 16	ЧА 12	АХ 10	ЭТ 9
ОД 47	ЛЕ 33	АР 26	ИМ 20	ЕП 16	ЫЕ 12	ЦИ 10	АЯ 9
ПО 46	ОГ 32	ТР 26	ВС 19	ЫХ 16	ОЧ 12	ЯТ 10	СН 9
ОР 43	ВЕ 32	ДА 25	ИИ 19	ПЕ 15	ТЫ 11	КУ 10	УЮ 9
ПР 41	ИН 32	СЕ 25	МО 19	АЗ 15	СП 11	ДУ 10	ЗО 9
ОЛ 41	ТЕ 31	ЛА 25	АК 19	ЕЙ 15	ЬС 11	КР 10	ОО 9
СК 40	АС 31	КА 24	ИК 19	ОЕ 15	БЫ 11	ТУ 10	3,750
ОТ 39	ОН 30						

2–d.—*Frequent digraphs in Russian plain text whose reversals are also frequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

ОВ 84	ВО 58	РО 56	ОР 43	ГО 50	ОГ 32	ВА 35	АВ 35
НО 67	ОН 30	ТО 56	ОТ 39	ОЛ 41	ЛО 30	ЕЛ 35	ЛЕ 33
ЕН 63	НЕ 34	НА 54	АН 45	ЕР 39	РЕ 37	ЕТ 33	ТЕ 31
НИ 58	ИН 32						

2–e.—*Frequent digraphs in Russian plain text whose reversals are rare or infrequent, accompanied by their frequencies from the table of 5,000 digraphs (Item 2–a).*

ПР 41	РП 2	СК 40	КС 3

2–f.—*Doublets occurring in Russian plain text, arranged according to their frequencies from the table of 5,000 digraphs (Item 2–a).*

НН 24	СС 11	ЕЕ 7	ММ 3	ЛЛ 2	ДД 1	РР 1	ЯЯ 1
ИИ 19	ОО 9	ВВ 5	АА 2	ТТ 2	КК 1		

[1] The 24 digraphs above this line compose 25% of Russian plain text.
[2] The 66 digraphs above this line compose 50% of Russian plain text.

2–g. *The 24 digraphs appearing 100 or more times as beginnings of words in 10,601 words in Russian plain text, arranged according to their absolute frequencies.*

ПР... 470	РА... 250	ГО... 169	ОБ... 146	ДО... 120	КА... 110	МЕ.. 107
ПО.. 405	НА... 246	СЕ... 167	ДЕ... 137	ОТ... 115	ПЕ... 110	ВС... 101
ЗА... 292	СО... 220	СТ... 161	НЕ... 122	ЭТ... 111	ТО... 108	МА.. 101
КО... 287	ВО... 179	ВЫ.. 159				

3–a.—*The 69 trigraphs appearing 100 or more times in 67,850 letters of Russian plain text, arranged according to their absolute frequencies.*

ОГО. 318	ТЕЛ. 188	ТОР.. 152	ПРИ. 137	РОД. 128	РОВ.. 116	ЧЕБ. 104
ЕНИ. 295	НОВ. 181	ЛЬН. 151	РЕД. 137	КОГ.. 123	СТИ. 115	ИНА. 103
СКО. 270	ЕЛЬ. 176	ПОЛ. 149	ЕТС. 135	АВО. 119	ИЛИ. 113	ТВО. 103
СТВ. 267	ОВА. 169	ЛЕН. 146	ННЫ. 135	ПЕР. 119	АСТ. 112	АБО. 101
ОСТ. 260	ОРО.. 167	НИХ. 145	ОВЕ. 134	ТВЕ. 119	АНА. 111	ИСТ. 101
ПРО. 233	СТР.. 165	НИЕ. 143	КОВ. 130	ЗАВ.. 118	НЫЕ. 110	ТРА. 101
СТА. 217	ЕСТ. 159	НИЯ. 143	ННО. 130	ВАН. 117	ОЛЬ. 110	ВЕТ. 100
ОВО. 204	АНИ. 158	КОМ. 139	СОВ. 130	КОД. 117	ПОС. 110	ОВС. 100
ВОД. 203	СКИ. 158	ИТЕ. 138	ПРЕ. 129	НОИ. 117	СТО. 110	РАЗ.. 100
ЕНН. 198	ТОВ. 158	НОС. 138	НОГ. 128	ЕРЕ.. 116	ЕГО.. 104	

3–b.—*The 20 trigraphs appearing 50 or more times as beginnings of words in 10,601 words in Russian plain text, arranged according to their absolute frequencies.*

ПРО. 205	ПРИ.. 95	ПОС.. 81	ВЫП.. 73	ПОД.. 61	СТА... 59	ГОД... 51
ПРЕ. 116	СОВ... 87	ПЕР... 78	РАБ... 72	БОЛ.. 60	РАЗ... 53	ГОР... 50
ЗАВ. 108	КОЛ... 84	ПОЛ.. 74	НАР.. 71	РАй... 60	КОН.. 52	

4.—*The 58 tetragraphs appearing 50 or more times in 67,850 letters of Russian plain text, arranged according to their absolute frequencies.*

НОГО.. 114	СОВЕ... 87	ЕЛЬН... 78	ВЛЕН... 68	ПРОИ... 60	ОИЗВ... 54
ТЕЛЬ.. 111	АВОД... 86	СТВО... 78	СКОй.... 66	РОИЗ... 60	КОТО... 53
ИТЕЛ.. 107	ЗАВО... 85	ИЧЕС... 76	СТАВ... 66	ОТОВ... 59	ННЫХ.. 53
КОГО.. 99	СТВЕ... 84	ОВЕТ... 74	АРОД... 65	ВЕТС... 56	ВОДС... 52
НОСТ.. 98	ЛЬНО... 83	ЧЕСК... 74	ЕЛЬС... 65	ТОРО... 56	ЕТСК... 52
ЕНИЯ.. 97	ПОЛН... 82	АНОВ... 72	ЕСТВ... 64	АТЕЛ... 55	ОТОР... 51
ЕННО.. 95	СКОГ... 82	ОСТА... 70	СТАН... 64	ГОТО... 55	ВСКО... 50
ЕНИЕ.. 88	ЕННЫ.. 81	СТРО... 70	НАРО... 63	ЕТСЯ... 55	ОЛХО... 50
ПРЕД.. 87	АБОТ... 80	ВЕНН... 69	ОВАН... 62	СТВА... 55	СКИй... 50
РАБО.. 87	ЛЕНИ... 80	ТВЕН... 69	СТРА... 61		

5.—*Average length of words in Russian plain text*=6.4 letters.

APPENDIX 6

CLASSIFICATION GUIDE TO CONCEALMENT SYSTEMS

CLASSIFICATION GUIDE TO CONCEALMENT SYSTEMS

I—LITERAL COVER TEXTS

1. *Systems of measured placement.*[1]

   a. Constant keying.

   (1) {1st, 2nd, 3rd, final, etc.} {letter, digraph, syllable, etc.} of {every, every other, etc.} word.

   (2) {1st, 2nd, 3rd, final, etc.} {letter, digraph, syllable, etc.} of {1st, 2nd, 3rd, final, etc.} word in {every, every other, etc.} {line, sentence, paragraph, etc.}

   (3) {1st, 2nd, 3rd, final, etc.} {letter, digraph, syllable, etc.} of {1st, 2nd, 3rd, etc.} word before or after {every, every other, etc.} {punct. mark, capital letter, etc.}

   (4) every {2nd, 3rd, 4th, etc.} word

   (5) {1st, 2nd, 3rd, final, etc.} word of {every, every other, etc.} {line, sentence, paragraph, etc.}

   (6) {1st, 2nd, 3rd, final, etc.} word of {every, every other, etc.} {line, sentence} containing an {odd, even} number of {letters, words}

   (7) {1st, 2nd, 3rd, final, etc.} word before or after {every, every other, etc.} {punctuation mark, capital letter, etc.}

[1] In all the systems in this paragraph, there must also be considered the treatment of the system counting ***backwards*** from the end of the cover text, with the secret text itself being backwards.

(8) word letter digraph syllable etc.	after	every every other etc.		"stopping point" when text is decimated (using either words or letters as single units) at an interval of 2, 3, 4, etc.

b. Cyclical keying. A cyclical numerical key may be employed with any of the systems listed above.

c. Aperiodic keying.

(1) A system wherein a long sequence of digits from a prearranged source is employed to denote the intervals between successive significant elements.

(2) An autokeying system wherein the first letter of hidden text is the first letter of the correspondence, this letter determining the interval to the second letter of hidden text, which in turn determines the interval to the third letter, etc.

2. *Systems employing grilles and diagrammatical overlays.*

a. Single and multiple position grilles. Either type of grille may be employed to select individual letters, digits, words, or phrases.

b. Diagrammatical overlays may consist of a diagonal line from one corner of a page to another, a "V" or some other simple pattern, a geometrical figure, or folds in paper.

3. *Systems employing physical indicators.*[3]

a. Shaded letters.

b. Breaks in the text.

c. Squeezing and expansion of the letters of words.

d. Variations in style of handwriting of words or letters.

e. Variations in spacing between words.

f. Letters or words in relation to punctuation marks, diacritical marks, or capital letters.

g. Misspellings.

h. Misplaced capital letters.

i. Elongated tails on words pointing to significant letters.

j. Blotted words, syllables, or letters.

k. Pinholes used to designate letters or words.

l. Variations in crossing of "t's" or dotting of "i's" used to point out significant letters.

m. Impressions.

n. Underlinings.

4. *Additional systems, applicable only to concealment of encrypted text.*

a. Structural features of the literal cover text may be employed to convey the elements of a cryptogram. Such systems are especially applicable when a message has been encrypted in a Morse code, Baconian, or Trithemian system.

(1) Dots on "i's" and crosses on "t's" may be used as Morse symbols.

(2) The number of letters, vowels, consonants, or syllables in each word, or the oddness or evenness of this number (or its length, mod 3) might be used to convey a single element in a Baconian or Trithemian system.

(3) Vowels or other letters might be categorized so as to be used as component elements in a Baconian or Trithemian system.

---

[3] In all the systems noted in this paragraph, the indicators might point out the actual letters or words involved, or they might be used as focus points to designate letters or words immediately preceding or following, or for that matter the indicators might refer to textual elements in the lines immediately above or below, etc.

b. If a cryptosystem is designed so that the cryptograms it produces look like something other than cryptograms, then such cryptograms are innately disguised. Also, any cryptograms composed exclusively of digits may be given a cover or form to make them appear to be something other than a cryptogram. Among these are the following:

(1) Notations of the moves or plays in a game.
(2) Shopping lists.
(3) Page of price quotations.
(4) List of nationwide temperatures.
(5) List of phonograph record numbers.
(6) List of pattern numbers.
(7) Page of mathematical computations.
(8) Serial numbers or commercial numbers.
(9) Report of collections or donations.
(10) Report of insurance claims or policies.
(11) Financial statements.
(12) Knitting instructions.

II—PICTORIAL COVER TEXTS

1. *Pictures*. Photographs, sketches and other drawings may contain secret text, based on the following:
   - *a*. Shadings.
   - *b*. Dots and dashes.
   - *c*. Line placements.
   - *d*. Hidden words, letters, or digits.
   - *e*. Shorthand systems.
   - *f*. Position or number of particular objects.
2. *Maps*. The following features placed on maps may have code meanings:
   - *a*. Directions.
   - *b*. Specific locations or specific buildings.
   - *c*. Plots of land.
   - *d*. Contours.
   - *e*. Place names.
3. *Music manuscripts*. The following are among the elements which may have cryptographic significance:
   - *a*. Pitch of notes.
   - *b*. Rhythm and duration of notes.
   - *c*. Keys.
   - *d*. Modulation.
   - *e*. Instrumentation.
   - *f*. Title.
   - *g*. Lyrics.
   - *h*. Marks of expression.

III—ARTICLES AS COVER TEXT VEHICLES

1. Intervals between knots in a string
2. Selection and arrangement of candy in a box.
3. Arrangement of cards in a deck.
4. Selection and arrangement of stamps on approval sheets.
5. Arrangement of stickpin holes in a hat or tie.
6. Insect collections.
7. Wearing or sending particular items.
8. Identity and arrangement of wash on a line.
9. Positioning of drawn window shades.

IV—KEY CONCEALMENT

Many concealment systems require the use of a specific key to be included in the correspondence. The following are among the many possibilities for concealed specific keys:

1. The number of letters in the place of origin, signature, addressee, salutation, complimentary close, or some other part of the correspondence.
2. The normal alphabetical position number of a particular letter or letters contained in the place of origin, signature, etc., of a letter.
3. The presence of a particular letter or word in some part of the correspondence.
4. The number of pages, paragraphs, sentences, lines, words, capital letters, etc., in the first (2nd, 3rd, . . .) line (page, etc.).
5. The presence of digits in some part of the correspondence, such as serial numbers, date or time in heading, mention of a birthdate or other anniversary in a letter, etc.
6. The number or identity of underlined or italicized letters or words.

Appendix 7

COMMUNICATION INTELLIGENCE OPERATIONS

COMMUNICATION INTELLIGENCE OPERATIONS

1. Communication intelligence processes. The principal processes of communication intelligence operations are as follows:

a. Interception of communication signals or messages and forwarding raw traffic [1] to communication intelligence centers for study.

b. Radio direction finding and radio position finding operations; identification of transmitters and radio operators by means of radio fingerprinting and Morse operator analysis, respectively.

c. Traffic analysis, or the study of the external characteristics of communications, without recourse to cryptanalysis of the message texts.

d. Cryptanalysis or solution of the texts of messages.

e. Translation and emendation of the message texts.

f. Large-scale production or exploitation of communication intelligence, after the initial break-in.

g. Evaluation of information, yielding military intelligence.

h. Collation, correlation and comparison of communication intelligence with other intelligence sources.

i. Distribution of communication intelligence to consumers.

2. Interception, radio direction finding, and radio position finding.—*a.* Messages transmitted by radio can be manually copied or automatically recorded by suitably adjusted radio apparatus located within range of the transmitter. Some messages transmitted over wire lines can likewise be manually copied or automatically recorded by special apparatus suited for the purpose. Correspondents have no way of knowing whether or not radio transmissions are being copied by the enemy, since the interception does not interfere in the slightest degree with signals being transmitted. Interception of wire traffic is much more difficult than of radio, mainly because the equipment either must be located very near the wire line, or connected directly to it.

b. It is possible to determine, with a fair degree of accuracy, the *direction* of a radio transmitter from a given location and, by establishing the direction from two or more locations, it is possible to determine the geographical *location* of the transmitter. The science which deals with the means and methods of determining the direction in which a radio transmitter lies is called *radio direction finding;* the method of determining the geographical location of a radio transmitter, by the use of two or more direction-finding installations, is called *radio position finding.*

[1] *Raw traffic* is unprocessed intercepted traffic.

3. Radio fingerprinting and Morse operator analysis.—*a.* Radio fingerprinting is one of the valuable adjuncts of signal analysis, a communications-engineering sister of traffic analysis. Radio fingerprinting consists of the analysis of the characteristics of the emissions of an individual radio transmitter by means of oscillograms of the emitted radio waves. The oscillograms of the emissions of unidentified radio stations are compared with those of known transmitters or radio stations, and thus it is possible to equate different call signs or different frequencies which have been used by the same transmitting station. Radio fingerprinting is normally not considered conclusive in itself, but is correlated with other analyses or confirmations.

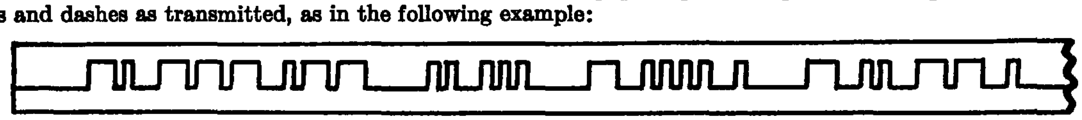
b. Another valuable adjunct of communication intelligence operations is *Morse operator analysis*. This analysis deals with the radio operators' characteristics when hand-sending is used; the analysis is based on the relative lengths and spacing of the dots and dashes composing the various Morse characters. It is a rarity when a radio operator will transmit a Morse character perfectly, i. e., make the dashes the correct length in respect to the dots, without any individuality (known as the "fist" or "swing") in the sending. Most operators do have certain individual characteristics or tendencies in the sending of certain Morse characters. In past decades, radio operators have identified characteristic "fists" of other operators based on the aural recognition of the rhythm of certain Morse characters. This art has been made more scientific through the use of actual physical measurement and through the assignment of a classificatory coding to the individualities present in the undulator-tape recording[2] of a Morse transmission. By matching measurements, individual radio operators may be identified, in spite of changes of call signs and other elements of the transmission.

4. Traffic Analysis.—*a.* A great deal of information of military value can be obtained by studying signal communications without solving encrypted messages constituting the traffic. The procedure and the methods used have yielded results of sufficient importance to warrant the application of a special term to this field of study; namely, *traffic analysis*, which is the study of signal communications and intercepted or monitored traffic for the purpose of gathering military information without recourse to cryptanalysis.

b. In general terms, traffic analysis is the careful inspection and study of signal communications for the purpose of penetrating camouflage superimposed upon the communication network for purposes of security. Specifically, traffic analysis reconstructs radio communication networks by: (1) noting volume, direction, and routing of messages; (2) correlating transmission frequencies and schedules used among and within the various networks; (3) determining directions in which transmitters lie, by means of radio direction finding; (4) locating transmitters geographically, by radio position finding; (5) developing the system of assigning and changing radio call signs; and (6) studying all items that constitute messages originated by operators and exchanged among themselves on a radio net.[3]

c. From a correlation of general and specific information derived by means of the foregoing procedures, traffic analysis is able not only to ascertain the geographic location and disposition of troops and military units (technically called "Order of Battle") and important troop movements, but also to predict with a fair degree of reliability the areas and extent of immediately pending or future activities. Traffic analysis procedures are followed to obtain informa-

[2] Such recordings take the form of a wavy inked line on a paper tape, serving as a visual representation of the dots and dashes as transmitted, as in the following example:



[3] Such operators' communications are termed "chatter" or simply "chat."

tion of value concerning the enemy, and to determine what information concerning our own forces is made available to the enemy through our own signal communications. Specifically, enemy military plans and operations may be revealed as follows:

(1) Unit movements and preparations for military activity may be indicated by rising and falling traffic volumes and changes in the structure of the network.

