~~TOP SECRET~~

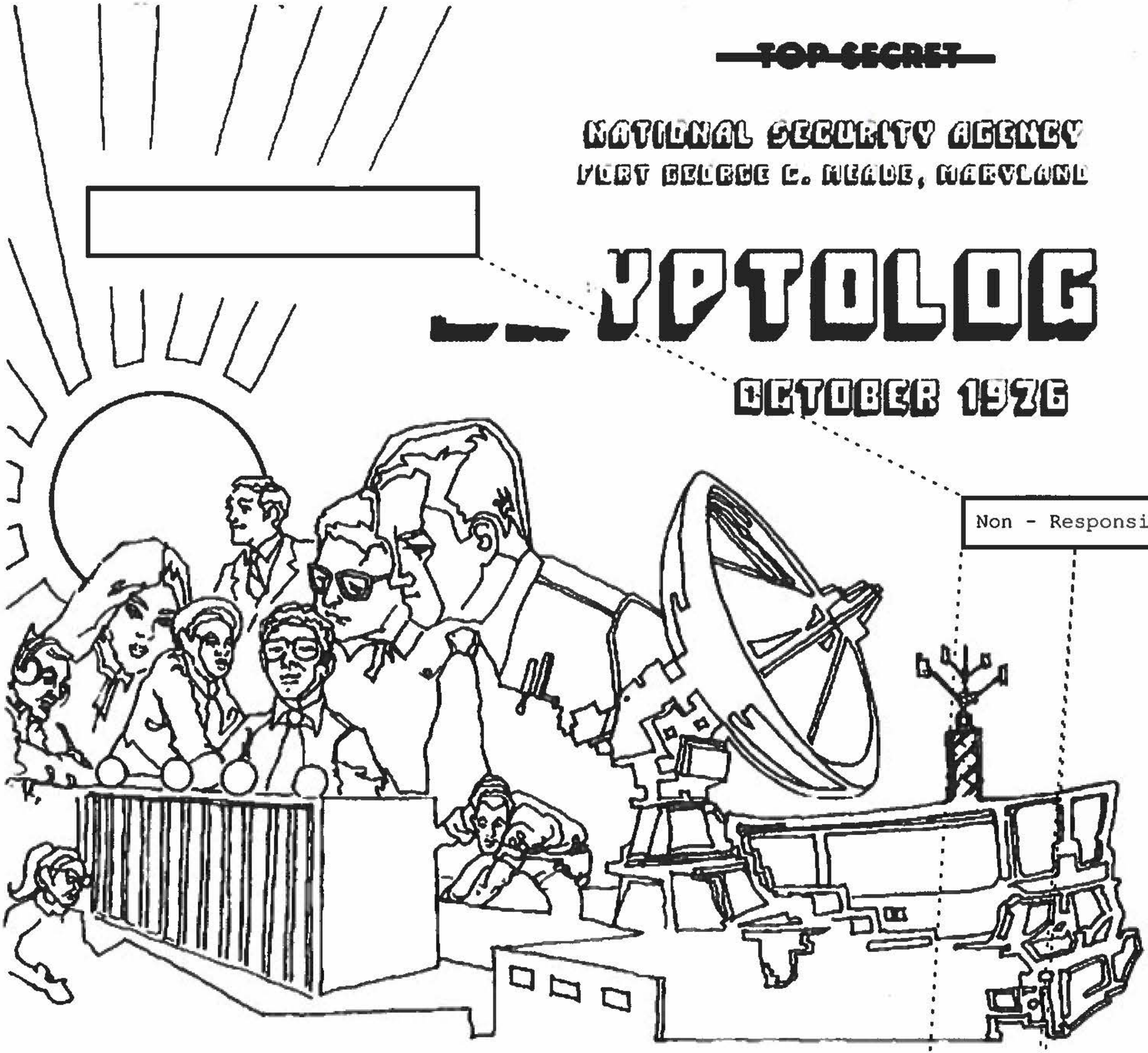
NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND

[redacted]

[redacted]YPTOLOG

OCTOBER 1976

Non - Responsive



~~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~~

Non - Responsive

Non - Responsive

Non - Responsive

Non - Responsive

Non - Responsive

Non - Responsive

Non - Responsive

MORE THOUGHTS ON "QUESTIONABLE" SIGINT

Fred Gerkens, F43

I have been following with interest the series of articles prompted by Vera Filby's "How Do We Know It's True?" (CRYPTOLOG, February 1976), since this topic was of prime concern during the time I spent in B31 and F47 working the North Vietnamese Air problem. We encountered a fair amount of communications deception, or "spoofing," on the part of NVAF fighter controllers. Most of these occurrences were poorly done and would not stand up to any intensive analytic scrutiny. Unfortunately, in many cases they effectively accomplished their purpose of disrupting U.S. air strikes. During LINEBACKER II, we were copying NVAF fighter (and missile) communications and passing the information to an Air Force GCI controller for use in protecting the U.S. strike aircraft. The timeliness required did not permit any second-guessing, and it was only due to the competence of the USA-523 operators that many of these attempted deceptions were recognized as they occurred. If there was any question or possibility that a MiG was actually reacting to the U.S. strikes, however, the information had to be used, creating not only security problems, but also causing some of our customers to begin to doubt the credibility of SIGINT. This led at one time to the deputy wing commander at Udorn (informally) demanding that TEABALL (our Weapons Control Center) be removed from the air because of its alleged unreliability.

