

THIS DOCUMENT CONTAINS CODEWORD MATERIAL

SECRET

CLASSIFIED BY NSA/CSSM 123 2 DEGLASSIFY ON 1 JUNE 2009

CCTTOLOG

Published Monthly by P1, Techniques and Standards, for the Personnel of Operations

PUBLISHER	WILLIAM LUTWINIAK
BOARD OF EDITORS	
Editor-in-ChiefDavid H	. Williams (3957s)
Collection	(8555s)
Cryptanalysis	(4902s)
Cryptolinguistics	(5981s)
Information Science	(3034s) P.L.
Language	(8161s)
Machine Support	(5084s)
Mathematics	(8518s)
Special ResearchVera R.	Filby (7119s)
Traffic AnalysisDon Tau	rone (3573s)

For individual subscriptions send
name and organizational designator to: CRYPTOLOG, P1

TEACHING COMPUTER SCIENCE TO LINGUISTS

hv	P16
υy	LIO

ELCITRA ERITNE SIHT
...UPMOC YB TES SAW
TES...UPMOC YB TES
MOCYBUPMOCYBUPMOCYB

P.L. 86-36

Consider the plight of the NSA linguist. He entered his chosen field to "ling", i.e. to work with foreign language material: to translate, to transcribe, or perhaps to analyze and report the significance of large amounts of text. He chose to deal with the fuzzy world of ambiguous meanings, of convoluted and unpredictable rules of grammar, and with the imprecision inherent in the transference of an idea from one language into another. He never cared much for the picayune rigor or the grubby technical details of engineering or the physical sciences — they just weren't appealing to him. He felt at home in that imprecise world of meaning that is so foreign to most Americans. And then he came to NSA.

Here the linguist must deal on a daily basis with computers. (As if it weren't bad enough that the phone company and his insurance agent used the darn things!) In fact not only does he have to *deal* with computers, he has to actually <u>rely</u> on them! More often than not, they

couldn't get the machine to print out your daily take separated in an ever so slightly different way.

The worst experience of all awaited the rare brave linguist who got involved in the design of a new computer system for his office. The project development people seemed to be a special breed of programmers whose incomprehensibility was matched only by their desire to document in a level of detail that baffled the minds of ordinary folk. Even though it is considered almost axiomatic that projects which don't intimately involve the proposed end-users from the very beginning are doomed to failure, the linguist finds participation in planning extremely difficult because of the "computer-ese" language barrier and because of the lack of understanding of language work by others. Once someone even asked a linguist on such a planning team if he really needed all 32 letters of the Cyrillic alphabet and couldn't he get along with just 25 or so because of computer limitations









provide his daily material for translation and store the older material. God forbid that he should have to actually enter his translation or transcription into them, for he has seen the words: ENTIRE FILE DELETED on more than one occasion after spending an entire day laboriously entering his work keystroke by keystroke!

But even worse than the computers themselves are their keepers: programmers - people who really have no comprehension of language work and who are always muttering about "saving bits" or something equally obscure, when all you really wanted to know was why you Because of this culture shock in going from the language world to the so-called electronic office and because of the tremendous improvements that are possible when linguists are included in the planning for the computer support for their work, the idea of an introductory (and terminal) course for linguists in computer applications to language processing was born. Such a course was developed and subsequently tested on two groups of linguists. The results of those experiences offer many interesting revelations about the nature of the linguistic point of view we that of computer science.

Before some of these experiences can be detailed, a brief explanation of the newly developed course is needed.

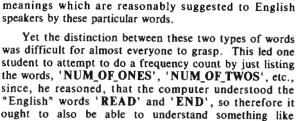
Credits

a) Doonesbury cartoons reproduced with permission. Copyright 1972, G.B. Trudeau/distributed by Universal Press Syndicate.

b) This article was prepared using the B7700 CANDE word processing system and a text composition system being designed and implemented by P16, and the final output of the article was done on the SEACO 1700 CRT phototypesetter in S3.

For a similar account of a linguist's first experience with computers, see Robert Wachal's article, "Humanities and Computers: A Personal View", North American Review, Spring 1971.

The course, CL-200, "Linguistic Applications of Computers", is both an introduction to computer science in general and to NSA-specific language projects in particular. The students experience the art of computer programming by learning enough about two quite different programming languages to write one small program in each language. In addition, they take tours of two computer operations areas in order to see some real machines in the flesh. The work in computational linguistics that is discussed includes the Agency's efforts in computer lexicography, computer scripting, speech processing, and automatic degarbling, as well as talks on the academic fields of machine translation of natural languages and artificial intelligence. The last portion of the course explains the Agency's project management system and the ways in which the proposed end-user can influence a new project to insure its success. A detailed outline of the course can be found at the end of this article.



names. 2 Variable names are unlike, for example,

programming language reserved words like 'READ',

DECLARE', 'FORMAT', 'PROCEDURE', etc.,

which have fixed and definite meanings to the computer;

"English" words 'READ' and 'END', so therefore it ought to also be able to understand something like 'NUM_OF_EIGHTS'!! It wasn't until all variable names in the course lectures and examples were changed to words that clearly had nothing to do with the semantics of the particular example (e.g. variables that held the frequency counts for certain characters had names like 'LION' and









Without a doubt the most difficult tasks for the students were the programming assignments. The reasons for this were not completely clear and often varied among the individual students, although two problems were shared by all. The first of these had to do with the choice of names for the variables in the programs. In high school algebra, for example, one usually prefaces a discussion of a problem with an explanation like: Let x be the number of apples that John bought. This is often not easily done in computer science, and even if it is, it is not sufficient for a large, complex computer program, as one quickly forgets what x was supposed to represent, or even what John was trying to do! Most professional programmers tend to choose names for the variables in a program that are at least somewhat suggestive of the meaning those variables have in that program. Hence a variable which denotes the position of a certain keyword within a section of running text might be named 'KEYWORD_OFFSET' or 'SUBSTRING_POSITION' if one were programming in PL/I, a language which allows very long, descriptive names, (or 'STRPOS' if one were using a more restrictive, inflexible language like FORTRAN which limits the lengths of names to six letters). Yet these expressions have no meaning to the computer. It merely stores the names in a table for future reference and sets aside a certain amount of computer memory to hold the values associated with those