(2) The military function of a network may be revealed by the characteristic traffic pattern which results from transmissions incidental to planning, supply, or transportation.

(3) Change of grouping, disposition of forces and fleets, and probable tactical developments may be manifested in the redeployment of the radio stations which serve military elements.

d. These very important results are obtained without actually reading the texts of the intercepted messages; the solution and translation of messages are the functions of cryptanalysis and not traffic analysis. However, the cryptanalyst is frequently able to make good use of bits of information disclosed by traffic analysis such as faults noted in message routing and errors in cryptography causing messages to be duplicated or canceled. Cryptanalysis can provide important information for traffic analysis, since the solution of messages often yields data on impending changes in signal communication plans, operating frequencies and schedules, etc. Cryptanalysis also yields data on specific channels, networks, or circuits which are most productive of intelligence, so that effective control and direction of intercept agencies for maximum results can be achieved.

5. Cryptanalysis. The most important steps of practical, operational cryptanalysis are listed below. These steps are more or less in the order in which they are followed, but in particular cases some of these steps may be interchanged, or omitted entirely.

a. The study of patent characteristics of message texts.

b. The study of any available collateral information, including that obtained from previous solutions.

c. The search for and study of indicators.

d. The determination of the type of cryptosystem used.

e. The separation of the traffic into groups of messages in the same or related keys.

f. The search for repetitions within and between messages.

g. The study of the beginnings and endings of messages.

h. The preparation of statistical counts of letters, groups, etc.

i. The reduction of the encrypted texts to simplest terms.

j. The test for probable words, stereotypes, isologs, etc.

k. The recovery of the plain texts.

6. Other Intelligence Sources. In addition to (1) traffic analysis and (2) cryptanalysis as means of obtaining information relating to communications, further data may be obtained (3) by the use of secret agents for espionage, (4) by the capture and interrogation of prisoners, (5) by the capture of headquarters or command posts with records more or less intact, and (6) by defection or carelessness on the part of personnel who handle communications. Of these six main sources, traffic analysis and cryptanalysis are the most valuable, due in great part to their reliability; they may be likened to "reading the innermost thoughts of the enemy." The amount of vital information they furnish cannot be accurately estimated as it fluctuates with time, place, circumstances, equipment, and personnel. For most effective operation, the results of both cryptanalysis and traffic analysis can be fitted together to yield a unified picture of the communications scheme. Therefore, if all transmitting stations can be located quickly and if all communications can be intercepted and solved, extremely valuable information concerning strength, disposition of forces, and proposed moves will be continually available.

7. Time needed for cryptanalysis and its dependent factors.—*a*. In military operations time is a vital element. The influence or effect that analysis of military cryptograms may have on the tactical situation depends on various time factors.

b. Of these factors, the following are the most important:

(1) The length of time necessary to transmit intercepted enemy cryptograms to solving headquarters. This factor is negligible only when signal communication agencies are properly and specifically organized to perform this function.

(2) The length of time required to organize raw materials, to make traffic analysis studies and to solve the cryptograms, and the time required to make copies, tabulate, and record data.

(3) The nature of information disclosed by traffic analysis studies and solved cryptograms; whether it is of immediate or operational importance in impending action, or whether it is of historical interest only in connection with past action.

(4) The length of time necessary to transmit information to the organization or bureau responsible for evaluating the information. Only after information has been evaluated and correlated with information from other sources does it become *intelligence*.[4]

(5) The length of time necessary to transmit the resulting intelligence (military, naval, air, etc.) to the agency or agencies responsible for tactical operations, and the length of time necessary for the agency to prepare orders for the action determined by the intelligence and to transmit such orders to the combat units concerned. The last sentence under (1) above applies here also.

c. Of the factors mentioned in *b* above, the only one of direct interest in this text is the length of time required to solve the cryptograms. This is subject to great variation, dependent upon other factors, of which the following are the most important:

(1) The degree of resistance of the system to cryptanalytic attack. This is dependent upon the technical soundness of the system itself, the technical soundness of the regulations and procedures governing the use of the system, and the extent to which cipher clerks follow these regulations and procedures.

(2) The volume of cryptographic text available for study. As a rule, the greater the volume of text, the more easily and speedily it can be solved. A single cryptogram in a given system may present an almost hopeless task for the cryptanalyst, but if many cryptograms of the same system or in the same or closely related specific keys are available for study, the solution may be reached in a very short time.

(3) The number, skill, and efficiency of organization and cooperation of communication intelligence units assigned to the work. Cryptanalytic units range in size from a comparatively few persons in the forward echelons to many persons in the rear echelons. Such organization avoids duplication of effort and, especially in forward areas where spot intelligence is most useful, makes possible the quick interpretation of cryptograms in already solved systems. In all these units, proper organization of highly skilled workers is essential for efficient operation.

(4) The amount and character of collateral information and intelligence available to the cryptanalytic organization. Isolated cryptograms exchanged between a restricted, small group of correspondents about whom and whose business no information is available may resist the efforts of even a highly organized, skilled cryptanalytic organization indefinitely. If, however, a certain amount of such information is obtained, the situation may be entirely changed. In military operations usually a great deal of collateral information is available, from sources indicated in par. 6, above. As a rule, a fair amount of

[4] Often referred to as *finished intelligence*.

definite information concerning specific cryptograms is at hand, such as proper names of persons and places, and events in the immediate past or future.[5] Although the exchange of information between intelligence and cryptanalytic staffs is very important, the collection of information derived from an intensive study of already solved traffic is equally as important because it yields extremely valuable *cryptanalytic intelligence* which greatly facilitates the solution of new cryptograms from the same sources.

8. Cryptanalytic research vs. exploitation.—*a.* In practical cryptanalytic operations, a distinction is drawn between the research or diagnostic type of cryptanalysis and the type of cryptanalysis required to exploit or process large volumes of traffic on a current or semi-current basis.

b. For example, in the case of solution of a code chart system, the cryptanalytic techniques involving the initial diagnosis of the nature of the system and the subsequent recovery of the syllabary and essential code values within the chart may be considered to constitute the *research* phase. The subsequent processing involving recovery of new coordinates after there has been a change of keys, and the decrypting and degarbling of current messages is considered to be the *exploitation* phase. The distinction between the two phases is not at all an arbitrary one, if only because the degree and type of background and skill required for the accomplishment of one differs from that required for the other. Furthermore, it is quite possible that during the research phase there might be a tendency to concentrate on only a limited amount of the available traffic in order to recover the required plaintext values—with perhaps those messages being set aside which are garbled or incomplete or which ostensibly would yield little help in analysis. This limited portion of the traffic would probably be worked on only by a comparatively small group of cryptanalysts. However, once the system is rendered exploitable, all current traffic would be worked on, and as much back traffic would be processed as might conceivably yield useful intelligence from the message contents or might be of technical interest or advantage to the operations of cryptanalysis or traffic analysis; for this exploitation phase, a larger group of persons would probably be required to maintain the processing on a current basis.

c. It may be seen from the foregoing that, generally speaking, cryptanalytic research is a prerequisite to cryptanalytic exploitation, and it is not extraordinary that this research phase can be of considerable duration, even to the point that many of the particular messages which are under scrutiny may no longer be considered as even semi-current. As a matter of fact, there may be instances when a system stays in the research phase indefinitely; this would be true in the case of high- and some medium-grade systems, and also in the case of other systems in which very little traffic is passed.

d. It is the very probability of a long research phase which makes it essential that all possible steps be taken to insure the continuous *exploitability* of cryptosystems. When continuity is maintained, simply a few messages may suffice to keep up with key changes. But, when continuity is lost, diagnosis "from scratch" is often required, even in the case of low-grade systems. This is true because the capability to effect current exploitation of messages after changes of key have occurred is most often based on the application of stereotyped phraseology to the messages. Often just knowing that dates or other sequences of numbers are likely to occur in certain messages or certain portions of messages is all that is required to keep current with key changes. Thus if, during a period when no attempt is being made to maintain cryptanalytic continuity, the users of a given system change the format of the underlying messages or

[5] In this connection, see the remarks on cribs and probable words in subpars. 2*d* (on pp. 3–5) and 49*c* (on pp. 82–83).

begin encrypting new types of messages in the system, the previously-used methods of exploitation may be rendered void—and any subsequent attempt to regain continuity may be very costly in time and methodology.

e. In practical communication intelligence operations there are many factors that interact in a way which may adversely affect the maintenance of cryptanalytic continuity. For example, a new system may appear which requires diverting the attention of skilled cryptanalytic personnel from an older system, or the proportion of time allotted to intercepting traffic on a particular net may be changed, or there might be a reduction in the amount of information usually received from traffic analysis (caused by changes in the enemy's call signs or radio procedures). However, it can generally be stated that once a cryptosystem is rendered exploitable, a "finger should continually be kept on its pulse"—be it ever so lightly—even if the underlying message texts carry little or nothing of *present* intelligence value. In the preceding sentence the word "present" was used advisedly; the possibility almost always exists that a system presently carrying no significant intelligence may subsequently be employed by the correspondents for messages of much higher intelligence content.

9. **Cryptanalytic records and reports.**—*a.* In practical cryptanalytic work the systematization of records and the maintenance of adequate files are of considerable importance. Likewise, the preparation of clear and concise reports, both technical and nontechnical, is a major facet of practical cryptanalytic operations.

b. All messages coming into the cryptanalytic section are assigned a reference number, and a log is kept of these messages showing pertinent data such as the call signs, the date and time of interception, the group count, etc. Duplicate messages (i. e., different intercepts of the same transmission, or intercepts of retransmissions of the same message) are stapled together and garbles are corrected. Other records and files are maintained for special studies; for example, there may be card files on the message indicators [6] which have appeared in the traffic, card files of keys used in past and current systems, etc.

c. Cryptanalytic reports fall into two main categories: (a) technical reports designed to give cryptanalytic personnel a summary of the cryptographic features of a system, with the steps which were taken to diagnose the system and effect a solution; and (b) nontechnical reports destined for intelligence consumers,[7] which reports consist for the most part of message decrypts. In the latter category either all decrypts may be furnished verbatim, or complete decrypts only of important messages (the rest of the messages being furnished in "gists" or in condensed form).

d. In technical reporting, clarity and detail are paramount.[8] A complete résumé of the diagnostic techniques employed in the identification of the system should be included, as well as a comprehensive outline of the steps taken to arrive at the initial solution.[9] It goes without

---

[6] In this connection, the location of groups of a message is designated by the terms A1, A2, A3 . . . if reference is made to the first, second, third . . . positions from the beginning of the encrypted text, and by the terms ZØ, Z1, Z2 . . . if reference is made to the last, penultimate, antepenultimate . . . positions of the encrypted text.

[7] These reports are invariably highly classified, because their dissemination is strictly controlled on a special distribution list of those who have a "need-to-know." This limited dissemination is absolutely essential in order to protect the information, and prevent drying up the source and negating the work of the many weeks, months, or even years that are represented by the fruits of the communication intelligence effort. When information derived through communication intelligence is included in military intelligence reports, it is disguised in such a way as to protect the source of the information.

[8] For an excellent exposition on the art of technical writing, see Joseph N. Ulman, Jr., *Technical Reporting*, New York, 1952.

[9] See also the remarks made in subpar. 47*f*, on pp. 78–79.

saying that close attention should be paid to *precise* cryptologic terminology in all descriptions of methods and techniques, so as to lessen the chance of ambiguity or possible misunderstanding by the reader. A cryptologic glossary should be freely consulted when the writer is not sure of the exact meaning of a term he is about to use.

e. In the next paragraph there is found an example of what may be considered as typical of a cryptanalytic technical report. Of course, there is no fixed, standard format for such a report, because the form and content of each individual report depend on circumstances at the time of writing. However, the hypothetical report in the next paragraph is intended to illustrate the amount of detail that might be included.

10. Illustrative example of a technical report. The following represents a hypothetical technical report on the cryptanalysis of a newly-encountered system:

(CLASSIFICATION) Copy No. ____
Special Distribution of ____ copies

REPORT ON THE SOLUTION
OF THE "CALOX" SYSTEM

5 January 19____

I—BACKGROUND

1. On 12 December 19__, a new discriminant, CALOX, appeared in the enemy's traffic. The discriminant appears in the usual position, the A1 group of the message.

2. Traffic analysis indicates that CALOX traffic is being passed on air defense nets. From the characteristics of the transmission of this traffic and associated procedures, it appears that CALOX is an administrative system rather than an operational system. It also appears that CALOX does not replace an existing system, but rather is a new system introduced for some special purpose. On the enemy's air defense nets, both cipher and code systems have been encountered.

3. CALOX traffic was segregated and logged in as received, together with the worksheet reference numbers assigned to all incoming traffic by the Traffic Handling Section.

II—PRELIMINARY ANALYSIS

4. The first step in treating the CALOX system was to complete the plain-component sequence on one of the messages, on the hypothesis of direct or reversed standard alphabets, using a strip board for this purpose. (The enemy has used standard alphabets in the past in one system, changing the juxtaposition of the components after the encryption of every few letters.) This disclosed nothing of significance.

5. Uniliteral frequency distributions made for each of the six messages intercepted on 12 December were flat; the average I. C. of 1.1 indicates that it is out of the question that the underlying cryptosystem is a monoalphabetic system involving single-letter cipher units. However, a rather odd manifestation in the distributions for each message was that C_c, D_c, H_c, L_c, and V_c were usually consistently predominating, while S_c, Y_c, and Z_c were consistently of very low frequency. No explanation of this phenomenon was forthcoming at the time.

6. In an attempt to disclose any similarity between the CALOX characteristics and those of another system, a check was made on previously solved enemy systems used on his air defense and other nets; this proved fruitless, as the uniliteral frequency manifestations of CALOX were unique to this system. A check was also made to find any possible isologs between CALOX messages and those of another readable system; however, this too proved fruitless, as did the examination of chatter associated with the CALOX messages in an attempt to reveal any clues as to the system or to uncover possible cryptographic service messages, etc.

7. Digraphic distributions were made of the messages of 12 December, but no unusual phenomena were visible. The observed digraphic ϕ approached that of digraphic ϕ random, and there was no evidence to support any matching of the rows or columns of the distributions if the hypothesis of a variant system with a small matrix were assumed.

8. Triliteral frequency distributions were made of each message to disclose repetitions; these repetitions were underlined in the messages, and a comparison was made of those repetitions occurring between messages of the same day. Many short repetitions of 3, 4, and 5 letters were disclosed, the number of these repetitions being considerably above that expected for random; however, no longer repetitions were uncovered, and the intervals between the repetitions had no common factors.

9. Every day's accumulation of traffic was examined statistically with a view to revealing possible key changes, and the phenomena in par. 5, above, continued. When on 19 December the predominant peaks and troughs no longer corresponded to the norms observed in par. 5, a change of keys was assumed.

10. A typical message in the first key period is given below:

LRZ DE CKS (Intercepted 17 December, on a frequency of 5600 kcs)

```
CALOX JOLDJ JLAPP DRELF QXEDZ   QIHFN WMGUH DMAYM IMNDY OMZCC
OMMYE HQDAH YEMNB VUGHD IMXOG   LDHUX MACJV VRNEK LCHEJ DZCDO
PRILM UGBOC DLXJL UBVVW TRAFX   LKNPA HSJNE HVCAC OQTHU FJVTH
DIKQW MCGIW HRMAF LKGBE FNPOG   JROGM WGUDM XJIJL BWEDK QCUMR
TJXAN BLTUR KMTOR CFIHV QCEKH   HUJNQ ATBVZ VNERI LHFOQ MLUMX
LAXVY HEQBX RIKRK YACSV LPOQP   NOBKU XGLED FNPAG JRRAB JLEBW
DKIQC MRADN VNURB TOPBH LIKLH   EPVTR BGYMA MYQWI PVLEM GLEGH
ODMXT DHONG XNXEL DWXGA LDIGB   GCILM ZQLAC LXODQ
```

III—THE SOLUTION

11. The following peculiar sets of similar sequences of cipher letters were noted during the examination of the 32 messages available in the first key period. The message reference numbers are given, together with the position in the message of the first letter of the sequence. (The position given is the *text* position, excluding the discriminant.)

	Msg No.	*Pos.*	
a)	60208	057	H F I J V T H O D K Q W M E C G A W H R
	61492	216	H F J V O T H D K Q A W M E C G W H A R
	60317	139	H U F J V T H D I K Q W M C G I W H R
b)	60317	123	N P A H S J N E H V C A C Q T
	62350	098	N P H I S J N O H V C C U Q T
c)	60317	184	J L B W E D K Q C U M R
	60317	291	J L E B W D K I Q C M R
d)	60295	114	T P I Q K Z H E H V P U V P B
	61007	253	T P O Q K Z H A H V P E V P B
e)	60943	147	H V G G A K W Q S O V R N
	62156	064	H V U G G K E W Q S V I R N

12. The behavior of the letters comprising these sequences indicates that A_c, E_c, I_c, O_c, and U_c most likely are nulls. On this hypothesis, evidence from the lengths of the repetitions now disclosed, and from the intervals between repetitions, indicates a digraphic grouping of the cipher text. A re-examination of the digraphic distributions reveals that there are no vowel-vowel contacts in the cipher text, except for combinations with Y_c. Furthermore, in retrospect it is seen that most of the cipher groups contain 1 or 2 vowels, never more; this significance escaped notice until the near-repetitions above were observed.

13. New digraphic distributions, omitting the 5 vowels, were made for the messages in the first key period. No matching qualities were manifested in the new distributions; but this time the observed digraphic ϕ very closely approximated the digraphic ϕ plain, thus it appeared that, in spite of the limitation of only 21 cipher-text letters remaining after the null vowels were discarded, the system was basically a digraphic system. (This would not exclude, however, a matrix containing a few frequent trigraphs or tetragraphs, etc.) Work sheets were now made for several of the best messages from the first key period, the messages selected being long ones that existed in more than one intercept copy so that garbles might be corrected.