This potential for causing the customer to lose faith in SIGINT because of our reporting of false information also raised serious questions in terms of our non-timely summary reporting. The problem really started during an earlier phase of the air war, when there was SIGINT evidence that U.S. operational commands had apparently overstepped imposed limitations. In some cases, intercepted enemy air surveillance tracking, which we knew to be valid, was used to show that U.S. aircraft had overflown restricted areas. Unfortunately, a feeling of animosity towards NSA grew in certain commands, and any admission that SIGINT could be "spoofed," even if we showed that we were able to separate the valid from the invalid fairly quickly, could have been used to discredit our reporting of all questionable incidents. I'm happy to say that the decision was made to publish the facts of the deception incidents, and I think our overall credibility was enhanced by our admission that we could be deceived for a short period, since it showed we were aware of the problem and not retreating behind a claim of infallibility just because we were using SIGINT.

Another very interesting form of deception was encountered which I feel should be included in any discussion of the credibility of SIGINT. That is deception aimed not at a potential enemy intercept operation, but that intended for the target's own chain of command. When MiG-19s were first introduced into the NVAF, a period of upgrade or conversion training was begun for a selected group of NVAF pilots. From various other sources we had a good idea of what the training program would consist of, and were gratified when SIGINT showed it proceeding according to our predictions. At this time we enjoyed an excess of redundant collection on NVAF flight activity (overlapping first-party "strategic" and tactical with third-party backup) and were confident that we were missing little if any activity.

Then, one day we intercepted a report from probably the unit's training officer to his headquarters giving a recap of the last month's training. The number of days and dates coincided with what we had logged from our daily collection (fair-weather training only at this point), but his figures for activity on those days were unrealistically high. Were we missing approximately 40% or more of flight activity by pilots in training, who were always under close surveillance and control from their instructors, with doubly or triply redundant collection coverage? This seemed unreasonable, and after going over the whole situation many times, I came to the conclusion, and published, that the training instructor was padding his figures.

Perhaps I should have coordinated the report a little more widely before putting it out.

Reaction from our customers was minimal, but repercussions internally were swift and strong, mostly along the line that I was "destroying the credibility of SIGINT." I could have merely published the data and let the customers draw their own conclusion, but I felt the comment was necessary. We at NSA, because of our constant exposure to it, should be the best judges of which SIGINT is valid, and which should be treated with reserve, and I personally believe it's the duty of a professional reporter to provide this type of evaluation to our customers.

I agree that the whole range of "questionable" SIGINT could use further study and discussion at NSA, not only perhaps to further define it and improve our capabilities to detect it, but also to bring out into the open what is often considered one of the "unmentionables" in our profession. Perhaps it will be less fearsome if more people understand it.

Non - Responsive

Non - Responsive

[redacted]

NSA-CROSTIC No. 5

By A. J. S.

The quotation on the next page was taken from a published work of an NSA-er. The first letters of the WORDS spell out the author's name and the title of the work.

DEFINITIONS	WORDS
A. Fingerless glove	70 91 134 174
B. Got rid of	32 49 68 153 2
C. Japanese statesman (1841-1909); the outstanding figure in the modernization of Japan	43 159 147
D. German postimpressionist painter of landscapes and animals (1880-1916)	81 108 5 152
E. Wan	71 89 96 162 128 88
F. British sculptor (1880-1959), born New York City of Russian parents; also, one of Kotter's Sweat Hogs (no kin)	26 114 62 166 54 120 140
G. Naomi's daughter-in-law	19 101 72 125
H. Proto-Malayan people of Philippines	93 171 10 34 112 1
I. Gem characterized by play of changing colors it displays	58 100 139 168
J. Lyric, usually of amorous character and adapted to musical setting	14 33 90 160 22 122 28 169

K. What the private investigator said when he located the mild-mannered son of the first president of the Republic of Korea, working at a shoe factory (son had left his job as a photographer for the Luce publishing empire) (11 words, suggested by song title)	6 64 20 146 31 51 73 99 115 130 136 154 167
	13 37 87 143 11 61 80 94 110 135 158 131
	170 57 124 113 82 118 44 150 163 77 137

L. Secret group of advisors to Charles II of England	12 95 84 133 149
M. Stringed musical instrument	9 106 76 155
N. Inferior devil	38 138 107
O. When told he should make one of these as a penance, the pilfering night watchman at the building-supplies warehouse said, "If you can get the plans, Father, I know where I can get the lumber!"	7 98 127 172 56 151
P. A lot of people make a mistake pronouncing this word (look it up!)	23 83 104
Q. U. S. government agency	55 39 21
R. Chinese family name	144 50
S. Group of related Indian tribes, Pacific Northwest coastal area	17 42 86 121 92 164 45