'BEAR', as opposed to 'NUM_OF_AS', and individual lines in the program were given labels like 'COW', 'DOG' etc.) that the students really caught on. While such a practice is at best poor for a professional programmer, it was almost mandatory for the linguist who would have read too much into the choice of a name otherwise. I really knew that this notion had been mastered when one of my students presented me with the frequency-count program abstracted here:

KAZOE: PROCEDURE OPTIONS(MAIN);
DECLARE (ICHI, NI, SAN) FIXED BIN(15,0);
HAJIME: READ FILE(SYSIN) INTO (TEXT);
IF SUBSTR(TEXT,1,3) = 'END'
THEN GO TO OWARI;
MODORI: IF SUBSTR(TEXT,I,1) = '3'
THEN SAN = SAN + 1;
GO TO HAJIME;
END KAZOE;

2. The fact that the computer does not understand English can make it rather tolerant of the inadequacies of some programmers. In a reasonably large program I wrote a few years ago that dealt with dictionary retrievals, I was quite proud of the clear and descriptive names I had chosen and the fact that they made the program so much easier to understand. It wasn't until the program was completely finished that I found out that there was something not quite right about some of the names I had chosen, names like: 'NUMBER_OF_RETREIVALS' and 'RETREIVAL_TIME'!! Since my misspellings were at least consistent, they were perfectly "understandable" to the machine!

This program uses quite descriptive names, but in Japanese!! Since clearly the computer could not read Japanese, this student showed that he realized what the descriptive names were really for: the people who will read the program, not the computer.³

The other problem in programming that was shared by all linguists was the difficulty in the reduction of large difficult tasks to a number of small easily accomplished actions. This type of problem-reduction approach to problem-solving is that often used in high school geometry ("If I can just prove that triangle ABC is congruent to triangle DBF, then I can do the whole problem.") and in symbolic integration ("Let's try integration by parts and see if we can get two integrals that are easier to solve."). This division of a complex problem into smaller, more rudimentary pieces is imperative if one is to write a program to solve that problem. One must reduce the given problem to small, primitive problems that can be easily programmed, often in a single line.

One student was so overwhelmed by the complexity of programming one problem and accounting for all the various possibilities that the first step of his solution was: "The programmer manually checks through the input and removes the following special cases..."! Other students just couldn't get enough of a handle on a small part of the problem to even begin, but when helped in that first step were able to complete it without too much difficulty.

It is possible that this is just a difficulty encountered by every beginning programmer, though I don't remember this happening at all among the students in my first programming class, a class made up entirely of students in engineering and the physical sciences. It may be that this, in fact, points to a basic difference in the approach of a linguist. They perhaps do not analytically dissect their work, as the engineer must often do, but rather approach the whole problem intact, perhaps slightly more attentive to one particular portion at any one time, but nevertheless keeping the entire problem in focus. This is sort of a gestalt approach as opposed to an analytic one, but an approach that may be required for language work.

This idea that linguists may have a different way of attacking problems is also supported by their reactions to the two programming languages that are taught in the class. Without exception, the students felt much more "comfortable" with SNOBOL, a language that is strange to many professional programmers. SNOBOL is unlike the more standard programming languages like FORTRAN, PL/I, ALGOL, etc. and in fact was designed for the processing of text in complex ways. SNOBOL allows one to "look for" complex patterns in running text and to easily manipulate the text portions it finds. For example, the

following one SNOBOL line formats arbitrarilary spaced English text and outputs it on the line printer:⁵

```
LOOPA TEXT (POS(0) SPAN(' ') NULL . SW) |

(ARB ' ') . SW SPAN(' ') |

(LEN(60) . OUTPUT NULL . SW ' ') |

(ARBNO(BREAK(' ') ' ')) . OUTPUT

ARBNO(NOTANY(' ')) . SW POS(60) |

(LEN(1) ARB) . OUTPUT

NULL . SW RPOS(0) = SW :S(LOOPA)
```

The linguists seemed to find SNOBOL much closer to their way of thinking than the more standard (i.e. to most programmers) programming language PL/I. Since it has been hypothesized that one's natural language limits or restricts one's thinking patterns⁶, it should also be equally possible that given a choice between two artificial languages, one would feel more "comfortable" using the one that least disturbed his natural thinking patterns. To the engineer this might be one of the so-called "algorithmic" languages like PL/I or ALGOL; to the linguist it might be a totally different sort of language like SNOBOL.

But by far the most interesting observation about the linguists was the imagination and enthusiasm with which they approached a foreign field. One principle which was expounded upon often was that of Day's Law, a property of some programming languages. A programming language is said to satisfy Day's Law if anything that a programmer would ever reasonably want to do, can be done easily and directly in that language. PL/I is an example of a language that satisfies Day's Law, and FORTRAN is an example of a language that clearly does not! The students were told that PL/I satisfied Day's Law and many were able to discover some actual PL/I syntax by guessing! One student theorized that PL/I ought to have a way to assign the same value to two different variables simultaneously and guessed the syntax:

$$I, J = \emptyset;$$
 (Correct!)

and another student felt that there must be a way to assign an initial value to a variable when it is declared, and guessed the correct syntax of the PL/I keyword INITIAL. Unfortunately there were also a number of incorrect guesses (though not incorrect applications of Day's Law), most notably when one student decided that array references really shouldn't have to be numbers, but ought to be arbitrary character strings. 8 This belief led the student to try the syntax:

5. This SNOBOL program was written by	R61,	in
response to a challange by another SNOBOL programmer.	7	

6. This is the Sapir-Whorf hypothesis in linguistics and an entirely artificial 'natural' language has been designed to test it's validity. For more information concerning this hypothesis and this new language, see James C. Brown's LOGLAN1: A Logical Language, 3rd edition, The Loglan Institute, 1975.

7. Days's Law is named in honor of its proposer, former Agency employee.

This is approximately true in PL/I if one considers PL/I structures
heterogeneous arrays with non-numeric subscripts, and is definitely
true in SNOBOL with the SNOBOL table; see T.W. Pratt's
Programming Languages: Design and Implementation, PrenticeHall, 1975, pp. 64-67.

