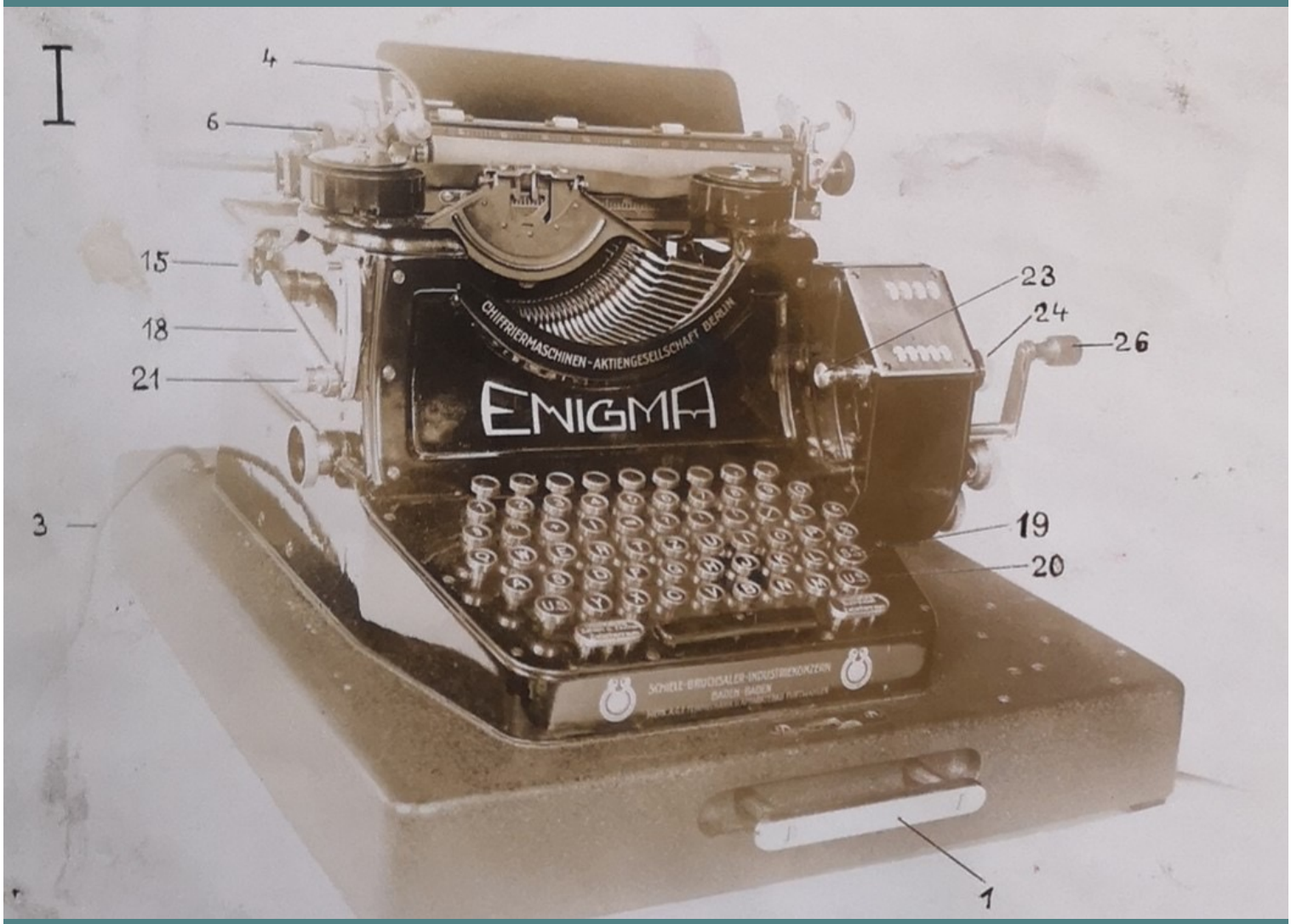
HistoCrypt 2024



7th International Conference on Historical Cryptology

25–27 June 2024, Oxford/Bletchley Park

Proceedings of the 7th International Conference on Historical Cryptology HistoCrypt 2024

Editors
Michelle Waldispühl and Beáta Megyesi

Published by:

NEALT Proceedings Series 53
ISSN 1736-6305 (online)
ISBN 978-99-1683-384-1
https://hdl.handle.net/10062/98422
https://doi.org/10.58009/aere-perennius0084
D-Space at Tartu University Library
ISSN 1736-8197, eISSN: 1736-6305

SPONSORS


ARQIT


BAE SYSTEMS


boxxe
Tech Solutions
Making
Tech Human™


FORTINET®


SECTRA


The
History
Press

Preface

The program committee is delighted to present the proceedings of the 7th International Conference on Historical Cryptology, HistoCrypt 2024. The conference will take place in Oxford and Bletchley Park, UK, from June 25-27, 2024.

Following the tradition of previous HistoCrypt conferences, HistoCrypt 2024 will address all aspects of historical cryptography and cryptanalysis. It is inherently cross-disciplinary, incorporating work from various fields such as mathematics, history, history of ideas, computer science, AI, computational linguistics, linguistics, and image processing. The conference's topics include, but are not limited to, the use of cryptography in military, diplomacy, business, and other areas, analysis of historical ciphers using modern computerized methods, unsolved historical cryptograms, mechanical encryption machines, the history of cryptography, the roots of modern ciphers in historical cryptology, linguistic aspects of cryptology, the influence of cryptography on the course of history, cryptology's connections to neighboring fields of study, and teaching and promoting cryptology in schools, universities, and the public.

The scientific program was compiled by an international scientific program committee, consisting of researchers in cryptology, history, intelligence, language technology, and linguistics. The program committee welcomed submissions in two tracks: *regular papers* up to 10 pages (including appendices, excluding references) on substantial, original, and unpublished research, including evaluation results, where appropriate; and *short papers* up to 4 pages (including appendices, excluding references) on smaller, focused contributions, work in progress, negative results, surveys, tutorials, or opinion pieces.

This year's conference received a record high of 42 submissions, a 75% increase compared to the average of 24 submissions in previous HistoCrypt conferences from 2018 to 2023. The submissions came from authors in European countries including Austria, Germany, Estonia, France, Hungary, the Netherlands, Norway, the United Kingdom, Slovakia, Spain, Sweden, Switzerland, as well as from Israel, Japan, and the United States.

The program committee aimed to compile a high-quality program with a wide variety of topics by conducting a double-blind review process. Each submission was evaluated by at least three expert reviewers in the corresponding field. The reviews were synchronized and, if necessary, thoroughly discussed among the reviewers and area chairs of the program committee. The final selection was based on their recommendations and discussions. In the end, we accepted 18 regular papers and 14 short papers for publication, resulting in a total of 32 papers (76% acceptance rate). Additionally, three abstract contributions were accepted for the poster session. All accepted submissions are included in this volume, organized alphabetically by the surname of the first author. The paper publications are presented in the first section, while the abstract contributions for the poster session are in the second section of the volume.

For the conference in Oxford, we have invited three keynote speakers who have graciously accepted our invitation. They are: *David Kenyon*, research historian at Bletchley Park and author of the book *Bletchley Park and D-Day* published in 2019; *Sarah Mainwaring*, Deputy Head Cyber Defence and Risk at the UK Ministry of Defence; and *Daniel Shiu*, Chief Cryptographer at ARQIT and former UK's Head of Crypto-

graphic Design and Quantum Information Processing at the National Cyber Security Centre (NCSC).

Furthermore, we are pleased that *Robert Hannigan*, former Director of GCHQ and current Warden of Wadham College, Oxford, will deliver the Bletchley Park lecture.

I would like to express my utmost gratitude to the numerous colleagues whose voluntary contributions made this conference possible. In particular, I want to extend my heartfelt thanks to the area chairs of the program committee: Carola Dahlke, Bernhard Esslinger, Benedek Láng, Beáta Megyesi, and Dermot Turing – for your wise advice, support, tireless efforts, and impressively quick response times! I also want to give a special thank you to Beáta Megyesi for her additional assistance in compiling the proceedings volume. Furthermore, we are deeply indebted to the 34 members of the extended program committee for generously dedicating their valuable time and effort to provide constructive and collegial feedback during the review process. Some of you took on an extra workload this year due to the high number of submissions – your dedication is truly appreciated. Lastly, I would like to express my gratitude to Arno Wacker for managing the conference website.

Our greatest debt is owed to the local organization team: Dermot Turing, Jacqui Garrad, Richard Benham, and Giuliana Forestieri. You have put together an extremely inspiring event, carried the burden of the local organization, and also managed the ticketing system. Thank you.

Lastly, I would like to extend a huge thanks to all the authors who have made these proceedings possible. Without your contributions, this volume would not exist. I wish you all an enjoyable and enlightening reading experience!

Oslo, May 3 2024

Michelle Waldispühl
Program Chair of HISTOCRYPT 2024

Program Committee

- Michelle Waldispühl (Program Chair), University of Oslo, Norway
- Carola Dahlke (area chair), Deutsches Museum, Germany
- Bernhard Esslinger (area chair), University of Siegen, Germany
- Benedek Láng (area chair), Eötvös Loránd University (ELTE), Hungary
- Beáta Megyesi (area chair), Stockholm University, Sweden
- Eugen Antal, Slovak University of Technology in Bratislava, Slovakia
- Richard Bean, University of Queensland, Australia
- Paolo Bonavoglia, Mathesis Venezia c/o Convitto “Marco Foscarini”, Italy
- Claire Bowern, Yale University, United States
- Sara Castro, US Air Force Academy, The United States
- Camille Desenclos, Université de Picardie Jule Verne, France
- Jörgen Dinnissen, Independent Researcher, The Netherlands
- Ekaterina Domnina, Moscow State Lomonosov University, Russia
- John Dooley, Knox College, United States
- Magnus Ekhall, Private Researcher, Sweden
- Karwan Fatah, University of Leiden, The Netherlands
- Giuseppe De Gregorio, University of Basel, Switzerland
- Otokar Grošek, Slovak University of Technology in Bratislava, Slovakia
- Mihály Héder, Budapest University of Technology and Economics, Hungary
- Eckehard Hermann, University of Applied Sciences Upper Austria, Austria
- Olga Kieselmann, Universität der Bundeswehr, Germany
- Nils Kopal, University of Siegen, Germany
- Harald Lampesberger, University of Applied Sciences Upper Austria, Austria
- Dominik Landwehr, Independent Researcher, Switzerland
- George Lasry, DECRYPT and CrypTool projects, Germany
- Philip Lavender, University of Gothenburg, Sweden
- Angelo Marcelli, University of Salerno, Italy

- Jakub Mirka, The State Regional Archives in Pilsen, Czechia
- Ingo Niebel, Private Researcher, Germany
- Anne-Simone Rous, State Palaces, Castles and Gardens of Saxony
- Sondre Rønjom, The Norwegian National Security Authority and University of Bergen, Norway
- Klaus Schmeh, Independent Researcher, Germany
- Betsy Rohaly Smoot, Independent Scholar, United States
- Gerhard F. Strasser, The Pennsylvania State University, United States
- Satoshi Tomokiyo, Cryptiana, Japan
- Fredrik Wallin, FRA, Sweden
- Frode Weierud, Crypto Cellar Research, Norway
- Pavol Zajac, Slovak University of Technology in Bratislava, Slovakia
- René Zandbergen, Independent Researcher, The Netherlands

Local Organizing Committee

- Dermot Turing (Local Chair), Kellogg College Oxford, UK
- Jacqui Garrad, The National Museum of Computing, UK
- Professor Richard Benham, National Cyber Awards, UK
- Giuliana Forestieri, Kellogg College Oxford, UK

Steering Committee

- Carola Dahlke (Chair), Deutsches Museum, Germany
- Benedek Láng (Vice Chair), Eötvös Loránd University (ELTE), Hungary
- Richard Bean (Secretary), University of Queensland, Australia
- Dermot Turing (Member), Kellogg College Oxford, UK
- Camille Desenclos (Member), Université de Picardie Jules Verne, France

Contents

Can Artificial Intelligence Solve the Mysterious Anagram From the Church of the Poor Clares in Bratislava?

Eugen Antal
Slovak University of
Technology in Bratislava,
Slovakia
eugen.antal@stuba.sk

Pavol Zajac
Slovak University of
Technology in Bratislava,
Slovakia
pavol.zajac@stuba.sk

Abstract

A mysterious anagram was found in the Church of the Poor Clares in Bratislava, but as far as we know it has never been successfully solved. The anagram contains 81 symbols, including specific diphthongs AE, CH, and GY. Unlike other anagrams typical of that age, the symbols are not ordered alphabetically. We suspect that a specific order of symbols is related to the original order of symbols in the plain text. Even with the suspected order of letters, the number of possible plain text candidates is too high to obtain the original text with standard methods. We examine alternative scoring methods based on modern AI text similarity to improve the quality of the candidate plain text candidates.

1 Introduction

This article focuses on a mysterious anagram related to the Order of Poor Clares. In the following sections, we briefly introduce the main historical facts about Poor Clares and the Church of the Poor Clares in Bratislava, which may be related to the anagram. We present the details and analysis of the anagram and approach how we tried to solve the anagram. We also discuss various artificial intelligence (AI) methods we believe can lead to solving this mystery.

2 Poor Clares in Bratislava and a Mysterious Cryptogram

The Poor Clares (the second Franciscan branch, also called the Order of Saint Clare) is an enclosed order of nuns in the Roman Catholic Church. They came to Bratislava from Italy in the thirteenth century. In 1297 Andrew III of Hungary donated (Szyllaba, 1944) an abandoned church and monastery to Poor Clares. They started to use them later at the end of the thirteens century after a reconstruction. In the second half of the fourteenth century, a church tower was built. The monastery burnt down in 1515, the church was also damaged by fire in 1590. The reconstructions started in 1637 by Péter Pázmány and were finished in 1640 by Losy Imre (Szyllaba, 1944).

In 1529 the Poor Clairs had to escape from Bratislava because of the Ottoman Empire's (close) siege of Vienna. None of the previous nuns returned to Bratislava. In 1541 new nuns of the order of Poor Clares were sent from Óbuda to Bratislava (Szyllaba, 1944). The nuns brought with them various treasures (gems, gold, etc.) and valuable relics (statue, altar, head of Queen Elizabeth, and other remains, Elizabeth's golden crown, etc.) of Queen Elizabeth - the mother of Louis I of Hungary (Louis the Great). In 1618 the nuns of the Dominican Order came to Bratislava and were merged with the Poor Clares (Szyllaba, 1944). These nuns also didn't come with empty hands. They brought treasures (altar, golden chalice, Saint Mary's picture with gems), and valuable relics (gold plated silver box with the head of Saint Margaret, and other remains) of Saint Margaret of Hungary. In 1660 a Saint Margaret statue was donated by György Lippay to Poor Clares. In 1714 some of the nuns moved to Buda. They took the part of treasures and relics with them (Szyllaba, 1944).

In 1618 a church bell was created by Georgius (György) Arnold and installed in the church. In 1700 the church tower was damaged by an earthquake. The Church tower was reconstructed by Abbess Balassa Eva Borbala in 1702.

The Order of Poor Clairs was dissolved in 1782 by Joseph II, Holy Roman Emperor. Most of the remaining relics and treasures were taken away. Based on Szyllaba (1944), in 1782 József Dankó a bricklayer from Bratislava was entrusted to empty

a place in one of the crypts[1], where Balassa Eva Borbala was buried. Several relics, documents, and books were hidden there. Later the bricklayer also carefully walled it. Unfortunately, there are no available direct archival documents to prove this information.

