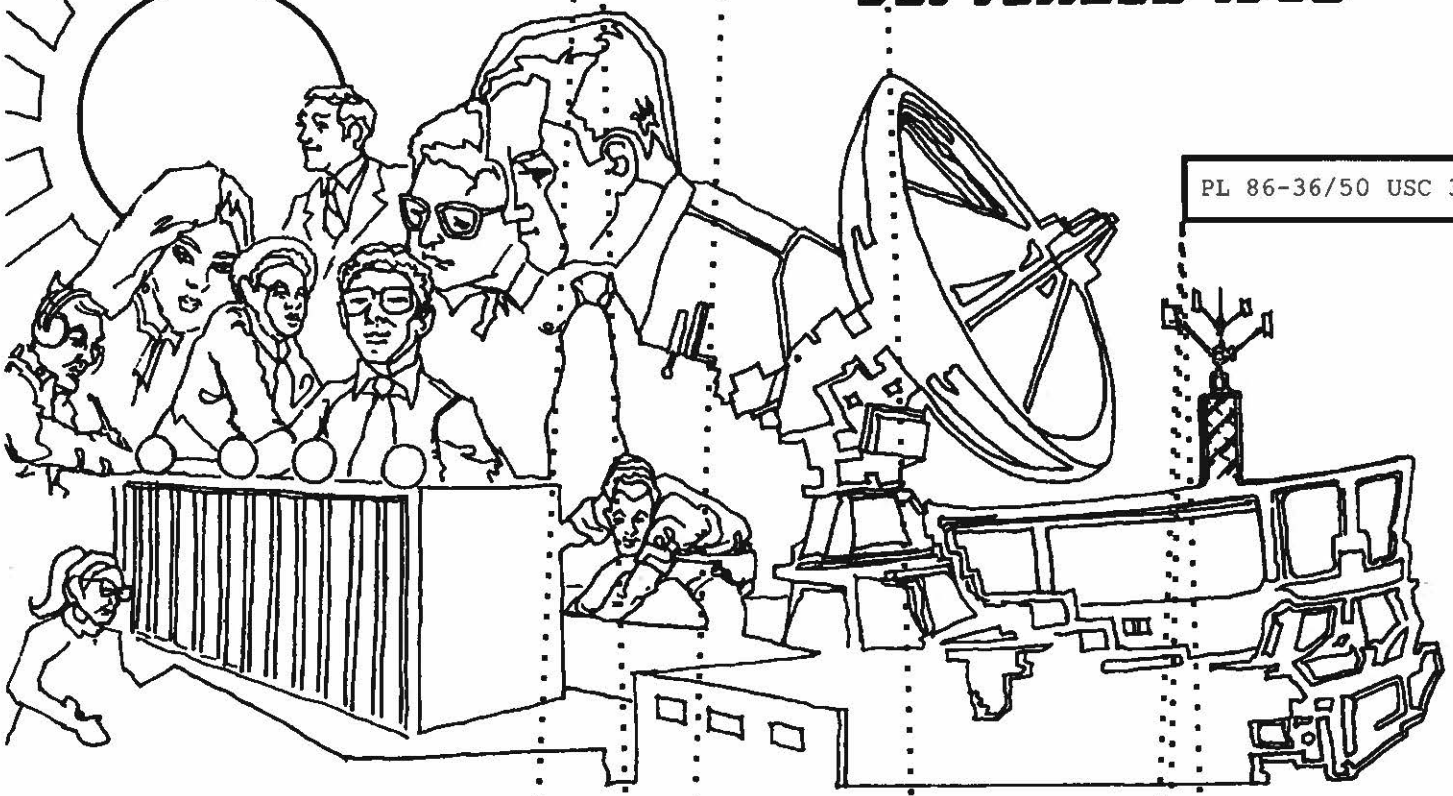
Non - Responsive

~~TOP SECRET~~

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND

CRYPTOLOG

SEPTEMBER 1976



PL 86-36/50 USC 3605

~~THIS DOCUMENT CONTAINS CODEWORD MATERIAL~~

~~TOP SECRET~~

Classified by DIRNSA/CHCSS (NSA/CSSM 123-2)
Exempt from GDS, EO 11652, Category 2
~~Declassify Upon Notification by the Originator~~

~~TOP SECRET~~

CRYPTOLOG

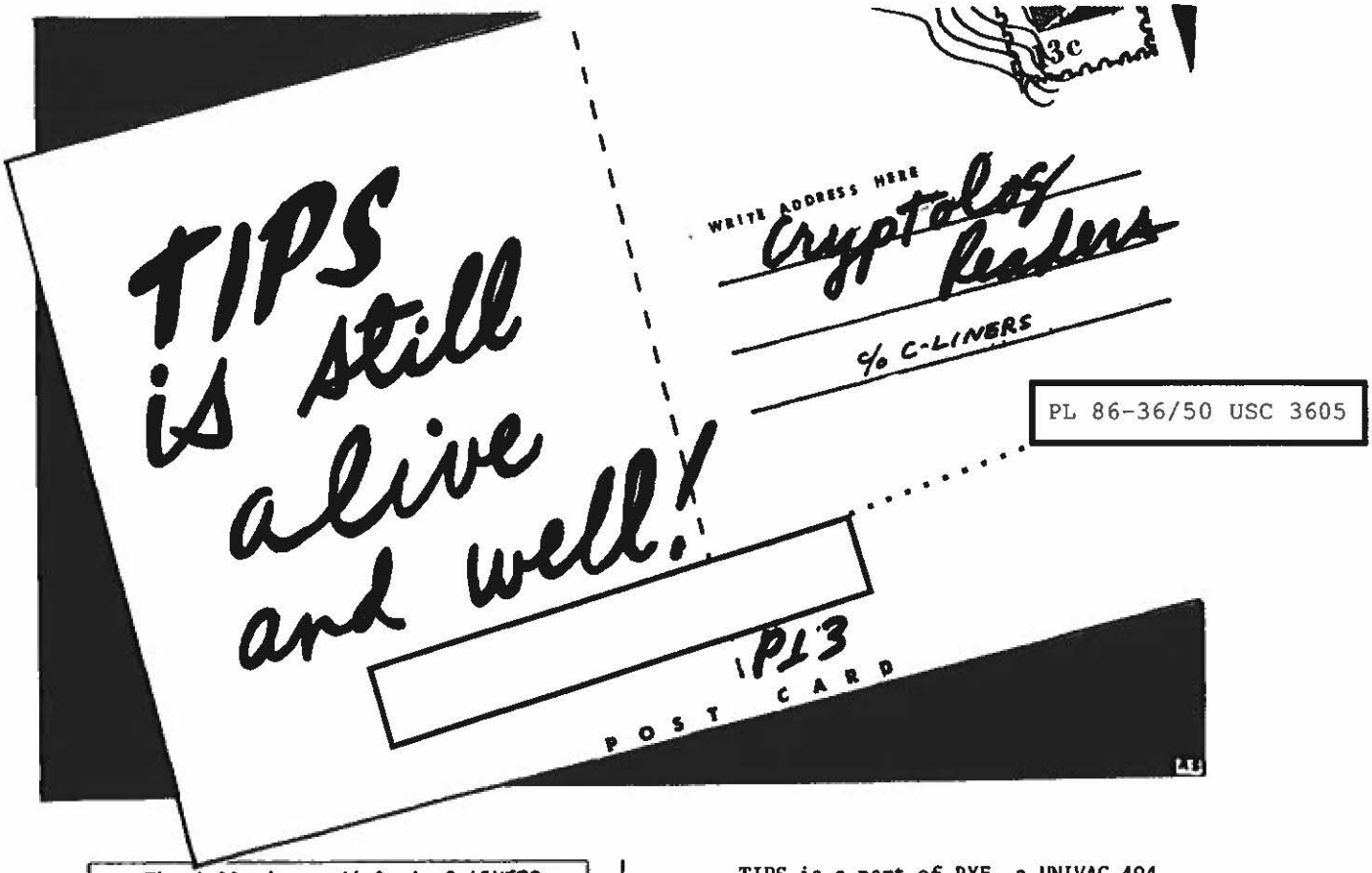
Published Monthly by P1, Techniques and Standards,