T. Sentence ending used by Yul Brynner in "The King and I" (abbrv, abbrv, abbrv)
46 67 119 142 156 8 18 63 53

U. British term of endearment (also used when addressing someone whose name you've forgotten)
29 16 60

V. Richest man in Rovaniemi -- "the —— of luxury"
47 157 15 75

W. Books printed in the 15th century
36 52 97 148 165 66 109 161 59 126

X. Take eagerly
78 116 123 4 48 69

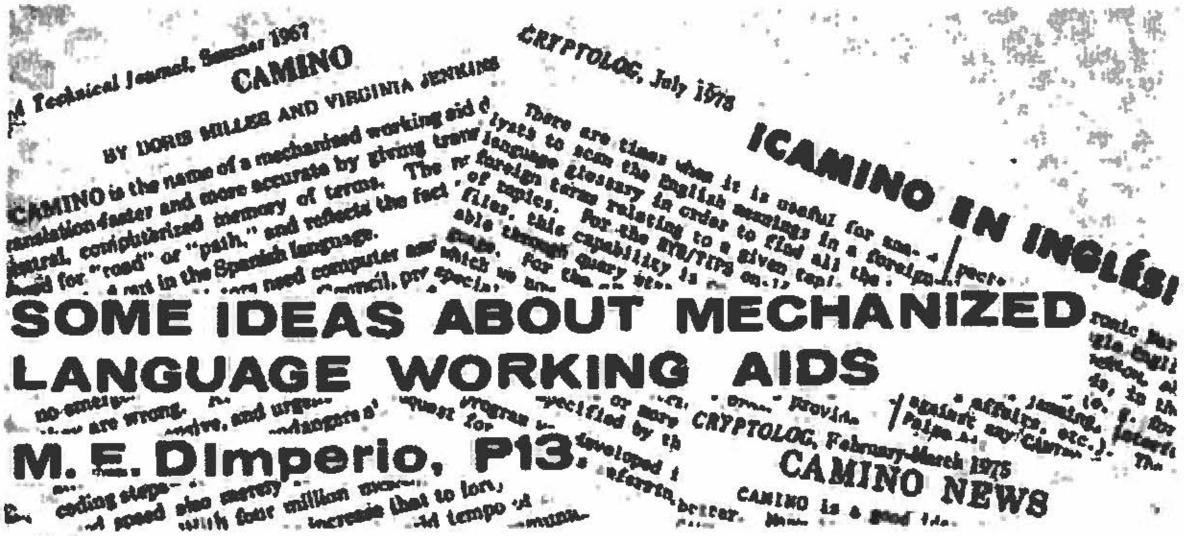
Y. Produce a design by lines corroded by some chemical agent
65 35 25 132

Z. Point of a pen (pre-ballpoint)
24 40 141

Z_1. Important component in computer-graphics system (abbrv)
27 79 102

Z_2. Dilute perfume, introduced c. 1709 in Germany (3 wds)
3 41 85 117 103 105 74 30 129 145 173 111

1 H	2 B	3 Z_2	■	4 X	5 D	6 K	7 O	8 T	9 M	■	10 U	11 K	■	12 L	13 K
14 J	15 V	16 U	17 S	18 T	19 G	■	20 K	21 Q	22 J	23 P	24 Z	25 Y	26 F	■	27 Z_1
28 J	29 U	30 Z_2	31 K	32 B	■	33 J	34 H	35 Y	36 W	37 K	38 N	39 Q	40 Z	41 Z_2	42 S
■	43 C	44 K	45 S	46 T	47 V	48 X	49 B	50 R	51 K	52 W	53 T	54 F	■	55 Q	56 O
57 K	58 I	59 W	60 U	61 K	62 F	■	63 T	64 K	65 Y	■	66 W	67 T	68 B	69 X	70 A
71 F	72 G	■	73 K	74 Z_2	■	75 V	76 M	77 K	78 X	79 Z_1	80 K	81 D	■	82 K	83 P
■	84 L	85 Z_2	86 S	87 K	88 E	■	89 E	■	90 J	91 A	92 S	93 H	94 K	95 L	96 E
■	97 W	98 O	99 K	100 I	101 G	102 Z_1	103 Z_2	104 P	■	105 Z_2	106 N	107 N	108 D	109 W	110 K
111 Z_2	■	112 H	113 K	■	114 F	115 K	116 X	117 Z_2	118 K	119 T	120 F	121 S	122 J	■	123 X
124 K	125 G	126 W	127 O	128 E	129 Z_2	130 X	■	131 K	132 Y	133 L	134 A	■	135 K	■	136 K
137 K	138 N	139 Y	140 F	■	141 Z	142 T	143 K	144 K	145 Z_2	■	146 K	147 C	148 W	149 L	150 K
■	151 O	152 D	153 B	154 K	155 M	156 T	■	157 V	158 K	■	159 C	160 J	161 W	162 E	163 K
■	164 S	165 W	166 F	167 K	168 I	169 J	170 K	171 H	172 O	173 Z_2	174 A	■	■	■	■ A.J.S.


Technical Journal, Summer 1967
CAMINO
BY DORIS MILLER AND VIRGINIA JENKINS
CRYPTOLOG, July 1975
¡CAMINO EN INGLÉS!
CRYPTOLOG, February-March 1975
CAMINO NEWS
CAMINO is a good idea

SOME IDEAS ABOUT MECHANIZED LANGUAGE WORKING AIDS

M. E. Dimperio, P13

This brief paper is intended to present some general comments covering my experience with the CAMINO Natural-Language Machine Files and other computerized natural-language working aids. It has seemed to me that these remarks might be of some value or interest to others concerned with mechanized language aids, whether as users or as designers.