^{3.} It is not at all so clear that a computer could not "read" Japanese. To the computer, there is no real difference between the string of letters 'BEGIN' and the string 'HAJIME'. It just so happens that most computers are built to recognize only one of these words as denoting the meaning "start". There would be no real problem, in theory, in programming the computer in German, Japanese, or even Arabic. There are, however, some political problems. For more details, see Theodor Sterling's article, "Computers in Developing Nations: A Cautionary Tale", Communications of the ACM, Volume 29, Number 12 (December 1977), pp. 971-972. Nilsson, Nils J., Problem-Solving Methods in Artificial Intelligence, McGraw-Hill, 1974, chapter 4.

UNCLASSIFIED

DECLARE COUNT('Ø':'9') FIXED;

to build an array which was indexed with the literals '6', '1', ..., '9'! Unfortunately such a statement has nothing like that effect.

The enthusiasm and optimism of the students during the entire course and in the programming assignments in particular was indefatigable. Two particular cases immediately come to mind. One student's first attempt at one of the programming assignments resulted in a list of diagnostic errors that was about ten times as long as the total length of the program being tested! Yet the student was only slightly daunted and completed the assignment with only a few additional runs. Another student, upon discovery of a small syntactic error in his program, was then so confident of his work that he was then willing to wager that the program would definitely be completely successful on the very next run! Much to the student's dismay it took many more runs before the program accomplished anything even close to the required task. Had I collected on all our 5¢ bets, I would now have a small drawerful of nickels!!

Course Outline CL-266 Linguistic Applications of Computers

A) INTRODUCTION TO COMPUTER SCIENCE

Hardware, Software, PL/I, SNOBOL, NSA Computer Complexes, Computer Operations, Current and Future Trends in Programming Languages.

B) COMPUTERS AND TRANSLATION

Computer Lexicography, Machine Translation and Machine-Aided Translation, Computer Scripting, Speech Processing and Computer Aids to Transcription, Artificial Intelligence.

C) COMPUTERS AND CRYPTOLINGUISTICS

Sorting, KWIC Indices, Automatic Degarbling Schemes, Mathematical Modeling of Languages, PTAH.

D) LINGUISTIC PARTICIPATION IN THE PLANNING OF A FUTURE COMPUTER SYSTEM

Required Agency documentation (SCP Series), Linguistic Input, Case Study: Project THISTLEDOWN, Proposal Evaluation.

SOLUTION TO NSA-CROSTIC NO. 25 (CRYPTOLOG, May 1979, by D.H.W.)

"[Once] some illiterate [soul] concludes that "type" means "type of," the step to "this type thing" is immediate. Since it is well known that no error is stupid or vulgar enough to guarantee that it will not become respectable, we refrain from rending our garments. But we submit that at this period English this type writing is not appropriate to this sort journal."

Editorial, NSA Technical Journal, [January 1958, by Sydney Fairbanks, reprinted in CRYPTOLOG, April 1979]

SIGN LANGUAGE

(Reprinted from Q*R*L)

International road signs are being seen more and more frequently in the United States. For the benefit of readers who may not be familiar with them, here are a few examples.



Trouble When Windy!!



You Are Now Entering Johnnie Appleseed Country!!!



Barber



Snake Crossing!!



- Bookie Joint!!



You are on a



Illiterate natives hereabouts



Watch out for returning boomerangs!!



Only One Out of Six Will Make It!!!



Caution: Drunker Drivers Ahead!!!



Stunt Drivers
Practicing!



Beware of



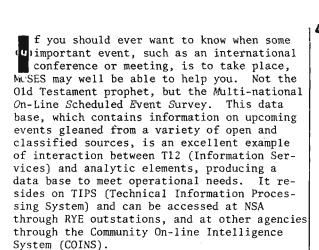
Unfriendly Spear-throwing Natives Hereabouts



Watch Your Step!!!!!

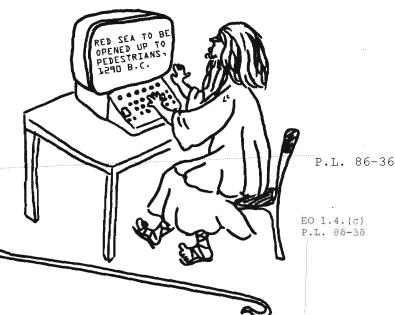
The Story of Moses •

G51



gather information on future events and put it into the data base as they find it, assigning a one-up number to each item. The way we used to get the MOSES information was to have an analyst submit queries through a RYE outstation at the end of each month. Those queries would then produce listings of selected items in item number order. But these listings were hard to use because they were not organized either by country or by event. To circumvent that problem the listings were reduced by hand to a list of events arranged by beginning dates. The list was then duplicated and distributed throughout the division.

(v) Developing those chronological lists by hand was quite a time-consuming task. This came to a halt after a reorganization shuffled away the capability to compile them. As a result, the original computer listings were delivered directly to the analysts who had become used to getting the nice chronological listings. The first time the raw computer listings were brought into our area we found out how Pharaoh's army felt that day in the Red Sea: MOSES had just closed a sea of paper over our heads!



(v) In that form, the MOSES extract was too unwieldy to use. If anyone wanted to find out when a given meeting was going to start, he might have to read through an inch-thick stack of paper before finding his answer, usually on the last page. Nevertheless, the MOSES information was still useful, so we asked for an extract: we wanted it in an easy-to-read format, sorted by host country and by the beginning date of the events.

The final format of the extract was quite usable, even if not ideally so. It was usable because it allowed people to look for their desired information quickly. After all, the main reason an analyst needs to know when a meeting is to start is to permit him to decide whether or not to publish a piece of information, which has just become available, relating to some event. For instance, if a

the instructions are history, not intelligence, and will not normally be published.

able to the person making the publish-or-not decision consists of its name and where it is to be held. With those two items, he can then quickly thumb through the revised version of the extract to the proper country, and then scan down the first word of the narrative descriptions of the events, and find the needed date.