After 1782 the church was used by other people, also for other purposes such as a library (Szyllaba, 1944) or as a museum (Fiala, 2001). In 1851 the church was renovated. In 1895 various fundraising events (also a sound competition by Frigyes Dohányi and Mózes Gaál) were held for the restoration of the pipe organ. Archduke Friedrich (Frigyes) and his wife Isabella also donated money to the Church of the Poor Clares in Bratislava (Szyllaba, 1944).

In 1901 a copper plate was found (Szyllaba, 1944) in the crypt of the demolished sacristy. This plate contains a cryptogram, the content of which is still unknown. Based on (Fiala, 2001) the plate could be dated to the eighteenth century. However, there was no official date estimation performed. It can be connected to donations or reconstructions of the church and also to the treasures hidden in the crypt.

2.1 Copper Plate With the Cryptogram

The copper plate that was found in the crypt (see Figure 1 and the Appendix) consists of 84 letters, from which the first three letters D.O.M. - the initial heading most probably stands for the phrase *Deo Optimo Maximo*. The cryptogram therefore consists of the remaining 81 letters, split into 9 rows and 9 columns.

Some letters are diphthongs and are specially printed to be able to distinguish them from other letters. The letters are ordered specially (see Section 3 for more details). Based on the letter ordering and on the letter frequencies (Section 4) we assume that the mysterious message is a Latin anagram. Please note, that some diphthongs such as GY are not common Latin, however, they can be also part of Hungarian names (such as György, Frigyes directly related to the church, and Poor Clares) or Hungarian city names. If we assume that the rumors about hiding the treasures in the crypt where the Abbess Balassa Eva Borbala was buried are true, we can guess some parts of the plain text to contain names and titles, e.g.: Balassa Eva abbatissa (which could explain some deviations in letter frequencies). It is not clear, whether the plain text also contains a date in Roman numerals. If there is a date, it cannot contain the letter X. A potential candidate is the Church tower reconstruction year, 1702 encoded as MDCCII can be constructed from the letters of the anagram.

The copper plate is currently deposited in the Bratislava City Museum, with number F1-2819. The material of the plate is copper. It is a thin plate with approx.[2] height 164 mm and width 108 mm. The weight of the plate is 120 grams. More detailed statistics of the anagram are presented in Section 4.

Figure 1: The mysterious anagram (From the collections of Bratislava City Museum, Slovak Republic)

3 Preliminaries and Notation

Let $\mathbf{m}$ denote a sequence of N letters $(m_1, m_2, \ldots, m_N)$ from alphabet A. We will call it an original phrase in short. In the examples, we will typically write the letter with punctuation given by spaces, but from the mathematical and computational point of view, spaces are ignored.

Let us denote the index set $I = \{1, 2, \ldots, N\}$. Let π be a permutation on I. Anagram of $\mathbf{m}$ is

[1] The church and the monastery had a total of three crypts.

[2] The plate does not have the same width (also height), there is a difference of a few millimeters.

any sequence $\mathbf{m}_\pi = (m_{\pi(1)}, m_{\pi(2)}, \ldots, m_{\pi(N)})$. Let $\leq$ denote a natural order of letters in the alphabet A. An anagram $\mathbf{m}_\pi$ is normalized, if $m_{\pi(i)} \leq m_{\pi(j)}$ for each $i \leq j$. Note that if some letters in $\mathbf{m}$ are repeated, there are multiple permutations π that produce the normalized anagram of $\mathbf{m}$.

To solve the anagram, we want to recover the original phrase, given $\mathbf{m}_\pi$. However, due to the composition of permutations, any $\mathbf{m}_{\pi'}$ can be potentially considered a valid solution of the anagram unless we have a method that can identify the original $\mathbf{m}$. The identification can be based e.g. on the grammar structure of the message, or related to a specific meaning of the message. We will use the term "anagram solving" in a weaker sense: our goal is to find any $\mathbf{m}_{\pi'}$, which satisfies some criteria, e.g. syntactic and semantic consistency with a target language and/or expectations of the meaning of the message.

For each sequence $\mathbf{m}$, we can define a function $pos_\mathbf{m} : A \to I \cup \{\infty\}$ as follows: If $x \in \mathbf{m}$, $pos(x) = \min\{i;\ m_i = x\}$, otherwise $pos(x) = \infty$. Any sequence $\mathbf{m}$ thus defines a specific order $\leq_\mathbf{m}$ of letters from the alphabet A: $x \leq_\mathbf{m} y$ is and only if $pos_\mathbf{m}(x) \leq pos_\mathbf{m}(y)$. An anagram $\mathbf{m}_\pi$ is ordered, if $m_{\pi(i)} \leq_\mathbf{m} m_{\pi(j)}$ for each $i \leq j$.

Simply said, the ordered anagram preserves the order of letters in the original phrase. An example of the ordered anagram of the phrase `MYSTERIOUS ANAGRAM` is `MMYSSTERRIOUAAANG`. In the ordered anagram, the prefix of the original phrase is partially visible, which can help in solving the anagram. Moreover, if we consider ordered anagrams, the space of potential solutions is reduced, because only some of the $\mathbf{m}_{\pi'}$ are ordered.

4 The Analysis of the Anagram

The first step of the solving process involves the study of the material and preliminary analysis. The anagram consists of $N = 81$ symbols, out of which $m = 20$ are distinct. Namely, the symbol counts n_i (in order) are:

```
{'p':  2, 'l':  3, 'v':  6, 'm':  4,
 'b':  5, 'e':  7, 'a': 12, 'c':  3,
 's':  7, 't':  8, 'n':  4, 'o':  3,
 'i':  5, 'ch':  1, 'r':  3, 'ae':  3,
 'd':  2, 'f':  1, 'gy':  1, 'h':  1 }
```

The number of symbols is too low to determine the plaintext language with statistical significance. However, from the context of the anagram, we suspect that the language of the inscription is (medieval church) Latin. This hypothesis is based on the artifact location and dating, as well as the initial heading D.O.M. (Deo Optimo Maximo). Inscription features special diphthongs AE, CH, and GY, with only AE being native to Latin. We suspect that CH and GY can either be local spellings of original Latin words (e.g. CHRISTUS, AEGYPTUS). Another hypothesis is that these letters are parts of the names of places or people.

4.1 Key Space

If the order of symbols is not significant, it is possible to construct C anagrams from these symbols, where:

$$C = \frac{N!}{\prod_{i=1}^{m}(n_i!)} = \frac{81!}{2! \cdot 3! \cdots 1!} \approx 2^{285}.$$

The formula for C is known: we count all possible orders of all symbols (numerator) and divide out the orders of repeated letters (denominator).

If the order of symbols is significant, the formula can be derived recursively as

$$C_{j-1} = C_j \cdot 1 \cdot \frac{\prod_{k=1}^{n_j-1}(k + \sum_{i=j}^{m} n_i)}{(n_{j-1}-1)!}.$$

Here C_j is a number of possible sequences of symbols from j-th symbol up to m-th. We must to prefix these sequences with a $(j-1)$-th symbol (only 1 choice). Then we can place any of the remaining $n_j - 1$ of the same symbols between already placed $(k + \sum_{i=j}^{m} n_i)$ symbols. Because of repetitions, we need to reduce this count by a factor of $(n_{j-1}-1)!$ (the number of repeated symbols except the first one on a fixed position).

In comparison to the general anagram, we are restricted in the placement of the first symbol in each group. The closed (non-recurrent) formula thus becomes

$$C_1 = \frac{N!}{\prod_{j=1}^{m}(n_i!)} \cdot \frac{\prod_{j=1}^{m-1} n_j}{\prod_{j=1}^{m-1}(\sum_{i=j}^{m} n_i)}$$

Thus, under the hypothesis that the order of symbols in the original anagram is significant, the search space is $C_1 \approx 2^{232}$. This space is thus too large to be efficiently enumerated and computationally explored by exhaustive search.

5 Basic Anagram Solver

In order to solve the anagram, we need more sophisticated methods than an exhaustive search.

One option is to use meta-heuristic algorithms that explore the space of 2^{232} symbol permutations and try to maximize a likeness to Latin text using some suitable fitness function.

An alternative approach is to focus on complete words and phrases instead of symbols. We try to concatenate textual elements that can be contained in the original anagram while verifying the validity of the solution. Basic validation is provided by the order and number of used symbols. An intelligent search should also focus on textual criteria such as Latin grammar and textual meaning.

5.1 Text Preparation

The preparatory phase for the search is the identification of individual words and short phrases that can be a part of the anagram.

In the initial experiments, we use a simple corpus obtained from two sources:

1. Latin Bible translation (Vulgata) (Bible, 1976) — this was used for initial experiments due to the potential religious context of the anagram,

2. Georgius Agricola : De re metallica/Liber VII (Agricola, 1556) — this is due to the fact that the text potential starts with the word PLVMBVM (plumbum, lead), and a potential partial decryption used in Google search pointed us to Agricola's treatise,

3. The Bull of Pope Alexander VI on the canonization of St. Clare of Assissi (Pope Alexander IV, 1255) — for Latin terms related to St. Clare,

4. List of personal Hungarian names from 16th century (van Nijmegen, 2002), and surnames of Hungarian nobility (de Zepetnek, 2010) — to have candidate names for the suspected author of the anagram. In our experiments, we only use names that contain the 'GY' digraph.

Before the search, the text was processed and split into individual words and short phrases (to prevent meaningless small words from complicating computer search). The processed phrases were then stored in a dictionary-like data structure for easier processing during the search.

5.2 Beam Search

It is computationally impossible to examine all possible (ordered) anagrams of our cipher text. Instead, we can use an intelligent search. The first algorithm we tried is greedy depth-first search with cut-off (a beam search): Start from some phrase, identify potential continuation, score them, and then continue recursively with the continuation candidate that has the highest score.

During the search, we keep track of the current partial phrase, remaining fixed-order symbols, and free symbols. Fixed-order symbols can only be used in a given order, while free symbols can be combined in any order. We have a list of words and partial phrases we evaluate as potential continuations of the given partial phrase. First, we check whether the new partial phrase can be concatenated with the previous one while remaining a sub-anagram of the target. For each potential continuation, we again obtain new remaining fixed order symbols and free symbols and assign a score that is used in the search.

The score assigned to the continuation candidate should reflect its suitability to help the final search. In the basic version of the search, we try to score continuation candidates relative to the purposes of reconstructing the ordered anagram, regardless of their meaning. After preliminary experiments, our score consists of three components:

1. if the candidate uses a new fixed order character, we assign it a flat bonus of 20 points (to try to fix the anagram order as soon as possible),

2. each character of the candidate counts as 1 point (preference to longer continuous continuations),

3. each character that remains in the set of free characters gives 10 points to the score (preference to more freedom in latter search stages).

Typically, we have a high number of potential continuation candidates in each step. To reduce the complexity, we only examine at most a fixed number $N = 10$ of candidates on each level (a so-called beam search), in the order based on their score. If the number of candidates is higher, we select N candidates with a roulette selection, where the probability of selection is directly proportional to the score. If there are no viable continuation

candidates, we use a simple backtrack. Similarly, we backtrack a level, if we examine N candidates at a given level.

5.3 Experimental Results

The beam search described in the previous section can successfully identify ordered anagrams of the target anagram on a single PC within a few minutes. Typically, the algorithm can cover almost the whole phrase up to a small number of letters that cannot be combined further and require excessive backtracking. In such a case, we simply restart the algorithm, but depending on the parameter N (branching of the beam search), we can just let the search finish on its own. The character of the search indicates, that a stochastic search, such as hill-climbing or evolutionary heuristics, would also be suitable for the purpose.

The textual quality of the found ordered anagrams is not good and reflects the selection of the corpus material. Examples of the solutions provided by the simple search:

- `Plumbum eum paululum castas et Saba necnon tibi et Abisai Nachor eboris caedet Arfad testae Agy hastae`
- `Plumbum levam clama est natus putabant assensu se nobis iacio Bebai charta Cretae additae Fogy Aether`
- `Plumbum melle accessu suas et nomina et citato es satiata peius Chobar barbae tradi fundae Bagy Aethan`

We remark that our goal is to estimate the feasibility of solving a specific anagram. A more thorough systematic future research of the proposed methods is needed to evaluate their overall suitability to solve ordered anagrams in general, and to optimize the parameters.

6 Anagram Solver Enhanced With AI Text Scoring

The critical part of the algorithm from Section 5 is the computation of the score for the candidate. The original score tries to maintain the largest possible freedom of choice for the continuation candidates and prefers longer phrases (from the corpus) to accelerate the search. However, grammar structure or meaning does not affect the score.

The rapid progress in natural language processing (NLP) gives us ready-made tools to use in text evaluation and generation. Language models can be used to predict continuations of phrases and evaluate the consistency of existing (partial) phrases. The main problem in our research is to connect NLP processing in an efficient way with a complex stochastic search within a large space of potential anagrams. If we focus on text generation, the pre-trained predictive language model focuses on text predictions that are not related to our anagram. It might be possible to train a custom language model conditioned on our anagram symbol distribution, but we believe that this would require a dedicated research and a large number of computing resources.

For our purposes, we focus on simpler language models that can evaluate text structure and language similarity. NLP models can analyze the structure of the sequence, and identify grammar sections such as noun phrases or verb phrases, as well as grammatical categories (number, case, gender, etc.). Language similarity score can identify meaningful continuations based on the similarity of the extended phrase with its predecessor. We can also use a language similarity score to compare the partial candidate phrase with some hypothetical phrase we suspect is encoded in the target anagram.

There are three main places of the NLP tools in our anagram solving:

- In the pre-processing phase we can analyze the potential corpus. We can identify larger meaningful blocks of text that can be used as candidates for continuation (such as noun phrases, and verb phrases). We can also add extra details to candidate sub-phrases (such as grammatical categories) that can be taken into account when scoring continuations.
- Directly during the continuation candidate scoring. Note that we need to evaluate a large number of continuation candidates even during the restricted beam search, thus we need to use only very basic and fast NLP functions in this phase of the search.
- In the final evaluation of the candidates. If our search algorithm can quickly provide a

large number of potential whole ordered anagrams of the target anagram, the NLP tools can then evaluate the result and identify the most promising plain texts.

In our experiments, we use the language model `la_core_web_lg` (Explosion, 2023), with the `spaCy` library (Burns et al., 2023) that provides NLP functionality.

6.1 NLP Preprocessing

The main problem with using NLP for corpus processing is that medieval Latin is a very complex language. It is not clear how to properly incorporate the grammar into scoring functions, as Latin is quite flexible with word order, and the meaning can be significantly changed by declensions. E.g. a possible start of the sentence in our anagram can be both 'plumbea capsula' (nominative), but also 'plumbeam capsulam' (accusative). The choice of the declension in the first part of the phrase will then influence the rest of the sentence structure.

It is also not clear whether the plain text is a single sentence, multiple sentences, or whether it is even written with a proper grammar structure. Construction of the algorithm that takes Latin grammar into account would require the involvement of a language expert, and it might be easier to just construct a suitable anagram by hand.

Even if we were to take Latin grammar into account, it does not guarantee that we produce a meaningful result. As an example, we have created a simple modification of our basic algorithm that alternates 'noun phrases' and 'verb phrases' (as identified by NLP processing of the corpus). Under these restrictions, the algorithm can still find the ordered anagram of the target, with a similar performance to the basic algorithm (but with a longer pre-processing phase). Unfortunately, the results are still non-sensical:

- `plumbum evacues Asub captum eas tenentem nocte solito lis sana Achis bibi Abrae datae far gyrat hae dat`
- `plumbum evacues Paulus metent colono mactatis Ananias bibe Becher stabat taedae risi fas gyrat hae da`
- `plumbum evacues aucupes celata notos meos mansi Achias intrant Bebet bibit Laed Safat aer gyrat hae da`

6.2 NLP in Candidate Scoring

In our case, a tight beam search (with low N) provides only a few candidates for the whole ordered anagram. However, if we set N higher, the search takes too much time and produces a large number of candidates with a similar beginning. It is thus more suitable to focus on using AI scoring already during the search, and not leave it to the end.