for the Personnel of Operations

VOL. III, No. 9 SEPTEMBER 1976

~~TOP SECRET~~

Non - Responsive



The following article in C-LINERS (C Group Machine Processing Information Bulletin), Vol. 3, No. 9, Issue No. 30, Spring 1976, is reprinted here with the kind permission of C-LINERS Editor David J. Williams.

As was the case with Mark Twain, reports you might have heard about the demise of TIPS -- NSA's Technical Information Processing System -- have undoubtedly been grossly exaggerated.

True, TIPS is showing signs of aging. After all, she has been around since the mid-Sixties. To some, such longevity should qualify her for some kind of geriatric support. To others, notably some folks in C11 (the Information Systems Division of C), the old girl is still very much alive. Admittedly, a young and more glamorous replacement is being sought. Nobody knows when this rival will be embodied (or "em-machined") but she is coming and plans are being made for her arrival.

At this stage in her career, then, one feels it would be a good time to record a few random thoughts about TIPS. A few words of background information may be in order for those not in the category of C Old-timers. First, a more or less official definition:

> TIPS is a part of RYE, a UNIVAC 494 remote-access, machine processing system at NSA. The TIPS system encompasses the hardware devices, software executive routines, conventions, communications package, and data bases in support of the quick-turnaround, on-line, information storage and retrieval capability within RYE. (See Section 4, of forthcoming USSID 703, Technical Information Processing System (TIPS), for general information about this system.)

Under the TIPS concept the data files are stored on CDC-606 tape drives. They are connected to Honeywell DDP-516 mini-computers, which execute the queries submitted by the various users. A query is a series of simple statements written in TILE (TIPS Interrogation Language). Normally, it might consist of just a retrieval command and a display statement. Queries are entered remotely from a teletype terminal, or some other peripheral device, like a Raytheon CRT or a Bostic paper-tape reader.

Chances are that, as a RYE user, you've already "interfaced" with the most common of these input devices, the lowly teletype. The manufacturer is the Teletype Corporation of America, and the most common terminal type (for RYE) is the ASR (Automatic Send and Receive)

Model 35, or simply "Mod 35" to the old RYE hands. (Incidentally, you might well impress some of the data systems newcomers at the Agency by reminding them that all teletypes are teleprinters but not all teleprinters are teletypes -- especially if you've heard them talking about "xeroxing" their output, with a small "x"!)

"ASR" means you can input a paper tape in the reader while at the same time receiving page-print at 72 characters a line. The reading speed certainly isn't the fastest by today's standards -- 10 characters per second, or about 100 WPM. By contrast, the Bostic units (each controlled by an ASR-35 terminal) can input at a rate of about 300 characters per second and punch paper tape at a speed of 105 characters per second.

The Mod-35s are flexible, however, and easy to operate. There are nearly 100 of them in operational NSA spaces (including FANX) hooked into RYE. This figure doesn't include the units belonging to C system organizations. Note also that it doesn't include the considerable number of Mod-35s connected to TIDE (Time Dependent System) housed in a similar U-494 main frame. The breakdown for ownership of the RYE-connected units is as follows:

OGA

A - 24		G - 13	
B - 12		P - 2	
C1 - 2		R - 2	
C5 - 2		V - 8	
E - 10		W - 15	

The above list doesn't include the non-NSA, service intelligence organizations, i.e. AFINAR (Air Force Intelligence and Research), USASRD (U.S.A. Special Research Detachment), and NFOIO (Naval Field Operations Intelligence Office), each of which has a terminal and can access selected TIPS/COINS files.

Remote SIGINT users, like the [redacted] and the USAFSS at Kelly AFB, Texas, have terminals connected to TIPS. They operate like regular TIPS customers, i.e., they are linked directly to the RYE master machine, and from there to the TIPS data bases to which they have been given access. [redacted]

Incidentally, if you have come aboard the Agency fairly recently, you may be confused about the relationship of TIPS to COINS. TIPS is the NSA mode of the Community On-Line Intelligence System (COINS) -- a network of intelligence-community computers which have been in place since 1967 in either pilot-experimental or final-operational mode. COINS currently interfaces with the NSA system through a "store-and-forward" switch, housed in a PDP-11 main frame and located at the DIA. In turn, COINS interfaces with the so-called IDHS (Intelligence Data Handling System), which links a number of remote customers indirectly to TIPS, the NSA system. These include:

- Air Force Intelligence (AFIN), Pentagon;
- Naval Ocean Systems Center (NOSIC), Suitland, Maryland; and
- FICEURLANT, Norfolk, Virginia.

They also include such very remote customers as:

- Air Defense Command (ADC), Colorado Springs;
- European Command (USEUCOM), Vaihingen, Germany; and
- organizations subordinate to the Pacific Command (PACOM), such as:
  - PACAF, Hickam AFB, Hawaii;
  - CINCPAC/IPAC HQ, Camp Smith, Hawaii;
  - COMUSKOREA, Yongsan, Korea.