In considering the design of a natural-language dictionary (as in any data management system, however sophisticated or rudimentary), there are three obvious kinds of things that need to be done:

a) building the file from external sources;
b) maintaining the file by additions, deletions, and corrections; and
c) querying, displaying, or extracting data from the file for day-to-day operational use.

The attention of most computer specialists has centered on the tasks under heading *c*: the ways of looking up, rearranging, manipulating, and displaying data from an existing file. This is, admittedly, the most interesting area conceptually, and provides the most scope for inventiveness in software design and programming techniques.

Unfortunately, however, at least in my experience, the real problems involved in the design of natural-language working aids revolve around areas *a* and *b*, and especially area *a*: the initial building of the glossary file. The querying or displaying of data from a completed file has rarely presented any problems in the CAMINO files or any other mechanized working aids I have worked with. Even when an on-line querying facility was available for the CAMINO RYE/TIPS files, it was used primarily (in fact, almost solely) by the file executives themselves for file maintenance. It was so little used that two of the three files involved were eventually removed from the RYE/TIPS system for that reason. Most file users have been quite happy to get a copy of the printout, so long as it was of a convenient size and shape for storage and use (i.e. a compact 8x11" document rather than a cumbersome machine listing on size 12 paper).

I was, myself, surprised and disappointed when I first discovered that our RYE file users (as distinct from the executives maintaining the files) had been so reluctant to use the extensive and well-designed on-line query capabilities provided through TILE. Our CAMINO Committee meetings were regularly attended by a TIPS representative, who warned us a year or more in advance that we were not using the capability sufficiently to retain it. With this in mind, we made a concerted effort to set up specially-planned courses in TILE simplified and adapted for the CAMINO user, and to encourage users to use the RYE terminals. In the last analysis, it was evident that it was simply much *easier* and more convenient for each linguist to use a printout. The printout was right on his desk; no one else competed with him for access to it; it was always ready to be opened and used. He already knew how to consult a printed dictionary, and had no new "languages," procedures, or rules to learn and get used to. The RYE terminal, in contrast, was "down the hall" or in another room. It was also likely, during the period before TIDE applications were given their own system, to be either malfunctioning or very slow in its response. In sum, I can hardly blame the linguists for using the handy printout; for them it was an intelligent, cost-effective choice for that time and that purpose.

Let's take a closer look at what the analyst who wishes to build a mechanized language glossary has to do. He must select, from a wide variety of potential sources, just which terms, from what kinds of documents (traffic, transcripts, collateral, in-house publications, 3x5 card files, published dictionaries, handwritten notes, etc.) he will include in the machine file. He must collate the selected data in such a way as to resolve discrepancies in representation, format, grammatical information, etc., and to provide a clear, consistent, and useful document or display for operational purposes. In choosing the data and designing the form of the aid, he must think of the whole range of potential users of his file: new trainees as well as senior analysts; cryptolinguists and reporters as well as translators and transcribers. He must get the data into machinable form in some way, and, as the file grows, he must examine successive sorted printouts, check for errors and discrepancies, and get them corrected.

In maintaining the file once it has been established and has completed its initial rapid growth, the file sponsor must see to it that the file continues to be responsive to all his users and their needs, in spite of frequent reorganizations of Agency personnel, changes in missions, and new developments in the external world of international events. He must make a continuing effort to collect contributions of new terms, and solicit constant feedback from all file users, while searching for new sources of data that should be added to the file.

Now let's shift our attention to the question of how the tasks outlined above can best be accomplished at present and in the near future and why. To most computer-oriented managers, it seems obvious to the point of triviality that each sponsor who needs a mechanized language aid should do all the work himself, preferably at "a terminal" (unspecified). If he cannot do all the work himself, for whatever reason, he should get "somebody else" (also unspecified) in his own organization to do it for him. The hypothetical manager we are quoting sees no difficulty in any of this, and considers the problem solved by his advice.

In certain specific contexts, where on-line terminals have already (for a variety of historical or practical reasons) become the primary or only way of accessing and manipulating all data for a problem, it makes excellent sense to put language aids on line too. In these cases (e.g., the G TENNIS facility), the needs of one specific set of users are being amply met by a complete system design that includes all the working aids they require. There are still, however, many scattered users who are, and may remain for some time, outside of these advanced projects. For them a generalized or standardized conventional file-maintenance procedure, involving periodic updating and reissue of printouts, has been an effective and economical way of getting the job done. This situation is analogous to that of bookbreakers who do not require customized processing systems and who make use of the Standard Bookbreaker's Package. CAMINO is such a system: a "standard language-aid package," as it were, now applied to more than 20 languages.

For the many users not now served by a customized on-line Data Management System like TENNIS, the managerial advice quoted a few paragraphs back doesn't help much. The senior linguist who is the one competent to develop and maintain a dictionary is usually called upon to do many other, more operationally pressing things in every minute of his day. He must translate, listen to tapes, and report on traffic; teach courses; develop training materials and tests; answer frequent questions and requests for trouble-shooting from co-workers, and supervise the work of junior colleagues (to name only a few of the demands on his resources). The problem is especially acute for multilinguists who must portion their time over several different languages.