14. On 28 December the first message was solved; this was Message #60317 which was one of the longest, and which was copied by three different intercept operators. One more cryptographic idiosyncrasy of the CALOX system was now brought to light: that of the peculiarity of behavior of Y_c which had been previously overlooked. This peculiarity was that Y_c was always present in pairs, fairly close together; every Y_c was followed by another Y_c, with from 2 to 10 letters intervening. This Y_c turned out to be a number indicator, and the cipher digraphs between the indicators represented single digits.

15. From the original solution, an equivalent digraphic matrix was reconstructed with the consonant coordinates in normal alphabetical order, as shown below:

θ_c^2

θ_c^1	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
B					BY	CC		AY	CH								CE	CK	2		
C		TE	TI	TO											TH						
D						F6	FE	EN		ER	EW		EL	EY							ES
F											GE							FF	FR		
G	MA	LL				MB			ME	LR								MM			
H				AI			AT						1	AN	AD		AR		AC		
J									RE		QU			RA			RD				
K						VE				UL											
L		OL	OP	OU		PA		OM		OR									ON		
M		CO		DC						D4	DE			DG							DA
N			8		HI			GT				HE	GR								
P					NE						ND	NC	MO	NF			NI				
Q		RI				SA				RP						RS		SE	RM		
R	LB							Ø	LE												
S																					
T	EE					EF	EI					EA		ED							
V									IS		IL			IK				IV			
W					OB					NU											NV
X					ST		TB		TA		SS					SO	T				
Y																					
Z													YC								

Noting evidences of symmetry in the matrix, the matrix coordinates were rearranged to yield the primary matrix which is shown below, including values which were interpolated on the basis of likelihood and alphabetical sequence.

θ_c^2

θ_c^1	Q	C	K	X	S	D	M	Z	T	F	P	N	G	R	B	H	V	L	J	W
H	A1	AA	AB	AC	AD	AE	AF	AG	AH	AI	AK	AL	AM	AN	AO	AP	AR	AS	AT	AU
B	AV	AW	AY	B2	BA	BE	BI	BL	BO	BR	BS	BU	BY	C3	CA	CC	CE	CH	CI	CK
M	CL	CO	CR	CT	CU	CY	D4	DA	DB	DC	DD	DE	DF	DG	DH	DI	DL	DM	DN	DO
T	DP	DQ	DR	DS	DT	DU	DV	DW	DY	E5	EA	EB	EC	ED	EE	EF	EG	EH	EI	EK
D	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW	EX	EY	EZ	F6	FA		FE	FF
F	FI	FL	FO	FR	FS	FT	FU		FY	G7	GA	GE	GF	GH	GI	GL	GM	GN	GO	GP
N	GR	GS	GT	GU	GY	H8	HA	HB	HC	HD	HE		HI							
V												IL	IM	IN	IO	IP	IR	IS	IT	IV
R	IW	IX	JØ	JA	JE	JO	JU	K	KA	KE	KI	KO	L	LA	LB	LC	LD	LE	LF	LH
G	LI	LL	LM	LN	LO	LP	LR	LS	LT	LU	LV	LW	LY	M	MA	MB		ME	MI	MM
P	MO	MP	MR	MS	MT	MU	MY	N	NA	NB	NC	ND	NE	NF	NG	NH	NI	NL	NM	NN
W	NO	NP	NQ	NR	NS	NT	NU	NV	NW	NY	O	OA	OB	OC	OD	OE	OF	OG	OH	OI
L	OK	OL	OM	ON	OO	OP	OR	OS	OT	OU	OV	OW	OX	OY	P	PA				
J											Q	QU	R	RA	RB	RC	RD	RE	RF	RG
Q	RH	RI	RL	RM	RN	RO	RP	RQ	RR	RS	RT	RU		RY	S	SA	SB	SC	SD	SE
X	SF	SG	SH	SI	SL	SM	SN	SO	SP	SQ	SR	SS	ST	SU	SW	SY	T	TA	TB	TC
C	TD	TE	TF	TG	TH	TI	TL	TM	TN	TO										
K							UL									VE				
S																				
Z	YC																			

By comparison with other messages in the same period, and with messages in subsequent periods, it was possible to recover the values inside the matrix in their entirety, as follows:

A	AA	AB	AC	AD	AE	AF	AG	AH	AI	AK	AL	AM	AN	AO	AP	AR	AS	AT	AU
AV	AW	AY	B	BA	BE	BI	BL	BO	BR	BT	BU	BY	C	CA	CC	CE	CH	CI	CK
CL	CO	CR	CT	CU	CY	D	DA	DB	DC	DD	DE	DF	DG	DH	DI	DL	DM	DN	DO
DP	DQ	DR	DS	DT	DU	DV	DW	DY	E	EA	EB	EC	ED	EE	EF	EG	EH	EI	EJ
EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW	EX	EY	EZ	F	FA	FC	FE	FF
FI	FL	FO	FR	FS	FT	FU	FY	G	GA	GC	GE	GF	GG	GH	GI	GL	GN	GO	GP
GR	GS	GT	GU	GW	H	HA	HB	HC	HD	HE	HF	HI	HL	HM	HN	HO	HR	HS	HT
HU	HY	I	IA	IB	IC	ID	IE	IF	IG	IK	IL	IM	IN	IO	IP	IR	IS	IT	IV
IX	IZ	J	JA	JE	JO	JU	K	KA	KE	KI	KS	L	LA	LB	LC	LD	LE	LF	LG
LI	LL	LM	LN	LO	LP	LR	LS	LT	LU	LV	LW	LY	M	MA	MB	MC	ME	MI	MM
MO	MP	MR	MS	MT	MU	MY	N	NA	NB	NC	ND	NE	NF	NG	NH	NI	NK	NL	NM
NN	NO	NP	NR	NS	NT	NU	NV	NW	NY	O	OA	OB	OC	OD	OE	OF	OG	OH	OI
OK	OL	OM	ON	OO	OP	OR	OS	OT	OU	OV	OW	OX	OY	P	PA	PE	PF	PH	PI
PL	PM	PN	PO	PP	PR	PS	PT	PU	PY	Q	QU	R	RA	RB	RC	RD	RE	RF	RG
RH	RI	RL	RM	RN	RO	RP	RR	RS	RT	RU	RV	RW	RY	S	SA	SB	SC	SD	SE
SF	SG	SH	SI	SK	SL	SM	SN	SO	SP	SR	SS	ST	SU	SW	SY	T	TA	TB	TC
TD	TE	TF	TG	TH	TI	TL	TM	TN	TO	TP	TR	TS	TT	TU	TW	TY	TZ	U	UA
UB	UC	UD	UE	UG	UI	UL	UM	UN	UP	UR	US	UT	V	VA	VE	VI	VO	W	WA
WE	WH	WI	WL	WN	WO	WR	WY	X	XA	XC	XE	XF	XI	XN	XP	XT	Y	YA	YB
YC	YD	YE	YF	YG	YH	YI	YL	YM	YN	YO	YP	YR	YS	YT	YW	Z	ZA	ZE	ZI

It will be noted that the matrix contains the 26 letters, and 374 of the highest frequency digraphs. When encrypting numbers, the cipher value for 1 is the cipher equivalent of A_p, the cipher value for 2 is the $\overline{\theta\theta}_c$ for B_p, etc., to $\emptyset_p=\overline{\theta\theta}_c$ (J_p).

16. In the matrix coordinates for the first key period, the nonrandom phenomena in the grouping of the coordinate letters was noticed, suggesting that some systematic method for producing these sequences was used. It evolved that these sequences were derived by simple columnar transposition using the following rectangles:

For the rows:

```
H D R L C
B F G J K
M N P Q S
T V W X Z
```

For the columns:

```
Q S T N B L
C D F G H J
K M P R V W
X Z
```

Thus the key words for the first period are HYDRAULIC and QUESTIONABLY (with, of course, the vowels omitted) for the row and column coordinates, respectively.

IV. CONTINUITY OF KEY CHANGES; SUMMARY

17. Having solved the CALOX system for the first period (12–18 Dec), the second period (19–26 Dec) was easily solved by the discovery of a pair of cross-key isologs on 19 December; the third period (27–31 Dec) was speedily solved by means of a signature crib; while the fourth period (beginning on 1 Jan) had to be solved by the general method of digraphic frequencies and digraphic idiomorphs. The row and column key words for the second period were COPYRIGHTED and DOCUMENT; for the third period, CHIMPANZEE and MANDRILL; but for the fourth period the same key word, MNTVD (Montevideo?), was used for both the row and column coordinates. The coordinate sequences were derived by simple columnar transposition, as in the first period.

18. If the enemy has found that two different sequences for the row and column coordinates is too inconvenient cryptographically and therefore continues to use the single keyword procedure started in the fourth period, a statistical technique has been devised for establishing the identity of some (or even all) of the letters of the coordinates, based on a consideration of the relative frequencies of the ciphertext letters. This technique is founded on the fact that in a single keyword procedure the combination of row 19 and column 19 of the basic

matrix will yield a low frequency cipher letter, as will the combinations of row 20-column 20, and row 9-column 9; on the other hand, the combinations row 17-column 17, row 5-column 5, row 13-column 13, and row 14-column 14 will yield high frequency cipher letters. With a single keyword procedure being used, the following is the expected *descending* frequency order of the twenty row-column combinations:

17 5 13 14 1 8 18 15 4 3 12 16 6 11 7 10 2 9 20 19

Even if two key words are employed for the coordinates, a modification of the statistical method is feasible, in those instances where any difficulty might be encountered in a new key period. The statistical techniques and the methods of their employment will be described in a later report.

19. No trouble is anticipated in keeping current with key changes in the CALOX system; traffic should be readable now on the first day of a key change. If the enemy used another set of 5 letters as nulls, instead of the vowels, the new nulls can be identified by searching for and examining near-repetitions, as shown in par. 11. A similar procedure would be used to identify a new number indicator, even though solution would not be impeded by this latter factor.

20. The traffic analysis report on the CALOX traffic gives complete statistics on the links on which CALOX is found, as well as a detailed summary on the number of messages intercepted, etc.

Appendix 8

PRINCIPLES OF CRYPTOSECURITY

In this text the student has gained a background in the elements of cryptanalysis, one of the subdivisions of communication intelligence operations. It is only proper that he should also be given an insight into cryptosecurity, one of the subdivisions of communication security. This should be done, not only to assist him in his cryptanalytic studies by making him conscious of the particular weaknesses inherent in each cryptanalytic problem which he solves, but also to inculcate in him the basic ideas of cryptosecurity so that if he is ever called upon to perform communication security duties, he will be conversant with its principles.

PRINCIPLES OF CRYPTOSECURITY

1. **Communication security.** As has been stated earlier in the text, communication security is the protection resulting from all measures designed to deny to unauthorized persons information of value which may be derived from communications. The components of communication security are: (a) *transmission security*, that component of communication security which results from all measures designed to protect transmissions from interception and traffic analysis; and (b) *cryptosecurity*, that component of communication security which results from the provision of technically-sound cryptosystems and their proper use. It is this latter component, cryptosecurity, which has the most direct application to cryptanalytics and with which this particular appendix deals.

2. **Degree of cryptosecurity required of a system for military use.**—*a.* The ideal cryptographic system for military purposes would be a *single, all-purpose* system which would be practicable for use not only by the largest fixed headquarters but also by the smallest troop unit in the combat area, and which would also present such a great degree of cryptosecurity that, no matter how much traffic became available, all in the same key, the cryptograms composing this traffic would resist solution indefinitely. Such an ideal system, however, is beyond the realm of possibility so far as present methods of cryptographic communication are concerned; in fact, a multiplicity of systems must be employed, each more or less specifically designed for a particular purpose. Of each such system, *the best that can be expected is that the degree of security be great enough to delay solution by the enemy for such a length of time that when the solution is finally reached the information thus obtained has lost all its "short term," immediate, or operational value, and much of its "long term," research, or historical value.*

b. In actual practice, cryptosystems are compiled for specific uses, and may be thought of in terms of high- and low echelon, as well as in terms of high-, medium-, and low-grade.[1] Thus, a cryptosystem may be compiled specifically for use by military attachés, or by diplomatic representatives, or by major military commands, or by secret agents, or by units in the assault phase of an amphibious operation, etc. The design of each one of these cryptosystems is governed by the relative security necessary for the type of traffic in question, by the ease and use and distribution required, and by such similar factors.

3. **Fundamental practical requirements of a cryptosystem for military use.**—*a.* Military cryptosystems must meet certain fundamental requirements of a practical nature because of definite limiting conditions in present military signal communication means and methods. Chief among these requirements are (1) reliability, (2) security, and (3) rapidity of operation. Their relative importance is in the order named.

b. As applied to a cryptographic system or device, reliability means that the cryptograms produced by the sending or originating office will be decrypted accurately, and without ambiguity

[1] See footnote 2 on p. 210.

by the receiving office; that the cryptographic system, whether a book, machine, or device, will be on hand and in good working order, available for instant use; and that when used it can be expected to be operative as long as needed. Reliability is of first importance, and it implies simplicity; usually, the more simple the system, the more reliable it is. *Security*, the protection afforded by a sound cryptographic system, and *rapidity*, the speed with which messages can be encrypted and decrypted, are requirements which generally conflict with one another in varying degrees according to circumstances.

c. Communication personnel must be governed by general principles, subject to existing circumstances, rather than by rigid regulations. Maximum security at all times should be the goal, but in messages exchanged among the higher headquarters some speed may be sacrificed to meet greater security requirements, while in messages exchanged among the lower headquarters security must often give way to greater speed requirements. For this reason various cryptographic systems must be available to meet varying types of situations.

d. Specific requirements which should be met by a cryptographic system for general military use are set forth below:

(1) Cryptograms must be in a form suitable for tranmission by standard telegraphic equipment and methods. This requirement eliminates any system which does not produce cryptograms composed of characters readily transmitted by a telegraphic system employing either the Morse or the teleprinter alphabet. Cryptographic systems using Arabic numerals are not so desirable as those using letters because the Morse signals for numbers are longer, except when "cut" numbers [2] are used, and are more difficult for the average American telegraph or radio operator to handle. Systems which produce cryptograms composed of mixtures of letters and figures, or of letters, figures, and punctuation signs, and which must be transmitted by Morse telegraphy are unsuited for practical usage. However, where such intermixtures are produced automatically by the cryptographic mechanism and are transmitted, received, and deciphered automatically, as in certain cipher teleprinters, their use is permissible. In order to be suitable for economical Morse telegraphic transmission, the cryptographic text must be capable of being arranged in regular sets of characters for these reasons: first, it promotes accuracy in telegraphic reception (since an operator knows he must receive a definite number of characters in each group, no more and no less); and secondly, cryptanalysis is usually made more difficult when the length of the words, phrases, and sentences of the plain text is not apparent. The usual grouping is in sets of five characters, although occasionally other groupings may be made in special circumstances. Such grouping is not necessary in cipher teleprinter systems.

(2) Regular channels of signal communication can carry only a limited volume of traffic. Their most efficient operation demands that the smallest number of characters actually necessary to convey a given message be transmitted. Therefore, the cryptographic text should be no longer than its equivalent clear text. In an exceptional case, the cryptographic text may be longer than the equivalent clear text, but a system in which the cryptographic text is twice the length of the equivalent clear text is useful only if it is of outstanding merit and suitable for certain restricted or special use. No system in which the cryptographic text is more than twice the length of the equivalent clear text is practicable for military usage. Most of the cryptographic systems in current use produce cryptograms which correspond in length with that of the original plaintext message or are somewhat shorter.

(3) General requirements of reliability and speed are that the operations of encrypting and decrypting be relatively simple and rapid. For use in the combat zone, operations must

[2] "Cut" numbers are abbreviated Morse signals for numbers, as distinct from the usual Morse number characters consisting of combinations of 5 dots and dashes each.

be capable of being performed under difficult field conditions and must not require the remembering and application of a long series of steps or rules. They must be such as to reduce the mental strain on the operator to a minimum. Complex processes requiring several distinct steps are not suited to use in the combat zone, but occasionally systems involving only two steps, if each step is simple and rapid, may be practicable for military usage.

(4) Cipher devices or machines for field use must be light in weight, rugged in construction, and simple in operation, requiring the services of only one operator.

(5) The system must be such that errors, which invariably occur in cryptographic communications, can be corrected easily and rapidly by cryptographic technicians. A system is impractical if frequently it is necessary to call for a repetition of the whole transmission, or for a rechecking of the original encryption.

(6) When cipher machines are employed, it should be possible to send in the clear the indicators to designate the particular settings of the machine for a specific message, without endangering the security of the cryptosystem. It might be noted that in almost every case of cipher machine usage observed to date, there is usually a quite complex method of enciphering or otherwise disguising the message indicators, in order to enhance security by preventing enemy cryptanalysts from correlating the intercepted messages together on the basis of their relative positions along the keying cycle. However, the encryption of indicators is not only time-consuming, but is also subject to errors in either the encryption or the transmission; and these indicator errors often make necessary the transmission of *corrected versions of the message texts* which, to a cryptanalyst, is a consummation devoutly to be wished.

(7) Again, if cipher machines are used, it should be possible to change the *internal machine settings* with a fair degree of facility and speed. Furthermore, the *checking* of the setting performed by the cipher clerk should be possible in a convenient manner, short of going through the entire procedure a second time. The cipher machine should yield either *page copy* (such as that from a teleprinter) or a printed tape, which may then be glued to a message form. In the event of a power failure, it should be possible to operate a cipher machine by a manual procedure.