(v) The problem then exposed was that all too often the event we were interested in was not even on the list; the format was usable, but the data base was incomplete. To remedy this a series of tetragraphs was devised,

CONFIDENTIAL

to be used in the TAG (Topics and Area Guide) After a few months experience with the line of every piece of product discussing the new easy-to-read printouts, user analysts date of any future meeting, conference or decided that calling T1221 (on extension similar event. 3258s) was even better, for a very good reason: up-to-the-minute information was Now, whenever an analyst gets advance available, rather than what had been on notice of an upcoming event, the information hand back on the first of the month. can be spread through use of one of the -SKD It's surprising how few linguists and analysts even know about MOSES.* Perhaps this article will help spread the word around. The MOSES folks can then use the SOLIS system and "pull" on the -SKD tetragraphs to get easy, rapid access to that advance notice. Once MOSES "knows" about it, anyone in the intelligence community can easily find out. *(c) Another useful file, also maintained Obviously, one of the most important sources of MOSES inputs is the translators and the checkers who keep -SKD in mind as they write excellent source of expansions of acronyms or review TAG lines. and abbreviations. P.L. 86-36

DOCID:	4019666	CONFIDENTIAL EO 1.4.(c) P.L. 86-36
:		
!		
		\

DOCID: 4019666 CONFIDENTIAL

TO START A SUBSCRIPTION TO CRYPTOLOG, OR TO CHANGE THE ADDRESS ON YOUR PRESENT ONE, CALL 3957s.

June 79 * CRYPTOLOG * Page 9

CONFIDENTIAL

- 1977bo - 8. - -

P.L. 86-36

TRAFFIC ANALYSIS

OF THE FUTURE (**)

P14



hat did you do in the war, sir? I was a traffic analyst. Oh, how interesting. What kind of cars did you analyze?

Various Agency analytic disciplines are on the decline. Take a look at the table in the January 1978 CRYPTOLOG article entitled "The Changing Face of NSA." It shows a sharp decline in numbers of people assigned in certain job fields over the previous five years. Of the Agency's analytic fields, traffic analysis showed the largest percentage drop. Then take a look at the latest Quarterly Management Report; traffic analysis has been reduced still further. Where are we going? What is happening?

(U) I'm going to stop here and make my prediction, a double-barrelled one.

There will come a time when all of what we now call traffic analysis will be centralized in a single organization in DDO (or whatever it will then be called).

This step will be taken to bring together and husband the few remaining TA skilled people, and to develop under a single line management authority, the unified technical data system and software system to support TA tasks worldwide.

(u) I don't know when this will happen. But I'm pretty sure that it will happen.

(u) Of course, I was wrong about Dewey in '48...

EO 1.4.(c) P.L. 86-36

How Are Your Stamina?

Sydney Fairbanks

Reprinted from The NSA Technical Journal of January 1959

One of our readers, stung by some peculiarly noxious idiot's-idiom that had crept into an official communication, called up the other day to ask, almost tearfully, if we thought he could volunteer to write all the memos put out in the Agency. No, we told him, the suggestion would probably not be well received; but we too, we confessed, had had daydreams of a similar czardom, lightened in our own case by the imposition of a scale of penalties. Omissions of the definite article ("subject memorandum is reprinted in referenced document") would call merely for confinement to barracks; references to the "overall picture" would involve a substantial fine; statements as to "the capability of the facility to become operational transmission-wise on a continuing basis" necessitate a painless beheading; and naturally anyone writing "the reason why this is so is because of the fact that" will be hanged, drawn, and quartered. Beyond these we progress to actual errors: "We hope you shall"; "like he did"; "this is a new one, and which"; "oil the bearings, such as we did yesterday"; and, of course, our friends "(comma) however", "this type thing", and "they are writing and notify the contractor." We are sorry, but we cannot tell you the penalties for these.

There are also the people who write, "this phenomena is noted in more than one media, and the discoverer is worthy of several . kudos,"-but we are becoming a common scold. Actually the matter of foreign plurals is not quite so simple as the purists would have us believe. "Data" which started life as a proper little plural is rapidly becoming a collective singular, and anyone adopting a holier-than-thou attitude about it should be asked how his stamina are this morning. Back formations of singulars are even more confused. A "tactic" or a "statistic" has no more right to exist than a "mathematic" or a "calisthenic"; but it does. A man joining the commandos should no more become a commando than a man joining the troops becomes a troop; but he has. And so...we suppose...from a purely scientific point of view...one has to admit the possibility that a time may come when something called a "kudo" can exist naked and unashamed; but not, we hope, until we are dust before the doors of friends, or radioactive matter a-blowing down the night.

We too can be scientific on occasion. A short while ago we lamented the absence of a good portmanteau word of spontaneous rather than deliberate generation. In accordance with Somebody's Law, which says that as soon as you say there is no such animal a perfect specimen trots around the corner, we came for the first time on the word "meld" embedded in a very distinguished matrix. Not the verb used in pinochle, which is of course German melden, to declare, but one meaning something like "amalgamate." Webster's International knew it not, and we were tempted to throw it out, but we have a weakness for monosyllables. To make a long story short, Webster's New World Dictionary of the American Language gives: "meld, v.t. and v.i. (merging of melt and weld) to blend, merge, unite." Sinking the classicist in the collector, we left it in the copy where we found it, and tiptoed away.

But this does not mean that we are prepared to accept "irregardless."

NSA - Crostic No. 26

WORDS

DEFINITIONS

Bu D.H.W.