Now active on the TIPS system are about 40 separate SIGINT files, not including about 15 support files for accounting, user-aid, and test purposes. They are managed by owners spread across A, B, C, G, P, V, W, [redacted]

Some of these 40 TIPS files are on the "technical" side. Admittedly, they are designed more for the SIGINT producer than for the user. [redacted]

The second major category of TIPS files could be termed the "intelligence" type. These are files available to both NSA users and the intelligence community through COINS. One big subset of the "intelligence" files comprises [redacted]

The TIDE file carries the latest 20 days of

EO 3.3b(3)
EO 3.3b(6)
PL 86-36/50 USC 3605

EO 3.3b(3)
PL 86-36/50 USC 3605

intercept; the TIPS files, about 90 days of data, 21 or more days old.

[redacted]

OGA

A version of SIRE, the SIGINT Requirements file managed by V1, is off-loaded to TIPS for the benefit primarily of remote SIGINT producers and collectors. (SIRE is actually maintained through the SOLIS system.)

[redacted]

At this writing, the volume of data on the TIPS system totals about 300 million characters. This volume is more than we once thought could be accommodated. If you are a potential owner of a new TIPS application, this shouldn't necessarily discourage you, unless you are thinking of submitting a large file for consideration. A number of current files are scheduled to be removed from the system. This should free some space for future applications. Call C11 for more information about available space.

To many potential TIPS/COINS users at NSA, a perennial problem has been, "Where can I find a convenient RYE outstation from which to input my queries?" Terminals are apt to be located mostly in machine rooms that may or may not be handy to your office. Admittedly, this is far from the ideal situation. Part of the problem is that Mod-35s tend to be a little noisy and not conducive to a serene office atmosphere. Perhaps one solution would be a telephone call to your friendly machine support unit (e.g., G8 for G files, or B42 for B applications), which could make the query for you. In any event, C113 (5203s), which functions as a TIPS customer support unit, in addition to staffing the NSA COINS Subsystem Manager's Office, can guide you toward the nearest outstation, or make a query for you.

Another recurring problem has to do with the terminal equipment itself. Like all of us, the Model-35s are aging; at the same time they are occasionally a prey to gremlins and mysterious poltergeists. Maintenance service, however, is normally prompt. Call RYE operations, 5935s, to enter a complaint and arrange for servicing.

Do you need a beginning or refresher course in TILE (TIPS Interrogation Language)? Formal training is offered aperiodically by the National Cryptologic School (MP-175). If you are interested in such a class, call E21 (8555s) or C11. The COINS Project Management Office does a good deal of informal training in TILE, as well as the languages used to access COINS files at [redacted] DIA, through its user support group. You can reach them at 1108s. C113 can also arrange informal training/briefings for potential users, or analysts who would like to have a refresher in any aspects of TILE.

At this writing the long-term future of TIPS is hazy. We are still in the era of TIPS I. Will there be a TIPS II? No one is saying, and presumably a final decision has not been made. At the moment one assumes that the TIPS II machine, or whatever system replaces her, will be strongly interactive -- with all the features, and high overhead, that this capability entails. But one hopes, however, that the time-honored batch mode will not be scrapped completely. A lot of us old-timers (and some younger ones who have grappled with day-by-day general processing at NSA) like to point out a simple but noteworthy fact: the remote-processing, batch-mode systems of yesterday are still here and still performing prodigies of labor for the Agency. One of these is SPECOL (Special Customer Oriented Language), [redacted] but extremely effective IS&R (Information Storage and Retrieval) and data-processing system for several different machines, including the 360/370 world. RYE/TIPS is another workhorse, limited to a UNIVAC main frame but reaching out as far as Europe and the Pacific to perform its IS&R and data-processing role. Note that both SPECOL and TIPS are more than IS&R systems. They are data-processors as well as retrievers, able to extract, sort, compute, format, and output information in many different ways, and for many different kinds of users. Can you say the same for *your* interactive system?

These random thoughts have not been aimed at disparaging the young DBM systems we are now ogling. Assuredly, they have a bright future at NSA, although their outlines are a bit murky yet. But let's not be in too big a hurry to divorce ourselves from the old batch-mode systems that have served us so well for a long time. They deserve at least a glass or two raised in tribute.

. . . And: Still More Comments! (On the AG-22/IATS Question)

[redacted] W309

Reprinted from C-LINERS, Vol. 3, No. 9, Spring 1976, Issue No. 30

Cecil Phillips' article "Musings About the AG-22/IATS" in issue No. 28 of C-LINERS [and reprinted in CRYPTOLOG, March 1976] caught my attention because of the recent work I have done on data bases built from AG-22/IATS (and, not too long ago, FF STRUM). I do not think that anyone would argue that computer records built totally automatically from AG-22 and IATS are of low quality because of poor planing or a lack of sophisticated programming. The quality of the copy itself is obviously to blame and I do agree with Mr. Phillips that the place to solve the problem is at the source; however, I disagree that more operator tags and less traffic is the solution.

I cannot recall any change or addendum to Morse copy instruction, within the last 20 years, aimed at lessening the amount of non-intercept data an operator must manually enter into his log (traffic service). In fact, the trend has been in the opposite direction. The implementation of AG-22 brought tagging; diminishing intercept resources demanded that we narrow our requirements and specifically define the collective objectives, and now COPES is with us. The Morse operator is, and always has been, overburdened and it seems to be some accepted rule of collection strategy that a position must be assigned more cases than it is humanly possible to cover. There is something drastically wrong with this rationale.

Most of our Morse operators from the Cryptologic Services are placed on the job (on OJT) just out of school and complete their military enlistments with less than 24 months on position. There is absolutely no comparison between U.S. copy (from a quality standpoint) and that of second- and third-party sources where genuine professionals are universally employed. I'm not trying to put down our operators; on the contrary, I think they do an exceptional job with the limited amount of experience they have on the average, but leaving important line-by-line copy decisions (i.e., summarizations) up to them, as Mr. Phillips has suggested, is not at all practical. Weak reception, high-speed transmissions, signals buried in QRM/N, coverage of both ends, etc., require intense concentration during copy by the operator. Intercept of transmissions under such adverse conditions is common and would hardly lend itself to a form of selective abbreviated copy. Gisting that which is the usual also requires that the experienced traffic analyst (or cryptanalyst) do it, especially if enciphered procedures are used.