We estimate the language quality of the continuation candidates with the use of just the function for language similarity. The original idea was to use language similarity between the examined candidate and its extension. We start the process by selecting a suitable phrase prefix manually and then follow the text similarity score (the N best candidates, without the roulette selection). Unfortunately, this scoring system often leads to repetitions. Furthermore, the scoring based on a language model does not properly handle proper names (Biblical figures, Hungarian names, ...).

An example solution provided by using a simple beam search following the text similarity score:

- *`plumbea capsula`* `massa massa massa Becbuc Tabeel Benennon Atita Choo Iturei taeduit Frigy taedae Ruth`

An alternative scoring system can be based on a text similarity of the partial anagram to some hint phrase (context). The authors of this article do not know Latin language. However, we can still use LLM AI assistants for machine translation of English prompts, along with phrase and word suggestions and grammar consultations. In our experience, the main advantage of AI assistants is that they can quickly provide and modify hint phrases without excessive search of dictionaries and thesauri. However, the AI assistants lack logic and are not trained in related tasks. E.g., the assistants were repeatedly failing to provide words containing restricted sets of letters, or letters in some prescribed order. However, due to the exploratory nature of our research, we have not conducted systematic experiments in the area.

An example of the hint phrase suggested by artificial intelligence (under our guidance) is the phrase `Capsula plumbea cum sacris reliquiis in sacristia abscondita est. Laus tibi, Christe Iesu.` Using language

similarity to this phrase, our search algorithm provides the following result:

- *plumbea capsula tenet* velet committant et Sabias sanat sociabis Schaur Boma aeneae Adae Dub Frigy Hur

The structure of the anagram is better, but the phrase is again meaningless. The ending of the phrase is saturated with names from our list of Hungarian nobility names.

7 Conclusions

Artificial intelligence is rapidly evolving and can help in tasks related to natural language processing even for researchers that are not familiar with the language. However, in the current state, there are still some tasks that require an expert and cannot be easily automated even with the pre-trained language models.

One of the tasks that AI assistants are good at is the recognition that a given anagram does not form a meaningful phrase. Thus, our mysterious anagram can potentially be solved by AI by feeding final phrases to language analysis. However, there are too many candidates, even if we restrict the search to phrases from the corpus. In our manually conducted search with a larger corpus from the Latin library collection (Carey, 2023) (obtained through a corpus importer of the Classical Language Toolkit (Johnson, 2021)), there are 4 suitable starting words of the anagram ('plumbum', 'plumbeum', 'plumbea', 'plumbeam'), and an abbreviation ('p.p.'). However, there are 5023 possible continuations of 'plumbea', 68415 continuations of 'plumbea capsula', and more than 200 thousand continuations of 'plumbea capsula tenet' (with only 19 out of 81 anagram symbols used). A simple evaluation of language similarity for these continuations takes more than 2 hours on a PC.

A more viable strategy might be the application of an algorithm that works by improving, connecting, and correcting several parts of the text with semantics and context. We have done this process manually, producing a working hypothesis:

- *plumbea capsula tenet*
  vvvvmmeeeaaasttnnn oooi
  CHartAE abbatissAE balassAE
  rr fGYh
  abdita est, MDCCII

Symbols 'o' and 'i' must be used before 'CHartAE' in that order, similarly symbols 'f','GY', and 'h' before the ending. The hypothesis has the structure: *leaden box that contains* (something related to) *documents of abbatissa Balassa* (place of hiding) *were hidden, 1702*. Notice that this process requires a lot of backtracking and changes, and as we cover more letters in the anagram, it becomes harder to correct mistakes or construct meaningful words from the remaining symbols.

Even though the space of possible (ordered) anagrams is very large, with the help of Latin experts it might be possible to recover the original phrase (or at least a correct Latin phrase that makes sense in the correct context). Employment of artificial intelligence for this task might be possible, but it would require constructing and training a custom model that can handle specifics of anagram solving (such as taking into account the number and order of symbols, and specific historical context).

Acknowledgments

This research was supported by grants VEGA 1/0105/23, and VEGA 2/0054/24.

References

Georgius Agricola. 1556. *De re metallica/Liber VII*. Wikimedia Foundation. https://la.wikisource.org/wiki/De_re_metallica/Liber_VII.

Bible. 1976. *Vulgate*. Oxford Text Archive. https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/0319.

Patrick J. Burns, Nora Bernhardt, Tim Geelhaar, and Vincent Koch. 2023. spaCy Project: la_core_web_xx. https://github.com/diyclassics/la_core_web_lg.

William L. Carey. 2023. The latin library. http://thelatinlibrary.com/about.html.

Steven Tötösy de Zepetnek. 2010. *List of Historical Surnames of the Hungarian Nobility*. Library Series, CLCWeb: Comparative Literature and Culture. http://docs.lib.purdue.edu/clcweblibrary/nobilitashungariae.

Explosion. 2023. Industrial-strength natural language processing in python. https://spacy.io/.

Anton Fiala. 2001. Tabuľka záhadnosti v zbierkach mestského múzea v bratislave. *Múzeum*.

Kyle P. Johnson. 2021. Classical language toolkit. http://cltk.org/.

Pope Alexander IV. 1255. *Clara claris praeclara*. Franciscan Archive. https://franciscan-archive.org/bullarium/clara.html.

Dr. Emil Szyllaba. 1944. *A pozsonyi Klarissza-templom története*. Bratislava : Litera könyvnyomda és könyvkiadóvállalat.

Walraven van Nijmegen. 2002. *Hungarian Personal Names of the 16th Century*. Society for Creative Anachronism, Inc. https://heraldry.sca.org/names/hungarian/.

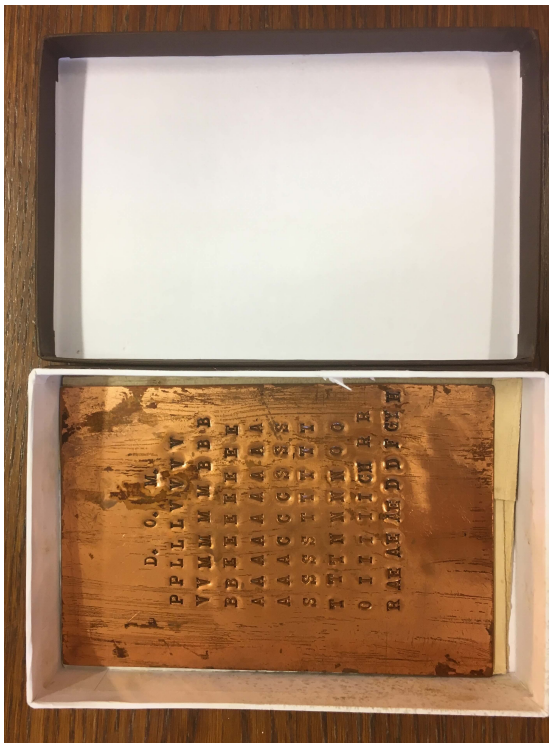
Appendices

Figure 2: Copper plate with the anagram (From the collections of Bratislava City Museum, Slovak Republic)

(a) Front side

(b) Back side

Figure 3: Front side and back side of the plate (From the collections of Bratislava City Museum, Slovak Republic)



(a) Deposited in a box

(b) Weight measurement

Figure 4: More pictures of the plate with the anagram (From the collections of Bratislava City Museum, Slovak Republic)

The Use of Volvelles in Two Early Modern Cryptography Manuals

Corinne Bayerl
University of Oregon, USA
bayerl@uoregon.edu

Abstract

This paper examines the form and function of volvelles (rotating paper discs) used to represent cipher systems in Giambattista della Porta's *De furtivis literarum notis* (1563) and in the French translation of Trithemius' *Polygraphia* by Gabriel de Collange (1518, transl. 1561). I analyze the use of volvelles in cryptography handbooks within the larger context of their overall function in the Early Modern period, and I identify the factors that may account for an increasing use of volvelles in succeeding editions of della Porta's and Trithemius' works in the second half of the 16th century. Collange's expanded version of Trithemius' manual and della Porta's numerous amended editions of his own handbook indicate that changes in the material representation of cipher systems correlate with an increased level of public knowledge about encryption methods.

1 Introduction

In the 2015 Folger Library exhibition *Decoding the Renaissance: 500 Years of Codes and Ciphers*, the cipher wheel was called "the most iconic image in the history of cryptography" (Folgerpedia 2015). While the exhibit did not include any actual handheld encryption device, it prominently featured the first extant instructions for building a cipher wheel using two copper plates (Alberti, 1568), as well as 16th-century cryptography books with moveable cipher wheels. As part of books, such moveable wheels are called volvelles, constructed out of superimposed layers of paper in the Early Modern period. Prestigious, expensive books were sold with fully mounted rotating discs, whereas in cheaper editions the readers themselves had to cut out parts of the apparatus and attach it to a page in the book, which is why volvelles are viewed today as an early example of interactive book design. (Baird 2004, Karr Schmidt 2011, Ellison 2017, Crupi 2019)

When examining the use of volvelles in Early Modern cryptography manuals, it is helpful to take into consideration their overall prominence in science books of the period. Thus, volvelles abound in 16th-century treatises on astronomy and nautical navigation and are by no means specific to books on cryptography. Recent scholarship in the history of science has drawn our attention to the fact that printed illustrations heavily influenced approaches in newly formed scientific disciplines (Kusukawa 2012). Similarly, studies in the field of book history have argued that volvelles should be regarded as instruments of science, especially in cases when the books' authors had a firm grip on their creation, either by closely collaborating with the printer, or by fabricating the volvelles themselves, as did astronomer Peter Apian when he used his own press to print the visually stunning *Astronomicum Caesareum* (1540) that included no fewer than 38 octagonal papercut instruments (Crupi 2018). While Apian remains a rare case of an author actually maintainng full agency over the material form of his own printed book, the same desire to control the printed form of their work may be discerned among Early Modern cryptographers.

In the following, I will explore possible factors for the increased prominence of volvelles in two famous 16th-century cryptography treatises:

Giambattista della Porta's *De furtivis literarum notis*, and the French translation of Trithemius' *Polygraphia*, which contains volvelles that do not form part of the original text. By exploring possible factors for the increased use of volvelles in these two books, I seek to answer the broader question of how changes in the material form of cipher systems correlate with the changes in public knowledge of encryption methods.

2 Della Porta: Fighting Contraband editions with Improved Volvelles

Della Porta's handbook *De furtivis literarum notis* is a particularly compelling text with respect to the use of volvelles in cryptography manuals. A look at the successive editions of this bestseller reveals that in the second half of the 16th century volvelles had become a prominent feature of science books in general, and of cryptography manuals in particular. Within della Porta's lifetime the book was published in seven editions, starting with the initial version in 1562. In 1591, two contraband editions appeared in London in 1591—one of them with a false imprint of Naples to make it appear authentic, and both of them featuring new, visually stunning volvelles.

After the publication of the contraband editions, della Porta felt compelled to react. He reissued several revised editions of the book in 1593, 1602, 1603, and 1606, and notably improved the usefulness of the cipher wheels for his readers. He advertised his new editions by pointing out that he had included new examples and instructions, and he also retitled the text as *De occultis literarum notis,* probably in order to distinguish it from the contraband edition that circulated under the original title.

In striking contrast to the baroque, sexually charged decoration of the cipher wheels in the 1591 contraband editions, the volvelles in the 1593 edition that La Porta oversaw draw the reader's attention to the three circles of letters, numbers, and symbols. The visual language of these volvelles clearly marks cipher wheels as instruments to be used rather than being merely ornamental features to be enjoyed.

The changes that della Porta introduced after—and most likely in reaction to—the publication of the two contraband editions include obvious changes not only in the appearance but also in the status and usefulness of the book's cipher wheels:

Figure 1: Reader-assembled volvelle in contraband edition of *De furtivis literarum notis* (1591). Courtesy of Folger Library

Figure 2: Reader-assembled volvelle in revised edition of *De furtivis literarum notis* (1593). Courtesy of Folger Library

While the number of volvelles within the book remains unchanged at three, and the reader still needs to assemble these themselves, della Porta expands and enhances his instructions for assembling and using the cipher discs. This improvement is already emphasized in the printer's preface where della Porta is favorably compared to other cryptographers like Trithemius because of his capacity to provide his readers with clear explanations.

Under the title *Qua ratione ad scribendum instrumento uti possumus* [In this way we can use the writing tool], the reader finds a 40-line, step-by-step explanation of how writer and receiver of a secret message should proceed when using the first of three cipher wheels (della Porta, 1593). Precision is of utmost importance in the use of the volvelle as a functioning encryption tool, della Porta warns his readers: If the wheel is not rotated correctly, *qui scribit & qui nuncium accepturus est,* facile decipietur [both the person who writes and the person who is about to receive the message, will be easily deceived].

Compared to the first edition of *De furtivis literarum notis,* della Porta strengthens the interactive aspect of the manual, emphasizing the reader's agency in not only reading, but 'handling' the material. In contrast to the contraband editions of the manual, the revised editions present paper volvelles as highly efficient encryption tools for readers who, guided by a knowledgeable author, learn how to avoid errors when putting volvelles to use.

The changes in design of the cipher wheels that della Porta introduced between 1562 and 1593 do not correlate to a change in his proposed encryption methods. Instead, his expanded instructions to his readers detailing how to assemble and use cipher wheels reflect the fact that his cryptography manual had become a bestseller at European book fairs and that the author needed to protect it against forgeries by increasing its usefulness in the eyes of the reader.

3 Trithemius / Collange: Converting Substitution Tables into Cipher Wheels

In contrast to della Porta, Trithemius did not include cipher wheels in his *Polygraphia libri sex* (1518). Not the earliest, but the earliest *printed* Western work on cryptography, Trithemius' *Polygraphia* is known both for introducing a polygraphic system on the basis of cipher lists (books i through iv) and for establishing a polyalphabetic substitution method on the basis of a *tabula recta* (book v). Without doubt, Trithemius' focus on cipher lists and the *tabula recta* rather than cipher wheels was a conscious choice. We know from the earlier *Clavis steganographiae* (written c. 1499, published posthumously in 1606) that Trithemius was perfectly familiar with cipher wheels, so the encryption methods described in *Polygraphia* did not rely on their use. As Maximillian Gamer points out in his commentary to *Polygraphia,* the lack of cipher wheels does not in itself constitute a step back in the evolution of Trithemius' work on cryptography (Gamer 2022).