Α.	Site	of	baseball'	s	Hall	of	Fame	
----	------	----	-----------	---	------	----	------	--

B. Dog star (comp)

C. Impending

D. Act of investiture of a monarch

E. NY town, center of winemaking activities

F. Uncultivated; uncultivated person

G. Diana Rigg's role in The Avengers (2 wds)

H. "____, my dear Watson"—Holmes' reply when asked how one should refer to a saffron-colored devilfish (3 wds)

I. Person shunned for social or moral reasons

J. Any of more than 100 fundamental substances

K. Ringing of bells

L. Why, at the Sioux bride sale, the maiden on the skin of the African river beast cost as much as one on a bearskin plus one on a calfskin (5 wds, followed by Word M)

P. Just underway (racetrack jargon, 3 wds)

Q. Candy made by boiling sugar and butter

M. See Word L (11 wds)

N. Room

O. Volunteer State

18 165 257 76 96 216 46 65 210 187 233

<u>184 68 120 260 39 56 226 83 20</u>

<u>213</u> <u>189</u> <u>194</u> <u>209</u> <u>175</u> <u>232</u> <u>115</u> <u>155</u>

21 89 64 78 179 41 103 192 219 160

231 247 241 235 109 263 171 72 16 137 162 34

100 105 53 182 94 223 117

227 132 200 169 62 80 251 32

45 29 149 6 262 172 252 86 211 166 225 238

195 52

87 119 135 112 110

201 51 178 181 253 243 208

<u>176 1 245 152 207 254 196 11 236 147 4 93</u>

101 259 151 224

the skin of the African river beast cost as 140 104 22 258 146 180 36 2 131 85 126 122

91 10 244 61 204 156 215 163 255 99 30

14 48

<u>242 13 66 31 44 264 221 118 159 214 8 246</u>

205 108 121 153 190 144 97 234 125 130 81

49 170 228 58 183 74 230 177 167 35 203

88 139 54 3 116 174 106

220 229 138 256 133 124

47 59 248 26 71 17 212 55 33

63 38 28 114 134 168 77 98 113 92 261 42 27

<u>57</u> <u>239</u> <u>37</u> <u>79</u> <u>145</u> <u>199</u>

R. Women of a Moslem household

S. Overture by Beethoven (op. 84, 1810)

T. Compass point

U. Tanzanian farewell to Miss Day (2 wds)

V. Resourcefulness; fantasy

W. Filling to the brim with a liquid (2 wds)

X. Assault; crime (Brit. sp.)

Y. Uppermost part of a building

 $\overline{173}$ $\overline{148}$ $\overline{136}$ $\overline{237}$ $\overline{158}$

95 9 90 102 193 69

 $\overline{107}$ $\overline{191}$ $\overline{123}$

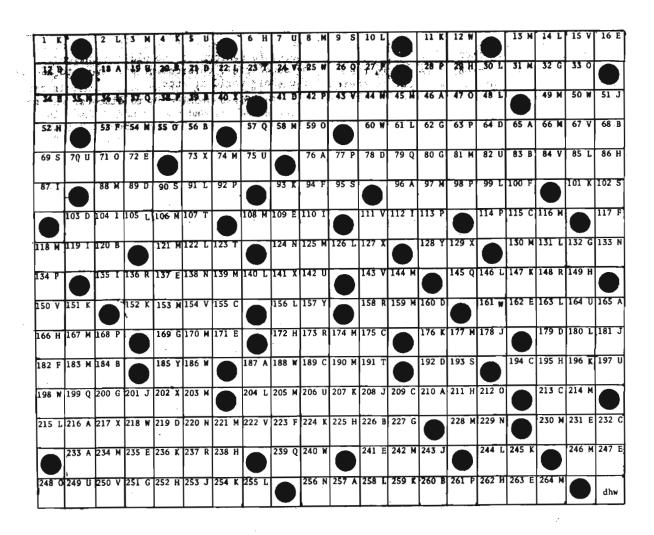
142 19 75 70 82 206 197 5 7 249 164

222 111 154 250 143 24 43 150 15 84 67

12 60 161 218 25 50 198 188 240 186

217 73 129 127 202 40 141

23 185 128 157



June 79 * CRYPTOLOG * Page 13

UNCLASSIFIED



articles* that the data standards program has encountered major problems. Mr. Pattie cites several important areas where the implementation of standards has been less than successful, and he says that the NSA Data Standards Center is just "scratching the surface." Mr. Murchison, in his final paragraph, says that "we sometimes feel that our job is impossible."

The root of the problem is not in the actual development of standards but in their implementation. Mr. Pattie blames the general user, at one point appealing for cooperation, and at another point** thundering: "Until we are authorized to deny computer use to those who refuse to accept Agency standards we can have little effect."

He believes that operational necessity
is really "operational convenience," and
backs him up: "Usually a sponsor
cries 'unacceptable' just because he does not
want to go to the trouble of re-programming."

This analysis of the cause of the problem is totally wrong. Most people suffer to a greater or lesser degree from lack of standardization and would do anything reasonable to avoid this. But standards are only justified if they bring tangible benefits. If the pain caused by implementation exceeds a reasonable threshold, then any benefit to the user will be cancelled out and we will be reduced to the untenable position of trying to enforce standards for their own sake. The rejection of "operational necessity" is the rejection of reality.

* "NSA Data Standards Center," by Mark T. Pattie, Jr., CRYPTOLOG, November 1978, and "Comment on 'Data Standards Without Tears,'" by CRYPTO-LOG, March 1979.

** Letter to the Editor, CRYPTOLOG, October 1978. To illustrate my point let us assume that NDSC is given "teeth" and can actually deny a user access to the computer if he does not adopt standards. To start with, the idea of enforcing standards only on new applications is doomed to failure because there are few really new applications. Everything we do is built on what we have done before. The new is tied inextricably to the old, and must therefore conform to the old conventions. To be successful, enforcement will have to address the whole world at once.

The first problem is to discover who is using standards and who is not. There is no easy way of doing this and the whole plan would probably founder at this point. But let us assume that a lengthy study is able to ferret out the facts. It would undoubtedly reveal that the whole Agency (even NDSC!) is guilty of incomplete standardization.

P.L. 86-36

The next problem is to determine exactly what we are going to enforce. A Groundhas at least 1000 data elements, the NDSC centralized file of data elements and data field definitions, lists only 200 standards. Shall we re-program to cater for the 200, and then, every few months, when a new standard comes along, re-program all over again? Or shall we suspend all operations until the other 800 are standardized? In the latter case, calculating from the fact that it took seven years to develop 200 standards, we will be waiting until the year 2006!