There are a number of important problems in this Agency today where "non-message" data is all that is "readable." Messages are either high-grade (unsolved) or practice, and hold little or no analytic value. Copy instructions for certain groups passing only practice sometimes require only the first line or so since we anticipate these texts will be repetitions from recovered pages. The chatter, whether the messages are of value or not, may contain routing instructions, frequency and schedule references, authentication, message precedence, etc., each of which is essential in determining communications procedure, identifications, and, most important, network continuity. We have even had callsigns, selected in a particular order from a recovered page, used as time indicators in missile-launch countdowns. Order of callup has also been used to establish continuities. International Q and Z codes have been used for purposes other than originally intended. There are many examples of Q's and Z's being used to indicate precedence and authentication or other uses peculiar to a target's need. Abbreviated plaintext chatter, prevalent on many problems, also requires special attention. Full copy of chatter, less the strings of V's, repeated calls, etc., is important to any analysis to be performed on the material.

Analysis of the collected data fluctuates according to need. When continuities are good and callsigns projectable, less emphasis is placed on the study of message externals. There are many traffic analysts here today who have never worked on a major target complex where the callsign system in use was unknown. Traditional TA for some of them is an unknown art that will be painfully rediscovered when these callsign systems change. We have enough trouble with routine changes (within a recovery system), so let's not lose the remaining means for reestablishing the continuities when callsigns can no longer be used as the primary lead to an identification or continuity. There is a great deal more to TA than callsign lookup, and it lies in the detailed study of all communications data, especially in the chatter. We should not omit or attempt to summarize this kind of information at the point of copy.

Extra intermediate edit steps to fix the traffic copy are no solution either. Manual

intervention to correct an automated process seems to me to be a step in the wrong direction. We will always need people to prepare the input (operators) and others to study the output (analysts). Mechanization should join these two functions; if we need to fix the system, let's do it on either or both these ends, not in the middle. Middlemen editing and/or correcting large volumes of traffic with strict deadlines to be kept contibute very little to the process, since we are frequently forced to accept anyone from any specialty who happens to be available to do this job. Re-identification (except on search material), as part of a maintenance process performed back here, perpetuates, rather than corrects, a bad practice. When intercept identifications must be changed, this means that the wrong target was copied as mission. It is our responsibility to see to it that the operator gets the right information to acquire and identify his target properly in the first place. After-the-fact case corrections are useless to him if the SOIs, i.e., callsign periods, are short in duration. Any other type of correction alters the original copy, which should not be done for a multitude of reasons. File maintenance, if it's worth doing, also requires quality control, which is turn requires time and resources.

The value of manually correcting traffic for data-base storage is questionable, since the case analyst is usually through with it at the close of the work day whether it is properly formatted or identified anyway. He has logged the important items from his material, updated his net diagrams, etc., and will probably never need the traffic again for the types of specific work he is required to do (e.g., TEXTA, in one of its many forms, which is, in turn, machined). I think it is best that we continue to get the best possible operator copy and retain all of it and the identifications (his, any intermediate machine re-idents, and the final) that are placed on the stored copy. Under no circumstances should we change the original version of copy or summarize any part of the data that does not follow a predictable pattern, such as valid messages and chatter, even if the instructions (SCOLs, etc.) require no more traffic than is necessary for identification. If we need to de-dupe, let the programming handle it.

So, this brings about another question: If the case analyst doesn't need a traffic data base in the normal conduct of his daily duties, who does? Management sometimes, for collection studies from data not available anywhere else, but this certainly does not justify a data base. The real user is the research analyst studying larger masses of both identified and unidentified traffic. Here is where new ideas and analytic concepts are born. It is for this reason, and perhaps this reason alone, that we maintain a number of these data bases. I think we are justified in doing this for just this purpose, but I would be hard put to support my conclusion. I suspect that I would have no case at all if I had to provide supporting evidence based on past usage of the data bases or the applications of special research, beyond case analysis, which is often sacrified in favor of continuing projects and the fulfillment of day-to-day commitments.

The potential uses of these data bases through readily available programs should bolster the TA imagination and create new approaches if properly publicized and if adequate indoctrination is provided, at least enough to get us beyond the customary case and data retrieval (with follow-on sort and list) that is usually requested by the average analyst. In spite of the more obvious shortcomings of formatting and processing traffic for these data bases, I feel they are valuable whether corrected or not and that we should promote the use of the numerous facilities and software that we have at hand in exploiting them. Getting this done is another gigantic problem in itself that needs the attention of management in particular.

The problem of formatting traffic for data-base storage is an old one with quite a history. I can remember my efforts (some 15 years or so ago) to have field analysts edit traffic for electrical transmission rather than prepare the detailed (complicated) TECSUMs/MATSUMs of the day. The idea didn't make it, but the same general concept of editing for data-base formatting I tried to sell is now the responsibility of the AG-22 operator, called tagging.