4. Fundamental assumption of military cryptography.—It has been seen that every good cryptographic system combines two more or less separate and distinct elements: a basic or unchangeable method or process, which is termed the *general system;* and a specific or variable factor which controls the steps under the general system and is termed the *specific key.* The secrecy of any military cryptographic system must be entirely dependent upon the specific key because it *must be assumed that the enemy is in full possession of all the details concerning the general system.* This assumption is warranted by the whole history of military cryptography and is based upon the two following considerations which all experienced cryptanalysts regard as valid. In the first place, in military cryptography there are more prolific sources from which to obtain information concerning cryptographic methods than there are in the isolated methods used by private individuals. In fact, by one means or another, the enemy can sooner or later come into possession of full information regarding the general cryptographic system. In the second place, within a very short time the number of messages available for study becomes so great, and the inevitable blunders in the handling of communications have become so numerous that a solution by detailed study can always be made by the enemy, with a consequent possible disclosure of the general system. If a cryptographic system adopted for military use were such that messages properly encrypted in that system could be solved easily (without having the specific keys applicable to the messages) once the underlying methods became known, the entire system would have to be changed, a new system devised, and thousands of persons in the military service trained in its operation. This, of course, would be impracticable. It is assumed that the enemy has knowledge of the general cryptographic system, its cipher devices, instruments, or machines. Only cryptographic documents which are given a limited distribution can be kept secret from the enemy, but they can be kept secret only for a variable length of time before they must be changed. These changes, as a rule, do not affect their method of usage. In cipher systems, the specific key must be susceptible of easy and rapid changes by prearrangement between correspondents. In systems for use by secret agents or very small military parties in the theater of operations, the key may be an easily remembered word, phrase, sentence, or number; it must not require the carrying of written notes on the person. In systems for use by commanders of large and intermediate or even small headquarters in the theater of operations, the specific key may be in the form of written memoranda, paper tapes, and the like. Generally, the specific key must be the same throughout a given period of time for all the members of an intercommunicating network, or at least only a very limited number of specific keys must be in simultaneous effect; otherwise confusion and delay are inevitable. As a consequence of this requirement, the enemy may intercept a good many messages all in the same specific key. A cryptographic system for military use must conform to all requirements of practicability set forth in par. 3, above, and to the foregoing statements concerning the specific key; the system must be such that it is practically impossible for the enemy to solve any message quickly enough to make the information obtained of real or immediate value in the tactical situation, even though he is in full knowledge of the general method of the system, possesses the cipher device or apparatus, if used, and may have available for study one thousand or more cryptograms sent on the same day. There is no single cryptographic system yet known which fully meets all these requirements, and in order to provide the necessary degree of security for a large army several different types of ciphers and cipher machines, as well as small codes for front line use, must be employed simultaneously.

5. Fundamental rules of cryptosecurity.—*a*. Failure to observe the fundamental rules of cryptosecurity often makes possible the solution of cryptosystems by enemy cryptanalysts. These rules apply to the originators of messages to be encrypted, as well as to cryptographic personnel. It is desirable to indicate the following points:

(1) Stereotypic phraseology must be avoided, especially at the beginning and ending of a message. *The known or suspected presence of stereotypic phraseology constitutes the basis of many methods employed in cryptanalysis;* in some cases, indeed, the only possible method of solution makes use of the presence of stereotypic phraseologies, or, as they are often called, *cribs*. Operating instructions for currently authorized cryptosystems prescribe the application of measures which effectively reduce the dangers of stereotypic phraseology to the security of those systems; however, as an added precaution, routine reports which inevitably are stereotyped to some degree should be sent by agencies of signal communication not susceptible to interception.

(2) Special care must be taken to see that the messages are clear and concise. If a message is ambiguous or incomplete, unnecessary confusion results and the accuracy of the cryptographic operation is brought into question.

(3) Messages should be shortened by the deletion of unnecessary words. Conjunctions, prepositions, repetitions of words, and, especially, punctuation should be reduced to a minimum. When punctuation is necessary, it should be spelled out, either in full or in abbreviated form. Numbers should also be spelled out. Where letters of the alphabet must be used, as in certain symbols designating types of equipment, it may be necessary to represent these letters by their authorized phonetic equivalents, where it is essential

that there be no possibility of error. Such spelling out however, should be held down to a minimum.

(4) Authorized abbreviations should be used whenever practicable.

(5) Regulations regarding the manner of indicating addresses and signatures should be carefully followed.

(6) Regulations governing the security classification of messages (TOP SECRET, SECRET, CONFIDENTIAL) must be observed at all times.

b. Much of the success which attends the efforts of cryptanalysts is based upon ignorance and carelessness on the part of cryptographic personnel. Rarely are cryptographic blunders the result of willful violation of instructions; but if cryptographic personnel realize that, by carelessness or ignorance, their own lives and those of thousands of their comrades are jeopardized, they will be more attentive to rules set up for their guidance. The most important of these rules are as follows:

(1) *Questionable messages.* Never encrypt a message which, in the opinion of the cryptographer, violates any of the provisions or regulations relating to the drafting of messages, until the question has been referred to and passed by someone with authority to change the message.

(2) *Mixing plain and cryptographic text.* Never allow cryptographic text with its equivalent plain language to appear in a cryptogram, and never mix plain and cryptographic text, *except in messages where such mixtures are specifically permitted.* This includes punctuation and abbreviations of any description. Such messages afford valuable clues to the enemy. As a general rule, if a message is to be encrypted at all, it should be completely encrypted.

(3) *Text of messages.* Never repeat in the clear the identical text of a message once sent in cryptographic form, or repeat in cryptographic form the text of a message once sent in the clear. Anything which will enable an alert enemy to compare a given piece of plain text with a cryptogram that supposedly contains this plain text is highly dangerous to the safety of the cryptographic system. Where information must be given out for publicity, or where information is handled by many persons, the plaintext version should be very carefully *paraphrased* before distribution, to minimize the data an enemy might obtain from an accurate comparison of the cryptographic text with the equivalent, original plain text. To paraphrase a message means to rewrite it so as to change its original wording as much as possible without changing the meaning of the message. This is done by altering the positions of sentences in the message, by altering the positions of subject, predicate, and modifying phrases or clauses in the sentence, and by altering as much as possible the diction by the use of synonyms and synonymous expressions. In this process, deletion rather than expansion of the wording of the message is preferable, because if an ordinary message is paraphrased simply by expanding it along its original lines, an expert can easily reduce the paraphrased message to its lowest terms, and the resultant wording will be practically the original message. It is very important to eliminate repeated words or proper names, if at all possible, by use of carefully selected pronouns; by the use of the words "former," "latter," "first-mentioned," "second-mentioned"; or by other means. After carefully paraphrasing, the message can be sent in the other key or code.

(4) *Plain texts.* Never send the literal plain text or a paraphrased version of the plain text of a message which has been or will be transmitted in encrypted form except as specifically provided in appropriate regulations.

(5) *Keys.* Never repeat in a different key or system, without paraphrasing, an encrypted message which has once been transmitted, unless specifically authorized by the appropriate authority.

(6) *New cipher keys.* Never transmit a new cipher key by means of a message encrypted in an old key.

(7) *Addresses or signatures.* Never place encrypted addresses or signatures at the beginning or end of the encrypted text. Bury them in the body of the message.

(8) *Identifying information.* Include in the address of an encrypted message only the minimum information necessary for the message to reach the headquarters for which it is intended.

(9) *Replies.* Never reply in the clear to an encrypted message.

(10) *Short titles.* Never use short titles as system or message indicators in encrypted messages.

(11) *Dummy letters and padding.* Never use dummy letters or padding unless their use is specifically authorized.

(12) *System indicator.* Never encipher, encode, or disguise in any way the system indicator, unless specifically authorized.

(13) *Notations.* Never place on the encrypted copy of a message any notations about the system or the subject matter of the message.

(14) *Work tables.* Never allow unnecessary materials such as books, documents, or papers to be on the work table during the process of encryption and decryption.

(15) *Filing messages.* Never file encrypted messages and their equivalent plain texts together. Work sheets must be destroyed by burning.

(16) *Check for accuracy.* Encrypted messages should be checked for accuracy by decrypting the message before transmission. Whenever practicable, this should be done by a cryptographer other than the one who originally encrypted the message.

(17) *Safeguarding material.* Observe all rules of physical security established to safeguard the cryptographic material and message translations. Utmost care should be taken to prevent the loss or unauthorized sight of the codes or lists of cipher keys in use. It is possible to photograph an entire code in two or three hours. Mere continued possession of the cryptographic material is, therefore, no absolute guaranty that it has not been compromised by photography or some other method of reproduction. The only absolute assurance of its not having been compromised is that it has *never* left the possession of the person into whose care it has been entrusted or the safe in which it is kept when not in use. Even if knowledge that a code or cipher has been compromised follows immediately after such compromise, the amount of time and the difficulties involved in notifying all concerned and distributing new cryptographic material are so great that serious damage is caused by the delay and interruption in communication, not to speak of the danger resulting from the enemy's reading the most recent messages in the compromised system.

(18) *Reporting compromise.* Finally, it must be realized that the compromise or capture of cryptographic material is a most serious matter. If there is any reason to suspect that such material or related documents have been compromised, higher authority should be notified by the fastest means possible. Not only is such material available to the enemy for reading current and old messages, but also the cryptanalytic data afforded thereby become most useful in working on similar systems to replace the compromised one. The failure to notify higher authority promptly, if compromise is suspected, may jeopardize the lives of thousands of soldiers and is therefore more serious than permitting compromise to take place.

if it could have been avoided. Regulations for reporting compromise should be carefully observed at all times.

6. Remarks on cryptosystems proposed by inexperienced persons.—*a.* The student has been exposed during the study of this text to cryptosystems which at first glance seem to provide a fair or even moderately high degree of security. Some of these systems may be found in Chapters VIII–XI; namely, certain systems in the category of variant systems, polygraphic systems, cryptosystems with irregular-length ciphertext units, and systems employing syllabary squares and code charts (especially if variants are incorporated into the ciphertext elements). However, the student has also been exposed during the course of his study to the methods of *solution* of these seemingly complex systems, and he has undoubtedly realized that for almost every cryptographic poison there is a cryptanalytic antidote. With many of the systems described herein, it is quite possible that solution based on a single message or a very few messages might be impossible. However, the indecipherability of a single, isolated message in a particular system is certainly no proof of the comparative security inherent in that cryptosystem, nor does it indicate the degree of the worth of that cryptosystem for practical communications.

b. From time to time interested individuals submit to various governmental departments cryptosystems proposed for military use. These individuals, often having but a meager knowledge and background in communication security principles and practices, usually fall into one of two categories: either well-meaning amateurs, or downright "cranks". Sometimes these individuals are motivated by the apocryphal rumor that the U. S. government allegedly is offering a large sum of money (usually "one million dollars") to the person who invents "an indecipherable cipher" (or a reasonable facsimile thereof). It should be appreciated that, at the present stage of development of the arts of cryptography and cryptanalysis, it is unlikely that an amateur would be in a position to make a significant contribution by devising a new cryptosystem that meets the theoretical and practical considerations as stated in the foregoing paragraphs.

c. If, after understanding all the requirements of military cryptosystems as stated in the preceding paragraphs of this appendix, an individual intends to submit a cryptosystem for consideration and evaluation for military use, he should be prepared to lay bare the details of his system and to encrypt a considerable amount of plain text. As a guide to the requirements that any amateur inventor should be prepared to meet, it is appropriate to cite the following set of rules which had been established by one governmental agency to permit the testing of amateur systems under conditions approximating operational situations:

(1) The inventor must submit the complete details of the cryptosystem, including any basic tables, charts, keys, etc., that may be used in the system. Then, using the sample material submitted, he is required to illustrate the step-by-step encryption and decryption of a short message. This is to ensure that there is a clear and thorough understanding of the workings of the cryptosystem.

(2) The inventor must submit a minimum of three different encryptions of a plain text of at least 750 letters in length, enciphered in three different unrevealed specific keys. This requirement permits determining the susceptibility of the cryptosystem to solution through the exploitation of isologs.

(3) The inventor must submit a minimum of twenty messages of at least 100 plaintext letters each, all encrypted in one unrevealed specific key. This permits determining the vulnerability of the cryptosystem to solution based on a volume of text. If found necessary, the inventor might be asked to submit additional messages encrypted in the same key, in case it is felt that the messages first submitted are insufficient for solution.

APPENDIX 9

PROBLEMS—MILITARY CRYPTANALYTICS, PART I

The problems in this appendix are grouped into ten sections, paralleling the sequence of the text, with scopes as follows:

Section A—Fundamental principles
Section B—Uniliteral substitution with standard and mixed cipher alphabets
Section C—Multiliteral substitution: miscellaneous matrices; Baconian and Trithemian systems; elementary Baudot systems
Section D—Multiliteral substitution with variants
Section E—Polygraphic substitution: four-square and two-square matrices
Section F—Polygraphic substitution: Playfair cipher systems
Section G—Polygraphic substitution: large tables
Section H—Monoalphabetic substitution with irregular-length cipher units: monome-dinome systems and others
Section I—Syllabary squares and code charts
Section J—Miscellaneous monoalphabetic substitution systems; concealment systems

The portion of the text which should be read by the student prior to solving the problems in each section is indicated in the section heading.

This set of problems is also available as a separate publication in a loose-leaf book of ten lessons. This book, entitled "Problem Book—Course, Military Cryptanalytics, Part I", contains the cryptograms which for the most part have been arranged in proper worksheet form, obviating the necessity of recopying; and frequency distributions are also appended to reduce the amount of time spent on the purely clerical labor incidental to solution.

* * * * *

DO NOT WRITE ON OR OTHERWISE DEFACE THESE PAGES!

DO NOT WRITE ON THIS PAGE

PROBLEMS—MILITARY CRYPTANALYTICS, PART I

A. Fundamental principles

(embracing Chapters I–IV, inclusive)

1. *a*. What four things were thought by Captain Hitt to be essential to cryptanalytic success?
   b. What six additional elements are also highly desirable?

2. *a*. Define the terms "cryptology", "cryptography", "cryptanalytics", and "cryptanalysis."
   b. What are the essential differences between substitution and transposition?
   c. Differentiate between a code and a cipher system.
   d. Explain the difference between the terms "general system" and "specific key."
   e. Distinguish between monoalphabetic and polyalphabetic substitution.

3. What four fundamental operations are involved in the solution of practically every cryptogram?

4. In the solution of cryptograms involving a form of substitution, to what simple terms is it necessary to reduce them in order to reach a solution?

5. Is it always necessary to determine the specific key in order to reconstruct the plain text? Explain.

6. Indicate the language in which you would expect the plain text of the encrypted portion of the following message to be written. Give reasons for your answer.

```
From: João Fialho, Rio de Janeiro.
To:   Gualterio Costa, Lisbon.

    Com referência ao seu telegrama. NSM NRJPN INJ PMVCOCEN
VNPSN PMBMPCEN QMT JBCVCJ IJUM DTGAJ LTMCPN KPJUCEMIVCNP PMHMQQN
UMIVCHMISJQ SMPVMCPJ SPCHMQSPM.
```

7. *a*. The letter *E* represents what percentage (to the nearest *whole* percent) of the letters in English telegraphic text?
   b. What are the four most frequent consonants in English telegraphic text?
   c. What are the five letters of lowest frequency in English telegraphic text?
   d. What are the four most frequent digraphs in English telegraphic text?
   e. Account for the discrepancies between frequencies of letters in English literary text and English telegraphic text.

8. What three facts can be determined from a study of the uniliteral frequency distribution?

9. In the following extract from a speech given during World War II, each dash indicates the omission of a letter. Complete the text by writing the necessary letters over each dash to form appropriate words.

```
    "Washington's Birthday is a most a p _ _ _ _ _ _ _ _ _ _ occasion
for us to talk with each _ _ _ _ _ about things as they are _ _ _ _ _
and things as we _ _ _ _ they shall be in the _ _ _ _ _ _.
```

DO NOT WRITE ON THIS PAGE

"For _ _ _ _ t years, General Washington and his _ _ _ _ _ _ _ _- _ _ _ Army were faced c o _ _ _ _ _ _ _ _ _ _ with formidable _ _ _ _ and recurring _ _ _ _ _ _ _. _ _ _ _ _ _ _ _ _ and equipment were lacking. In a _ _ _ _ _, every winter was a Valley Forge. Throughout the _ _ _ _ _ _ _ _ _ states there existed selfish men, jealous men, _ _ _ _ _ u l men, who _ _ _ _ _ _ _ _ _ _ that Washington's _ _ _ _ _ was hopeless, that he should ask for a n _ _ _ _ _ _ _ _ _ peace.

"Washington's _ _ _ _ _ _ _ in those hard _ _ _ _ _ has provided the _ _ _ _ _ for all Americans ever since—a model of moral _ _ _- _ _ _ a. He held to his _ _ _ _ _ _, as it had been charted in the Declaration of Independence. He and the _ _ _ _ _ men who _ _ _ _ _ _ _ with him knew that no man's life or _ _ _ _ _ _ _ was secure, without freedom and free i _ _ _ _ _ _ _ _ _ n s.

"The present _ _ _ _ _ struggle has _ _ _ _ _ _ us increasingly that _ _ _ _ _ o m of person and _ _ _ _ _ _ _ y of property anywhere in the _ _ _ _ _ depend upon the security of the rights and obligations of liberty and _ _ _ _ _ _ _ everywhere in the world.

"This war is a new _ _ _ _ of war. It is _ _ _ _ _ _ _ _ _ from all other wars of the _ _ _ _, not only in its methods and _ _ _ _ _ _ _ but also in its geography. It is warfare in terms of every c o n _ _ _ _ _ _, every _ _ _ _ n d, every sea, and every a _ _ _ _ n e in the world. The _ _ _ _ _ oceans which have been h e r _ _ _ _ _ in the past as our _ _ _ _ _ _ _ _ _ _ from attack have become _ _ _ _ _ s s battlefields on which we are _ _ _- _ _ _ _ _ _ _ being challenged by our enemies."

10. *a.* In the following examples the words of sentences have been transposed. Rearrange the words to make plain text.

(1) AT NOTHING REPORT THIS TIME TO

(2) ARTILLERY SECTOR BARRAGE NORTHWEST HEAVY IN

b. In the following examples the letters of several words of each sentence have been transposed. Rearrange the letters to make good words that will give intelligible plain text.

(1) Eight SESTYODRER have DTPADERE to join SAKT REOFC

(2) ABELNU to contact ATTAINBLO on my right AFKLN

c. In the following examples the words of each sentence have been transposed and, in the case of several words, the letters have also been transposed. Reconstruct the plain text.

(1) OLANG RIDGE TANK GIMNOV EHOTISL EAST NOMLCU

(2) DOWN MEYEN OFANERTON SIX THIS OTHS SNEALP

d. In the following examples, the letters of each word of each sentence have been rearranged in the order in which they appear in the normal alphabet. Reconstruct the plain text.

(1) ADELY AACKTT CDDEEHLSU OT CCEEMMNO AT EGHIT HIST GIMNNOR

(2) ADEEIILMMTY NOPU CEEIPRT ADHIRTWW OT AADEEGNPRRR IINOOPST

DO NOT WRITE ON THIS PAGE

e. In the following examples the plain text has been broken up into groups of five letters and then in each group of five the letters have been rearranged in the order in which they appear in the normal alphabet. Reconstruct the plain text.