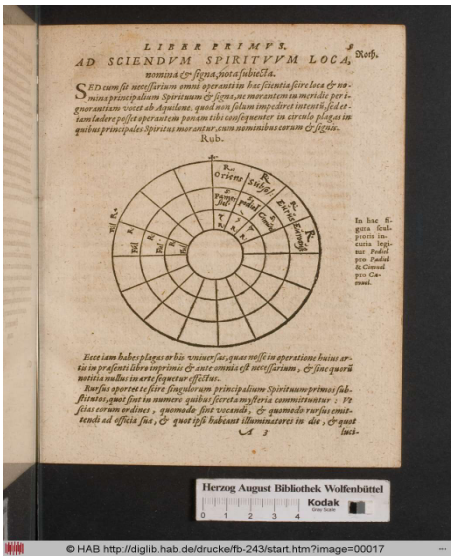


Figure 3: Printed, two-dimensional cipher wheel (not a volvelle) in Trithemius' *Steganographia* (1606). Courtesy of Herzog August Bibliothek, Wolfenbüttel

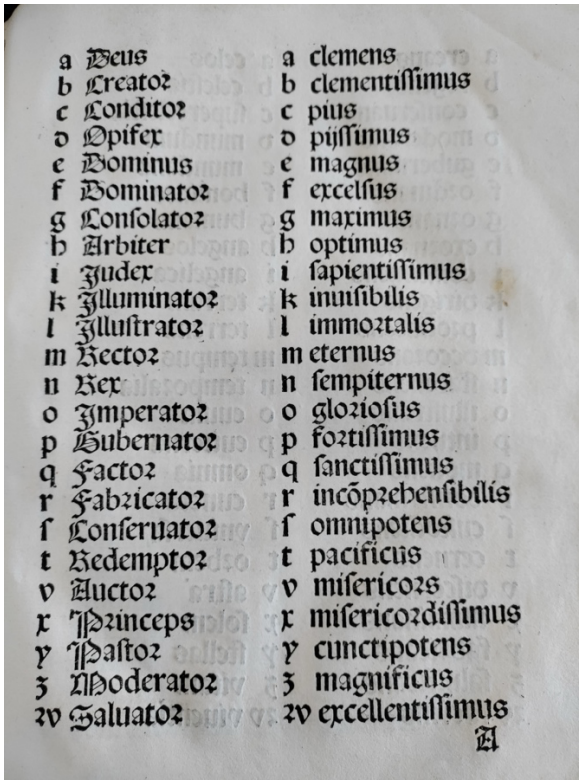


Figure 4: Latin cipher list in Trithemius' *Polygraphia* (1518), with each word in a column standing for one letter in the alphabet. Courtesy of Herzog August Bibliothek, Wolfenbüttel

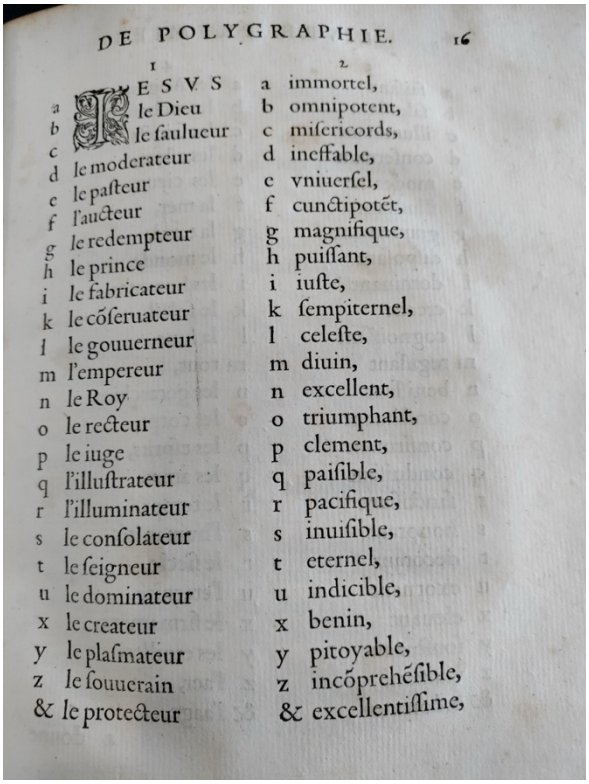


Figure 5: French cipher list by Collange in *Polygraphie, et Universelle escriture Cabalistique de M. J. Tritheme Abbé* (1561). Courtesy of Herzog August Bibliothek, Wolfenbüttel

One of the factors that helped further the fame of Trithemius' *Polygraphia* in 16th-century Europe was a widely circulated French version of the text by Gabriel de Collange, which proposed vernacular equivalents for the Latin cipher lists of books i through iv (see figure 5).

Yet it was not so much the translation of the cipher lists but rather an entirely new, 110-page long addendum to Trithemius' treatise that turned this French edition into one of the most coveted illustrated books of the Renaissance. In the last part of the work, entitled "Figures et tables planisphériques," de Collange suggests that Trithemius' tables may easily be converted into cipher wheels, an insight that reappears shortly later in della Porta's handbook.

De Collange's translation is a stunning example of an interactive Early Modern cryptography manual: The in-folio volume includes thirteen intricate, readily assembled volvelles, twelve of which are based on Trithemius' *tabulae rectae*.

Each volvelle is divided into 12 sectors forming a wheel with 12 spokes. The reader can rotate the disc under a vertical fixed strip that contains the two parts of the alphabet in their ordinary sequence (a to m, n to z; see figure 6). By turning the disc to a specific section against the fixed vertical paper strip, the reader is then able to find a letter that is part of the cipher text. De Collange explains that in order to create an equivalent of Trithemius' three *tabulae rectae* (containing 24 alphabets each), it was necessary to create twelve cipher wheels, with one wheel facilitating the representation of six alphabets.

Considering their choice of title—*Polygraphie, et Universelle escriture Cabalistique* (1561)—it would be worth exploring in detail what may have motivated the French translator Gabriel de Collange and the publisher Jacques Kerver to emphasize the occult aspects of Trithemius' work. While Kerver is known to have published important works in the history of secrecy and

occultism, de Collange's general interest in these matters is based on conjecture, not evidence. He is said to have translated Agrippa of Nettesheim' *De occulta philosophia* into French, but none of the several manuscript translations attributed to him shortly after his death in a biographical encyclopedia entry seem to have survived in French archives or libraries (Grudé 1584).

Whatever their motivation, it is clear that de Collange and Kerver introduce volvelles in their edition of Trithemius' *Polygraphia* in order to underline its proximity to occultism, as the volvelle's visual design clearly shows. Avid readers of Trithemius not only enjoyed the aesthetic qualities of the volvelles but understood their function as well, as we may glean from traces of their use, such as the handwritten corrections by Duke August the Younger (aka Gustavus Selenus) on one volvelle in his edition of de Collange's translation. (Strasser 1982)

4 Conclusion

This analysis of succeeding editions of della Porta's and Trithemius' works in the second half of the sixteenth century accounts for the increased prominence of volvelles in best-selling cryptography manuals. On the one hand, the growing demand for 'moveable', 'interactive' science books and the pressure on authors and printers to outdo contraband editions by improving the design of the legitimate edition(s) clearly contributed to the decision to prominently feature volvelles. On the other hand, the frequent inclusion of printed cipher wheels in cryptography manuals is also indicative of an evolving knowledge of encryption methods among a larger readership. Guided by detailed instructions on how to (not) use volvelles, Early Modern readers were enabled to gain experience in the practical use of the cipher wheel as important cryptographical tool.

Figure 6: Printer-assembled volvelle in *Polygraphie, et Universelle escriture Cabalistique de M. J. Tritheme Abbé* (1561). Courtesy of Bibliothèque Nationale de France

Figure 7: Detail view of printer-assembled volvelle in *Polygraphie, et Universelle escriture Cabalistique de M. J. Tritheme Abbé* (1561). Courtesy of Herzog August Bibliothek, Wolfenbüttel

References

Leon Battista Alberti. 1568. *Opusculi morali.* Bartoli, Venice.

Peter Apian. 1540. *Astronomicum Caesareum.* Apianus, Ingolstadt.

Davis Baird. 2004. Thing Knowledge. University of California Press, Berkeley.

Gianfranco Crupi, 2018. "Apianus e le volvelle del Cielo." *Paratesto* 15 (2018): 31-47.

Gianfranco Crupi. 2019. "Volvelles of knowledge. Origin and development of an instrument of scientific imagination (13th-17th centuries)", *JLIS.it*. 10 (2), 1–27.

Giambattista della Porta. 1563. *De furtivis literarum notis vulgo. De ziferis Libri IIII*. Scotus, Naples.

Giambattista della Porta. 1591. *De furtivis literarum notis vulgo. De ziferis Libri IIII*. John Wolfe, London.

Giambattista della Porta. 1593. *De occultis literarum notis* [new title for *De furtivis literarum notis*]. Lazarus Zetzner, Montbéliard

François Grudé, sieur de La Croix du Maine. 1584. *Bibliothèque Françoise*. 6 vls. Vol. 1. Abel l'Angelier, Paris, 111-112.

Katherine Ellison. 2017. *A Cultural History of Early Modern English Cryptography Manuals.* Routledge, Abingdon.

Folgerpedia. 2015. *Decoding The Renaissance: 500 Years of Codes and Ciphers exhibition item list.* https://folgerpedia.folger.edu/Decoding_the_Renaissance_exhibition_item_list#Disks_and_Volvelles

Maximilian Gamer. 2022. "Kommentar". *Die Polygraphia des Johannes Trithemius nach der handschriftlichen Fassung. Edition, Übersetzung und Kommentar*, vol. 1. Brill, Leiden, 133-207.

Suzanne Karr Schmidt and Kimberly Nichols. 2011. *Altered and Adorned: Using Renaissance Prints in Daily Life.* Yale University Press, Art Institute of Chicago.

Suchiko Kusukawa. 2012. *Picturing the Book of Nature: Image, Text and Argument in Sixteenth-Century Human Anatomy and Medical Botany.* University of Chicago Press, Chicago.

Gustavus Selenus. 1624. *Cryptomenytices et Cryptographiae Libri IX.* Stern, Lüneburg.

Gerhard F. Strasser. 1982. "Die kryptographische Sammlung Herzog Augusts: Vom Quellenmaterial für seine 'Cryptomenytices' zu einem Schwerpunkt in seiner Bibliothek." *Wolfenbütteler Beiträge* 5:83-121.

Johannes Trithemius. 1518. *Polygraphia libri sex.* Haselberg, Basel.

Johannes Trithemius and Gabriel de Collange (transl.) 1561. *Polygraphie et universelle escriture cabalistique*. Kerver, Paris.

Johannes Trithemius. 1608. Steganographia. Bernerus, Frankfurt.

What Encryption Errors Can Reveal: Cross-Cipher Errors in Mary Queen of Scots' Letters

Norbert Biermann
Universität der Künste Berlin
mail@norbertbiermann.de

Satoshi Tomokiyo
Cryptiana
verlat@hotmail.com

George Lasry
The DECRYPT Project
george.lasry@gmail.com

Abstract

In the recently deciphered letters from Mary Queen of Scots, a large number of systematic encryption errors were found and attributed to confusion as a result of concurrently using at least one other cipher key to communicate with a different recipient. In this paper, we further analyze such cross-cipher errors in those letters and identify additional cipher keys involved. This analysis also reveals valuable insights on the secret communications of Mary, Queen of Scots. We employ several techniques including statistical analysis, which may be applied to the analysis of encryption errors in other collections of historical enciphered documents.

1 Introduction

Lasry et al. (2023) found an unexpectedly high number of encryption errors in Mary Queen of Scots' ciphered letters to Michel Castelnau, the French ambassador to England, and concluded that a large part of them is explainable by the assumption that the encipherer unintentionally[1] used symbols from another cipher which was in use at the time, most probably to communicate with another recipient. They provide a hypothesis to explain this cross-cipher phenomenon (op. cit., Appendix B):

> When a cipher secretary starts using a new cipher he is not yet familiar with, it is likely that at first, he will be looking up the correct graphic symbol(s) for each letter of the alphabet he wants to encipher, in the cipher table. But after a while, the secretary is likely to start memorizing symbols representing the most frequent letters of the alphabet so that enciphering becomes faster, without always having to consult the cipher table. If the same secretary needs to encipher another letter on the same day with a different cipher, and if he continues to rely on his memory rather than on the cipher table to encipher the most frequent letters of the alphabet (e.g., i or e), he may subconsciously recall the symbol for that letter from the wrong cipher table, which he may have used recently to encipher another document. This theory is also consistent with such errors being systematic, and since the secretary might not be aware of them, he repeats them throughout the document he is currently enciphering.

We assume that the likelihood of such errors, which we call *cross-cipher errors* (CCEs), was increased by the fact that Mary lived in captivity and her secret letters not only had to be written in cipher, but the entire process of writing and sending them had to be concealed, putting her secretaries under mental stress. Moreover, the coming and going of secret messengers dictated the cryptographic work rhythm: Sometimes, a secretary was forced to work in a hurry or all night long so as not to miss the messenger's departure.[2]

[1] The opposite assumption, that such errors were made deliberately to increase the security of the cipher is worth discussing, but there are several points against it. One is that those errors were quite often corrected by the encipherer in ways that cannot be explained by cryptological sophistication, such as striking out the wrong symbols or simply overwriting them with the correct ones.

[2] Nau, Mary's French secretary, writes in a postscript to a letter to Archbishop Beaton that he was up all night deciphering letters that had arrived late in the evening (Labanoff, 1844, v, 13, *J'ay veillé toute ceste nuict pour deschiffrer voz lettres et aultres qui furent apportées hier au soir bien tard*). In the letters to Castelnau, we find wordings such as *la haste du (*or *de ce) porteur* (F89, F113) or *je suis si pressée du partement inopiné de ce porteur* (F34), and, again to Beaton, *ce qui suit est escript en fort grande haste* (Labanoff, 1844, v,

Lasry et al. identified the cipher between Mary and her ambassador to France, James Beaton, Archbishop of Glasgow, as the source for a large part of the cross-cipher errors in Mary's letters to Castelnau. The authors also hinted at the possibility that other ciphers were involved in this intriguing cross-cipher phenomenon.

In this paper, we confirm this latter hypothesis by systematically analyzing thousands of cross-cipher errors in Mary's letters to Castelnau, and we present various insights that this research produced.

In Section 2, we introduce the ciphers Mary used to communicate with Castelnau and Beaton, and use examples to demonstrate how cross-cipher errors manifest themselves. This reflects the state of knowledge as described in Lasry et al. In Section 3, we provide statistics of the systematical survey of all potential cross-cipher errors in the examined letters from Mary to Castelnau. We then introduce in Section 4 two additional ciphers we found to have caused such errors. In Section 5, we use different methods to corroborate our findings. Further intriguing insights resulting from our analysis are presented in Section 6. We conclude our examination in Section 7.

2 Two ciphers – occasionally mixed up

In this section, we describe two ciphers Mary Queen of Scots used, one with Michel de Castelnau and the second with James Beaton, and demonstrate CCEs, i.e., how symbols of the latter cipher occasionally appear unintentionally in the letters written in the former cipher.

16th century nomenclators were typically composed of:

- An alphabet section with one or more symbols (homophones) representing each letter of the alphabet
- A nomenclature with symbols representing common words, parts of words, persons, or places
- Special symbols like *nulls*, which should be ignored, a *deleter* symbol indicating that the last symbol should be erased, a *repeater* symbol for doubling the last symbol, or symbols for punctuation marks.

22). Note that we follow the naming convention outlined in Lasry et al. (2023, 109) to refer to individual letters discussed therein (F89 etc.).

Both ciphers we introduce in the following are based on this structure, and they have a further similarity in that they use *diacritics*, i.e., marks that alter the meaning of specific symbols when added to them.

2.1 Mary-Castelnau cipher (MC)

This is the cipher Mary Queen of Scots used (at least) between 1578 and 1584 to communicate with Michel Castelnau, as reconstructed by Lasry et al. (2023). All newly discovered letters to the French ambassador (op. cit.) are written in this cipher which we refer to as the *Mary-Castelnau cipher* (MC). Table 1 shows an extract of the cipher key.