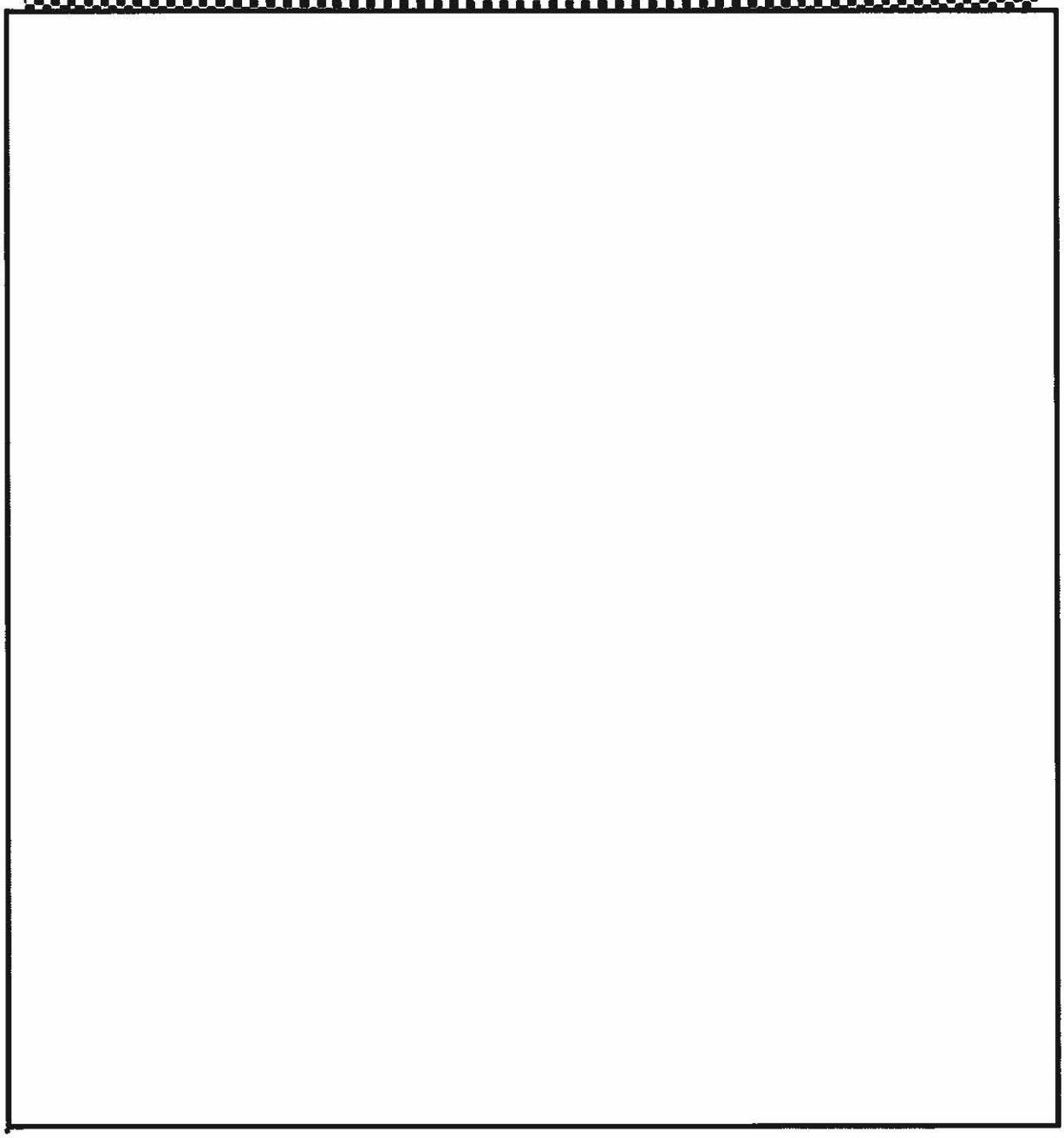
The formatting of Morse traffic has been tied to a number of vehicles over a long period of time and evolution, from TECSUMs to MATSUMs to STRUM to AG-22. All of these have had a measure of success, but now that we have managed to automate forwarding directly from the intercept position, our concentration should still be on intercept and the improvement of copy. I believe that placing the formatting responsibility almost totally on the operator through tagging will divert our collection from these objectives and generally degrade traffic quality. We should be able to do without some of it or substitute multifunctional and/or automatic tags and develop programs, with the objective of doing a better job of data-base formatting, that are a little less operator-dependent.

Surely we have the expertise on hand to do this. We have a pretty good system working for us now, not just in AG-22/IATS, but in a variety of other areas, e. g., automated callsign projections, machine decryptions, radio wave propagation, cryptanalytic diagnostics, etc. If we can "teach" the machines to do these things, there is no reason why we can't get on with the task of formatting traffic acceptably with follow-on programs doing the specialized work. I do know that edit programs have been written to "fix" message tests from existing data bases, not only for decryption, but also for indexing,

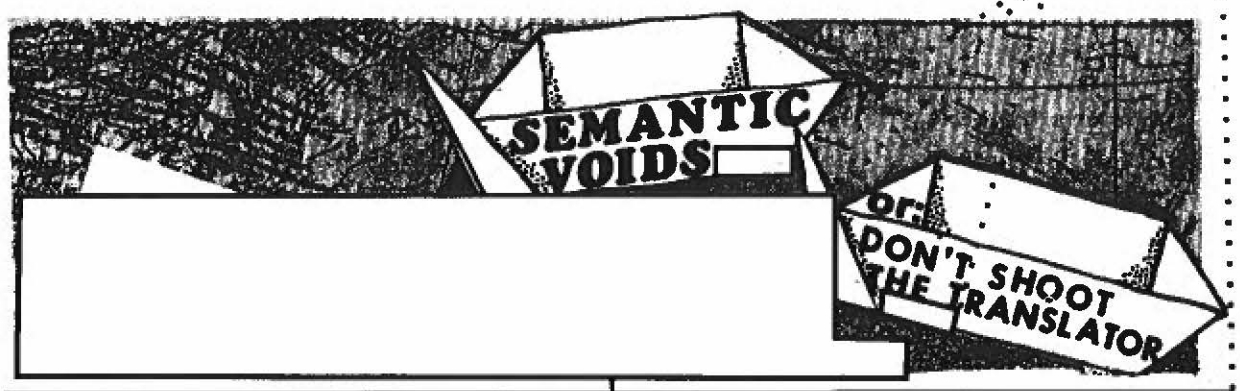
diagnostics, etc., as well, and the cryptanalyst has been generally satisfied with the product. There have been programs developed that scan chatter, find enciphered address groups, decrypt them, and place them in sort field (key) locations so they can be listed in an orderly fashion and in context. We should be able to expand on these techniques and do a fairly good job of mechanizing the editing and formatting of AG-22/IATS "take" with minimal tagging. I contend that if we retain selective retrievability in these data bases so that we can get back to this material through identifications (case and terminal) for the specialized processing that needs to be done, that is sufficient.

Successful traffic analysis depends upon traffic quality. Let's do something that will assist the operator in making that product more accurate, complete, and at the same time make his job easier. It's time the machines were put to work to serve us.