Both Mr. Murchison and Mr. Pattie are, therefore, wrong in suggesting that the general user is the villain. There are facts of life which make enforcement impossible in a large, complex organization, irrespective of the attitude of the users. I am personally in favor of standardization; yet when I had the job of establishing standards for a project, I was forced by operational necessity to be satisfied with partial standardization. If a standard exists I will adopt it. But if it does not exist, what do I do? How much time can I spend tramping around trying to find out what

P.L. 86-36

other people and projects are using? Inevitably, I invent my own names, abbreviations and codes just to get my job done before the deadline.

If we cannot blame the general user, can we blame NDSC? Of course not. As Mr. says, the development of standards is inevitably "a long and rigorous process." NDSC is staffed by able people who are dedicated to their work to an extent rarely found. No one could do better. The truth of the matter is that once non-standardization has reached a certain extent in a complex organization it is not possible to enforce standardization without causing unacceptable damage. Non-standardization is self-perpetuating. During the time it takes to standardize one data element, a dozen new non-standard uses can evolve, so that the standardization process never catches up. It is a vicious circle which can be expressed thus: Lack of standardization causes an increased lack of standardization.

Once the truths of the situation are recognized then we can open our minds to reasonable alternatives. I firmly believe that the Data Element Dictionary/Directory is the best alternative we have. My personal experience of a DED/D is that it mitigates the adverse effects of existing non-standardization by tying different naming conventions into a single data definition. But it could do much more than this. If all applications are suitably described in the Directories, then everyone will have easy access to the conventions used by other people. describes it very aptly. The Directories will show "what the current usage of data fields is along a wide spectrum of different Agency applications. Exposure to this usage will gradually lead us towards the necessary standardization." If the Directories tell me that a certain data element has six different names and no standard, then I will surely adopt one of the six, thus preventing a seventh from appearing. The process will not of itself bring about standardization, but it will certainly slow down the everincreasing tendency to non-standardization. It will reverse the vicious circle.

took me to task for using the word "magic" in my first article, and indeed I ought to withdraw it. I am tempted to let it remain because it does help to express the extent of the breakthrough in a very difficult area. NDSC will still have the labor of establishing standards. No machine can do that. But the DED/D will fill the gap where there are no standards, and will act as a positive force preventing the proliferation of non-standard uses, as well as providing NDSC with a mine of information as input into their processes. And all this without any additional sweat, toil or tears on behalf of the NDSC.

There is, of course, labor involved in the acquisition and loading of data into the Directories, but this is not NDSC's work. It is spread over the whole work force and does not bear very hard on any particular individual or office. I have sat on a terminal loading such data and can vouch for the fact that it is not very exciting, but neither is it very hard. Large, complex P.L. 86-36 files can be described and loaded within a few days, not much compared to the length of time it takes to standardize one data element. Other things must be done, such as making the DED/D well-known and easily available to everyone, and encouraging its use by providing a variety of services. It is all well within the realm of the possible.

The fact is that the DED/D is a new tool which benefits everyone without any additional effort (documentation has to be done with or without it, and, in fact, your data elements may be loaded by someone else with overlapping interests). It is not even necessary for all data from all areas to be loaded before the benefits are felt. There is no enforcement, no need for "teeth." There are no adverse side effects. Nobody can lose, least of all NDSC.

I vote we give it a try.

...Mr. Pattle Replies

suppose that if the Editor would allow it we could go on throwing barbs at one another for months or even years, but there is really nothing to be gained from that. I would propose that we bring a halt to the exchange, and I would like to apologize if I have hurt any feelings or stepped on any toes.

P.L. 86-36

What really bothers me is not the name calling or the obfuscation but the apparent failure of Agency personnel to understand what we have been trying to say about our work in the first place. ______ just happened to put on paper some of the things we have been hearing all along. In spite of our letters and articles on the subject people still do not seem to understand what the role of the NDSC is.

We do not, and I repeat, not, impose data standards on anyone. When higher authorities (DoD, NBS, and the like) agree to a standard, then NSA has to abide by that or request an exemption. We help to coordinate action on those terms. Our main job with the Agency is to work with those who want to propose a cryptologic standard, coordinate the proposal with others, and get Agency agreement to call it a standard.

It often takes a long time to do this, partly because the people we deal with are involved with data standards only as an add-on to their normal duties. We appreciate the support we do get but we recognize that data standards may not have as high a priority in their lives as we would like.

We do not expect to standardize everything in the Agency, nor do we think that necessary. Where programs have limited applicability and use, it is enough that they work. When the applications or uses are more widespread, standardization becomes a worthy goal.

It is costly to standardize and it is fairly easy to cite such costs as a reason

to keep going our individual ways. I wonder if anyone has ever given any thought to what it costs when we do not standardize.

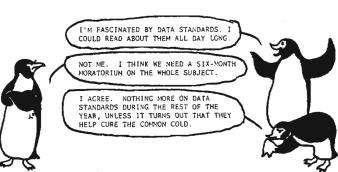
We have never spoken out against the use of Data Dictionaries/Directories. Our own PANDOWDY file is a small but

useful example of one. Our main concern is that these dictionaries seem to be proliferating when there ought to be more of an attempt to incorporate them into a single one for all of DDO to use. It appears to be a repeat of the problems encountered by users of COINS files. By not standardizing we force users to make costly searches of all possible files (directories?) to make sure nothing is overlooked.

Perhaps it might be useful to close with an illustration of the difficulty we have in the area of data standardization, even when we are talking about something that has long been standardized. This past March we received two annexes to a USSID for comment; in each there is a line with a non-standard date (DDMMY) instead of the standard YYMMD, which was implemented in 1971.

How could someone, eight years later, put something like that into a program? Why wasn't this noticed by someone else long before it got to the NDSC? I can only assume that no matter how many data standards exist, there will always be people who will want to go their own ways,

and we will continue, as before, to be accused of getting in the way of operational necessity.



Letters to the Editor

[In the May issue of CRYPTOLOG we asked if anyone could identify a sixth language of the Soviet Union which is not written in Cyrillic, the other five being Armenian, Georgian, and the languages of the three Baltic republics.]

To the Editor, CRYPTOLOG:

I believe that the sixth language you are looking for is German. There are almost a million Soviet citizens whose primary language is German. They are the survivors and descendants of the people of the Volga German Republic who were forcibly resettled to remote parts of Soviet Central Asia in 1941, where they still live in village communities and cooperatives, speaking their own language, running their own schools, and, despite odds, maintaining a cultural integrity.