(1) ORSUU ABIMR AEHNS ENSUV ADKOR ADEGM EEINN EMNVY EELSS S

(2) AEIRR ACNNO AINSS CEEPR AORST ILLRT EEMRY ACELP EMNST DERST DEOY

11. Using cross-section paper prepare a uniliteral frequency bar distribution of the letters of the following paragraph:

"The shortest and surest way to live with honor in the world is to be in reality what we would appear to be; all human virtues increase and strengthen themselves by the practice and experience of them."

12. Determine the class to which the cipher systems, which were used in enciphering the following messages, belong:

a. ORANA THPNO SKTCD MEEES CERAE RNUSA ETLGD AYECA

b. DHJJK QOAHR XKSOF HPQGA PPHLA DIADE HJROA MAHQA

c. ROLEH KBWFZ CQCPZ NVJWZ MIVEQ EPCIN OJSJU YMWQB

13. Which of the following substitution ciphers are monoalphabetic?

a. UJKLW EUVKL FSPAQ PHTKR DZNGL SELYN XYXBX JDATU
WEUZG WFVXM MNZAY AOSGU DCLGI OEWJE IFOKM KNWAP
KOIEV AROEV WSCWN SBCYX

b. HUPYP XXAEP AFGZP VGLHA SLXHU SXXAY PWKAS LHPRH
ALOBA XPLVS WUPJP OBSHU HUPGF XGKPH PVSWU PJOPZ
SVPYS MPOAX ULSLP CGNJX

c. GXYVL ZXMXS LOZGR WEJLX PWTKZ GMXLW QIVZW QBRXK
KTDVL MXAEX VHMXA LOTLY TKDWX GBQKQ LWZXG RTYYZ
KTOXG AWXLQ LOZGR XVWGQ

14. The following messages were enciphered monoalphabetically. Determine in each case whether the cipher alphabet used was a standard or mixed alphabet and if standard, whether direct or reversed.

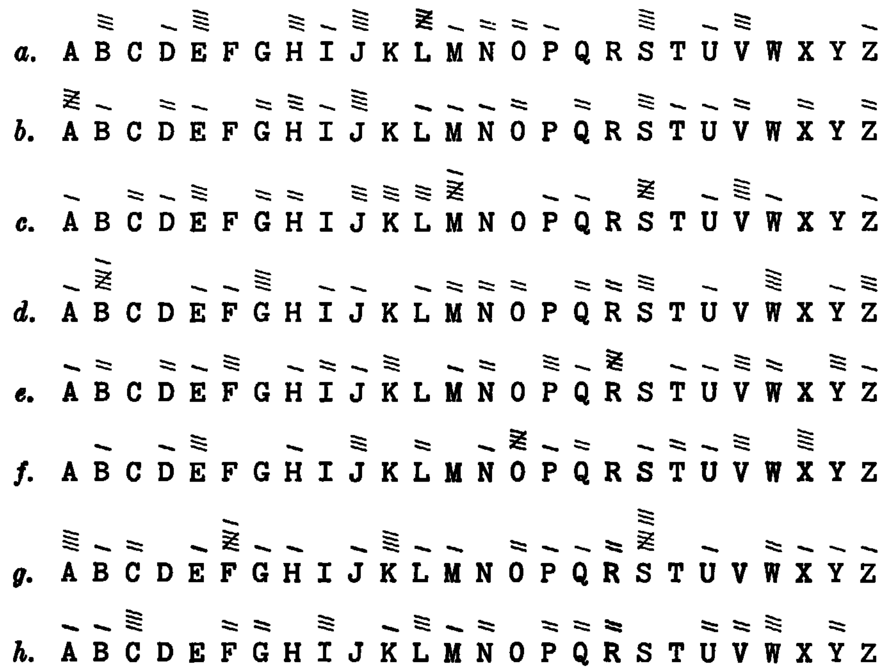
a. ANVOR LOUNQ RLEZW ZHNEZ WZBOR ZKYLF AOZSO ONORF
PJZPP LDZDN LRZLB LABWZ HNAPO WQHOO RZIZU

b. ESPAP LVDLY OECZF RSDTY ESTDO TDECT MFETZ YBFTN
VWJTO PYETQ JTELD OTCPN EDELY OLCON TASPC

c. PYHYL XOLWY JJVYX OILYR YQYPJ KNYLK YHYLC PAYAC
LYXIR QYJVO ZKOXC PCREK UKUPJ IUJUO PRIAS

DO NOT WRITE ON THIS PAGE

15. Derive the ϕ_p, ϕ_r, ϕ_o, Λ_p, Λ_r, and Λ_o for each of the following distributions, and evaluate the [monoalphabetic] goodness of ϕ_o and Λ_o of each in terms of "good", "fair", or "poor", entering these data in a diagram copied from the one given below. On the basis of the foregoing, decide which distributions are most probably monoalphabetic and which are most probably nonmonoalphabetic, indicating your decision by a check (✓) in your diagram; in the case of those not clearly belonging in either of these categories, check "decision suspended".



								Goodness of ϕ_o			Goodness of Λ_o			Decision		
	N	ϕ_p	ϕ_r	ϕ_o	Λ_p	Λ_r	Λ_o	G	F	P	G	F	P	mono.	non-mono.	susp.
a.																
b.																
c.																
d.																
e.																
f.																
g.																
h.																

DO NOT WRITE ON THIS PAGE

16. From the intercepted traffic of three intercept stations operating in the same sector of the front, the following code messages were selected for study by a member of the cryptanalytic section at GHQ. They are undoubtedly three versions of one enemy message, but there appears to be a number of differences, due no doubt to operating difficulties at the several stations. Study the messages and reconstruct from them the actual code text sent by the enemy station.

I. Time intercepted 1612 by HS   W F F   DE   L D C

GR 35 BT

NR17	DYBIE	DUFTO	AMEJA	KIBON	SGCOY	FOBAK	DODLA	LUFYD	KAWAL
APAYN	CODAP	KEDUR	JOPID	JENOX	MEHAZ	LOGIS	KUTEG	EVAUK	IPBEM
KEHZA	HOBWE	AVDUZ	FOFA_	EMCOZ	EGBLO	DOFYO	ENC__	MAWEN	_____
_____	_____	_____	_____	_____					

II. Time intercepted 161Ø by MR   M F F   DE   L D C

GR 35 BT

NRØ_	DYBIE	BUFTO	AMEJA	KIBON	IPKO_	F_BAK	DODLA	LUFYL	KAWAL
APAYN	_____	__DUA	__PID	JENOX	NEHAZ	LOGIS	KUTEG	EVAUC	IRBW
KEHZA	SOBWE	VADUZ	FOFET	EMCOZ	EGBLO	DOFYO	AECDA	MAWEN	___OM
EMCOZ	ACFAH	LOFIR	Ø935						

III. Time intercepted 1612 by YG   W F F   DE   L D K

GR __ BT

NR17	DYBIE	DUFTO	AMEJA	KSBON	IPCOY	___A_	DO___	LUFYL	KAWAL
APETYN	CODAP	KEDUR	WOPID	JENOX	MEHAZ	LOGHKUTEG		EVAUK	IPBEM
KEHZA	HOBWE	AVDUZ	FOFET	EMCOZ	EGBLO	DOFYO	ENCOA	MAWEN	MAWEN
EXFOM	EMCOZ	ACFAH	LOFIR	Ø935					

B. Uniliteral substitution with standard and mixed cipher alphabets
(embracing Chapters V and VI)

1. *a.* What is the first step one should take in attempting to solve an unknown cryptogram that is obviously a substitution cipher?

b. If this step is unsuccessful and the cryptogram is obviously monoalphabetic in character, what type of cipher alphabet may be assumed to have been used?

2. *a.* Name two methods of solving monoalphabetic substitution ciphers involving standard cipher alphabets.

b. In the solution of a substitution cipher by completing the plain component sequence involving reversed standard alphabets, what are the successive steps?

c. Why do monoalphabetic cryptograms involving standard cipher alphabets yield such a low degree of cryptosecurity?

3. What are four characteristics of vowels which permit their classification as such in monoalphabetic substitution ciphers involving mixed cipher alphabets?

DO NOT WRITE ON THIS PAGE

4. *a.* What two places in every message lend themselves more readily to successful attack by the assumption of words than do any other places? Explain.

b. What is meant by the "probable word method" of solution?

5. *a.* What is meant by the word pattern "A B C B A D B"?

b. For each pattern given below, indicate one good English word that contains the pattern:

(1) A B C B A D B

(2) A A B A

(3) A B C D A

6. Give two reasons why the enciphered text of a military message is generally divided into groups of five characters, prior to transmission.

7. Solve the following cryptogram and indicate the specific key ($A_p = \theta_c$):

```
JMQVS  QZXIF  FMZSL  IZMLZ  CEMEB  FQOME  MDXYQ  OZCYY  XJMZI  VMZIY
OQWYI  DKYMV  MZMNQ  EQKMX  CCWZB  CYIXI  CDYYX  CBZQI  FZCQN  HWDOX
ICDJQ  YPMMD  YMVMZ  MFSNQ  EQKMN  QDNEW  OJMAW  IBEMD  XNMYX  ZCSMN
YXCBU  MQZME  CVIDK  CWZXZ  CCBYX  CZMQZ  BCYIX  ICDYY  XCBZQ  FYXCD
```

8. Solve the following cryptogram, and indicate the specific key:

```
WXLMK  HRXKL  ATOXU  XXGHK  WXKXW  MHIKH  VXXWT  MHGVX  MHTKX  TPAXK
XLNUF  TKBGX  TVMBO  BMRAT  LUXXG  KXIHK  MXWLM  HITVD  GHPEX  WZXXX
```

9. Solve the following cryptograms, and indicate the specific keys:

a. QHHYL YDWQJ JMEFC

b. YXSED YFSXU HWXUS

10. The following badly garbled cryptogram was intercepted. Reconstruct the original plaintext message, resolving the errors and omissions, and indicate the specific key:

```
HUVSH  UDSU-  EKHCU  IEQWU  DK-RU  HOXHU  UUYMX  JIU-U  DTQJU  TEDUA
YNTUS  -----  IJEFY  DIJKH  SJYE-  IOQLU  RUUNY  IIKU-  JEQBD  IKRHE
TYDQJ  -SECC  Q-TIJ  EYDYW  YQJUK  DYJJH  QYDCD  WFHEW  HQKIK  DTUHJ
XAFHE  RYIYE  DIEVF  QHQMH  QUXJ-  EEV-F  -SYQB  THTUH  IDMCR  UHIYT
```

11. *a.* Construct a triliteral frequency distribution showing one prefix and one suffix of the letters of the cryptogram below.

b. Prepare a condensed table of repetitions of digraphs and trigraphs appearing more than twice, and include all repetitions of longer polygraphs.

c. Using the data obtained in *a* and *b* above, complete the solution of the cryptogram and recover all keys:

```
UBSYB  VXRPN  CGUMZ  XGPNP  CUBQP  UXXFZ  XBNBM  IGVRP  NVXUY  RXGND
FBZHI  ZUXGL  LBUIB  MQLZR  BMBNX  VGNOP  PABAZ  UBZPN  BCGHB  MGLBV
NPUXF  BZVXP  CDUBB  NHGLL  BVXPQ  QFPXP  DUZQF  GRUBR  PNNZG  VVZNR
BMGVV  GPNVN  BDZXG  HBEBR  ZYVBP  CZAHB  UVBOB  ZXFBU  RPNAG  XGPNV
```

DO NOT WRITE ON THIS PAGE

12. Solve the cryptogram below, suspected to contain the probable word "BLOCKADE"; recover all keys.

```
LCTCE LUZOD UCREA WZUSN FZXDY DRTLD SDRZS DEUCM UZZKZ UDCDV
TQTXD AOYZC ZWYDX PTVZD SCMZZ RZAQL LDECM ZURXD TLCMT LWZZR
ZSSZX CZVLC DOUDX PZCWT UUTHZ SUDAD EUFZL LZYLX DRCNR EZLCD
MTUTL LMDLC NYZLM DUZOD LNCND RLTRV MTLVT ATHZV UTNYY NRZLX
```

13. Solve the following cryptogram and recover all keys:

```
JZDFV WHEDZ VHWDS YKTWD OEDZD EDSEC CWHHW EDZTE XXWSZ VNZVZ
SPFJK VZTYP HJDWO LJWDP VPWTI REDZE XEKVF PJVEY HHJEF EDZFV
WHEDZ VHJPJ ZHJLP JXEKV JLTWM WHWED WHWDM WSWDW JREXI YKZCE
KDJPW DCEMW DONZH JJEPJ JPSBE KVFEH WJWED HNZHJ EXXPW VJEND
HJEFS EDXWV CPJWE DVZGK ZHJZT
```

14. Using the sequences recovered in Problem 12, solve the following cryptograms and indicate the specific keys:

a. URJJR XQUQX KSARB BETOI

b. FDLDY XZUMU EUFPN DVOFE ALYRW UMLJX AFDYE XEKQP
DOYCV REUAX

15. The following cryptograms, enciphered with random cipher alphabets, are in bona fide word lengths. Solve them.

a. HY ARVJZGHAROT VK CGKMMGKHZM LKUG LKUG OROE HOZ
EMVHFSRMJROT JEHZPUHGVEGM RO MCJKKSJKUME

b. RGRQRU TDSPYURDP ZFTAVDRC AYCFO JO DRZYUUFSPPFUZR
TFADYGP

c. CDGWDSA LCAUMMDCR BUCD YV DVDJR IYSUAUYVS LZCYSS CUTDC

16. In solving several unrelated monoalphabetic cryptograms, the following cipher alphabets were reconstructed. Recover the key words from any five of these alphabets.

a.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: N L W P F R T H S Y D Q A K V E B M X G C O Z I J U
```

b.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: Z Q X P E O N M W L K J H G F D B V Y U T R I C S A
```

c.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: P Q E R V M O Z W U T H A X B C D F S Y G I J K L N
```

d.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: C D G P V Z K H Q L A E I J N S W U B F M O T X Y R
```

DO NOT WRITE ON THIS PAGE

e.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: L B E K D G R M F A X S N H C Z T O I Y U P J V Q W
```

f.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: A U Z J T X H S W G R M B N O C I Q F E K Y P D V L
```

g.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: C K V E B O Y F D P Z G Q H S I T L W N J U R A M X
```

h.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: L M C P O Q I J H R S N T B D E U G V K A W X Y F Z
```

C. Multiliteral substitution: miscellaneous matrices; Baconian and Trithemian systems; elementary Baudot systems

(embracing Chapter VII)

1. Solve the following cryptogram and recover all keys:

```
DTLRW  EOEOE  WHRRW  RLAWH  WADED  AWRLE  LEORR  EWTOR  WAOHW  HORLE
LRWAR  RRRWH  WAWHO  EORLE  LEWRW  AWHOH  LRLEL  RWAOH  OELRO  AOAOE
LRORR  EOAOA  WHWTW  HWAWA  WRWAW  HDERT  OEWHW  HREOR  OARTO  ELROR
REWRW  EWAOH  DEWRL  RWAWA  WRWAW  HDEDA  LRLRW  AWHOA  DELRL  TLTLR
OAWRD  EWRLR  WAOAL  RRARA  LRWEO  EDERT  OEWHR  RWRLA  WHWAD  EDAWR
LELEO  TWHOE  WHWHW  ARALR  OEOHW  HREOT  DTORR  EREWR  DEWRL  RWAOR
LEORO  EDEWR  LELEW  HOEDT  OAWEL  TLTLR  OEDEO  ADELR  LTOHL  RLELR
WAWHL  EOTWH  WAWAW  RWARR
```

2. This message was sent by the Fifteenth Infantry. Solve it and recover all keys.

```
CYAON  XCNNO  CNAOA  OOGON  NGBYO  XOXRO  CGNYR  OANRE  AGROO  XAOAN
AXAXA  GANAG  CNROO  XOXBY  ANAGC  NBECX  BNBXC  GROON  COREC  NAYBG
CEONN  OAOOG  RONON  ORORE  OONGB  YOXOX  RYAGA  XBYAN  OGCNA  OOYOG
NOOXC  YNXOG  AOANC  NAGRE  AGBYO  GNOAO  BOAOC  NCGAG  CNONB  OCNAO
OYCOO  EONNO  AOOGR  ONONG  RONOA  GCNRE  AOOXR  XAEBY  ANBO
```

3. Solve the following cryptogram and recover all keys:

```
RGGPE  EGRRG  GPESG  RRGPP  GEPRG  ERGGS  ASGRR  RGSAE  PPGPG  APPRA
EAESG  RRGPP  GERAP  RGSRE  GPARG  PGSPP  GPRGR  AEAPP  PSPGA  RPEGA
RRRGG  PRRRE  PGPPR  AEARS  PGPER  GARPE  GARRR  GGPRR  RPAEG  SGAAP
GPPPR  AEPES  GPRAG  PRAPE  PRPRA  EGRGP  RAGAG  PGPRR  GPRRG  RASAS
GPRRG  RGSPP  GPAEG  ERSPG  RGGSR  EPPGR  GGGSP  PGRPG  GAPGR  SREPG
ASPRG  SGAGE  RREAE  SGRRG  RRRPG  SPPPP  GSAEG  RPGGA  EPRGG  PEEGR
RAGRP  PGRPG  GAARG  SRARP  GPGPG  AGSPE  ESPGR  GGRER  GPRRR  PGERG
GPAGG  RASGP  GAPPG  SAEAR  PAEPR  GGPPR  AEGER  GGPEE  GPRAP  PGPRR
```

DO NOT WRITE ON THIS PAGE

4. Solve the following cryptogram and recover all keys:

AACAA	BBBAA	ABAAC	AABAB	BACCA	ABCCA	ABAAB	CAACC	AAAAB	BAABA
AAAAB	BBAAB	ABBAB	CCAAB	ABAAB	AACBB	AACBC	BAAAB	BBABC	CACBB
BBBBC	ACABB	AABAA	BCAAC	ACABB	CAACA	ABAAB	BBCAA	ABAAB	ABAAB
AABAB	BACCA	AAABB	ACCAA	BBCCB	CCAAB	BACCC	CABBA	ABCBC	ACAAC
AAACA	CBCAB	AAAAC	ACCBA	ABAAC	ABABA	AACBC	BCCCB	AABAA	CABAC
CBAAB	AACAB	A							