Table 1: Mary-Castelnau cipher (MC), extract

a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	x	y	z
o	q	ω	//	c	∂	θ	ß	ɑ	ʃ	n	ŋ	×	ȝ	+	y	ſ	ɛ	ʌ	ı	ɬ	ß	ꝛ
		π	4	7	4	?		9	□	\	℘		2		ʁ	∮	ж	⊥	т			ǂ

q,	advis	ω·	ent	n/	monsieur
\,	affaires	ȝ	est	e·	on
ɑ·	ant	ꝛ.	et	℘.	par
ɓ·	com	ω,	faire	?	plus
6·	con	c·	ion	7	pour
y	de	9·	ite	ɬ.	que
н,	depesche	ɛ	je	π,	service
∮	des	ж	la	2	tout
ʃ	dict	ʌ	les	∂	vous
+·	en	ı	leur	9	voz
4.	endre	ç	mais		
K	Monsieur de Mauvissiere				
! delete last symbol		*4 repeat last symbol*			

2.2 Mary-Beaton cipher (MB)

Apart from the newly discovered letters which are held by the Bibliothèque nationale de France, most of Mary's letters in ciphertext are found in TNA SP,[3] BL,[4] and SCA JB.[5] Of these, SP 53/22,

[3]The National Archives, State Papers. The letters we used are from SP 53/10 and SP 53/18.

[4]The British Library. The letters we used are from Cotton MS, Caligula C III.

[5]The University of Aberdeen, Scottish Catholic Archives, Archbishop James Beaton's Papers. Mary's letters in SCA JB are mainly those addressed to James Beaton, and their plaintext is printed in Labanoff (1844) as "Déchiffrement. — Collection du docteur Kyle, à Preshome." Bishop Kyle deciphered the letters in cipher and provided his decipherments to Labanoff (1844, i, 399).

SP 53/23, and SCA JB 3/4 include collections of keys. Several of these ciphers were used between Mary and James Beaton. The particular cipher that we refer to as the *Mary-Beaton cipher* (MB)[6] is found in TNA SP 53/23 no. 38 (dated 1577) and the verso of no. 34 (crossed out), as well as SCA JB 3/4 p. 40 (no. VIII).

Based on extant letters, MB was used at least from 20 February 1576[7] to 10 September 1582.[8]

Table 2 shows an extract from the MB cipher key.

Table 2: Mary-Beaton cipher (MB), extract

a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	x	y	z
//	—	∧	⊼	+	∨	⌝	⌞	ε	⊓	⊣	∼	ϕ	ǂ	ȝ	β	ω	∝	ϑ	♪	ρ	ꝑ	☾
			⋇															⊤				

o. advis	ϑ est	ɥ on
s. affaires	ȣ et	s par
⊛ ant	m faict	ȣ plus
ɲ com	ȣ. force	o pour
ŋ con	☋ ite	n que
ȣ de	ⅩⅠ je	2. service
\. depesche	y les	ȣ vous
ȣ des	⊣ leur	e voz
☋ endre	ȝ mais	
⊣ ent	ɔ. monsieur	
ö Monsieur de Glasgo		
⫴ *repeat last symbol*		

2.3 Examples of cross-cipher errors

All the examples given below are taken from Mary's letters to Castelnau, which means they should be decrypted using the MC key (Table 1).

Example 1.

n e c e s s a s r e
x c ω c ε ϑ o ε s c

Any contemporary decipherer would have tacitly corrected this[9] to the French word *necessaire*. The symbol ε appears in this word with two different meanings: First, it stands for *s*, which is correct according to MC, but it was also wrongly used to encipher *i*. For the latter case, we assume that the secretary unconsciously made use of the Mary-Beaton cipher in which the symbol ε is assigned to the letter *i*. We denote this specific cross-cipher error as ε=*i*, indicating that in MB, the cipher that causes this CCE, the symbol ε represents *i*.

We distinguish between *uncorrected* and *corrected* errors. In many cases, an encryption error was spotted and corrected by the secretary enciphering the letter. Such corrections could have been made by crossing out the wrong symbol or by overwriting it with the correct one. In such cases, it is often difficult to determine what the corrected (wrong) symbol looked like, so we have not included them in the analysis presented here. But most often, to invalidate a wrong symbol, the *deleter* symbol was used, and written right after the original (deleted) symbol which is still legible, and therefore useful for our analysis. While our first example represented an uncorrected error, Example 2 shows corrected ones:

Example 2.

Monsieur de M a u v i s s i e r e
n/ ȣ ! y ŋ o ı ϑ ε ! a ε ε a c s c

Here,[10] the scribe made two errors while enciphering *Monsieur de Mauvissiere*. The first error is using the symbol ȣ to encipher *de* (ȣ=*de*). This symbol appears only in the key of MB, not in MC. The second error is the same one as in example 1 (ε=*i*). This time, however, both erroneous symbols have been invalidated and corrected.

Both cross-cipher errors presented so far, ε=*i* and ȣ=*de*, can be attributed to the Mary-Beaton cipher. This is the case for a large number of errors appearing in the letters to Castelnau, but is not true for all such errors, as demonstrated in Example 3:

Example 3.

m on d r o i c t
ŋ e· + ! // s ȝ ε ω ∧

Here,[11] we assume that the secretary made – and corrected – the cross-cipher error +=*d* in *mon droict*. This time, MB cannot explain the error: although the symbol + is also part of MB, it stands for *e* in this cipher. This is why we assume that there was another, unidentified cipher also used by Mary's secretary at the time, in which + enciphers *d*, causing this specific cross-cipher error.

Errors like those in our three examples recur systematically and quite often. For example, in

[6] It is called "chiffre de Nau" in the endorsement of a letter to Beaton (18 September 1581, SCA JB 2/5/7, printed in Labanoff (1844, v, 254)). MB may have been given to Nau by Beaton when Nau came over from France to enter Mary's service in 1575. Before this cipher came into use, another cipher (SCA JB 3/4 No. II) called "chiffre de Raulet" (Labanoff, 1844, v, 89) was used.

[7] SCA JB 2/3/13

[8] BL Cotton MS, Caligula III, f. 426

[9] F87, line 10

[10] Beginning of F87

[11] F87, line 14

a single letter (F87), we observe +=*d* 7 times, and ɛ=*i* no less than 64 times. This means they cannot be purely random errors which are expected to some extent in any historical ciphered letter.

3 Systematically analyzing cross-cipher errors

In the previous sections, we summarized and exemplified the findings described by Lasry et al. (2023, App. B). In this section, we record all potential cross-cipher errors that occur in the letters from Mary to Castelnau (not only those we can assign to MB), drawing new insights from initial statistical observations.

3.1 Determining potential cross-cipher errors

For our examination, we systematically compared deciphered and edited versions of the letters with the raw transcripts of the symbols. Non-matching symbols found in this way were considered potential CCEs unless another plausible explanation for their appearance could be found, such as:

- The secretary missed a letter in the middle of a word while enciphering
- The secretary forgot to add the correct diacritic near a symbol, or wrote a wrong or misplaced diacritic
- What seems to be an error may be an acceptable spelling of a word according to the – sometimes inconsistent – orthography of 16th-century Middle French

Similarly, we took into account errors corrected by the secretary only if other plausible causes of enciphering errors could be excluded.

3.2 Overview of the results

Examining all the 57 ciphered letters from Mary to Castelnau, we counted 2,740 occurrences of potential cross-cipher errors. The examined letters count up to a total of around 176,900 symbols, of which the 2,740 potential CCE occurrences make up a percentage of 1.55 – a remarkably high ratio.

Table 3: Cross-cipher errors by criteria

	uncorrected	corrected	Σ
intrinsic symbol	1,487	911	2,398
extrinsic symbol	80	262	342
Σ	1,567	1,173	2,740

Table 4: Cross-cipher errors by occurrence

CCE	occur.	cipher	CCE	occur.	cipher
o=*i*	525	?	ſ=*t*	17	MB
ɛ=*i*	234	MB	ƀ=*vous*	17	MB
∧=*c*	161	MB	c=*c*	16	?
ı=*e*	105	?	⊤=*c*	15	?
+=*e*	73	MB	⊤=*t*	15	?
ɑ=*a*	68	?	∠=*t*	14	?
⊤=*r*	68	MB	♪=*u*	14	MB
//=*a*	67	MB	s=*par*	13	MB
ω=*r*	67	MB	ƒ=*n*	12	MB
ȝ=*s*	64	?	ı=*s*	12	?
s=*s*	55	?	ɑ=*n*	10	?
ȣ=*et*	48	MB	ŧ=*o*	10	MB
m=*r*	48	?	ƒ=*r*	10	?
x=*[x2]*	44	?	✕=*r*	10	?
⊼=*d*	43	MB	o=*n*	9	?
ȝ=*de*	42	MB	++=*[x2]*	9	MB
ŋ=*con*	39	MB	6=*d*	8	?
o=*pour*	36	MB	+=*d*	8	?
∝=*s*	33	MB	ϑ=*est*	8	MB
✕=*e*	32	MB	—=*l*	8	?
o=*o*	31	?	ȝ=*p*	8	MB
∧=*i*	29	?	⊬=*leur*	7	MB
∧=*n*	28	?	ȝ=*r*	7	?
ꝑ=*p*	28	?	ꝏ=*on*	6	MB
c=*a*	25	?	4.=*que*	6	?
∼=*m*	25	MB	//=*s*	6	?
+=*u*	25	?	x=*s*	6	?
s=*p*	23	?	o.=*advis*	5	MB
ı=*a*	21	?	—=*b*	5	MB
ȝ.=*dict*	20	?	ɛ=*e*	5	?
+=*r*	20	?	⊬=*l*	5	MB
ꝓ=*y*	20	MB	ɛ·=*la*	5	?
ȝ=*n*	19	?	e=*n*	5	?
n=*que*	18	MB	|=*r*	5	?
//=*u*	18	?	ƈ=*z*	5	MB
4=*t*	17	?			

Table 3 shows how these occurrences distribute

across criteria corrected/uncorrected and whether the wrongly used symbol is part of MC (intrinsic) or not part of it (extrinsic).

One can see that CCEs with intrinsic symbols are much more frequent than those with extrinsic symbols. This suggests that an overlapping vocabulary of symbols between two cipher tables was likely to lead to confusion and increase the occurrence of cross-cipher errors. Errors involving intrinsic symbols were more likely to be overlooked by the secretary, while the majority of CCEs with extrinsic symbols were fixed during the encryption process (probably because an extrinsic symbol stands out visually, usually causing the secretary to spot it and correct it right away).

The CCEs can be attributed to over 140 distinct errors. To exclude random errors, in our analysis, we ignore those with less than 5 occurrences across all 57 letters, which reduces the above number to 71. These 71 errors, which we consider to be potential cross-cipher errors, are listed in Table 4, sorted in descending order of occurrence. The errors that can be attributed to the Mary-Beaton cipher are marked (MB in the third column).

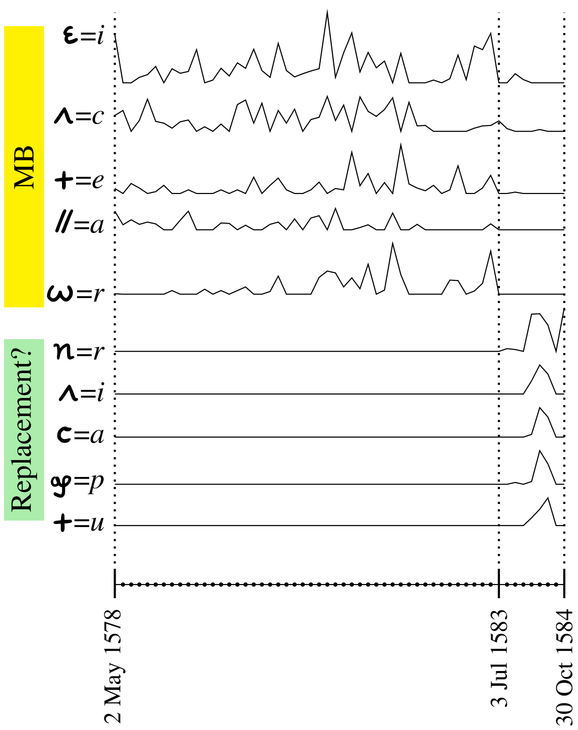
4 Identifying other ciphers causing CCEs

In this section, we present three other ciphers, two of which we found to cause cross-cipher errors in Mary's letters to Castelnau, matching the vast majority of the 71 CCE candidates shown in Table 4.

4.1 Late Mary-Beaton cipher (LMB)

When we examined the incidence of CCEs per individual letter, we observed that the CCEs attributed to the Mary-Beaton cipher practically disappear from a certain date (3 July 1583). Specific other CCEs, on the other hand, only seem to begin to appear after this date. Figure 1 illustrates this phenomenon showing the frequency of selected CCEs over time, per letter from Mary to Castelnau.

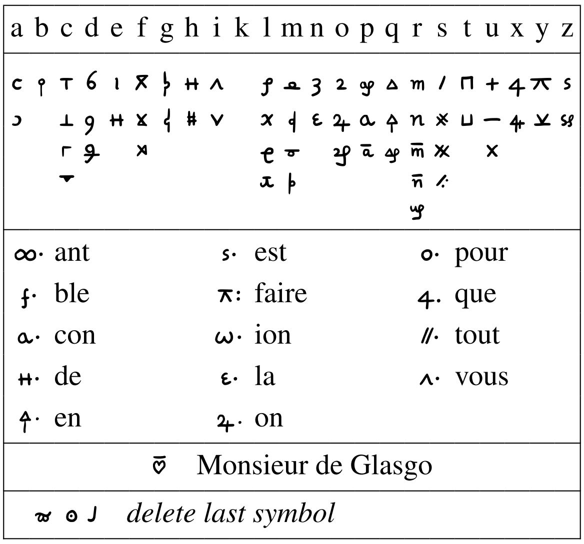
Figure 1: Frequencies of selected CCEs per letter



This led us to suspect that the cipher between Mary and James Beaton was replaced from this point onward by a new cipher which the newly-appearing CCEs may be attributed to. Indeed, we found such a cipher between Mary and Beaton, perfectly matching the CCEs that occurred after 3 July 1583. We refer to it as the *Late Mary-Beaton cipher* (LMB). The key of this cipher, however, could not be found in archives, and rather, it was reconstructed from three extant letters from 1586 enciphered with it.[12] The reconstructed key is given in Tomokiyo (2023).

Table 5 shows an extract of the cipher key of LMB, and Table 6 lists CCEs that we can attribute to LMB.

Table 5: Late Mary-Beaton cipher (LMB), extract


a b c d e f g h i k l m n o p q r s t u x y z

ant · ble · con · de · en · est · faire · ion · la · on · pour · que · tout · vous

Monsieur de Glasgo

delete last symbol

[12]TNA SP 53/18/60: Mary to Beaton, 12 and 16 July 1586, and to Monsieur de Mondevis, Cardinal de Laurea, 30 June 1586, printed in Labanoff (1844, vi, 362, 381, 347).

Table 6: CCEs attributed to LMB

CCE	occur.	cipher	CCE	occur.	cipher
ı=*e*	105	LMB	ȝ=*n*	19	LMB
m=*r*	48	LMB	⊤=*c*	15	LMB
∧=*i*	29	LMB	/=*s*	12	LMB
ϑ=*p*	28	LMB	6=*d*	8	LMB
c=*a*	25	LMB	4.=*que*	6	LMB
+=*u*	25	LMB	ε·=*la*	5	LMB

In the following, we provide examples of encryption errors that can be explained by LMB.

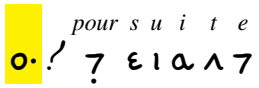
Example 4.

r est e b l t s s e m ent

This[13] should read *restablissement*, but symbols from LMB are used for the wrongly enciphered *a*, *i*, and *m*. Only the last error, involving an extrinsic symbol for *m*, has been spotted, deleted with a deleter symbol, and corrected by Mary's secretary.