Suspected by Stalin of being potential fifth columnists, these people were segregated by sex, age and family status, and shipped off to the Asian boondocks. They were not put into prison camps, but were simply set down on the steppes and semi-deserts and left to fend for themselves. Those who survived did so through the assistance of the local tribesmen who taught them how to build sod shelters and survive the first shocks of resettlement.

In 1964 these Germans were officially rehabilitated, which restored to them certain rights of citizenship, including the right to serve in the armed forces and to hold membership in the Communist Party.

With regard to the language used by these people, in his book *National Languages* in the USSR: Problems and Solutions, the author, M.I. Isaev, states that "Soviet Germans use the same literary language that is employed in the German Democratic Republic and West Germany..."

Senior Linguist A Group (SLAG)

P.L. 86-36

To the Editor, CRYPTOLOG:

(U) I would like to complain about the recent changes in the library. Basically, what has been done is to move all mathematics and engineering volumes to FANX and leave all computer science entries at Ft. Meade. This inconveniences many people, since there are people interested in computer science at FANX and many mathematicians at Ft. Meade. At a small cost in effort the library could at least have left some of the duplicate mathematical items (of which there are many) at Ft. Meade and sent some of the duplicate computer science works to FANX.

(u) It simply is inadequate to be told that we at Ft. Meade can always order from FANX, and vice versa. For example, I often need an immediate answer to a question in, say, linear algebra or statistics. The time lost by many people will certainly reduce the efficiency of the Agency's technical effort.

(u) Even more helpful than merely splitting the duplicates would have been to try to cover the content of various topics at both branches. Thus if we have 25 works on dynamic programming, none of which are duplicates, it surely would have been possible to divide them such that the basics of the field could be found at both locations. For this option, of course, the library would have needed technical advice. As I have pointed out to the library's management a number of times, there are mathematicians in many parts of the Agency ready to lend a hand.

(v) I am afraid that the library has taken the easiest (to them) route, without fully considering the needs of the users.

The Chief of the Library replies:

the people in A who return to Ft. Meade to discover that many services they hoped to regain have been moved to FANX to make room for them.

(U) Dividing a library collection is not a simple task. The card catalog shows the location of books. If the collection had been divided as proposed, there would have been no way to know where anything was, since changing all records immediately would have

(U) I can understand the frustration felt by

changing all records immediately would have been out of the question. Under the system we chose, users can be informed of the location of a particular class of books regardless of what the catalog indicates. Moreover, in recent years very few duplicates have been purchased, and since most requests are for current books, dividing duplicates would not have provided a great service.

(v) Gradually we can change the location of particular books and the appropriate records as use shows where the demand is. Further, new books will be purchased for both collections regardless of class if the need is there and the budget and space permit. We are already doing this for books in computer science.

Chief, T121

P.L. 86-36

EO 1.4.(c)

P.L. 86-36

AND FROM AN OVERSEAS READER...

To the Editor, CRYPTOLOG:

article, "Let's Not (s-ecq) Lose Our TA Skills" (CRYPTOLOG, March 1979) made an untimely appearance here The week before it arrived, several of the people here, due to return to NSA in the near future, had attended reassimilation briefings, during which the overstrength skills problem was not only raised but was also given as a rationale for the reassignment of some to a different career field. Some of the moves were in the out-of-TA direction! It's easy enough to fall out of touch when headquarters is an ocean away, but it's even easier to be confused by the apparent contradiction between the M3 view of TA as an overage skill, and the view shared by

that TA is a field with an impending shortage crisis. Perhaps some of the 50-odd traffic analysts now in excess could be used to head off the coming shortage. The concern about the effect of transfers out of TA and the lack of new blood is shared by many. This feeling was expressed here recently by visiting managers, who stated a need for analytic talent to work in the rapidly expanding world of multichannel communications.

P.L. 86-36

(v) What all of this may boil down to is the ever-present problem of individuals holding the title but not doing the analytic job, and the apparent inability of management to cope with this issue. Good traffic analysts are hard to find, and so may be talent in other overstrength skills; the key may be quality vs. quantity, another problem that's always with us. If nothing else, article may make some managers realize that skill balance by the numbers is not an end-to-all-your-troubles elixir for reducing the workforce, and that talent returning from overseas should not be regarded as a magic

ingredient for such a brew.

Kathy Bjorklund, F8313

SECRET SPOKE

7he Baltic Encoders Anthony Reiskis



In this article, which originally appeared in Q*R*L (Quarterly Review for Linguists) for February 1974, the late Mr. Reiskis postulated a

P.L. 86-36 EO 1.4.(c)

DOCID: 4019666	SECRET SPOKE EO 1.4.(c) P.L. 86-36
	(v) There is a viable alternative which would be more in keeping with the "off-the-shelf" approach, and which would thereby be more

June 79 * CRYPTOLOG * Page 19

P.L. 86-36

EO 1.4.(c)

SECRET SPOKE

economical in dealing with the type of threat postulated here.

The following article appeared in *The Wall Street Journal* for December 1974, and was later reprinted in various NSA publications. It seems appropriate to publish it once again, as a postscript to Mr. Reiskis' article.

THE NAVAJO CODE TALKERS

Navajos weren't the first Indians used by the U.S. military to confuse foreign enemies; Choctaws transmitted orders by telephone for the Army infantry in World War I, and early in World War II Comanches were employed in similar activity in the European combat zone. But the Choctaws and Comanches conversed in their native tongues. The Navajos, on the other hand, developed a special coded alphabet of 38 symbols plus an auxiliary vocabulary of 41 other terms. It's been described by anthropologists Henry Dobyns and Robert Euler as "absolutely unbreakable."

Skilled as the Japanese cryptographers were, it's doubtful whether they would have understood Navajo even if there had been no attempt to disguise it. At that time it was virtually an unwritten language and even today few non-Navajos have succeeded in mastering its complex glottal sounds and vowel tones.

