Translation Tool Cracked A 1700s Code

By JOHN MARKOFF

Using statistics-based translation techniques like those used by Google Translate, a team of Swedish and American linguists appears to have cracked one of the most stubborn of codes: the Copiale Cipher, a hand-lettered, 105-page manuscript that appears to be from the late 18th century.

Discovered in an academic archive in the former East Germany, the elaborately bound volume of gold and green brocade paper holds 75,000 characters, a perplexing mix of mysterious symbols and Roman letters. The name comes from one of only two non-coded inscriptions in the document.

The first 16 pages have been deciphered and turn out to be a detailed description of a ritual from a secret society that apparently had a fascination with eye surgery and ophthalmology.

Uncertain of the original language, the researchers faltered before following their hunches. First, they assumed the Roman characters and not the abstract symbols contained all of the information.

But when that approach failed, they figured that the code was what cryptographers call a homophonic cipher — a substitution code that does not have a straightforward correspondence between the original and encoded information. And they decided the original language was likely German.

Eventually they concluded that the Roman letters were so-called nulls, meant to mislead the code breaker, and that the letters represented spaces between words made up of elaborate symbols. Another crucial discovery was that a colon indicated the doubling of the previous consonant.

The researchers used language-translation techniques like expected word frequency to guess what a symbol might equal in German.

"It turned out that we can apply a lot of those techniques to code breaking," said Kevin Knight, a computer scientist at the Information Sciences Institute at the University of Southern California, who worked on the project.

The work is being praised by other experts. "Cracking the Copiale Cipher was a neat bit of work by Kevin Knight and his collaborators," said Nick Pelling, a British software designer and a security specialist who maintains Cipher Mysteries, a cryptography news blog.

The work has value to historians who are trying to understand the spread of political ideas. Secret societies were all the rage in the 18th century, Dr. Knight said, and they had an influence on both the American and French Revolutions. He recently shared the decoded Copiale text with Andreas Onnerfors, a historian at Lund University in Sweden and an expert on secret societies.

"When he saw the book and the decoded version, he was very excited about it," Dr. Knight said. "He found a political commentary at the end that talked about the natural rights of man. That was pretty interesting and early."

Modern examples of challenging ciphers include the communications the Zodiac killer sent to the police in California in the 1960s and '70s, and the "Kryptos" sculpture, commissioned for the C.I.A. headquarters, which has been only partly decoded.

A major challenge of the code-breaking world is the Voynich manuscript. Comprising 240 lavishly illustrated vellum pages, it has defied the world's best code breakers. Though cryptographers have long wondered if it is a hoax, it was recently dated to the early 1400s.

With a University of Chicago computer scientist, Dr. Knight this year published a detailed analysis of the manuscript that falls short of answering the hoax question, but does find some evidence that it contains patterns that match the structure of natural language.

"It's been called the most mysterious manuscript in the world," he said. "It's super full of patterns, and so for somebody to have created something like that would have been a lot of work. So I feel that it's probably a code."

Found in Translation

How Kevin Knight, a computer scientist, and his team decoded the Copiale Cipher.

FIRST ATTEMPT

The team initially thought the cipher's meaning was contained in the Roman letters (*N b x*) and tried to decrypt them into 80 languages. It turned out that the characters represent blank spaces and were used to confuse decoders.

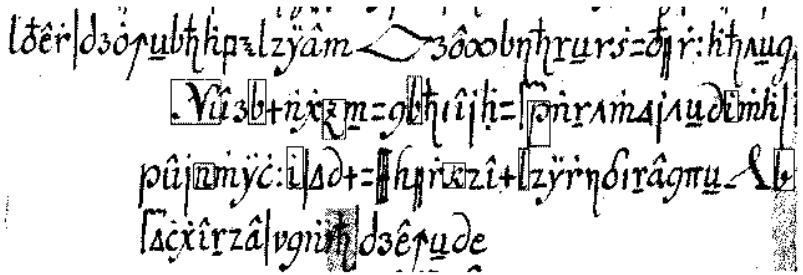
SEARCH FOR A GERMAN 'CH'

The researchers started over and tried to decode all the symbols into German, since the manuscript had been discovered in Germany. Their first step was to find regularly occurring symbols that might represent the common German pair *ch*.

SUCCESSFUL TRANSCRIPTION

After finding a potential *c* and *h*, they used patterns in German (*t* most often follows *ch*, for instance) to decrypt the cipher one step at a time. The team was unable to interpret the large symbols (), which it thinks refer to the names of "doubly secret" people and organizations.

ORIGINAL TEXT



GERMAN TRANSCRIPTION

versprochen habe die  rey ihnen zuverrahten
Er mag nun hierauf antworten was
er will so muss er dem diriginenden 
folgendes nachsprechen

ROUGH ENGLISH TRANSLATION

to disclose the  to them.
He may answer here whatever he pleases
so he must repeat the following which the conducting 
is saying

THE NEW YORK TIMES

A description of a secret society's ritual is finally revealed.