Example 5.

pour s u i t e



This word[14] should read *poursuite*. The first symbol, o·, was wrongly taken from LMB to encipher *pour*. In this case, the secretary spotted the mistake, erased the wrong symbol with the deleter symbol, and wrote the correct 7 from MC.

Example 6.

des e d v o u e r de oit c e

This[15] should read *desadvouer de tout ce*. Three symbols are wrongly taken from LMB, of which only the last one is corrected. Without knowledge of the LMB cipher, the decipherer has no chance to conclude that // · is meant to encipher *tout*.[16]

4.2 Mary-Aubigny cipher (MA)

Still, the LMB cipher could not account for large numbers of other potential CCEs, including o=*i* which accounts for 525 occurrences, from a total of 2,740. We conducted a systematic search for another cipher that could explain those unattributed CCEs, and found a promising candidate, which we refer to as the *Mary-Aubigny cipher* (MA).

[13] F96, line 26

[14] F42, line 20

[15] F96, line 48

[16] This letter had been leaked to spymaster Walsingham and is printed in Labanoff (1844, v, 458). Indeed, we read there only *désadvouer de ce*, implying that Castelnau's secretaries ignored the symbol // · they could not make sense of.

The table for this cipher is the second of the two ciphers written on SP 53/22 f. 21. The endorsement in one word on the verso is illegible, but this cipher must have been used in Mary's correspondence with Esmé Stewart, sixth Seigneur d'Aubigny and Duke of Lennox, because the cipher symbol assigned for "daubugny" (O:) is the same as the enclosure marking for the Duke of Lennox in the letters from Mary to Castelnau.[17] Table 7 shows an extract of the key, and Table 8 lists the CCEs that we can attribute to it. Note that

- ı stands for the letter *e* in both LMB and MA, so the CCE ı=*e* is included in both Table 6 and Table 8
- the symbol ∠ (Table 8) does not appear in the cipher key (Table 7). However, we assume that it is a half-finished version of 4, both symbols enciphering *t*.[18]

Table 7: Mary-Aubigny cipher (MA), extract

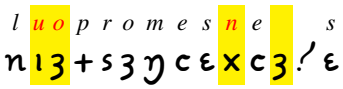
a b c d e f g h i k l m n o p q r s t u x y z
z ϟ c q ı # ◻ ⊓ o ȣ n ϑ ∧ v s ⧺ 3 4 6 7 8 9
q· de
O: (Monsieur) d'Aubigny
! *delete last symbol* x *repeat last symbol*

Table 8: CCEs attributed to MA

CCE	occur.	cipher	CCE	occur.	cipher
o=*i*	525	MA	s=*p*	23	MA
ı=*e*	105	MA/LMB	+=*r*	20	MA
3=*s*	64	MA	4=*t*	17	MA
x=*[x2]*	44	MA	c=*c*	16	MA
∧=*n*	28	MA	∠=*t*	14	MA

Example 7.

l uo p r o m e s ne s



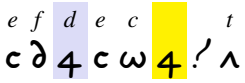
This example[19] should read *les promesses* but includes no less than four symbols wrongly taken

[17] For enclosure markings, see Section 5.2. Moreover, next to the name "Daubugny", there is a special mark apart from a cipher symbol. The same mark is found in several other cipher tables in SP 53/22 next to the names of the respective correspondents such as Châteauneuf (f. 22), Cherelles (f. 23), and Claude Hamilton (f. 33). The first clue that attracted our attention to Aubigny was a symbol later added to represent Cavaillon, who was a secretary to the Duke of Lennox (Bossy, 2001, p. 91, fn. 17).

[18] In the letters from Mary to Castelnau, ∠=*t* is always a corrected error, implying that the secretary spotted the mistake before finishing to write the symbol.

[19] F165, line 24

from the MA cipher, only the last being corrected by the secretary.

Example 8. 

This[20] should read *effect*. The first error is not a CCE: We would expect the repeater symbol 4 from MC, to repeat the previous *f*. Instead, we have the wrong but similar-looking 4, uncorrected. The second error is a CCE: The second 4 is an MA symbol for *t* that has been deleted and corrected by the secretary.

4.3 The Nau brothers' cipher (NN)

The key (reconstruction) is found in TNA SP 53/23 no. 43.

This is a candidate we have considered as a potential source of CCEs, but dismissed after further analysis. We mention it here to illustrate that a set of potentially matching CCEs may not lead to a decisive conclusion that a specific cipher is causing CCEs. On the one hand, several likely CCEs, ∧=*c*, ı=*e*, ſ=*p*, and 4=*t*, could have been explained by the concurrent use of the NN cipher table. On the other hand, we also found that ∧ can be better explained by MB, and the other three are covered by MA. There is therefore not enough evidence to conclude that those CCEs in Mary's letters to Castelnau can be attributed to NN.

5 Corroborating our findings

We validated our hypotheses by statistical analysis, by examining the mentions of enclosed letters intended for other recipients, as well as by analyzing CCEs in the opposite direction, i.e., CCEs in Mary's letters to Beaton (encrypted with the MB cipher) that can be attributed to the MC cipher.

5.1 Statistical analysis

Our main assumption was that the secretary who enciphered the letters to Castelnau was affected not only by his level of concentration at the particular hour of the day or night but also by the specific mixture of other ciphers he may have used on that certain day (or week) to encipher letters to other recipients. Accordingly, to validate our CCE hypothesis, we expected certain sets of CCEs to occur together in particular letters. Thus, in one letter the errors induced by a certain cipher would predominate, resulting in an individual CCE "fingerprint" of each letter. Indeed, we found a high correlation between specific CCEs, and using a K-means clustering analysis, those correlations were found to fairly match the set of CCEs attributed to the MB, LMB, and MA ciphers (see Section 4). Those findings are described in more detail in Appendix A.

5.2 Evidence from mentions of enclosed letters to other recipients

Mary's secret letters, after being folded and sealed, were usually marked externally with a cipher symbol that stood for the recipient. This would allow Castelnau to identify which letters were addressed to him, and to whom the other letters should be forwarded. A ciphered postscript in the letter addressed to Castelnau, usually written by Nau, Mary's secretary, would list the names of the recipients of enclosed letters, together with the marking symbols identifying each recipient. During the relevant period 1578–1584, this external marking symbol was often the one representing the name of the recipient in the cipher used to encipher the letter. Thus, the letters to Michel Castelnau de Mauvissière were usually marked with the symbol K, which stands for *Monsieur de Mauvissiere* in MC (cf. Table 1).[21]

Based on those postscripts, we can confirm that the MB, LMB, and MC ciphers were indeed used concurrently with the MC cipher. For instance, F87 from Mary to Castelnau, enciphered with MC, ends with *Le chiffre cy-encloz marqué* 8̈ *est a Monsieur de Glasgo* (the enclosed ciphered letter marked 8̈ is for the Archbishop of Glasgow), and the marking symbol represents the Archbishop of Glasgow (James Beaton) in the MB cipher (cf. Table 2).

In the 57 letters from Mary to Castelnau, we find 20 mentions of "Monsieur de Glasgo" in conjunction with the symbol 8̈ from 2 May 1578 to 1 June 1583 and three further mentions of the same person associated with the symbol 8̄ from 3 July 1583 to 3 September 1583 (cf. Table 5). Both symbols correspond to James Beaton, the archbishop of Glasgow, in the MB cipher and the LMB cipher, respectively, and the latter marking symbol appears at the same time we estimate the MB cipher was replaced by the LMB cipher (cf. Section 4.1).

Similarly, the symbol for d'Aubigny in the MA

[20] F98, line 2

[21] See, for example, the verso of F87, `https://gallica.bnf.fr/ark:/12148/btv1b9059908w/f122.item`

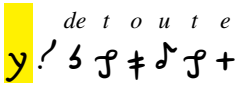
cipher, 0:, is used in association with enclosures for "le duc de Lenox" (i.e., d'Aubigny) twice in letter F225 from 31 December 1582 (cf. Table 7).

5.3 Reciprocal cross-cipher errors in letters from Mary to Beaton

Another way to corroborate our findings is based on the following idea: If such a large number of CCE occurrences in Mary's letters to Castelnau are caused by the concurrent use of the Mary-Beaton cipher, then we should also find the phenomenon in the opposite direction, i.e., in her letters to Beaton enciphered with the MB cipher, we should find cross-cipher errors that can be explained by concurrent use of the Mary-Castelnau cipher. And indeed, we found numerous occurrences of those kinds of errors in the letters from Mary to Beaton. For example, in the letter from 10 September, 1582,[22] we find two CCEs in the phrase *le principal autheur de toute l'entreprise*:

Example 9. *p r i n r i p a l*

Example 10. *de t o u t e*



Apparently, the wrong symbols for *c* (uncorrected) and for *de* (corrected) are taken from MC (although the dot in y:=*de* is missing).

6 New historical insights from cross-cipher errors

Based on our analysis of CCEs, we were able to gain highly interesting insights into Mary's secret communications:

- While the earliest of the newly deciphered letters from Mary to Castelnau is dated May 2, 1578, CCEs in letters from Mary to Beaton that can be attributed to MC (see Section 5.3) start to appear earlier. For instance, a significant amount of those CCEs is found in a letter to Beaton from 5 November 1577,[23] which shows that the Mary-Castelnau cipher was in use as early as this date.
- The CCEs attributed to the MA cipher show that Mary's secret communications with d'Aubigny were more extensive than previously known. The analysis of CCEs also shows that at the beginning of September 1582, Mary temporarily loses contact with d'Aubigny following the Raid of Ruthven and the pro-English coup in Scotland,[24] but that soon afterward the communication channel is re-established.[25]
- Previously, F308, one of the 57 letters presented by Lasry et al. (2023) could not be dated. We found several MB CCEs but no LMB CCE in this letter, allowing us to determine with high confidence that F308 was written before July 3, 1583.
- We can establish that the MB cipher between Mary and James Beaton was replaced by LMB in mid-1583 when MB CCEs disappeared and LMB CCEs began to show up, and based on the enclosure markings (see Section 5.2), we can even narrow down the timing of the replacement to between June 1 and July 3 of that year. This is remarkable as no letter encrypted with the new cipher before 1586 seems to have survived.

7 Conclusion

The systematic examination of ciphering errors and cross-cipher errors in particular in Mary's letters has led to valuable insights. Those errors were frequent, as she was communicating with multiple recipients, her secretaries using a different cipher for each recipient, and working under time pressure, so that multiple letters could be handed over to a trusted courier visiting Mary.

It is difficult to assess whether our method can be successfully applied to other collections of ciphered letters. Cipher secretaries around the world might also have had a stressful job, handling a high volume of communications using different ciphers at the same time, so those kinds of errors could have happened as well. If there is enough ciphertext material to analyze, it may be worthwhile to take a closer look at the errors they made, applying the techniques described in this paper.

[22] BL Cotton MS Caligula C III, f. 426, reproduced in Labanoff (1844, v, 309)

[23] SCA JB 2/4/3

[24] F229 of 2 September, 1582, in which hardly any CCEs can be assigned to MA. This fits in with her stating in the same letter that she would like to write to Aubigny if only she could (*j'eusse tres volontiers escript au duc de Lenox* [i.e., d'Aubginy] *si j'eusse eu aulcun moyen de ce faire*, line 10–11).

[25] E.g., F151 of 10 September 1582, with a significant number of CCEs that can be assigned to the MA cipher.

Acknowledgments

The work of one of the authors has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

List of Abbreviations

BL The British Library

CCE Cross-cipher error

LMB Late Mary-Beaton cipher

MA Mary-Aubigny cipher

MB Mary-Beaton cipher

MC Mary-Castelnau cipher

NN Nau Brothers' cipher

SCA JB The University of Aberdeen, Scottish Catholic Archives, Archbishop James Beaton's Papers

TNA SP The National Archives, State Papers

References

John Bossy. 2001. *Under the molehill: an Elizabethan spy story*. New Haven: Yale University Press.

Alexandre Labanoff, editor. 1844. *Lettres, instructions et mémoires de Marie Stuart, Reine d'Écosse, 7 vols.* London: Charles Dolman.

George Lasry, Norbert Biermann, and Satoshi Tomokiyo. 2023. Deciphering Mary Stuart's lost letters from 1578-1584. *Cryptologia*, 47(2):101–202, March.

Satoshi Tomokiyo. 2023. Ciphers of Mary, Queen of Scots. http://cryptiana.web.fc2.com/code/mary.htm.

Appendix A. Clustering analysis

We first examined the correlations between certain pairs of CCEs, by computing their frequencies in each of the 57 letters, and computing the correlation according to the Pearson correlation formula. Let f_{ik} be the frequency of CCE C_i in letter k, and $\bar{f}_i$ the arithmetic mean of f_{ik} across all k. Then for a pair of CCEs C_i and C_j, the Pearson correlation coefficient r_{ij} is computed as follows:

$$r_{ij} = \frac{\sum_k (f_{ik} - \bar{f}_i)(f_{jk} - \bar{f}_j)}{\sqrt{\sum_k (f_{ik} - \bar{f}_i)^2 \sum_k (f_{jk} - \bar{f}_j)^2}}$$

Some correlations were exceptionally high between CCEs we had attributed to a particular cipher, and could not be the results of purely random errors.

Next, we aimed at grouping all types of CCEs into subsets of highly correlated CCEs, likely to have been caused by the concurrent use of the same cipher (other than the Mary-Castelnau cipher). For that purpose, we employed a common clustering algorithm, k-means,[26] based on the frequency of CCEs in the 57 letters from Mary to Castelnau. We limited our analysis to the types of CCEs that occur at least five times in the letters, i.e., the 71 CCEs given in Table 4.

The distance between two elements is (inversely) measured using the Pearson correlation coefficient mentioned above. The number of clusters – the k in the k-means – was found to be $k = 3$ using the "elbow method".[27] The results are shown in Figure 2.[28] We observe the following:

- Cluster 1 is mostly composed of errors attributed to LMB, which are strongly correlated, or moderately correlated with one another.
- Cluster 2 is mostly composed of errors attributed to MA.
- Cluster 3 is mostly composed of errors attributed to MB.
- A few errors attributed to MB are also associated with cluster 2, but with a low correlation.
- Some errors for which we did not identify a matching cipher table (UC) are assigned to the various clusters, but with a lower correlation.