But rather than take a chance the Navajo code talkers improvised a system substituting clan names for military units, the names of birds for airplanes and fish for ships, plus a double alphabet when it was necessary to spell out proper names. The idea originated with a Navajo-speaking white man, Philip Johnston, an engineer with the city of Los Angeles who was raised on the Navajo Reservation where his father had been a missionary. During the first few months of the war, he suggested his plan to a high ranking Marine Corps officer. It was approved after five Navajos demonstrated its possibilities to Marine brass.

By April 1942, Marine Corps recruiters arrived at the reservation searching for Navajos who were physically fit as well as fluent in Navajo and English. The first group of volunteers, 29 youngsters from various boarding schools in Arizona and New Mexico, were sent to boot camp at San Diego. From there they were transferred to the Field Signal Battalion at Camp Pendleton, then assigned to Marine combat divisions throughout the Pacific.

Philip Johnston joined the Marines in the fall of 1942 and was put in charge of the code talker training program. Eventually some 320 Navajos served in combat under the program.

The code talkers served in many campaigns, usually in two-man teams conversing by field telephone and walkie-talkie to call in air strikes and direct artillery bombardment. Marine Corps archives contain ringing praise for the Navajos from commanders in the field.

The idea for a formal association of code talkers grew out of the 1969 annual reunion of the Fourth Marine Division Association, which honored several of the Navajos. Two years later the Navajo Tribal Museum, the repository for Philip Johnston's papers and other code talker memorabilia, sponsored a two-day reunion. Now the Navajo Code Talkers Association numbers more than 100 members.

C.A.A. NEWS

FRANCIS X. PORRINO

It's always hard to say goodbye to an old friend. Frank Porrino came to work at the Agency is 1955, after a tour with AFSS. At the time of death he was Chief of Al and President of the CAA. Quick, caring, good-humored, professional — these are some of the words that he brings to mind. If you ever worked near him you will remember his laugh, and the infectious way it spread to those around him.

FUTURE CAA PROGRAM NOTES

27 June. 1400 hours, Room 2W087.

A05: "Ockham's Razor—A Philosophical Approach to Problem Solving"

7 August, 0900 hours, Friedman Auditorium. Vera Filby, E41: "The Impact of the Privacy Act on SIGINT Reporting"

DID YOU KNOW?????

CIA has a collection of over 20,000 motion picture films and videotapes in its Pictorial Services Branch, Office of Central Reference. This collection dates back to 1940 and includes films and tapes in several languages. The collection ranges from foreign newsreels, documentaries and feature pictures to videotape recordings of foreign and domestic television programming of intelligence interest. The subject content is very broad, covering the entire spectrum of political, military, economic, biographic, scientific, technical, cultural and goegraphic area interests. There are tapes and films of world leaders, Soviet space flights, world-wide political and military events, military parades in Moscow and Havana, to name but a few.

These films and tapes, some of which may contain intelligence information not received in printed form, are available for NSA analysts to borrow or to view at CIA. The Pictorial Services Branch has the technical capability to play back virtually any video format. P.L. 86-36

For additional information, call

(NSA liaison at CIA) on 983-8215s
if you're at Ft. Meade, or 9983-8215s if
you're at FANX.

CLUB AND ORGANIZATION OFFICERS--

THE PAGES OF CRYPTOLOG ARE OPEN TO YOU FOR ANNOUNCEMENTS OF YOUR GROUP'S ACTIVITIES. SEND YOUR SUBMISSIONS TO P1, CRYPTOLOG, OR CALL 3957s.

UNCLASSIFIED



(Classified SECRET SPOKE in entirety)

in G9 has a problem. In the course of his work, he keeps turning up fragments of various published items, all of which are in English. Unfortunately, these scraps are so short that Bill and his colleagues have not been able to identify very many of them.

P.L. 86-36

The publications appear to come from a rather eclectic bookshelf. The subjects range from contemporary American fiction to religious mysticism to some rather dreary texts on communist political themes.

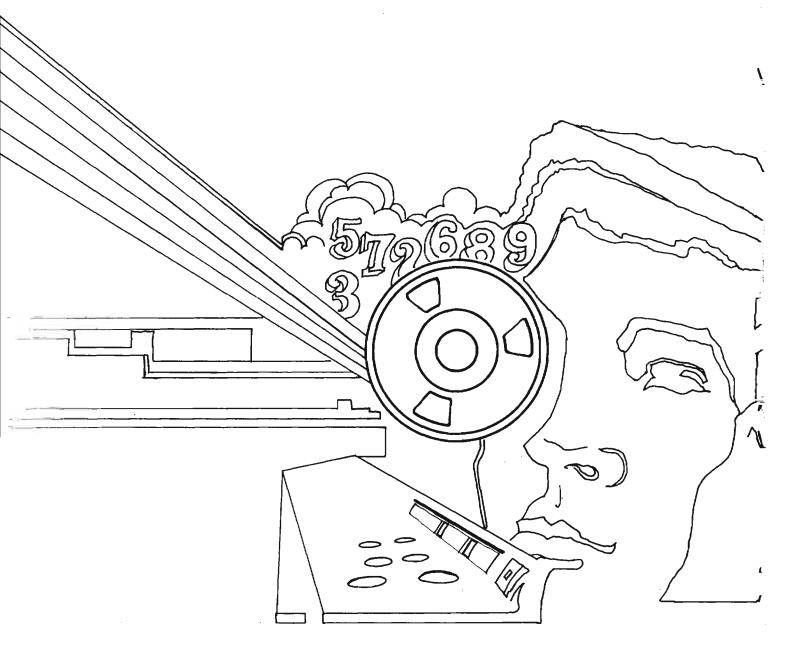
If you are a voracious reader with a wide-band approach to the selection of your reading materials, maybe you can help Bill identify some of his bits and pieces. If you'd like to give it a try (and if you have a green badge), call him on 4382s.

June 79 * CRYPTOLOG * Page 21

SECRET SPOKE

DOCID: 4019666___





THIS DOCUMENT CONTAINS CODEWORD MATERIAL



This document is from the holdings of:

The National Security Archive

Suite 701, Gelman Library, The George Washington University

2130 H Street, NW, Washington, D.C., 20037

Phone: 202/994-7000, Fax: 202/994-7005, nsarchiv@gwu.edu