5. Solve the following naval message and recover all keys:

11101	10333	12231	03023	33122	31000	06002	60610	15231	40424
24052	33206	03042	61122	33263	12334	11052	33011	00001	12200
20010	02600	06151	62611	13367	89310	62222	26050	41221	04101
30511	24230	52604	22221	21604	10151	10023	14122	30105	00113
50024	11111	33504	10131	42305	03042	60623	10360		

6. Solve the following cryptogram and recover all keys:

45264	56282	02523	29276	16145	23820	63216	52729	27212	60652
16729	47694	56529	02146	04161	25424	90692	12143	65026	45672
92325	61272	84543	04182	04221	67262	94523	41252	92945	23820
46272	34506	52921	63023	45646	74565	29082	21670	23456	12582
02947	27650	29210	23472	12543	65000				

7. Solve the following cryptogram and recover all keys:

05105	23804	91161	38349	22702	74491	16138	33834	92274	27505
31612	74492	16127	14914	92274	38216	12724	91161	27138	10523
84274	05405	23801	61491	16105	22713	80271	05227	44910	51052
05327	14921	60491	05227	10502	74163	38016	11653	85492	27405
20531	61494	49238	42713	82492	27427	20522	71380	49127	02714
91270	49149	12702	72273	05505	30522	74272	16127	13814	93052
49449	24910	52380	05149	23834	91492	27449	23823	82384	38105
23844	91050								

8. The following is a text in the Baudot teleprinter code enciphered by a simple machine employing five two-position switches which operate polarized relays. Each switch has the function of changing the polarity of its respective baud (a single "mark" or "space" impulse), if the switch is in the 'active' position. If the switch is in the 'inactive' position, the polarity of the baud is unaffected. The switch settings remain constant for each message. As an example, if switches 1 and 4 are active (x), and 2, 3, and 5 are inactive (o), then the word REPORT is enciphered thus:

Key:	xooxo	xooxo	xooxo	xooxo	xooxo	xooxo
Plain:	-+-+-	+----	-++-+	---++	-+-+-	----+
Cipher:	++---	---+-	+++++	+---+	++---	+--++

Solve the message and recover the switch settings:

+-+--	+-++-	+---+	++-++	++++-	+---+	-+++-	-+---	-----	-+-+-
+-+--	+----	+---+	--+++	++--+	++-+-	+-+++	+++-+	+++-+	--+++
++--+	+++++	--+-+	++-++	+-+--	+---+	-++-+	+++++	+-+--	++--+
+++--	+++--	--+-+	+++++	-----	+---+	---++	++--+	+++++	--+-+
+---+	+-++-	+--++	+---+	+--++	--+-+	-++--	+-+--	+-++-	+--++
+---+	-+++-	-+-++	++--+	-++--	++---	-++-+	-+--+	+-++-	-----
+++--	+--+-	-+-+-	+++-+	+++-+	+---+	+---+	-++-+	+-+--	-++--
-++-+	+-+--	--+-+	+++++	+---+	++-++	+-++-	+++--	+---+	-+++-
-+---	+-+--	-+-+-	+---+	++--+	+--++	---++			

DO NOT WRITE ON THIS PAGE

D. Multiliteral substitution with variants
(embracing Chapter VIII)

1. Solve the following cryptogram and recover all keys:

RADEK	EPEVE	TIBOL	AGODU	JOBEK	IBIJO	BUJAV	AMELA	BEKIR	EFEDO
VIJOS	ADOJE	KIBAM	OSACU	GEGEP	IBOKI	JUCEC	IMINE	POJUC	ERENA
BUBEK	ORADE	KETES	ETIJO	FAGOD	UDOJE	KIDIJ	OBUJA	CEBOF	OBABU
DALEJ	ONIDO	NABOB	EPIGI	METEC	OJOTI	SABOT	IDUMO	FABUN	ADUDE
TOGIB	ESEBU	GECOP	ATAKE	CENAV	AMOLO	MENAD	UDECE	BOFOD	ADUDA
LEBOS	IJOVA	DODET	INIDO	COFID	EVECI	BUDAL	EBOVI	DONAJ	OBEKI
VADUD	EKOGO	REMOP	ESARA	JEKAD	OPIRI				

2. Solve the following cryptogram, and recover all keys:

DRDDS	YDARA	RRSBY	ABTTY	ARHID	BTBAD	YYYBS	AAAHI	DATDH	RYBTD
RBRIA	IHHBT	DDIAA	IBBHA	YDTHY	AHIBA	YTYDY	YBDYH	SDDIS	BAAST
YDRHS	DSRYR	DTSRR	ARRYB	SABTT	YHRAI	DBIBA	DDYYB	SAHAH	IDATD
TSDBS	HYHDI	SDTTT	TYYHH	STYIS	BAAST	DDAHD	HYTRH	HIIDA	RSBBA
RIHBA	IHIRH	DBSHH	ARIDA	AIIBY	BDISI	DDYAB	BYTHH	IIYHT	YBSDD
YRSRR	IHHTD	DTTAA	IRYST	SHDHA	BAITI	YTAHH	YARAI	RHDIY	DDDYA
TBDTH	HSBAA	DTDDR	HYDDR	YBDHS	HSRDD	DASIR	IIDST	BDSIS	DTTBH
SHRIA	AHIBB	ISBIH	IRHAY	DBBAA	IDHSH				

3. Solve the following cryptogram and recover all keys:

99185	78212	28789	02504	15300	40614	57346	42072	15300	25744
84526	61181	87583	57831	14709	06847	30131	52186	92431	03035
20313	26418	57268	41206	34256	97290	78079	03129	57508	21953
31725	13610	86364	71867	26049	28230	08315	89088	87911	02082
31145	65731	88043	13066	47303	61899	20069	73121	55991	82010
28746	89041	69829	07831	86881	59126	92728	71443	20532	86492
47025	83510	96053	43785	06268	05092	68107	08192	18028	64947
07829	40669	15219	05610	40016	89015	35575	23260	47643	67106
55006	87845	52126	943						

4. This message is suspected of having an ending similar to Problem 3. Solve it and recover all keys.

22087	12919	83053	47658	05566	22622	35487	51378	58346	50207
71518	73596	10326	94547	81461	10114	67377	57935	30532	93746
60193	09466	49688	85798	84933	08628	90510	45303	84765	83157
42128	64936	79542	80938	24418	66379	08286	76866	94226	37166
83560	50758	95601	96226	48235	94038	15674	39242	62774	37954
69386	51682	10966	79757	48932	41353	29463	73265	12948	49568
83939	83775	79451	29784	53037	57695	31293	22149	17257	30069
86367	94519	77983	89597	93949	87242	59000	85044	27266	25706
9123									

DO NOT WRITE ON THIS PAGE

5. Solve the following cryptogram and recover all keys:

80713	06941	35696	80213	28061	37695	69680	91394	78800	25513
28096	91134	47713	68026	97695	13913	72502	56475	80280	88091
35802	25247	31341	39696	25525	12508	09132	47825	81314	74256
69525	51301	36477	13169	46966	90699	80247	46951	30801	80525
11378	04470	69213	11308	03477					

6. Solve the following cryptogram and recover all keys:

18905	52131	89011	04414	52131	34022	05518	92022	35156	19005
52240	55145	19020	21561	67189	08815	60110	44190	08801	11900
22055	05514	54044	15460	35832	53583	14303	41532	53474	15459
46035	83813	14280	27946	04603	14448	51628	03143	58404	33637
04044	15291	37031	43036	73730	72971	87296	73684	70757	26957
30572	71872	97075	72550	57261	76847	29729	60661	77186	51572
71871	85385	94572							

7. Solve the following cryptogram and recover all keys:

72109	19015	41776	04657	89925	96235	70368	62717	67091	83938
99294	88596	52368	62170	37091	22620	80735	96695	04627	17032
53136	77644	22537	12262	47907	38026	22703	88434	30196	04118
66826	27034	15596	84825	35230	46569	16375	84979	74893	10920
85780	73541	97477	67212	08479	35210	91365	78947	39865	97030
28334	15432	54516	59910	04639	82992	26541	09142	43430	28208
75852	33987	03712	25322	67217	58569				

8. The following cryptograms are suspected to be isologs. Solve them and recover all keys.

Message "A"

09728	23144	33987	73514	27769	10677	94418	99479	41948	66432
24374	48499	56758	47636	35546	81176	12242	30777	76194	15272
62644	85211	21361	71687	28759	72459	47047	20204	22145	53570
21377	58467	36166	13037	05358	25876	64403	33524	36847	98975
76679	83637	79946	05777	46243	95667	15086	47920	54391	27284
32060	43178	94367	66414	32190	15429	62648	60975	47915	66679
14422	70281	93894	71368	35325	27686	21707	79439	22000	

Message "B"

87560	77444	35211	41109	33772	89084	55415	78586	41056	35506
15844	48995	20110	23777	58199	19437	57052	62714	37174	88756
25154	11724	98779	72367	61813	38507	47890	68719	65521	08875
68548	81270	37609	17554	83811	72477	85433	50805	37598	60718
37306	17704	06159	62714	46551	69370	50945	58696	19561	70682
86600	23474	55377	71502	16576	41295	65052	00751	47289	33956
59497	38764	66574	72261	08560	73763	68350	48516	25000	

DO NOT WRITE ON THIS PAGE

9. The following naval messages are suspected to be isologs, containing the probable word "TASK FORCE." Solve them and recover all keys.

Message "A"

```
43022  83524  26060  98448  56175  57368  05544  54713  25748  18995
73211  78809  78230  46746  55566  38971  52835  54310  66179  30225
49705  63605  75310  83452  92351  03132  27998  93539  26288  11095
80473  12200  63369  42108  52097  11477  11306  68721  98883  68453
95650  15184  59749  92076  67000
```

Message "B"

```
77639  32338  96687  32583  16771  36033  25195  21007  61936  37147
94702  74323  91551  84030  23211  74696  15784  34746  34170  59391
35584  17645  65752  24915  07432  64598  99104  17307  66639  31127
90402  53353  77760  84479  75139  10388  02285  42214  80132  62568
27529  42875  07934  45455  20000
```

10. The following cryptogram is suspected to begin with the opening stereotype "REFERENCE YOUR MESSAGE. . . .". Solve it and recover all keys.

```
40162  42385  52104  83121  44422  37211  99099  42127  37912  77785
80116  44444  13378  77640  12255  50022  48883  78850  22287  84629
99920  06648  91253  20729  01331  81222  90051  99523  19391  41936
61045  48376  88311  15454  00022  05509  60615  57129  18859  20396
66603  14945  35079  88552  82411  08663  05032  28600  07722  55212
00080  00774  72883  40000
```

E. Polygraphic substitution: four-square and two-square matrices

(embracing pars. 64–70 and par. 73 of Chapter IX)

1. Construct a digraphic distribution for the cryptogram below. Solve the cryptogram and recover all keys.

```
MHEAX  PSOZP  LMHTX  NPRQU  EHHGR  HGRLC  HUZWV  ABDMD  WOHUZ  BXRDD
DURHG  RCGHO  SOZBX  NWSZO  RGBLP  CSOZP  ZCOCB  LBTQL  CPGRC  CLUSD
WSPRP  XMDYG  AMDFM  HILQH  CQYOI  QPFGF  NDBIP  FCCVA  LFCCC  QYOOZ
RTOQV  HCGIQ  KPDLI  YAOHA  CGHOP  FGFTD  CPAOH  ACVLF  CCCQY  OQUOD
CNOGC  NWAQC  HTMHK  QZGLC  HUPBG  TIDWQ  OFWGZ  OLGKG  DLZWV  AZBXR
AUVAD  HPFMH  PFZBO  HWLCG  HOLZS  OZBXN  WSZOI  UTAAO  GAHAE  UMHYO
LTEUM  HHULV  WQMZY  TZTZO  YGSOL  ZCPEU  ONMUW  GRTCG  HOMHM  SPROT
MFLFC  PKSRO  WZMNQ  UCPAO  HACVC  OMHDT  DUPFG  FNQYG  QDIBW  SZOYG
BBYOZ  NXRLW  HUIQW  SQDDC  ZBIBR  MSOEN  IARUD  WGRCG  HOMHM  SPRQD
DCLZT  NWLKN  PFXDU  TWAZO
```

DO NOT WRITE ON THIS PAGE

2. The following cryptogram is suspected to contain the probable word "REQUISITIONS". Solve it and recover all keys.

DIAFI QGYII KGITG CQCOV DEKUD MRIRN RONKU NKPDH CURHP AQQPL
ODCMY BHEMR QQEGL PQCIP OPLNS NDTDP UTRSE MRNOA QLODD AQCDZ
EMWKN KDCXP RKHTH EQRQU BPLPL KNQLE SRHFS QSOQR CMQLH LACTX
GKKFC MTTOB RPITB KMPHL BCIUM TZFRW ZRCKC MDCVC OAUAQ PRRIV
RQPKT PCEQH KPUCR ONKHO QQPLO DCMYB HEMRD IAFIQ ITIWS OEMDM
QTRQB KLGTB CEAYI ULKNX

3. The following cryptogram was enciphered by means of an inverse four-square matrix, wherein the cipher sections are normal alphabets (I=J) inscribed in straight horizontals. Solve it and recover all keys.

QUKFF TTOIP KIGBA DESFQ HDGLA GKIEE TORDG HSORH FOFVK IGORH
HDGUU UIPWF ONMDR IHBME SOBUM OFFUK GIGLA GOOMI GHPQG IFGER
GIION LQDFQ QKHDW GFGFG NUTOM HLPOK GGQXA HGQPY KTPZK TLPFV
KYKTN ERTUQ ITRKF GDNUS IDLUP DHAKB ODGTF ZIAFQ FFZDP IGHRI
QCFQA HGQMT FUAHG QPDDP TYLPE QANSD RTTLL CKIKA RIDCT OLUFA
RFUCT OWF

4. Solve the following cryptogram and recover all keys:

WBITH SSESS NAECP ARTRG DOCPR OTTGA TCENL ZRTOO OADPE HPVNI
FNEOT SOOWC ADWBS PENQC OVASB SVDDR NSROQ CCAWO OGENZ PQSWO
POPAR TRGSN ISCEO TNRRO QCZEW OOGEN ZPQSW BOPWQ RPIAH AECOG
HAEHZ TSQPO ITCNH ARTWP SUHSC AABSQ SSSQD ASGAZ IACWH AIEKN
RDSAL ENHBP NAACS QSSSQ DASSN ZIEKN ODCWH AWOPO AEKRT SMCAL
HW

5. The following two messages were intercepted on the same radio link half an hour apart, Message "B" being in answer to a request for a service. Solve the texts, recover all keys, and determine the cause of the cryptographic error involved.

Message "A"

DSZMC MGIQM ABVGE DSUXI TOSQO RNRSB PNQOH NTBLL QNQSS ICRYU
TQCCK GATFN YFVGE DCGNU MOLLN PSOSB NPSQO RNRSB PNCMM BRPOG
LLYXC MGIQM ABSON QLZLC FDYRV IORSB

Message "B"

OYRUP UKVTU WOIWL LNREV VDNBB ZYZNO ASTDH SVODE TSTYN VPZSR
VBPMF WWQXS SKIWL LPWYR CDDEY CNDNO YCNBB ZYZNO ASPUC OUCBW
DESFP UKVTU WONDY BDIDM XLSZI VBZNO

DO NOT WRITE ON THIS PAGE

6. The following cryptogram, suspected to begin with the words "AIR RECONNAISSANCE REPORTS", was enciphered by means of a four-square matrix with four different keyword-mixed sections. Solve it and recover all keys.

```
ICROI  KHCAA  OUTCI  HBONR  RDTTC  DFILP  MSTDX  ALBEA  OYTIP  CPFHC
QCRDT  MXBUS  CDMOE  BIVHE  GGLBX  ANDUS  KCSFE  AOYKC  PLHCU  SLOBT
NKNLI  GPFHC  NDGCI  XYFRR  HEGGO  WIRQI  ILIRN  BTTHG  TMOUS  CBTRD
USHSL  BSOAR  UFCSC  AEHCF  TSOEA  QHCTM  ARNRO  SRUOE  AQLAD  BXAIA
CTNKO  GSFUI  OECBT  YUSHS  ASTDH  CASCF  TSOEA  QBDIK  CSQIB  RNKLB
SOARP  NQEME  NRRDU  FARNB  UTRSG  CSCGC  NLSOB  DSFOR  USHSK  GIHAS
SFCKH  DBDTY
```

F. Polygraphic substitution: Playfair cipher systems
(embracing pars. 71 and 73 of Chapter IX)

1. Solve the following cryptogram and recover all keys:

```
UASKU  ASPKM  MRIOO  RIXYR  OROSM  SSDUA  RHLVB  RSAAK  SWAXS  ABGCW
PNCWX  SAUBS  NANMS  MVUYN  ARORH  GRARL  SAYNU  KSAOR  MYIAA  UBFMS
SDUNW  CASEN  CAHND  RFBNL  ASBUR  FVURO  MWBER  PNYIA  CUGRW  OSRXS
AUBUL  BUGDM  KSMRG  NHQDM  DUHUS  BWCAS  UBEOA  SADWB  NODKA  DFBRH
SPRWH  OUCAD  FTNOS  BAURD  OTMSS  DUNXB  VLUAK  SHQKS  MRUAR  HLVCU
CWFBN  LQOAR  ORHGR  ALVAL  UBUAC  ERBAD  EQYXO  EXHUM  SWPRF  SONUK
GNDRU  PURXW  RWQHF  RSDAS  NOADU  AWBNO  WSIGF  SLRUB  CAUOR  PASSB
NWBHN  PBSNA  UAOER  BRBRZ  ROKRO  RVRNU  USDRS  RSPRW  FRDA
```

2. The following cryptogram is suspected to contain the word "DIVISION". Solve it and recover all keys.

```
MPQKK  ASZQK  KAHXE  HLKYS  NDTPC  QOLNP  RCAHL  MSKND  YGQKD  URFQK
EHLKY  SNDTP  SAOES  YFRQP  FEYSM  OFDAF  RJRSD  URFRN  TPCQU  LLMSK
NDUDF  MJEHR  VNQKU  DECLF  AKBHI  YQVSM  FOSYD  Y
```