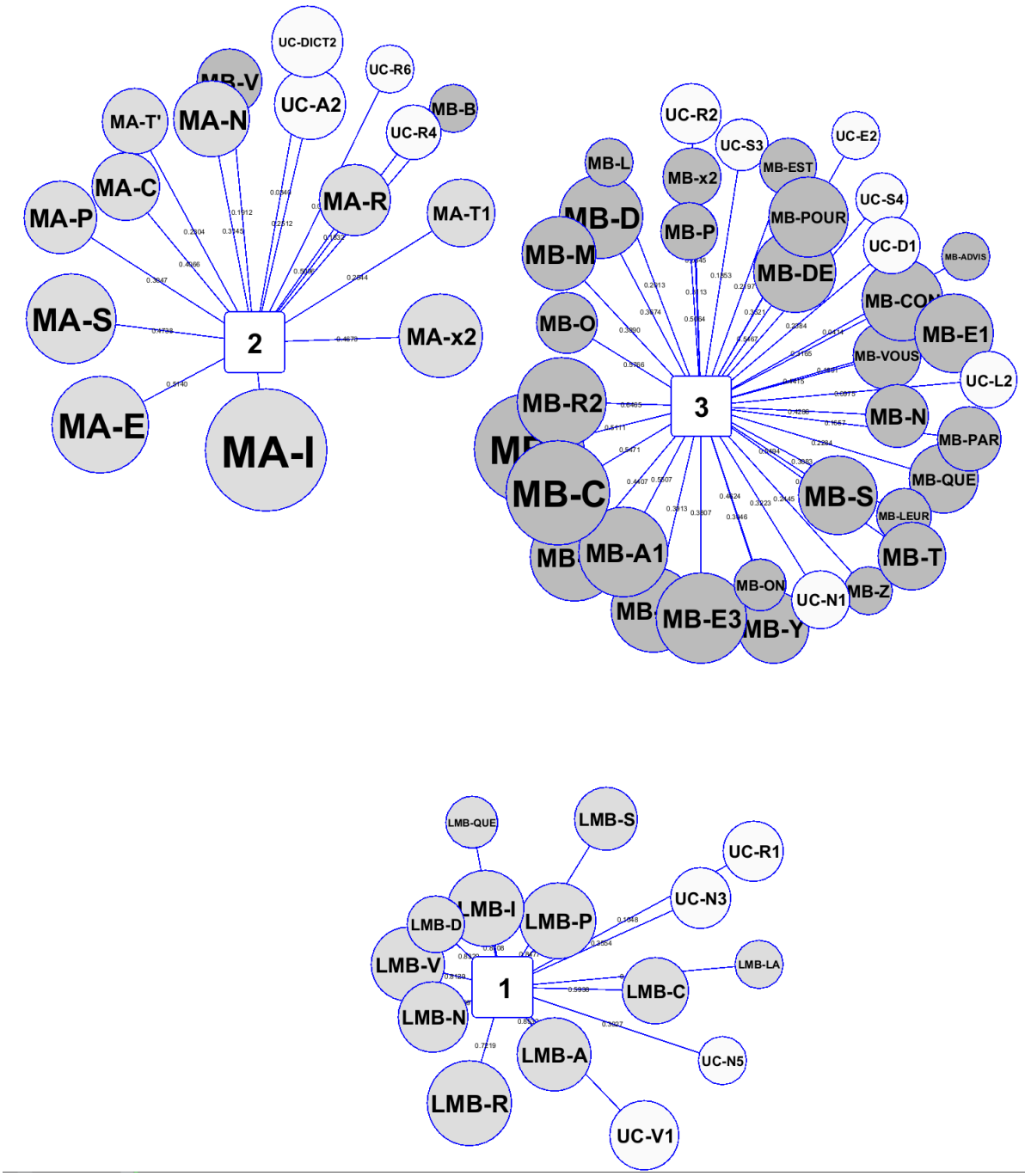
Those results provide solid statistical evidence for the identification of the MB, LMB, and MA ciphers as being the root cause of almost all those errors.

[26] *See* Wikipedia, *k-means clustering*, `https://en.wikipedia.org/wiki/K-means\_clustering` (as of December 1, 2023, 19:00 GMT)

[27] *See* Wikipedia, *Elbow method (clustering)*, `https://en.wikipedia.org/wiki/Elbow_method_(clustering)` (as of December 1, 2023, 19:00 GMT)

[28] We show each such CCE type in the plot using a colored circle, whose size is proportional to the number of occurrences of the error (using a logarithmic scale). We also tie each CCE to the cluster identified by the k-means algorithm, and the length of the edge between the circle and the cluster is inversely proportional to the Pearson correlation between the frequencies (in the 57 letters) of the error and the average frequencies of the members of the cluster.

Figure 2: Clustering of the cross-cipher errors.



The enigma of Lorenzo Ventura's cipher

Paolo Bonavoglia
paolo.bonavoglia@liceofoscarini.it
Former teacher, webmaster
at the "Convitto-Liceo Marco Foscarini"
I 30121 Venice / Cannaregio 4942

Abstract

The aim of this research was to find the algorithm used in the ciphers mentioned by Blaise de Vigenère in his treatise, where he states that in 1569, while in Venice, he learned about a steganographic cipher by a certain Lorenzo Ventura, similar to Tritemio's *Ave Maria*. It had been used by the bailo in Constantinople after Sultan Selim II prohibited him from writing his dispatches in cipher. Now that a collection of letters, notes, and handwriting examples belonging to Ventura has been found in the State Archives of Venice, initial findings emerge that confirm, at least in part, Vigenère's claims.

.

1 Introduction. Vigenère wrote ...

In 1569 Blaise de Vigenère was in Venice as an officer of the French embassy, and in his *Traicté des chiffres*[1] wrote[2]:

> In the year 1569, I was in Venice, the Turk Selim father of Amurath who reigns today, secretly making his preparations to invade the kingdom of Chippre; for fear that the Bayle of the Venetians residing in Constantinople, should not inform them of what could be foreseen, forbade that they no longer had to write by any sort of cipher, as well as in all patent and open packages, & in intelligible letter: while some people found themselves in difficulty, a doctor named Lorenzo Ventura introduced himself, who presented them with the following secret; to write whatever they wanted, especially all kinds of remarks, and in common writing, which had another meaning hidden underneath, as they pleased, subject to certain very advantageous conditions which he paid for his salary. What I wanted to allege to show what esteem and importance it is is artifice.

It is clear that Vigenère is describing a steganographic cipher, citing Trithemius Polygraphiae[3] and the Ave Maria cipher[4]. as the source for the *Ave Maria* cipher. However, a search of the Venetian archives reveals that soon after the sultan banned ciphers in 1567, the baylo Giacomo Soranzo wrote very brief dispatches with a long postscript in red ink[5]. Soranzo, in fact, used lemon

[1] (de Vigenère, 1587) p. 183.

[2] Ibidem, p.183; translated from the French original, 1587: *L'An 1569. que i'estois à Venise, le Turc Selim pere d'Amurath qui regne auiourd'huy, faisant sourdement ses apprests pour enuahir le royaume de Chippre; de peur que le Bayle des Venitiens residant à Constantinople, ne les aduertist de ce qu'il en pouuoit pressentir, defendit qu'ils n'eussent plus à fentr'escrire par aucune sorte de chiffre, ains à pacquets tous patents & ouuerts, & en lettre intelligible : de eproyeux se trouuans en peine, se presenta vn medecin nommé Lorenzo Ventura, qui leur presenta le secret cy dessus; d'escrire tout ce qu'ils voudroienr, surtoutes sortes de propos, & en escriture commune,qui eust autre sens caché audessous, tel leur plairroit, moyennant certaines conditions bien aduantageuses qu'il demâdoit pour son salaire. Ce que i'ay bien voulu alleguericy pour monstrer de quelle estime & importance est cest artifice*

[3] (Trithemius, 1613). Polygraphiae VI is the main work of Abbot Tritemio, also known as Johannes Trithemius (1462-1516), considered one of the founding fathers of modern cryptography. His first work, titled Steganographia, was the first cryptography book published in print, but accused of witchcraft, it was condemned by the Church and placed on the index of prohibited books, where it remained until 1900. Tritemio's most well-known ciphers are the Ave Maria and the Recta Tabula, of which something will be said in this article.

[4] Trithemius did not use the name *Ave Maria* for his cipher, a nickname that was introduced much later.

[5] *ASVe CCX Dispacci degli ambasciatori a f.2 3-6-1567* In the following these archive abbreviations are used: *ASVe*

juice as invisible ink, a method for which he was severely reprimanded by the Council of Ten[6] because the method was known even to the Turks. Among the archive papers, one finds a draft letter by the Council addressed to Selim, rejecting his ban using subtle arguments. It is not certain whether the letter was actually sent; however, Soranzo's dispatches reverted to being encrypted as before.

Perhaps Vigenère had misunderstood? It remained doubtful whether one of Ventura's ciphers had been employed at any time. It is obvious that a steganographic text, by its very nature, can be very difficult to detect, if used correctly, and Vigenère provides no details about this cipher.

2 Ventura's papers

A necessary premise: Ventura calls *manifesto* or *palese* the text that is visible and meaningful but hides a secret message. The real message is called "secreto" (secret). *Scontro* is a cipher sheet, *contrasegno* is a keyword or key-phrase.

These papers were found in *busta* 6 of the State Archives of Venice collection named *Cifre, chiavi e scontri di cifra* . . . , in a fascicle named *Lorenzo Ventura fisico*[7].

At the beginning of this file, we find a letter in which Ventura presents to the Council of Ten a new method of writing in code that is articulated in various ways, as we will see later on. He boasts that his ciphers cannot be decrypted without the most profound artifice of the various *scontri* (key sheets) and *contrasegni* (keywords or keyphrases) of his.

We will now see three examples of different modes, without instructions.

3 Example n.1

Among the sheets of this file we find one with this introduction[8]:

> A very easy and convenient method, which requires no keyword or keyphrase, but only the first and second letter of one's own name, or of the one who writes, or to whom it is written, and thus the first or last characters of the plaintext, and the entire complete phrase if desired. Indeed, the composition is somewhat forced because it is formed like that of the *caselle*. But what is of greater importance is that no perforated paper is required; everything necessary is provided . . .

The word *caselle* brings to mind Hieronimo di Franceschi's *cifra delle caselle*[9], a cipher that utilized small windows for arithmetic encryption and decryption operations. However, Ventura's method, mentioned in 1567, appears to differ significantly. It is more likely referring to Cardan's grid, also known as Richelieu's grid[10]. Both systems employed small windows cut out of cardboard to write the message inside. After removing the cardboard, the remaining space was filled with additional text to make the message appear plausible, and even misleading.

Indeed, at first glance, the first example looks like a Cardan's grid without a grid; the words of the true and secret text are hidden between the words of the *manifesto* i.e., the false text, as seen in Figure 4.

The first segment of the secret text *L'armata nemica è già ritornata* is hidden in the manifest in the middle, the following segments of the secret

= *Archivio di Stato di Venezia, CX = Consiglio dei Dieci; CCX = Capi del Consiglio dei Dieci*The *Consiglio di Dieci* = Council of Ten was a powerful organ of the Republic of Venice, responsible for state security, intelligence and also for cryptography.

[6]The letter is in *ASVe CX Parti Secrete filza 12 3-4-1567*

[7]Here *fisico* should be understood as physician, not physicist. Lorenzo Ventura is mentioned in Paolo Preto's book (Preto, 1994) on page 272. Preto's book, *I servizi segreti di Venezia*, focused on secret services and intelligence, with only a chapter dedicated to cryptography. It is rich in citations and references to the Venice Archives collections, but provides fewer details on cryptographic technicalities. There are few works specifically dedicated to Venetian cryptography, including essays by Predelli (Predelli, 1869) and Cecchetti (Cecchetti, 1869), Pasini's booklet (Pasini, 2019), which was republished in 2019 under my editorship, and Meister's chapter on Venetian ciphers (Meister, 1902), which is undoubtedly valuable from a cryptographic perspective. It is unfortunate that Meister's research only reaches up to the year 1550 and does not mention Ventura.

[8]English translation; here is the original 16th century Italian: *Modo facilissimo, et comodo, che non ha bisogno alcuno di scontro, ouer contrasegno, ma è basteuole la prima et seconda litera solamente del proprio nome di coluj che scriue, ouer a chj vien scritto, et così li primj , o ultimj caratterj del palese, et tutta la integra dicione se si vuole. Ben è uero che la composicion è alquanto forzata perciò che è formata come quella delle caselle. Ma quello che è di maggior importanza non ui si ricerca carta sbusata, ma qui u'è tutto quel che fa di mestierj nell'occorrenze sue.*

[9]For more information about this cipher, see (Bonavoglia, 2019) and (Bonavoglia, 2023), p. 166.

[10]See (Kahn, 1996) p.144-145. Original is in (Cardanus, 1553)

text are also confused, apparently in random order.

What is yet to be understood is how the sender could let the recipient know how and where to find the right windows. To locate them, a number could be used for the location of each window. How can the recipient know the location of the words or letters in the secret text? Clearly, this information should be present in the keyword, which in this case is the last word of the overt message, *satisfacione*. So, this word should contain information about the location of the virtual windows, but how?

4 Example n.2

As a second example, we see, in Figure 5, a very short message and a corresponding ciphertext (above, Manifesto).

It is immediately noticeable that there is no trace of the words of the secret text in the manifest message. So it must be a system that uses single letters instead of words.

Indeed, looking in the *manifesto* for the letters of the secret message we find them in good order, marked bold in the figure, until the letter ***m***; now after ***m*** we expected a ***p*** like in *tempo* but the only ***p*** in the manifesto is in the first line in *per*. Of course this is not the way.

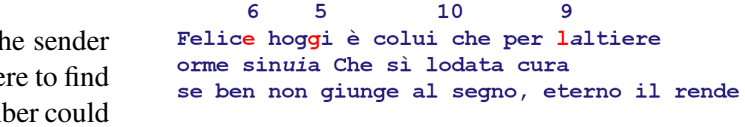
And how was the sender supposed to communicate the position of the first letter of the secret, and the one of the next letter and so on?. The keyword (contrasegno) here *FELICE* should contain this information in some way.

The first simple conjecture is that a letter of the key means a number, the ordinal inside the alphabet, so:

F	E	L	I	C	E
6	5	10	9	3	5

where 6 could mean sixth letter from the beginning, 5 mean the fifth letter from here, and so on.

Now, this rule initially works fine: the sixth letter is **E**, the following fifth is **G**, the following tenth is a space, and the following ninth is **L**. So far, we have **EGL**. By counting spaces as well as letters, this forms the beginning of the secret message, although the spaces should be ignored when completing the plaintext. However, in the subsequent part of the cryptogram, this rule no longer applies. Therefore, there seems to be something wrong; it's likely just a coincidence, and the general rule remains unknown



5 Example n.3

This example, visible in Figure 6, was the most puzzling of the three. Ventura boasted that he could reduce the secret message to a much shorter one, which is the opposite of what usually happens in steganography, where the ciphertext is often longer, sometimes much longer, than the original message. In this case, an Italian text of 977 characters is compressed into a Latin text of 32[11] words, totaling 206 characters, with a compression ratio of 4.7. While a Latin text is typically shorter than its Italian equivalent, it's rarely to such an extent. Modern compression algorithms can achieve similar ratios, but the resulting compressed file is usually a sequence of bits rather than a text readable in any known language.

And so the suspicion arises that Ventura invented a method or algorithm specifically tailored for this message. In other words, a method that does not apply to just any text, which is a prerequisite for any form of communication.

One of the most efficient compression algorithms used today involves converting the sequence of bits into a single number using a base much greater than 2. For example, the decimal number 10000, in base 2, is represented as 0010 0111 0001 0000, which is much bulkier; in hexadecimal, it's represented as 2710, which is more compact. With larger bases, the representation becomes increasingly compact. Of course, for this method to work, both correspondents must know the number N used as the base, as well as all the N signs used.

These considerations reminded me that algorithms of this type were well known and used since the 15th century: syllabaries and dictionaries, used by most nomenclators, were useful also to shorten the length of a message and the time necessary to write it, using an enlarged alphabet, one that had also tens of syllables and hundreds of words. They were naive compression algorithms after all.

On the contrary, Trithemius' *Ave Maria*[12] de-

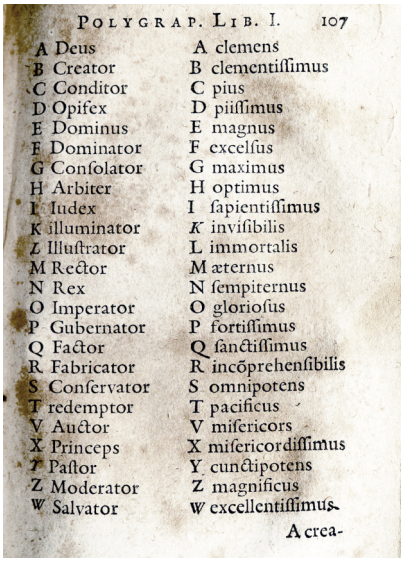
[11] Indeed, there are 31 clearly visible words and a closing sign barely readable; in the following, I assume 32.