Faced with all these immediate pressures, the senior linguist has little or no time, energy, or peace of mind left over to work on the dictionary at all, let alone do the extra initial work required to mechanize it. He certainly is not usually able to spend hours of his time sitting at a terminal (which he is not routinely using for any other purpose), keying in his own data or making line-by-line changes. Finding someone else to do these things within his own organization is also apt to be very difficult in practice. Everyone else who is qualified is just as busy, and just as unable to take time to sit at a console for hours on end, concentrating on the dictionary. Helpers who are not qualified almost invariably cause far more problems than they solve.

Our hypothetical manager, whom we have been quoting as a Devil's advocate, has an immediate reply, by which he again seeks to define away the awkward resource-allocation problem: "If the sponsor organization doesn't need the dictionary badly enough to assign the necessary resources to it, obviously it isn't really needed at all, so we can forget it." Again, I cannot accept this representation of the situation, however useful it may be in simplifying matters for the manager.

It sometimes seems to me that we have a strange "schizophrenic" attitude in our Agency about "research," or indeed any long-term activity that provides a cumulative background of knowledge and perspective on day-to-day problems. We know that we need this background and continuity when the chips are down, but we never feel able to assign to it the long-term commitment of resources which its development and maintenance requires. This ambivalence applies not only to dictionaries, but to continuity on cryptosystems and targets not currently active but likely to surface again at any moment. One of the useful functions that P1 has performed over the years has been to fill this gap wherever possible, and to maintain a cumulative and historical perspective on operational problems.

In my work with mechanized natural-language files, I learned some of the things discussed above the hard way. I began early, and have continued -- without notable success -- to grapple with the problem of devising a simple, economical, and effective technical solution for the difficulties of building and maintaining language files. I experimented with teaching file executives to use terminal systems (e.g., the LODESTAR and Burroughs Word Processing Systems) to maintain their own files, and soon discovered that few of them had the time to spare for keying in their own data. There were other unexpected flies in the ointment as well. Disasters behind the scenes in the processing system caused loss and destruction of stored files. The constant, arbitrary changes of software and "rules" (to which all modern computer systems seem to be subject) created a real problem for the occasional user such as the CAMINO file executive; he was apt to discover, on trying to bring up his file or run a program after a lapse of a few weeks, that nothing worked anymore and nobody knew why.

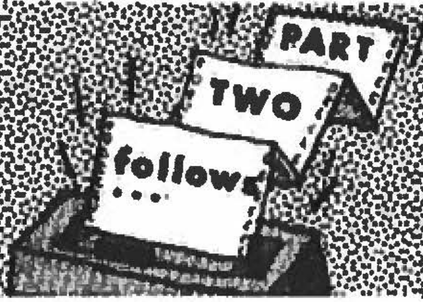
In closing, I would like to list a few summary comments regarding the philosophy underlying the present IBM-370 off-line CAMINO files.

- I believe that computers are not an end in themselves; as a computer specialist primarily interested in applications, I am here to serve a user who wants to perform an operational task; in this case, to perform specific linguistic and cryptolinguistic work in support of NSA missions.
- I am convinced that linguists and others working on NSA language problems need the best working aids they can get; they need not only the general, open-source desk dictionary, but also a broad spectrum of specialized working aids and glossaries which are related to a specific Agency target and mission, and can be developed only within the context of that task.
- I believe strongly that machine processing can be of real service to language analysts in developing and maintaining good working aids. The initial extra expenditures of effort and time are amply repaid by later advantages (e.g., the availability of one, up-to-date edition of the file to all potential users, and the relative ease of maintenance).
- Linguists and file sponsors frequently do not, themselves, have the time and energy (nor are they usually permitted the time by their managers) to create such a mechanized aid alone.
- I have felt that the best way I could help to fill these real needs, in the absence of any technical solution for the problems, was to do for the file sponsor all the machine-oriented things that did not require his linguistic expertise or his knowledge of the problem. Since I had a background in linguistics and a passing acquaintance with a number of natural languages, I could also help by doing some collating and editing of a relatively mechanical sort, and in planning the file so that it accomplished what the sponsor needed. With a generalized system like CAMINO, I could perform these services across-the-board and in an interdisciplinary way for many users, and thus free them for the skilled linguistic and analytic work they, and only they, are qualified to perform.
- For all these reasons, CAMINO was designed as a conventional file-maintenance system using a standardized header-trailer format. The format was so planned as to accommodate the widest possible scope of user needs. I have also made it a practice to help the file executives in any way I could as they were building their files, removing from them as much as possible of that initial burden.

I hope that the systems of the future will solve some of the problems mentioned in this paper. A good solution will become a more and more acute necessity in the next few years, as the Agency's growth is constrained and the volume of our problems expands. Language analysts will have even less time -- to allocate to many more activities -- than they now have. In particular, I hope that each analyst will soon have a terminal at or on his own desk, for his own use, and that the computer system will provide to each individual in the entire range of its users an interface that truly makes possible a "personal computing facility."