3. Solve the following cryptogram and recover all keys. It is suspected that this message is signed "WINTHROP COL INF."

```
4L654  LC31V  PV7WX  VZXB1  DS07L  4CW4K  0FRT4  L790L  HRYNM  RRMDQ
QV9R6  MCX4K  QF4N4  L790L  HROPE  4NRQB  4MXSW  NØENU  GCQX4  KØD51
NPZ54  RL4VQ  PFHN4  L790L  HREM8  X41ND  APZ14  NXCM4  RTP64  M5HFZ
C3R9Q  4CI2H  XZ481  ØL4YN  PQLMH  RT4PQ  BQRMD  3
```

4. The following cryptogram is suspected to contain the signature "CLINTON COURTNEY COL INF." Solve it and recover all keys.

```
LHQEF  IPXOE  OPYQE  NCOPC  MAGZH  EFQEL  MCOBY  FMEBO  DKDYS  YFQEL
EFDFH  EFQET  OLIEG  GYHLO  SLMLQ  LBPJZ  SPTFA  OFQRL  DQELN  HQEFF
MIBTZ  OFGWJ  FOSLM  DPYFQ  ELBLY  BYFQE  LYFQE  LLMSX  LEBIF  EFDMQ
ECFZN  LQEMF  XSLMT  OLIDY  OADSF  EOSQE  MLEFQ  RQILU  BPSTL  MFHLM
FIKQE  FFYFQ  ELTOL  LGDYD  TLQLB  PKFAO  FQROT  PCQEM  OSQEM  LMBLY
BGCGZ  FQRWO  OLSAP  SKLLM  ESZQR  LKMOE  SPSRN
```

DO NOT WRITE ON THIS PAGE

5. The following cryptogram is suspected to contain the probable beginning "PART ONE OF THREE PARTS." Solve it and recover all keys.

```
NBBVC  QKVHI  EBMMN  BBVII  BDLBK  LSFXV  RKCBV  MKRYF  QTBVR  HVIYP
YBVHB  ODBFT  XEGRR  LWYBV  FYIIK  TUCPH  MWPYF  YWRQA  IBLHZ  VGYUU
YCAVE  GHIRW  UVHPK  RBDDI  SYEAI  ZNTIQ  NYMWP  YFYIK  KTSYU  EQVXP
UVPTF  MRWIP  YCVDD  RFYGB  SMYCA  VEGHI  KGNYN  LTBIV  KRHFG  LFLFG
ABYDP  TIQNY  YFQTB  VIKNY  CMVHP  BPTLP  IZYUK  IEGUU  TQFYB  CYDPB
NRVYV  AYGOZ
```

G. Polygraphic substitution: large tables

(embracing pars. 72 and 73 of Chapter IX)

1. The following cryptogram is suspected to contain the probable word "RECONNAISSANCE". Solve it and recover all keys.

```
SACJJ  YROHT  KPLPD  OCVPS  LNPEG  NRPSP  FPLUQ  TLWPR  CJKLR  NQERO
CVMFS  ELZQZ  RRAOT  HSQPG  TLGLN  RQSUZ  KKKKJ  EMVNL  LUARQ  ESAMW
KKLPS  LAPPZ  QVKKP  BCJJY  RLCJH  ACOAR  BHLLJ  HQTRP  ASSLR  PSLNL
QJQTA  JNLIG  NRWXA  IHIYD  KKJEC  PYOSP  KOFBQ  TQPYP  NZSOA  MDZKR
FPSXP  KFJPR  OEAKC  EASLP  DOPBS  IAXSX  PBLPH  TWXRF  GZQTL  WPRCJ
JYAKH  TJYAA  NNSXC  BROWE  SARDL  LMLAX  AFYUN  CPKMS  NEQJQ  TAJ
```

2. The following cryptogram is suspected to begin with the opening words "WEATHER FORECAST FOR WEDNESDAY ONE THREE SEPTEMBER." Solve it and recover all keys.

```
OLTSH  MXDAW  AUSTZ  FAWOL  QJWDE  MKBSA  ODWEW  DTGYG  KCVEJ  YDYQG
HXFEK  KFPOR  IFODR  DMUOB  MHHGS  KVQVQ  ODQLA  WWRUS  KKDOC  AUSVQ
ODDOY  DVEMY  MHJYD  YMYHX  HMSTY  GEIVU  VQVCT  SKXXD  MXDOC  KSKLX
USMHB  ZLUCF  XWBQV  JOUJD  IWQFU  OUSNO  LHECW  DXJJM  VQAIY  DVEMY
MHVRV  EAPVQ  ZEHTK  WCKXW  BQHZB  RHPCJ  FSTWO  GZXIZ  AMAGJ  DYGFE
MHERV  EOUTW  WDJYD  YRKRG  WFTGA  WQWDO  ODVEJ  YTSWP  OZNTI  WHXYJ
BZKWC  KDQAM
```

3. The following two messages, intercepted on links known to be passing traffic enciphered by means of random digraphic tables, are believed to be isologs containing the probable signature "MAJ GEN CARTER WORTHINGTON." Solve the texts and reconstruct the fragmentary table.

Message "A"

```
CXJIG  ONBXJ  LVOPL  DXGOI  UTLIZ  VDMDE  AYAMC  XBDDZ  JXVKD  QIYIG
JOKWD  EIGJX  BROUL  NSRSC  DEUWQ  KVNLN  ZHYMI  QDWKS  ERAVZ  HLDRD
DEIQO  FQQHT  OFVBD  EPCJI  GSXJN  ZNNIG  OFNBS  RZHJU  TIAAG  PDZGP
XFDEK  WFHWX  MLPYR  NAYAM  ERAJU  ISXOW  UWQKV  NLNZH  YMAVH  WOWSC
JXOFQ  QMOSR  AYNRD  ZCOIS  SRZHH  TVFIQ  VNFHT  QUTHT  EXEVX  GIYIG
OFYRJ  LOFOF  IAIGB  TMOSR  AYNRH  FWXGD  PXOLC  OENSG  SLMGC  XIDVM
OPIHL  NZHTM  SLSCO  FGLIG
```

DO NOT WRITE ON THIS PAGE

Message "B"

CXJIG	ONBXJ	LVOPL	DXGOI	UISCX	JNZNN	IGERA	JOFZH	WCPMS	RAYNR
KQRIB	RLDYM	JOEDX	GTICO	VUQFU	OKMLI	JXENM	XQFNE	SCSRW	XLHHK
ENFWD	TSLDR	LVCOW	GZTIG	NUDMD	EAYPS	XQXQT	ILHYO	CXVFO	FJLOF
COHII	QPCDZ	CHNNI	YIGWG	CXBDK	CUTFJ	QFUOK	MLIJX	VUINX	SLIUW
RIDTS	LVFIY	IGIGS	CDEVF	TIJXF	LLNKJ	RTERS	XOWJI	AJJXS	RAYPS
NRQVG	PPSLV	JIDTE	EVFIY	IGSLD	JPSKS	JOSCD	EIQCH	LIERA	JVFTF
WXZZL	IJXFG	LICOL	VCXCC						

4. Solve the following cryptogram and recover all keys:

00331	50971	14347	26106	72173	14241	19506	23501	15006	45106
21410	72472	18919	20184	00189	06731	50975	30403	11539	32626
09192	35611	51861	22467	21207	10742	37235	11441	61153	93271
05529	31865	52009	06247	11411	50193	18651	61842	66274	47000
22380	53186	14107	22365	16189	00419	51914	79067	00839	70801
37105	18939	12623	43408	13327	30710	84274	40036	72234	03186
21152	40082	92011	12239	3284					

5. The cryptogram below is suspected to begin with the stereotype "REFERRING TO YOUR MESSAGE" or "REFERENCE YOUR MESSAGE." Solve the text and recover all keys.

MRADM	TGCIY	IYMFG	NNLSR	KQFBD	MDWII	DSZGN	MIJAG	OOLGI	DEVLT
DGCII	YDLCI	MMTJI	UPNMV	ZPLGI	DMYIT	IPOVG	IPTGO	PLMMC	HJPBM
RCDGK	FWJIH	CEEFM	DODSZ	TENDG	KFWIN	NMLEV	EZFTA	SDIPH	MTTDL
GTRQM	DMZUR	ODNPC	JNJGC	IIQMU	ZKLIY	NJNCW	QMZFV	ODNWG	PRGNL
CURPM	IADGI	VRGNR	FURVP	IFDUJ	TDLPO	JVRTD	AZMRI	IFXDG	GDHVV
ZPIQM	EMFJP	CSMKJ	LMMND	PQWYZ	BOZNI	JYIPJ	DMDYJ	PNIPE	MFJPC
SMKJL	FENGN	PWFNV	FJJIW	IGTTM	OTEOW	CRVWL	FELET	SZTNM	VRSMT
RTEQV	RVQJQ	MRVNO	VGIPR	MTKQX	GCJEL	CMZHP	RTLNM	LCRIY	RCZYG
PWXPA									

6. The following cryptogram is suspected to contain the probable word "AIRCRAFT." The encipherment is believed to involve a tetranome trigraphic system employing a matrix similar to that illustrated in Fig. 59 of the text, the ciphertext sections being composed of the dinomes 00–99 inscribed in the normal manner of writing, but the plaintext sections consisting of keyword-mixed sequences which differ from those in the text example. Solve the cryptogram and recover all keys.

06017	84949	12253	31747	60314	27082	40187	76111	02401	24548
27123	65681	68314	21499	45187	58917	33094	14388	43534	27719
15172	77432	49450	74872	60111	26641	45732	76760	53457	61959
45734	93043	77092	81943	22012	93875	27133	04690	40077	56561
29112	50735	76150	54214	98753	57002	08658	36161	32472	10717
77391	82533	17470	26419	56124	24707	77723	29634	91057	60779
55141	14531	43856	92871						

DO NOT WRITE ON THIS PAGE

H. Monoalphabetic substitution with irregular-length ciphertext units: monome-dinome systems and others
(embracing Chapter X)

1. Solve the following monome-dinome cryptogram and recover the original matrix:

78131	76784	31174	50078	76343	47807	41346	53334	01331	01799
78318	76441	31917	92478	74179	10834	76033	55723	40178	31347
46554	65323	41305	86131	34767	30345	77787	48763	77689	76072
76747	88123	11278	31788	76503	47753	17807	67921	07276	07310
17997	88878	74703	05323	15777	71034	76371	33764	47117	37607
88390	00666	33300	03985	79531	31533	78342	47800	17230	75560
34850	74547	83189							

2. The following monome-dinome cryptogram is believed to contain the probable word "DIVISION." Solve the text and recover the original matrix.

17832	00066	16927	80635	28420	04596	95220	01900	21500	40563
26746	12576	80705	88123	53921	31118	13281	29159	46465	61576
52844	90033	94526	59400	25284	30032	00457	80758	80707	00526
73941	20854	56640	59352	91625	97612	46977	89125	05945	22008
41401	51129	31702	91067	53763	59062	38071	67003	84670	04267
78579	20084	17919	60266	43595	65697	00036	12004	97616	87202
60045	70787	05971	26122	81200	19003	00841	76912	09599	72673

3. The following cryptogram was intercepted on a link which has been known to be passing traffic in two different monome-dinome systems, one involving a matrix of the type shown in Fig. 75 of the text, the other involving a matrix of the type in Fig. 77. Solve the text of the message and recover the original matrix.

47631	82870	14628	31274	12741	16263	16054	63152	84662	60736
97728	46198	46972	13808	46287	46364	83788	72846	60846	28738
27578	87073	18279	62736	97462	83107	36977	45636	26962	73168
62763	12138	08462	87316	06379	82647	28467			

4. The following messages, intercepted on a link known to be passing monome-dinome traffic, are believed to be isologs. Solve the texts and recover the original matrices.

Message "A"

94872	33935	61227	89316	23405	09079	43810	57678	93386	41999
83809	08334	94194	76279	99496	30576	79199	54343	57683	04186
07981	43349	83529	09638						

Message "B"

94378	11935	62887	39326	81405	09079	41320	57673	93136	41999
81309	03114	94194	76879	99496	10576	79199	54343	57631	04136
07982	43149	31589	09613						

CONFIDENTIAL

DO NOT WRITE ON THIS PAGE

5. The following messages are believed to be isologous monome-dinome ciphers. Solve the texts and recover the original matrices.

Message "A"

73507	09885	01652	37531	09804	39858	14983	12316	52371	12890
93312	42689	30741	59012	54398	50563	98460	77297	30415	65075
43098	13500	74379	06814	51983	12316	52371	13559	33124	39842
16361	80772	97056	29092	58145	15465	07901	10121	98617	56398
94163	84731	35039	04398						

Message "B"

36713	45807	18921	63867	55406	58179	56296	89216	37798	07485
62909	18085	43072	74292	56571	84650	14339	73640	72171	32564
58871	43063	74180	79875	62968	92163	77676	85629	06509	89612
34339	73484	97424	81798	72517	13747	74292	78017	08465	26896
80036	88716	74065							

6. Solve the following monome-dinome-trinome cryptogram and recover the original matrix:

61745	04120	43950	43238	65332	06382	01503	20682	61661	20436
53513	17150	68412	19203	16204	38543	12043	20150	35350	12335
45039	44171	20186	50929	78509	23850	46204	84739	45049	62065
82820	43532	01561	93231	65184	71533	53842	04541	62453	32043
85421	68564								

7. Solve the following uniliteral-biliteral cryptogram and recover all keys:

PVOYA	CKRTE	AUOOD	KNWOI	BKEIA	UBTAP	WOIDG	OBKNT	AENXB	TAEBG
YAEUI	ENLCT	EOBZF	HOOBL	YIEBG	UUONT	BXPXR	MIBKA	CWOIE	PKCGP
VAYEF	TEINM	PKSGE	YAODK	UEDLR	ZEYAN	GCWUY	AUPKP	MEOIA	CVPWY
RWOYC	WAPWO	IYAOR	WSVCH	EIRVC	KYYPK	OICKY	NWODH	RKDGE	AEBXU
ERXDM	EYABT	EUCWN	GRTDW	PHOAO	PGUNG	RKCVY	ONZBG	UENTX	

8. The following cryptogram, enciphered in a Playfair-type digraphic-monographic system, is suspected to begin with the probable stereotype "MORNING REPORT FOR MONDAY NOVEMBER TWENTY FIRST." Solve the text and recover all keys.

AQTIN	JFQHQ	PTLGP	TAQSK	IVATX	CJEHQ	PZKMR	ZGHYN	PNPPQ	QTDMK
MLRGP	TBWRZ	PZPRG	LVTPG	GAHHQ	MPGAY	QMHMF	KRRKQ	HQMKM	RJNPH
EJCMD	KZYSR	KQBCA	KQRYQ	MCQGG	AHHQN	PRYQM	QXGLV	QHJTN	MQKPD
AHCTM	KQVGG	AHHQT	AKQVP	KMRJN	PHEJC	MDKZY	SRKLV	LOCMX	CXKTP

DO NOT WRITE ON THIS PAGE

9. The following cryptogram was enciphered in a dinome-trinome digraphic system employing matrices similar to those in Figs. 90*a* and *b*, except that the internal numerical sequences have been changed. The message is suspected to end with the signature "VINCENT ANDERSEN COL INF." Solve the text and reconstruct the matrices involved.

71665	73330	13492	25221	39225	86765	01802	60940	44263	12514
47303	60733	96104	70273	72027	53072	85735	39518	42301	07824
22132	71923	51903	51663	92569	09402	78709	40353	01078	21946
95755	85962	42213	27197	65187	26752	74097	55734	86919	61182
81051	02719	85196	57392	20085	32536	75171	92577	63494	35234
45067	19349	22522	04714	41045	22216	57508	77537	16223	93144
24586	34944	82506							

10. The following cryptogram, based on a Morse code system, is suspected to begin with a spelled-out number. Solve it and recover all keys.

71430	62809	18592	35607	61572	04953	79012	87548	65983	04037
95327	30751	34904	56564	20813	01258	16408	97156	64597	60410
83159	34702	68032	95357	25173	02589	41582	60360	91754	

I. Syllabary squares and code charts
(embracing par. 80 of Chapter XI)

1. The following cryptogram is suspected to begin with the words "REFERENCE YOUR MESSAGE NUMBER THREE FIVE FOUR DATED ONE SEVEN SEPTEMBER." Solve the text and reconstruct the cryptosystem employed.

CRDSC	RJSIS	KSJTY	CLSKR	CQERB	YJQIR	KPKOJ	RCRSD	TFSDJ	TYCDP
FRDXB	RCSFS	JSCSJ	UFRJQ	IRKPK	UPDDO	FOXBP	LCRRB	DUKPI	SLSJS
KOKRG	OJNLP	KPCOF	SPFJP	CRRBF	TJNLR	AYJSH	TFOXB	QCKRH	TESDY
LRHUK	UPCFO	JNLRC	NLUFT	JSEOJ	SFXCU	YFESK	PJSHX	BRPFB	XXBFY
SCDPK	SKUCR	JWOCS	LRBKO	KRFXJ	ULPBR	QCKSK	VBQCP	JPCPJ	UJRCS
OEJRF	PJSHU	BOBRK	WPCRI	HVOBJ	PKRFY	DORDD	UKPDT	BPLPH	TFOXB

2. The enemy is using a system incorporating a 10 x 10 bipartite square within which there are inscribed letters, digits, and syllables. The row and column coordinates are invariable, but a different internal chart is used each day. The chart for 16 December was reconstructed and found to be based on the key word "PYRAMIDS" as follows:

	Ø	1	2	3	4	5	6	7	8	9
Ø	P	Y	R	RA	RE	RED	RES	RI	RO	A
1	1	AL	AN	AND	AR	ARE	AS	AT	ATE	ATI
2	M	ME	I	9	IN	ING	ION	IS	IT	IVE
3	D	4	DA	DE	S	SE	SH	ST	STO	B
4	2	BE	C	3	CA	CE	CO	COM	E	5
5	EA	ED	EN	ENT	ER	ERE	ERS	ES	EST	F
6	6	G	7	H	8	HAS	HE	J	Ø	K
7	L	LA	LE	N	ND	NE	NT	O	OF	ON
8	OR	OU	Q	T	TE	TED	TER	TH	THE	THI
9	THR	TI	TO	U	V	VE	W	WE	X	Z

DO NOT WRITE ON THIS PAGE

The next day the following cryptogram, suspected to contain the word "CROSSROADS", was intercepted. Solve the text, reconstruct the chart for the day, and determine the internal key word.

```
20885 05895 63623 06969 94155 89207 84736 03577 95613 87830
50669 41544 84621 21220 38314 26767 93520 18310 32027 91036
84729 44468 79547 86950 58955 87962 14305 07194 35877 99758
02021 07847 22975 88809 84530 30193 44787 97978 30507 95755
59540 85494 72897 26082 72944 49207 79820 51079 94101 00258
28975 80202 43690 59650 93505 89563 58071 12150 71943 58427
30548 75035 68
```

3. The following two messages were encrypted by means of the Aggressor code chart illustrated in Fig. 100 of the text, but with a different set of coordinates. It is suspected that both messages end with the signature "ENRIQUE VALENZUELA." Reconstruct the plain texts and recover the coordinates of the chart.