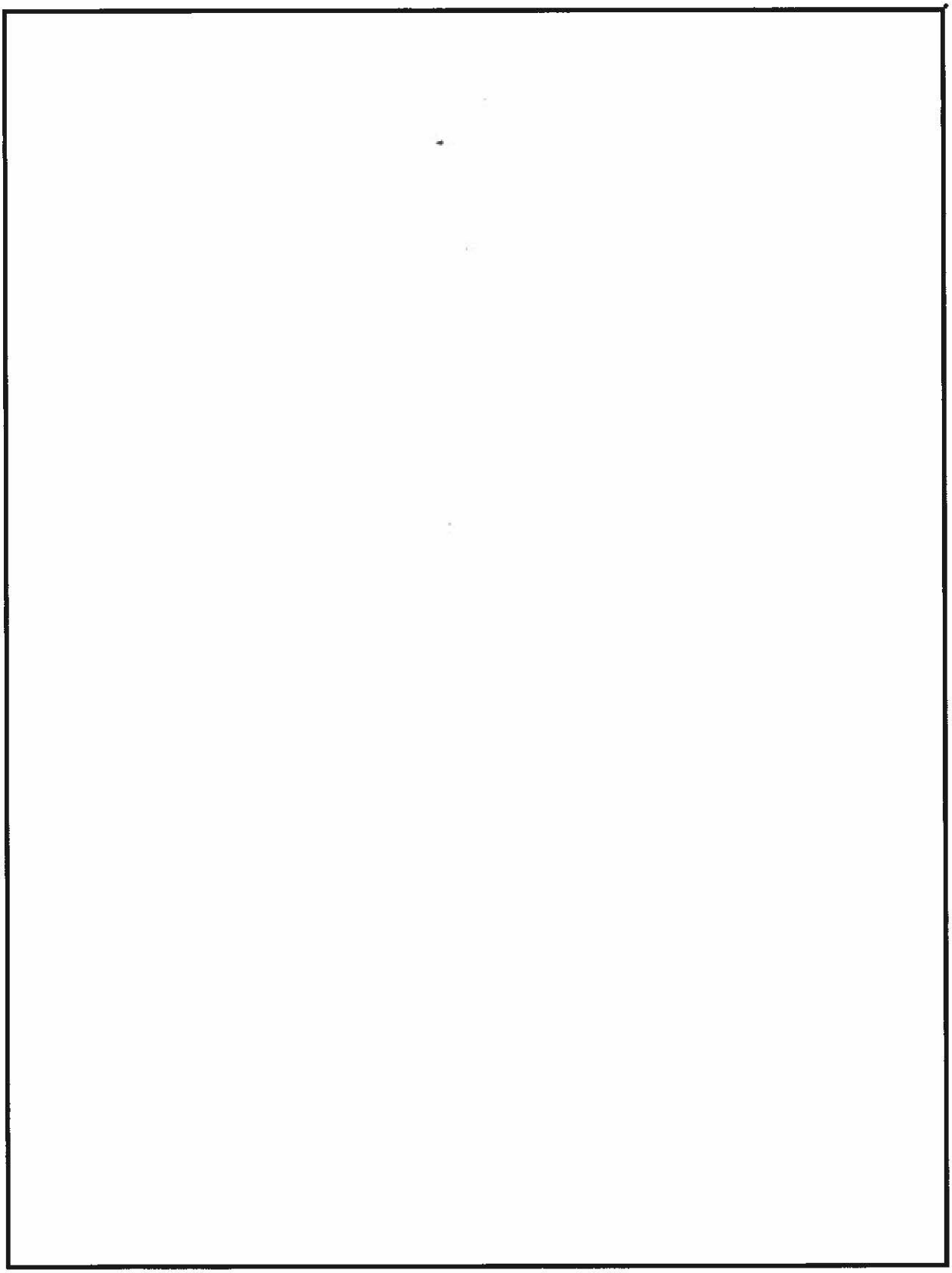
(SECRET [redacted])

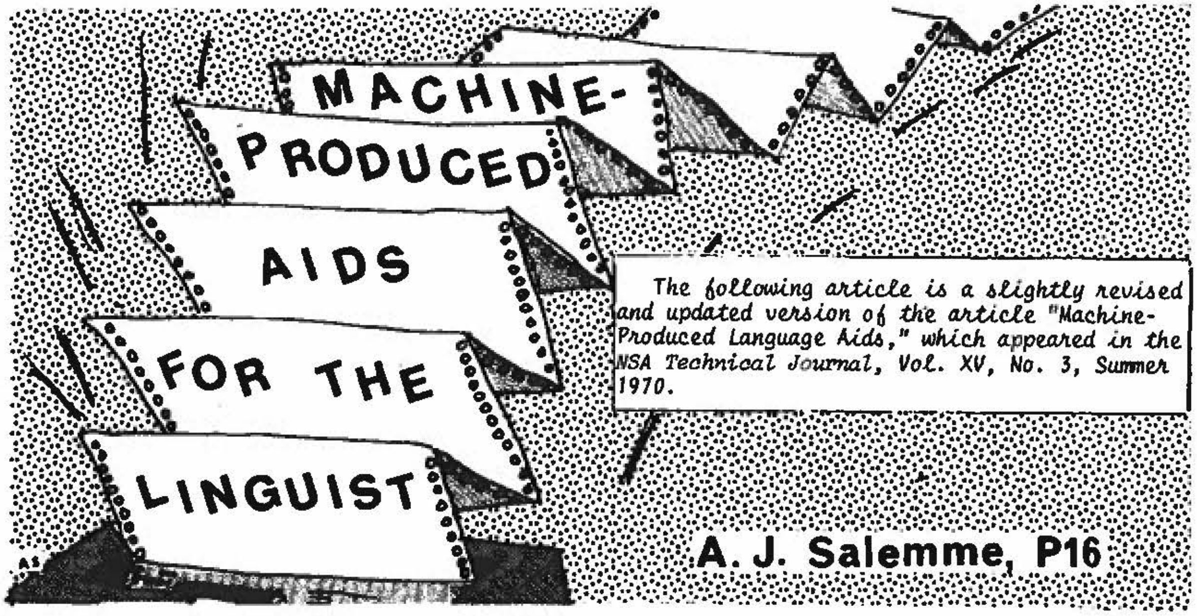


Non - Responsive



EO 3.3b(3)
EO 3.3b(6)
PL 86-36/50 USC 3605





MACHINE-PRODUCED AIDS FOR THE LINGUIST

The following article is a slightly revised and updated version of the article "Machine-Produced Language Aids," which appeared in the NSA Technical Journal, Vol. XV, No. 3, Summer 1970.

A. J. Salemme, P16

The COMINT linguist has at his disposal, in addition to general and specialized dictionaries, various types of language working aids that arrange individual words or connected text in ways that look unusual to the linguist[1]. The purpose of this article is to acquaint the non-linguist, and the linguist with no previous COMINT experience, with the COMINT need for those working aids and with some of their characteristic uses and limitations.

But before discussing these aids, it might be worthwhile to explain why language aids have been created to manipulate words in ways that seem to differ so greatly from the normal patterns of spoken and written language. Unlike Moliere's bourgeois gentleman, who was surprised to learn that he had been speaking prose all his life, there are probably few who are surprised to learn that normal spoken and written language is unidirectional. That is, people start speaking and, whether or not they have previously organized their thoughts logically, they produce their words one after the other in a definite, irreversible time sequence. Or they start writing and, depending upon the particular language, put the words down in a definite sequence from left to right, right to left, top to bottom, etc. This text, for example, has the words arranged the way we like them -- and it's not just because a typewriter couldn't be hooked up to type the words boustrophedonically.