A central issue is this matter of the "user interface": the language and set of conventions the user has to learn in order to build, maintain, and query a natural-language file (or, indeed, to do anything at all) on a terminal. Unless this interface is so designed as to be as simple and natural as possible, and, in fact, to be minimal or vanishing as an obstacle to the linguist, it will be an unfair imposition upon him to force him to use the system. The language should be so transparent as to be invisible, and the user *should have to learn only one such language* to do everything that he needs to do (not many, as at present: M-204, TILE, DBMS, etc., plus a plethora of operating systems, file descriptions, editors, control languages, etc.). The terminal and the system must be truly user-oriented, in a way that can come only from a careful, open-minded study of what users really do and what they really need, on their own terms, and not a study made by computer specialists who have no interest in or knowledge of anything but hardware and software, as if they were ends in themselves. If we do not have such user-oriented systems, I would still consider it cost-effective for the user, even in 1985, to turn his back on the shiny new machine and stick with his printout and his pencil.

MACHINE-PRODUCED AIDS FOR THE LINGUIST



The following is the conclusion of an article which started in the September 1976 issue of CRYPTOLOG.

A. J. Salemme, P16

Backward Listings

Backward listings (also called, at NSA, "backward dictionaries" or "rhyming dictionaries"[1], and, in the academic world, "a-tergo dictionaries") are "backward" in that, unlike ordinary dictionaries (which list entries that are alphabetized on the first, then the second, then the third, etc. letter), they list entries alphabetized on the last, then next-to-last, then third-from-last, etc. letter. Such listings, therefore, have the following appearance[2]:

OUCH
CROUCH
TOUCH
UGH
BOUGH
COUGH
LOUGH
ENOUGH
ROUGH
MEDIUM-ROUGH
THROUGH
SOUGH
TOUGH
EXTRA-TOUGH

[redacted]

It is easy to see from the above examples that the most commonly used term "backward dictionary" is, strictly speaking, a misnomer because dictionaries include either definitions (in the same language) or translations (into another language) -- that is, no true "dictionary" is a simple listing of words either in normal or reverse alphabetic order. The term "rhyming dictionary" is a double misnomer, since words ending in the same letters do not necessarily rhyme.

The examples demonstrate that backward dictionaries (to use the common term) while differing from ordinary dictionaries in omitting definitions or translations, are similar to ordinary ones in that they usually list just one form of each word (e.g. in English, the singular form of the noun, the verb infinitive, etc.; [redacted] Therefore backward dictionaries are not entirely useful when dealing with inflected forms (such as occur in English [redacted]). For example, a person looking in a backward dictionary for English words ending in -CHES must not confine his search to specifically that location (where he will find just a few technical or archaic plurals such as LEIOTRICHES and GRAMOCHES), but must know enough English to refer to the words ending in -CH and also those ending in -CHE, and must then be able to ascertain which of those words can have inflected forms in -ES. He will, in other words, look in the backward

[1] In pre-NSA days an irreverent cryptolinguist applied the term "B.A. dictionary" to an embryonic (hand-sorted!) backward listing of [redacted] words. When the term gained currency and he was asked at a formal meeting to expand the abbreviation, he was quick on his feet and came up with an acceptable substitute, "backward alphabetic." But that expansion never had the appeal of the original one.

[2] Or, rather, they *should* have this appearance. When one is discussing backward listings, one must make absolutely certain that the listener has a clear mental picture of what the speaker has in mind. The speaker may have to use his hands lavishly to explain, as when defining a circular stairway. Several years ago a completely printed and bound NSA publication containing [redacted] place names had to be scrapped because of a misunderstanding of what "backward" meant. The place names, instead of being alphabetized and *printed* from right to left, were alphabetized correctly but printed from left to right. The [redacted] place names printed in the form

[redacted]

were as useless to the [redacted] cryptolinguist as listings such as KROY WEN and NODNOL WEN would have been to a person desiring a backward listing of U.S. place names. No, it was pointed out to the originating element, there was absolutely no way that the incorrectly printed listing could be used -- not with mirrors, not with photographing the pages and reversing the image, not with nothing! The words in a backward listing *have to be* printed from right to left.

dictionary in vain for the forms CROUCHES, TOUCHES, HEADACHES, but will have to create them for himself.

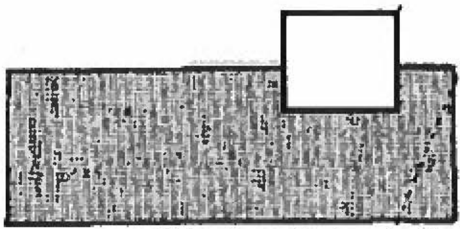
[redacted]

Another limitation is that, like an ordinary dictionary, a backward dictionary does not list all personal and place names. Therefore a person looking for *all* the English words ending, for example, in -INS, must not only look in the backward dictionary (under -INS and under -IN), but must also refer to backward listings of English personal names (if he wants to find PERKINS) and place names (if he wants to find PLAINS[3]), just as the [redacted] cryptolinguist must refer to a backward listing of personal names if he wants to examine all the [redacted]

One advantage that a backward dictionary has over an ordinary forward-listed dictionary is that its format enables the user to run his finger down the columns easily to look for *word patterns* without the optical distraction of intervening definitions or translations. Therefore, a person looking, for example, for a right-to-left pattern ....ABBA in a [redacted] backward dictionary can relatively easily find such words as:

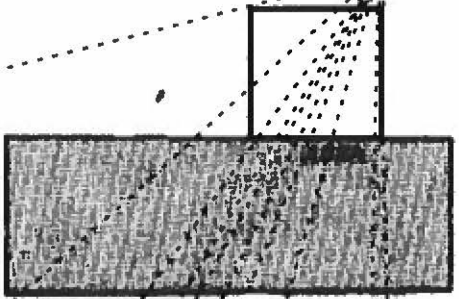
[redacted]

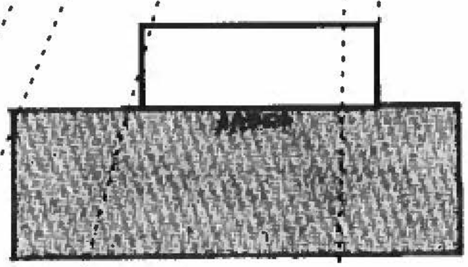
Such word patterns can be located easily by the use of a simple template (a card with a cutout, or, preferably, with the desired pattern written on the edge) that is run down the columns of words, as shown below:


[redacted]

Note the cutout revealing only the [redacted] (sounds like [redacted] doesn't it?)

As simple as this method is for finding word-end patterns, it still is time-consuming when one considers that, for example, [redacted] Moreover, the search for word patterns in backward dictionaries is even more time-consuming when one is looking for patterns *within* the words, rather than just at the end. For example, the search for the pattern ....AABCC...., goes relatively slowly until one finds such an occurrence as the following:


[redacted]


[redacted]

The tedious job of sliding the template down each column of the [redacted] backward dictionary, and then offsetting it one letter position and repeating the process, etc. in a search for word patterns could, of course, be performed by a simple machine process that uses the magnetic tapes already in existence for the backward dictionaries in several languages. It is, therefore, strange that word-pattern listings have not yet been produced from the many NSA backward listings [redacted] One virtue of any word-pattern listings so produced would be their consideration of a much more extensive vocabulary than the carefully preselected vocabularies that typify the existing word-pattern studies. That virtue, of course, would have to be weighed against the major virtue of the existing word-pattern studies -- [redacted] For any word patterns derived from a backward dictionary would necessarily reflect the peculiarities of the words

[3] When I wrote this article back in 1969, I used THE PLAINS, thinking of Ohio. [redacted]

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

in their dictionary form only, rather than in their variously [redacted]

Backward listings available at NSA: The backward listings [redacted] available at NSA are mostly NSA-produced, but also include a few publications produced by commercial publishers or the academic community. The languages currently represented are English, [redacted] Languages for which NSA backward listings of various types are planned in the future include [redacted]

Window Indexes

The term "window index" is applied to a type of linguistic reference aid that is usually called, in academic life, a "concordance"[5] and, in information-retrieval circles, a "KWIC [key-word in context] index." The typical format of each line of a window index includes message identification information (system, date, message number, etc.), plus the indexed word in alphabetic order in the "sort field," with as much preceding and following context as will fit into the left and right "context fields." The sample below was constructed from contexts taken from the now defunct *Washington Daily News* (DN) and from the *Washington Post* (WP). The second column in the sample indicates the date (year-month-day) and the third column indicates the location of the context (1st position -- newspaper section; 2nd and 3rd positions -- page; 4th position -- column).

The basic features and uses of a window index are apparent from the above sample. To discuss the features first, it is apparent, that, when one finds a word in which he is interested (say, ALABAMA) and wishes to have the entire context or contexts in which it occurs, he can either go to the original source (say, DN 690521 -091), or, if that source is not immediately available, he can recover the entire context by taking the first and last complete word in the window-index line, looking them up separately elsewhere in the sort field to get additional text, and repeating the process until the first and last words of the text are reached. To obviate this laborious process, window indexes are usually accompanied by "message prints" -- complete texts of the messages, sorted in some order (date-time, message-number, etc.) for instant reference.

A second immediately apparent feature of the window index is that, since one line of text is printed out for each indexed word occurrence, the size of the index will be inordinately enlarged if each and every word in the text were indexed. Therefore, to cut down the number of lines printed in the window index, a preselected list of nonsignificant but high-frequency words (in the above example, such words as AT, AND, IN, BY, THE, ON, INTO, WAS) usually is programmed to be excluded from the indexed alphabetic column (*but still retained in the left and right context fields*[6]).

```
DN 690521 -013   Y PROCEEDINGS.  AN    ABA SPOKESMAN IN WASHINGTON
DN 690521 -503   OING ON.  IN ADDIS    ABABA I LOOKED UP MARIO BUSC
DN 690521 -503    WAS GOING ON.  IN    ADDIS ABABA I LOOKED UP MARI
DN 690521 -206    100 FEET INTO THE    AIR, FROM SPREADING TO NEARB
DN 690521 -091   TING.  THIS COVERS    ALABAMA, GEORGIA, LOUISIANA
DN 690521 -211   VE MAIL REGULARLY.    ALTHO HANOI SIGNED THE CONVE
DN 690521 -012   ., TODAY ASKED THE    AMERICAN BAR ASSOCIATION TO
DN 690521 -252    FAIR PRICE TO THE    AMERICAN TAXPAYER.
DN 690521 -031                         APOLLO ASTRONAUTS THOMAS STA
DN 690521 -466   AT EVEN BEFORE THE    APOLLO LUNAR ORBIT IN DECEMB
DN 690521 -231   AWARE THAT CERTAIN    AREAS IN THIS CITY HAD BEEN
WP 690521 A011    WAS RESCUED BY AN    ARMY HELICOPTER.  W. BRUCE W
DN 690521 -052    QUIPPED, CAJOLED,    ASKED QUESTIONS AND TRIED TO
DN 690521 -012   AMS, R-DEL., TODAY    ASKED THE AMERICAN BAR ASSOC
DN 690521 -012   D THE AMERICAN BAR    ASSOCIATION TO DETERMINE THE
DN 690521 -031               APOLLO    ASTRONAUTS THOMAS STAFORD, J
```

[redacted]

[5] "An alphabetical verbal index showing the places in the text of a book or in the works of an author where each principal word may be found often with its immediate context" (Webster's Third Edition).