Message "A"

```
ATFHX LMKZU ALCOT GQTKQ RDKNK LELDR JPXTR OGYRJ UASGB WATHX
ODDVI XSRBD LICEX LJAVL TAETU YNVQL KHWDA LYGPT DILGY RLUEF
NYZRQ QLXSY UV
```

Message "B"

```
ALLQQ TMBVU ELUAT XXLPQ RJWGI MGGXC ATDRN PQTJB WDKZE QLAUG
WJALY XPLOZ SRLZA GYGXH FGYYU P
```

4. The following message was encrypted by means of the Aggressor code chart illustrated in Fig. 100 of the text, but with a set of *nonvariant* coordinates. Reconstruct the plain text and recover the sequences for the row and column coordinates.

```
EOBOJ OBOEO IOAYI TJODO EOAZJ OIQAU KUEOF QAQJO JZJQJ YEOJS
JOJTJ SEOAX ANAQL XJOBO FUDUI PEOHQ BQGQJ PASIR CZATJ OBWEO
JSJOH ZJPIZ HXKVE OFVJO LREOF VDTDS LSETL TJSKS BWJXE TMWJX
```

5. The following 30 messages, intercepted on low-level ground nets, have been selected for study because of homogeneous characteristics. Solve the texts and reconstruct the cryptosystem employed.

No. 1 INL DE ETN 1630 08 Aug 53

```
77051 93945 39891 95925 49323 52235 90004 59560 81550 59032
62009 20878 95503 60579 60925 89095 42930 52788 65002 1
```

No. 2 INL DE ETN 1633 08 Aug 53

```
77051 91945 39896 90925 49323 52235 97459 56037 62051 05011
00617 89507 05883 19513 05291 98945 95204 943
```

DO NOT WRITE ON THIS PAGE

No. 3 INL DE ETN 1636 08 Aug 53

77051 96945 39896 90925 49323 52235 97457 33215 55957 40582
92462 19550 70650 54595 28056 24107 11104 58861 95490 590

No. 4 SIA DE BYE 0805 10 Aug 53

77050 98945 09490 90925 39093 23522 35974 59599 00880 56068
90953 40578 32350 42392 85619 54605 90

No. 5 LQS DE QEI 1836 11 Aug 53

77057 92945 39496 90925 39393 23522 35900 04595 87055 67332
15559 53705 82924 67895 91052 70628 66009 18053 24

No. 6 ALU DE YUT 1644 12 Aug 53

77056 91945 39897 90925 39193 23522 35974 59522 05804 69624
95760 52795 08673 98205 45955 07505 71260 01834 2261

No. 7 QEI DE INL 1256 20 Aug 53

77051 99945 39195 90925 19093 23522 35974 59520 68054 62231
72567 25780 50905 15437 95460 52795 60970 59088 31555 035

No. 8 USO DE SIA 0917 21 Aug 53

77053 93945 09993 95925 19393 23522 35974 59525 05602 35295
68052 19510 05925 29078 38483 51000 95970 590

No. 9 OAY DE NBQ 2357 27 Sep 53

77059 96945 19695 90925 19299 68093 60174 04595 98055 78040
74630 09576 05279 50815 05468 01055 95398 2

No. 10 NBQ DE OAY 0010 28 Sep 53

77050 93945 09090 95925 19499 68093 60174 04595 50155 20510
50119 85909 09950 17819 05279 52169 05494 98964 94143 076

No. 11 OAY DE ALU 1008 29 Sep 53

77053 97945 39090 90925 19999 68000 52601 74045 95990 56234
55503 59523 05203 53106 60380 69541 05927 40179 04958 3

No. 12 USO DE TOB 2219 29 Sep 53

77056 99945 19193 95925 19999 68000 52601 27645 95606 33105
56219 55024 65054 59504 05560 19580 38930 52722 46639 527

No. 13 ETN DE SIA 2347 30 Sep 53

77051 93945 19697 90925 69099 68093 60127 64595 71081 40550
35002 79585 05371 79586 66110 54595 88058 72490 17569 624

~~CONFIDENTIAL~~

DO NOT WRITE ON THIS PAGE

No. 14 TOB DE BYE 2142 30 Sep 53

77054 99945 19396 95925 69099 05280 93601 74045 95530 50195
11605 70536 21229 54505 25370 46317 95860 50162 43909 09545

No.15 BYE DE YUT 1818 10 Oct 53

77058 90945 39493 90925 39095 06201 17404 59561 05791 95995
12057 24595 10059 27495 29054 59582 05789 74035 62005 490

No. 16 SIA DE USO 1629 17 Oct 53

77057 95945 39891 95925 39295 06201 12764 59544 60660 51390
95320 55636 21229 57513 05795 94964 94952 16905 59991 9

No. 17 ALU DE LQS 2001 19 Oct 53

77056 92945 39995 95925 39995 06200 50174 04595 02053 39541
05678 05000 90105 01198 78955 00905 45886 19535 05147 21055

No. 18 LQS DE YUT 0905 12 Nov 53

77053 98945 09990 90925 39192 60860 17404 59518 26059 09571
05560 19092 75664 08840 90169 56905 27190 33262 30958 0

No. 19 SIA DE ETN 1314 14 Nov 53

77053 98945 39693 90925 39797 05008 60127 64595 26350 55275
42323 59098 45953 34805 61459 58805 60689 09260 80761 69208

No. 20 INL DE SIA 1116 25 Nov 53

77051 97945 39393 90925 19592 61552 60174 04573 32155 59530
31030 67210 05927 44595 16057 25740 35809 89506 84970 590

No. 21 YUT DE BYE 1348 04 Dec 53

77055 98945 39697 95925 79671 36017 40459 53147 05459 53205
56019 13052 95610 54969 09459 51606 05198 39533 05470 144

No. 22 OAY DE INL 1956 06 Dec 53

77057 90945 39995 90925 89725 21360 12764 59560 81401 20572
96084 03695 58059 07895 90059 02892 72459 52805 26406 036

No. 23 QEI DE YUT 1710 07 Dec 53

77056 91945 39290 95925 29676 25260 17404 59551 32053 79559
05019 54605 63950 85705 45956 02805 71260 01834 2261

No. 24 EIB DE BQT 1207 05 Jan 54

77056 90945 39190 90925 59023 43531 55459 57260 05035 05080
95170 55497 92101 09589 05567 89519 05767 09662 0140

~~CONFIDENTIAL~~

DO NOT WRITE ON THIS PAGE

No. 25 NLE DE IAQ 1519 11 Jan 54

77057	92945	39593	90925	39390	23270	35315	54595	92415	40089
05508	06335	24921	09596	05262	19566	05866	09298	3596	

No. 26 QSN DE TNY 1034 12 Jan 54

77051	96945	39096	90925	39190	23435	32405	54595	50815	61205
72499	55805	90959	00516	09890	90988	31810	05436	80313	0

No. 27 YEO DE UTA 0932 23 Feb 54

77053	98945	09991	95925	19698	25212	40353	15545	95603	77305
72953	90566	35247	61400	00765	58292	46956	45057	0590	

No. 28 SOL DE AYS 1440 13 Mar 54

77051	93945	39796	95925	39696	03118	45959	25705	90805	39093
44953	93005	12506	01256	52311	05595	92056	02370	56	

No. 29 LUI DE OBU 1846 30 Mar 54

77056	90945	39497	90925	69096	03162	42459	50883	00719	40501
95532	70562	43909	05637	77321	21092	94953	80573	3272	

No. 30 BQT DE UTA 0503 31 Mar 54

77050	93945	09590	90925	69398	44062	42459	57905	01952	2

J. Miscellaneous monoalphabetic substitution systems; concealment systems
(embracing pars. 81–84 of Chapter XI)

1. The following two messages were intercepted on the same net during a morning offensive. Solve them and recover all keys.

Message "A"

ILPUR	PSBOK	YBDGY	YNCNP	VXHCJ	ROKFI	NPXVP	EWHLX	KWYXQ	BVYDB
YYZCX									

Message "B"

RBVYW	BYYZC	THEEB	MMEZI	ABXHW	SYXCK	NNBYM	SXQDY	AEPCX

2. Solve the following cryptogram and recover all keys:

GHUIP	GJHNI	MBSPR	IUSNS	SHPGB	RQSAJ	OOQRS	SIPQS	MHVJI	PQIVV
JFKIH	GSANY	PTKKD	YJKQX	ZVNPR	OQNJS	NUYEF	ENPZV	NPROQ	QTKLN
FFHON	QUPOQ	HUQTV	OTNDO	JTKLP	OWOJT	EJWNQ	EKJAE	IOUTI	QQGTB
ANNKZ	XVUKR	EMMQR	RTEII	KVPXV	UKRRE	MWKIK	VPXVU	DAEIP	RSQII
GWQLQ	UXVVX	VUATG	WQQVP						

DO NOT WRITE ON THIS PAGE

3. The following cryptogram is suspected to contain the word "REGIMENT." Solve it and recover all keys.

BUFWW	HEAGH	MITAJ	JSKLS	MNHIU	YUQAI	AUMQM	UWIZU	VOJFH	GUFMD
GQYKQ	LLSUI	YKCGS	WUFQU	UQKCF	IWHQK	OGSWI	YJYOY	SKXFQ	IADQB
YKJYS	VCLLC	SSHJW	YMJYS	XWYKB	JUKKU	PTPWI	IUQHK	XHYKB	JYCOF
YQENV	QICUY	BIRRW	PIYYN	PSWJE	PMIOP	UQARH	KYQCG	WPGEP	GAEPL
VGFHJ	FBBLZ	USSXS	RMYFQ	XTSHD	TTDYC	FXFQH	HLPKT	GUDXC	BTFJO
WHAMV	UWNHG	HMIQQ	NZHKF	JHKIU	SWWAL	HSIUM	HHAEZ	IIZWH	HPHWU

4. The following cryptogram is suspected to contain the word "BATTALION." Solve it and recover all keys.

RDITJ	OGDPF	QUMTD	IKRUI	ULHKI	MASKI	JMHQD	BBUGI	AUYBU	ILQUM
PDTTD	FKQUD	BBUGI	KFBKV	UQUIK	ZKQUH	KTDHT	ECMKI	KFBAQ	IKHIU
PDIKQ	KCLUI	KODFI	AULUW	TERDI	CDIDR	LUJQU	MJMVU	IRKET	IKHQA

5. The following two messages are suspected to be isologs. Recover the plain text and reconstruct the cipher alphabet involved.

Message "A"

QMSLG PWJRN NPHRE ENQOH MABGR PLNWJ RNNVH MAYTR GSLAP HMHXR
EFRGB ILNPH

Message "B"

ETOSR DMOTA DPCON NORXD TARSU TWGAS UASAS NYEXL IAHRI PLNSA
OFGSB EOELT

6. Solve any five of the ten innocent-text messages given below.

a. TO COVER AND HIDE YOUR FINANCES, I MAINTAIN STAFF TO BARTER FRANC EXCHANGE.

b. BEARER IS A FRIEND. I CANNOT WRITE MUCH. WILL YOU BE READY TO MEET A TRUSTED INTERMEDIARY SO YOU CAN GIVE ALL NECESSARY DOCUMENTS TO HIM NEXT WEEK? I HAVE NOTHING MORE TO SAY.

c. The following is an actual radiogram which was held up by a censor:

ES100 81 RADIO—WASHINGTON DC 210P JULY 1 1941

ROBERT C JOHNSTON III
CARE HELVETIA PALACE HOTEL LUGANO (SWITZERLAND)

MATERIALS ADEQUATE. CONTRACTS YOUR ASSISTANT RETURNED TO ME WERE OKAY AT CONSULATE. INSURANCE EXPEDITED QUOTATIONS ON CARGO RATES ALSO AS OF TOTAL POOLED SHIPMENTS. BECAUSE PRESENT HOLDING OF SHARES IS LINKED TO UNDECLARED ASSETS ATTENTION IS DIRECTED TO PROGRAM OF DENUNCIATION BY AGGRESSIVE MINORITY CHALLENGING DIVIDENDS. REALTY VALUES PREDICT SUSTAINED EXCEPTIONAL PROFITS HEREAFTER WITH DISCLOSURES AT MEETING ON SEVENTEENTH BREAKING UP RENEWED OPPOSITION.

REMSEN

d. The following message, poorly spelled, is shown just as it was written by the originator.

DISATISFIED WITH IMMIDIATE RESULTS OF YOUR AMALYSIS STOP ADITIONAL RENUNERATION TO ALL PERSENNEL WHO INDENTIFY COMPONANT PARTS OF THE ALLOIS OTHER THAN NICKEL OR COPPER.

e. CATALOG VALUES CANCELED UNLESS OFFER STRENGTHENED. COMPLEX SELECTION EFFECTIVELY STALEMATED. PSYCHOLOGICAL FOUNDATION UNWARRANTED.

f. REQUEST CONSIDERATION OF ONLY THE MOST USUAL PLACEMENTS OF ORDERS FOR BALANCED BUSINESS TRANSACTION. LET NOTICE BE GIVEN FOR A SUBSTITUTION AS SOME ARTICLES NOT EASILY CHANGED. PARTNER AGREEMENT IS EXTRA HELPFUL WHEN MAXIMUM.

g. The secret text in the following message was solved by Sherlock Holmes in one of Sir Arthur Conan Doyle's stories.

"The supply of game for London is going steadily up. Head-keeper Hudson, we believe, has been now told to receive all orders for fly-paper and for preservation of your hen-pheasant's life."

h. In April 1941, the French (Nazi-controlled) newspaper *Paris-Soir* printed the following poem by an anonymous contributor apparently extolling Adolph Hitler. A fairly literal translation is given beneath the original.

Aimons et admirons le Chancelier Hitler
L'éternelle Angleterre est indigne de vivre
Maudissons et écrasons le peuple d'outre-mer
Le Nazi sur la terre sera seul à survivre.
Soyons donc le soutien du Führer allemand
Des boys navigateurs finira l'odysée
A eux seuls appartient un juste châtiment
La Palme du vainqueur attend la Croix gammée.

(Let us love and admire Chancellor Hitler
Eternal England is not worthy to live
Curse and eliminate the people across the sea
Nazidom on earth will be the sole survivor.
Let us therefore support the German Führer
The seafaring boys will finish the odyssey
To them alone a fitting punishment
The palm of victory awaits the swastika.)

DO NOT WRITE ON THIS PAGE

i. The following telegrams were filed by the same originator to the same addressee on successive days.

Message No. 1

MONEY SENT TO NEW HAVEN IN BANK ON ARTS ACCOUNT. HE MAY HAVE DRAWN ALL AND GONE HOME BEING SO BORED. AS ALWAYS HE NEEDS A LOT OF CASH, YET IS NOT ABLE TO DO MUCH WITH THE MONEY SENT TO PAY HIS LAB SCHOOL BILLS.

Message No. 2

SENT FORTY DOLLARS BY CHECK TO HENRY AND ALICE AFTER THEY PAYED ALL ARTS OLD DEBTS STOP. I REALLY WANT AGREEMENT WITH THEM AND WE MUST MAKE ART STOP CHARGES AND BILLS ABOUT TOWN. SEND NO MORE CHECKS TO HIM OR CASH TO WASTE ALWAYS FOR HE SPENDS IT.

Message No. 3

MOTHER SENDS LOVE AND WAITS EACH LETTER AS SHE ALWAYS DOES, SO WRITE AS MANY AS TOM AND ELOIS DO TO HER. AS ALWAYS, JERRY.

j. The following innocent-text message is suspected to contain information on ships sailing from Boston.

AT NINE ELEVEN THIS MORNING FOLLOWING CONVERSATION SURPRISED REPORTERS: "RUSSIA, FRANCE AND ENGLAND ASSURED PREPARATIONS WERE NOW RAPIDLY ATTAINING SATISFACTORY ALLOCATIONS. HOWEVER, THEIR ALLOTMENTS VARIED SLIGHTLY, AND IN AGREEING PROBABLY WERE NOT SURE OF THE OUTCOME. UNLESS TERMS ARE QUICKLY DETERMINED, MANY SPECIFIC AGREEMENTS WITH EACH OTHER WILL HAVE TO BE MADE."

INDEX

EXPLANATORY NOTES

1. This index contains detailed entries from the body of the textbook and from prose passages in the appendices. An attempt has been made to suppress index entries which would lead the reader to nothing more than a definition of a term; if the reader desires only a definition, he should refer directly to the Glossary, Appendix 1.

2. The entries contained herein have been indexed both by paragraph number and page number. Under the heading "Paragraphs", the entries from the *body* of the textbook have been indexed down to sub-subparagraph wherever possible. Entries from the *appendices* are indexed at least by appendix number and, wherever practicable, by paragraph number within the pertinent appendix; for example, an entry from paragraph 6 of Appendix 8 is indexed in the "Paragraphs" column by the notation "App. 8: 6." Entries from *footnotes* are indexed by showing the paragraph location followed by a dagger [†] and the footnote number; for example, footnote 12 out of subparagraph 60*o* is indexed by the notation "60*o*†12."

INDEX

~~CONFIDENTIAL~~

LIST OF CHARTS

U. S. GOVERNMENT PRINTING OFFICE: 1956