The listener or reader perceives each string of words in its produced sequence (and then is able, of course, to scramble them at will in his mind). He does not usually listen or read by jumbling the words back and forth out of time or space sequence. Nor does he usually listen or read "backwards." Spoken or written sequences that are actually intended to be interpreted in a "backward" sense are usually contrived for artistic or comic effect. Such contrived sequences include palindromes ("Able was I ere I saw Elba"), typographic tricks (caption on a cartoon showing two people shouting to one another as their respective cable cars pass: "I said wait for me at the other end" and "bne rehto eht ta em rof tiaw dias I") or dialogue in animated cartoons (the hero, propelled at breakneak speed over a hazardous course, stops dead to exclaim, "What a buggy ride!", continues on his way until he crashes, and then is propelled backwards over the same course, during which he stops to exclaim, "Ride buggy a what!").

Except when creating or responding to such stylistic tricks, normal people (that is, those not suffering from some psychological or physiological impairment of the ability to produce or to perceive speech or writing) handle language strings in their usual sequence. But COMINT language specialists do not deal with language data that is "normal." In the everyday world a person on the end of a telephone line who does not hear a word perfectly can ask the person at the other end to repeat it. The NSA transcriber obviously cannot do so. In the everyday world an interpreter, seeing a confused expression on the face of the person for whom he is interpreting, can clarify the linguistic ambiguity right

PL 86-36/50 USC 3605

then and there. The NSA linguists has to deal with the ambiguities as they stand ("He *said* June, but I think he meant July"). In the everyday world a person who receives an important telegram with a critical word garbled can request a repeat. The NSA analyst can only hope that the intended recipient will be just as confused as he is and will request a repeat that is also made available to the NSA analyst. If he does not, the NSA linguist has to degarble the text as best he can, even if that involves his going out on a limb.

NSA linguists, accustomed to dealing with these and related problems that arise when listening in on foreigners' conversations and reading their foreign-language telegrams, know that often they cannot attack a foreign-language text in strict left-to-right order. Nonlinguists sometimes have the impression that, just like a kid who can't have his dessert until he has finished his carrots and string beans, the NSA linguist has some professional or moral obligation not to look at the second word until he has translated the first one, or to look at the third until he has translated the second, etc. But this is not true even in what might be called normal translation work, when, for example, a commercial (that is, non-COMINT) technical translator is translating a completely ungarbled text from a printed, open-source book or magazine -- when translating a technically complex or grammatically obscure sentence, he may indeed have to look ahead (to the next sentence, to the next page, or even to the next chapter) for clues that will resolve the ambiguities. And the situation is even more complex for the NSA linguist, who often finds that, because of message encryption, garbling, poor audio signal, etc, it is impossible to attack a written text in a strict left-to-right sequence [2] or to transcribe a radiotelephone conversation from the first syllable to the last on the tape.

The NSA linguist is a kind of scientist, in the sense that he can isolate the words he wants to examine and can subject them to any kind of test he wants, in order to extract their intelligence meaning.

[redacted]

The NSA linguist's job is analogous to that of the FBI specialist in the chemical-analysis or ballistics laboratory. But whereas the tests run by the FBI specialist weigh and measure the quantitative properties of physical objects, those run by the NSA linguist assess the qualitative properties of words. And words are "things" that keep changing. If an NSA linguist has a garbled 5-letter word, how does he decide which of the three obvious degarbles is the most likely to have been the word intended in the original text? If an NSA bookbreaker has an unrecovered value in a one-part code between letter M and O, how does he decide which are the best guesses? If an NSA research analyst has a message signed by [redacted] how does he decide which [redacted] it is? Or if he has, in his traffic, an abbreviation with 40 possible expansions, how does he determine which is the one that the message originator had in mind and that the message recipient will recognize immediately? (Incidentally, how, in dealing with intercept text written all in capital letters, did the NSA linguist recognize the abbreviation in the first place?) The answer to all these questions is the same: he made a judgment based on his thorough knowledge of the language with which he is dealing (not just a thorough "book-l'arnin'" knowledge of the language, but a thorough knowledge of it as it is actually used by the specific user, with all his specific educational, occupational, social, and other peculiarities), and based on his submitting of the language data to validity tests at all stages of intercept, decryption, translation, and interpretation. These tests that the NSA linguist subjects his language data to are just as valid as the chemical and spectrographic tests conducted by physical scientists, but the results of the linguistic analysis do not have a nice, solid scientific look to them. A typewriter expert can state in court, "The letter *m* in the examined document could not have been made by a Smith-Corona typewriter" and stand ready to back up his statement with enlarged photographs showing measurable distinctive features. But the NSA linguist who states with identical firmness, "The letter *m* in this word must be a garble for *s*" usually cannot support his findings as impressively, and might even sound downright shifty as he brings in such qualifiers as "usually cannot," or refers -- however modestly -- to his "years of experience," "feel for the language," or to "letter-frequency probabilities," "contextual incongruity," and other qualitative, rather than quantitative, proofs.

How, then, does the NSA linguist equip himself when attacking COMINT text in a particular foreign language? Obviously, the first step is to acquire newer and newer dictionaries and to augment them with operational language files and specialized glossaries. But these all list words in the normal alphabetic order

[2] See: "Right-to-Left Text Sorts Are Not Impossible," by [redacted] *CRYPTOLOG*, August 1974.