[6] Several years ago, [redacted] with little intelligence content that were to be "suppressed." When the linguist returned to his desk and recalled the cheerfulness with which the programmer promised to suppress the words, the linguist rechecked and was abashed to learn that the programmer was planning to suppress those little buggers into oblivion whenever and wherever they occurred.

A third feature of window indexes is that they contain the words as they actually appear -- note, for example, the typical *Daily News* spelling ALTHO. Therefore, the variously inflected forms of the same word will appear in different places in the window indexes (the occurrences of the [redacted] for example, may be separated by many pages from the same place name in the inflected form [redacted]). Moreover, any nonstandard spellings, misspellings, or typographical errors will be indexed as though they were valid words. A mechanically-produced window index of the *Daily News*, for example, would contain many entries such as . . . IT WAS REPROTED THAT. . . Similarly, a mechanically-produced (that is, unedited) body of [redacted] would contain a very large percentage of nonvalid words [redacted] spelling errors, run-together words, word segments, etc.). A [redacted] window index of unedited [redacted] might therefore contain invalid entries such as the following:

[redacted]

[redacted]

[redacted]

The uses that are listed above are certainly not arranged in any order of ascending or descending importance, since it is felt that all the steps involved in the collection, forwarding, decryption, translation, interpretation, and intelligence reporting of traffic can justifiably be called "the critical step."

Window indexes available at NSA: Since window indexes have such a wide range of cryptolinguistic and intelligence use, they have always been among the first machine aids requested by cryptanalysts dealing with a specific language. Therefore NSA and its predecessor and cooperating agencies have produced, over the years, a large number of window indexes in most of the languages of vital intelligence interest. In addition to differing according to language, these window indexes differ greatly according to [redacted] and the physical form of the output (many of the older, mainly pre-edited indexes are in the form of printouts that, while still rather bulky, can be referred to like an ordinary thick book, but many of the more recent, unedited indexes containing such a

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

large number of nonsignificant and/or garbled entries in the sort field that any printout would be physically and psychologically unwieldy, are printed on microfilm reels, which can be read on a viewer and partially reproduced as desired, or retained in the computer and called up by CRT or other programs.

Conclusion

While the types of language aid described in this article can be produced for probably all the languages of NSA interest, they have not yet (1976) been produced to the same extent for all such languages. For example, the Agency has produced the following [redacted] For a number of administrative-technical (rather than linguistic) reasons, linguists specializing in languages other than [redacted] do not have such a complete assortment of backward listings or other machine language aids available to them. The situation is analogous to the situation with regard to [redacted] contrast sharply with the small, individual hand files (and even mental files) in other language areas. Is it because some languages are more important to the Agency than others? Or that some are more difficult to degarble or decipher? Or that there are only a few thousand, rather than a few hundred thousand, place names involved in one type of traffic? Or that some language areas have a larger number of people available to edit materials for compilation into machine-produced aids to benefit themselves and any linguists who might come after them? Certainly *any* window index in language *x* is better than *no* window index in that language. And a window index of diplomatic text in language *x* and a backward listing of [redacted] in that language are even better when one is attempting to translate a garbled or partially recovered text in that language.

The argument that backward listings and window indexes (and, for that matter, language files) are extravagances that cannot be produced right now because our few linguists are too busy producing translations is in the same category as the argument that none of the currently producing (but linguistically undertrained) personnel in a particular language area can be spared for training that would appreciably improve the overall quality of the output, or for recording new, ambiguous, unknown, or otherwise recordable linguistic data that have occurred in traffic and might conceivably reoccur (perhaps on a day when the one person who might remember the specific message in which a particularly obscure word occurred six months ago is now on two-weeks annual leave).

But how does the file-building, aid-producing Ant convince the blithely translating Grasshopper that one of these days he's going to wish he had similar files and language aids? Especially when the Grasshopper tells the Ant, "It's *easy* for you to build *your* files and make *your* backward dictionaries, because you've got a highly organized society with hundreds of specialized helpers cranking out transcripts, [redacted] and translations, but I'm the only one I've got to rely on -- I do the [redacted] [redacted] transcription, translation, and reporting, and I'm *also* supposed to act nonchalant." It's an old problem and many of the Grasshoppers do realize that it is a big one. So let's not be too critical of their outwardly nonchalant air and let's hope that they can squeeze out a minute or two each day to produce the linguistic aids that we know they will need when the cold wind of winter starts to blow.

~~(TOP SECRET UMBRA)~~

[redacted]

Non - Responsive

Non - Responsive

[redacted]

Non - Responsive

Non - Responsive