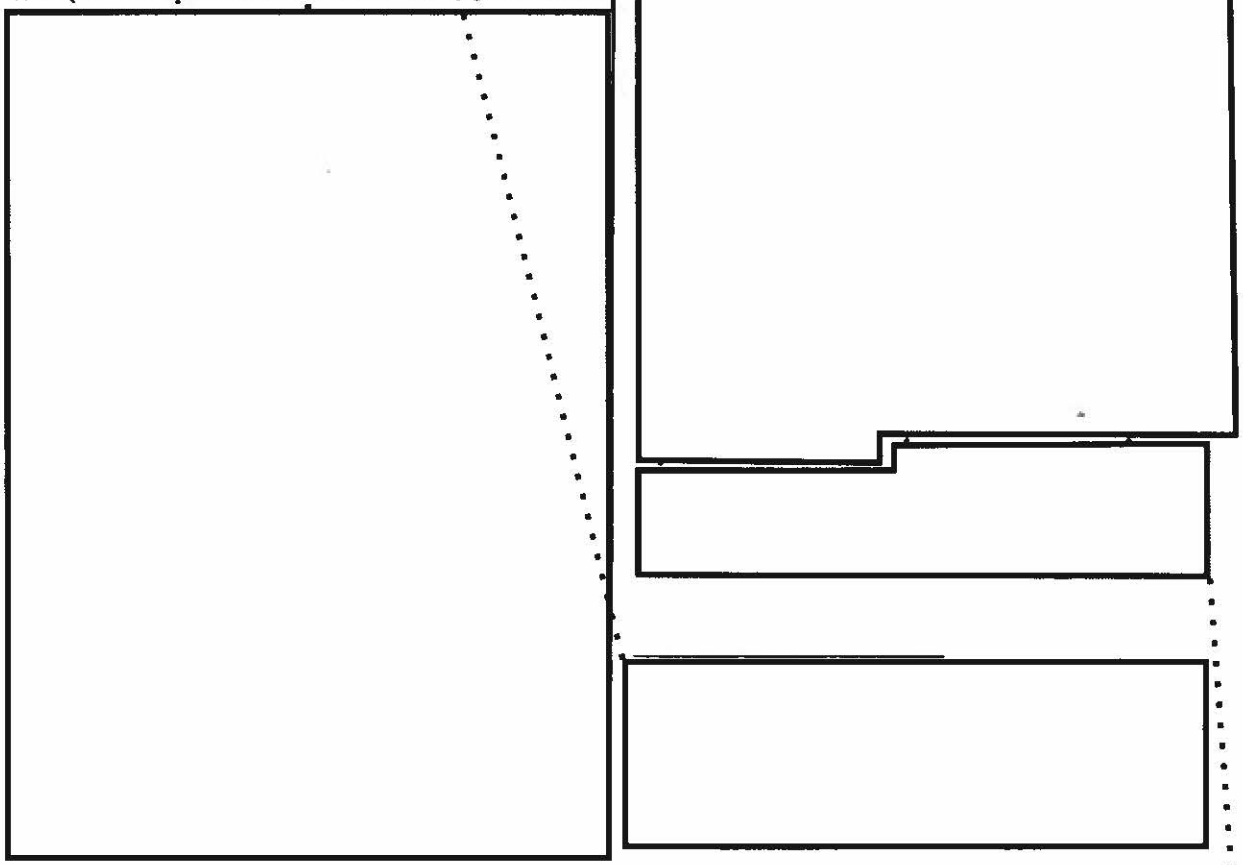
and are therefore of limited value to the cryptanalyst or linguist trying to cope with a word with its beginning missing, garbled, or cryptographically unrecovered. Therefore, NSA analysts specializing in language have felt the need to prepare various types of language aids that list words in other than normal dictionary order[3].

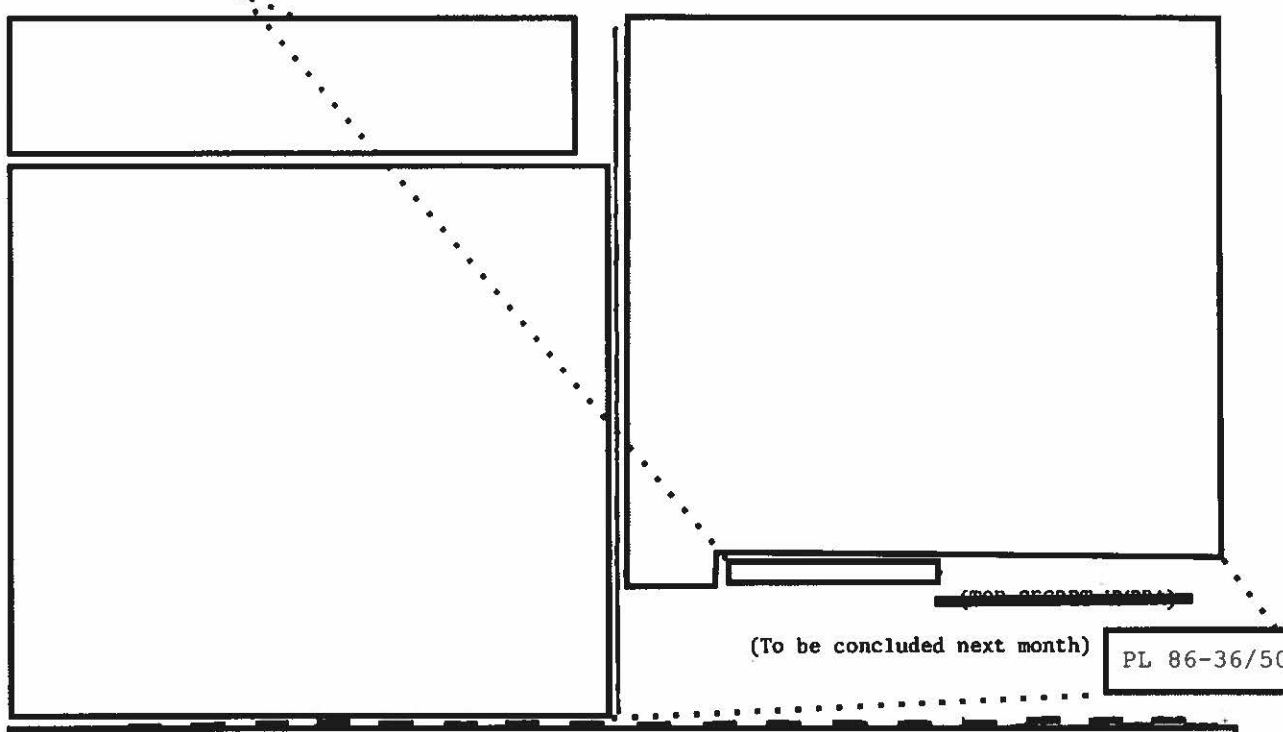
The aids fall generally into three basic categories:

- word-pattern listings,
- backward listings,
- window indexes.

Word-Pattern Listings

Word-pattern listings are listings usually of preselected words (sometimes including word phrases of two or three words) that are typical of a particular type of context and that have a particular pattern of letter repetition. For example, the English words AARDVARK, EEL, LLAMA, CANNOT, and WILL each contain a doubled-letter sequence that can be represented by the coding AA. The words LLAMA, COMPLETE, and MILITARY each contain a repeated letter with one intervening letter, which pattern can be represented by the coding A-A. The phrases CAN NOW and TO OPEN each contain a doubled-letter sequence that can be represented either by the coding AA (if the space between words is nonsignificant) or A-A (if the space between words is significant).





~~(TOP SECRET UMBRA)~~

(To be concluded next month)