[12] See (Trithemius, 1613) p. 107. The name *Ave Maria* is

scribed in the next paragraph, was well known and often ridiculed as a waste of paper and time. However, if used in reverse, could it significantly reduce the size of the message?

Is it possible that Ventura employed a reverse *Ave Maria*?

This idea may seem absurd for many reasons, yet it aligns remarkably well with this example, not to mention what Vigenère wrote in his treatise


POLYGRAP. LIB. I. 107

A Deus	A clemens
B Creator	B clementiſſimus
C Conditor	C pius
D Opifex	D piiſſimus
E Dominus	E magnus
F Dominator	F excelſus
G Conſolator	G maximus
H Arbiter	H optimus
I Iudex	I ſapientiſſimus
K illuminator	K inviſibilis
L Illuſtrator	L immortalis
M Rector	M æternus
N Rex	N ſempiternus
O Imperator	O glorioſus
P Gubernator	P fortiſſimus
Q Factor	Q ſanctiſſimus
R Fabricator	R incõprehenſibilis
S Conſervator	S omnipotens
T redemptor	T pacificus
V Auctor	V miſericors
X Princeps	X miſericordiſſimus
Y Paſtor	Y cunctipotens
Z Moderator	Z magnificus
W Salvator	W excellentiſſimus.

A crea-

Figure 1: The first page of the *Ave Maria* with two alphabets.

Firstly, for readers unfamiliar with this cipher, let's provide a brief explanation starting from Figure 1, where the first two alphabets of 24 letters appear[13]. In Trithemius' book, there are hundreds of such alphabets to cover messages of equivalent lengths. Each character is encrypted with a word chosen sequentially from the corresponding alphabet[14]. For instance, if the first character is **T**, the next word in the sequence is *Redemptor*, which becomes its cipher. Similarly, if the second character is **R**, the next word is *Redemptor*, resulting in a ciphertext beginning like this: *Redemptor excelsus*. Eventually, the ciphertext resembles a sermon or prayer[15], appearing unsuspected to any potential intercepting spy. Each alphabet provides roughly 24 interchangeable words, ensuring the creation of plausible texts such as sermons and prayers.

After examining the three components of the example—the fake Latin text, the key, and the true text in Italian—the broad outline method attributed to Ventura emerges as follows[16]:

1. Encrypt each word (or group of words) of the secret plaintext into letters of the alphabet, using a reverse *Ave Maria* in Italian.

2. Encrypt each single letter and the corresponding letter of the keyword into another letter, using a polyalphabetic cipher like Bellaso's or Recta Tabula with a *Bellaso contrasegno*[17].

3. Encrypt each letter obtained in the previous step into a word, using a cipher like *Ave Maria* in Latin.

A first way to synthesize this method is to use mathematical symbolism: let A be an *Ave Maria* enciphering function to substitute an Italian word (or group of words) x with a single letter a, and A^{-1} its inverse; let P be the polyalphabetic cipher that transforms a letter of the alphabet into another from the same alphabet; let B be another *Ave Maria* cipher in Latin. Finally, we have the formula:

$$y = B(P(A^{-1}(x)))$$

A second way is to use a flow-chart like diagram like the following:

never used by Trithemius but was introduced much later.

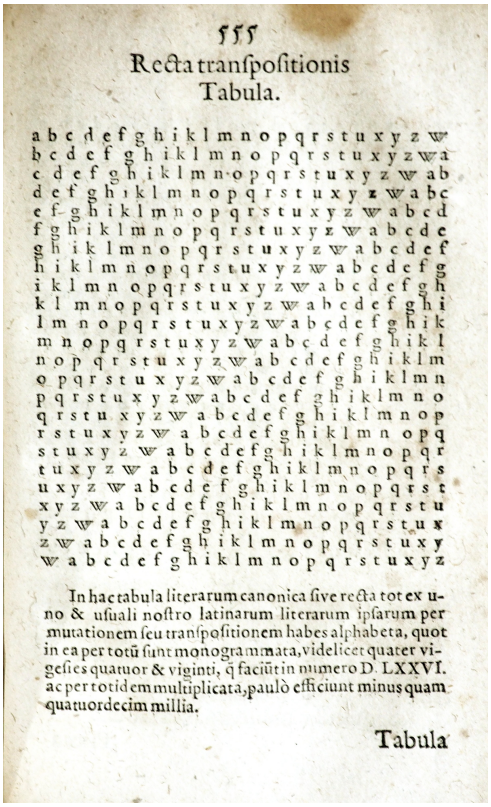
[13]This alphabet was used at the time for the German language: ABCDEFGHIKLMNOPQRS TVXYZW. Note that **W** is the last letter after Z, presumably added recently.

[14]It's important to emphasize that *Ave Maria* corresponds to a polyalphabetic cipher, meaning it's a one-to-one relation, not a one-to-many relation as sometimes misconceived. In fact, the encryption function takes as input not only the letter or word to be encrypted but also, essentially, the ordinal number n of the alphabet used, and outputs the encrypted letter or word, as well as the ordinal number of the alphabet to be used in the next step, which here is simply $n+1$ but could also be interpreted differently. Thus, both the procedure for encrypting and that for decrypting are unambiguous.

[15]Hence the name *Ave Maria*, as mentioned earlier.

[16]A very similar outline is found in one of the *Falso Scontro* cipher variants by Hieronimo di Franceschi, see (Bonavoglia, 2022), the most prominent secretary of ciphers in the late 1500s. Did Franceschi, who was about 30 years old, know Ventura or read his papers?

[17]I did not refer to it as Vigenère's tableau (see the original in French (de Vigenère, 1587) p. 50, and (Kahn, 1996) p. 149) because Vigenère published it in 1586-87, two decades later!.



fff

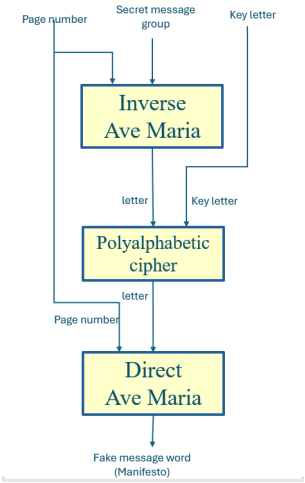
Recta tranſpoſitionis Tabula.

a b c d e f g h i k l m n o p q r s t u x y z w
b c d e f g h i k l m n o p q r s t u x y z w a
c d e f g h i k l m n o p q r s t u x y z w a b
d e f g h i k l m n o p q r s t u x y z w a b c
e f g h i k l m n o p q r s t u x y z w a b c d
f g h i k l m n o p q r s t u x y z w a b c d e
g h i k l m n o p q r s t u x y z w a b c d e f
h i k l m n o p q r s t u x y z w a b c d e f g
i k l m n o p q r s t u x y z w a b c d e f g h
k l m n o p q r s t u x y z w a b c d e f g h i
l m n o p q r s t u x y z w a b c d e f g h i k
m n o p q r s t u x y z w a b c d e f g h i k l
n o p q r s t u x y z w a b c d e f g h i k l m
o p q r s t u x y z w a b c d e f g h i k l m n
p q r s t u x y z w a b c d e f g h i k l m n o
q r s t u x y z w a b c d e f g h i k l m n o p
r s t u x y z w a b c d e f g h i k l m n o p q
s t u x y z w a b c d e f g h i k l m n o p q r
t u x y z w a b c d e f g h i k l m n o p q r s
u x y z w a b c d e f g h i k l m n o p q r s t
x y z w a b c d e f g h i k l m n o p q r s t u
y z w a b c d e f g h i k l m n o p q r s t u x
z w a b c d e f g h i k l m n o p q r s t u x y
w a b c d e f g h i k l m n o p q r s t u x y z

In hac tabula literarum canonica ſive recta tot ex uno & uſuali noſtro latinarum literarum ipſarum permutationem ſeu tranſpoſitionem habes alphabeta, quot in ea per totũ ſunt monogrammata, videlicet quater vigeſies quatuor & viginti, q̃ faciũt in numero D. LXXVI. ac per totidem multiplicata, paulò efficiunt minus quam quatuordecim millia.

Tabula

Figure 2: Trithemius recta tabula.



Now let's see if it is possible, consistently with the example provided, to reconstruct the method in detail, how the *Ave Maria* cipher type should be structured, how many pages, how many alphabets, and how to use them.

Let's examine two possible implementations:

5.1 Implementation 1

The first implementation assumes that the secret text is arranged in a single line of 32 groups. The cipher should consist, as in the *Ave Maria*, of as many rows as there are letters of the alphabet (24 in German or 20 in Latin and Italian). For simplicity, let's assume that the text in the example is all placed on the line identified by the letter **T**:

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

Figure 3: The original Vigenère table using the 20 letters Latin alphabet; given key letter E and text letter Q, the cipher letter is V at the intersection of row E and column Q (Or vice versa).

The cipher should consist, as in the *Ave Maria*, of as many rows as there are letters of the alphabet (24 in German or 20 in Latin and Italian), and assuming for simplicity that the text in the example is all placed on the line identified by the letter **T**:

S	…	S	…	S	…	S	…	T	…
T	Tenetevi	T	a tutta forza	T	et non ui rendete	T	à patto alcuno	T	hauerete subito
V	…	V	…	V	…	V	…	V	…

and the resulting ciphertext from this first step would now be **TTTTTT**.

The next step involves encrypting each of the letters obtained with another, using the keyword or *contrasegno*, which here is *SANCTVSMARCVSVENETVS*. For this purpose, one of the many polyalphabetic ciphers invented in the sixteenth century is required. Ventura uses the unusual *contrasegno*, which is the one used by Bellaso in his ciphers[18]. It is likely that he uses one of these. To simplify matters, I have used the well-known Vigenère table with an alphabet of 20 letters (see Figure 3), despite it being an anachronism as it was only published in 1586.

Secret	T	Tenetevi	T	a tutta forza	T	et non ui rendete	T	à patto alcuno	T	hauerete subito …
Key	S		A		N		C		T	…
Fake	P	Nunc	T	dimittis	I	seruum	Z	tuum	Q	Domine …

The ciphertext is now: **PSTIZQR …**, which in the next step must give Ventura's solution: *Nunc dimittis seruum tuum …*; this requires assigning

[18] See (Bauer, 2007) and the original (Bellaso, 1553).

the first resulting letter **P** to *Nunc*, the second **S** to *dimittis*, and so on.

5.2 Implementation 2

The second implementation hypothesis is suggested by the fact that the secret text is a sequence of eight sentences of similar structure, an imperative followed by an object complement or similar, a conjunction, and a subordinate sentence: it could thus be a cipher of only eight lines of four groups; each line is identified by a letter of the *contrasegno* (keyword) **SANCTVSMARCVSVENETVS** as follows:

A	Teneteui	a tutta forza	et	non ui rendete …
C	Hauerete	subito soccorso	di	vittouaglia dinari soldati et altre monicjoni
E	Non innouate	cosa alcuna	fin che non	hauete altre nostre per salute comune.
M	Defendeteui	alla muraglia	quanto podete et non	uscite fin che altro auisamo
R	Viuete	sicuri	che senza fallo alcuno	s'hauera Vittoria immortalissima
S	Scriuete	quanto più presto	per più strade se	ui occorrera cosa che ui sia di importanza …et bisogno Accioche subito proueder si possa
T	Sappiate	che si tenira sempre buon conto	de	la fede et sincerita che …hauete sempre uerso noi dimostrato
V	Perseuerate	donque al simile	che non	si mancherà di giusta ricompensa

Indeed, several combinations produce plausible texts, such as the following two, while others sound really bad.

A	Teneteui	M	alla muraglia	E	fin che non	S	ui occorrera …
M	Defendeteui	A	a tutta forza	M	quanto podete et non	V	si mancherà …

This first step produces **AMESMAMV** as an intermediate ciphertext to be overwritten with the letters of the keyword. This can be implemented using an 8x8 Vigenère table identified by the eight letters of the keyword. Finally, as in the previous implementation, the individual letters will be replaced by words from the fake text *palese*.

Overall, this example method 3 is an ingenious steganographic cipher, producing short fake texts from a limited set of true secret messages. This makes it impractical for long diplomatic dispatches; it can be used for short conventional messages, such as the famous Radio London messages during World War II. However, for this purpose, there are lighter methods available.

6 Conclusions, open questions

In conclusion, Vigenère did not misunderstand; indeed, he was quite accurate in mentioning Trithemius ciphers as the source of Ventura's ciphers.

Still, there are open questions:

- How did Vigenère learn about these Ventura ciphers in 1569, two years after this episode? Did he have some knowledge among the secretaries of ciphers for the Council of Ten? In 1567, Zuan Francesco Marin (or Marino) was the most prominent secretary for ciphers, the last great Venetian cryptanalyst.
- Were these ciphers actually used by the Republic for diplomatic or military dispatches? No evidence of such use has been found, up to now.
- The first mode is a Cardan Grid without holes in the paper, only virtual holes; but how to find their location using the *contrasegno*?
- The second mode looks similar but works on single letters rather than words. Same question as above.
- The third mode uses a reverse and a normal *Ave Maria* cipher with a polyalphabetic system. The outline design is clear, a detailed implementation

7 Acknowledgements

A special thanks to the entire staff of the State Archives of Venice for their assistance during my research.

Thanks also to ChatGPT 3.5 for the revision of the English text, that was done by paragraph.

References

Friedrich Ludwig Bauer. 2007. *Decrypted secrets: Methods and Maxims of Cryptology*. Springer, Berlin.

Giovan Battista Bellaso. 1553. *La cifra del sig. Giouan Battista Bellaso, gentil'huomo bresciano ...* G.B. Bellaso, Venezia.

Paolo Bonavoglia. 2019. The cifra delle caselle a xvi century superencrypted cipher. *Cryptologia*.

Paolo Bonavoglia. 2022. The Enigma of Franceschi's Falso Scontro. In *Proceedings of the 5th International Conference on Historical Cryptology HistoCrypt 2022*, Uppsala. Linköping University Electronic Press.

Paolo Bonavoglia. 2023. *La crittografia della Repubblica di Venezia*. Aracne, Roma.

Hyeronimus Cardanus. 1553. *De subtilitate*. Sebastianum HenricPetri, Basilea.

Bartolomeo Cecchetti. 1869. Le scritture occulte della diplomazia veneziana. In *Atti del Regio Istituto Veneto Tomo XIV Serie III*, Venezia. Istituto Veneto.

Blaise de Vigenère. 1587. *Traicté des chiffres ou secrètes manières d'escrire*. Abel L'Angelier, Paris.

David Kahn. 1996. *The codebreakers*. Scribner, New York.

Aloys Meister. 1902. *Die Anfänge der modernen diplomatischen Geheimschrift*. Ferdinand Schöningh, Paderbord.

Luigi Pasini. 2019. *Delle scritture in cifra usate nella Repubblica di Venezia*. Aracne, Roma.

Riccardo Predelli. 1869. Saggio di scritture in cifra usate dalla repubblica veneta (sec xvi-xviii) estratte dagli archivi veneti. In *Atti del Regio Istituto Veneto Tomo XIV Serie III*, Venezia. Istituto Veneto.

Paolo Preto. 1994. *I servizi segreti di Venezia*. Il Saggiatore, Milano.

Johannes Trithemius. 1613. *Libri Polygraphiae*. Lazari Zetzneri, Argentorati (Strasbourg).

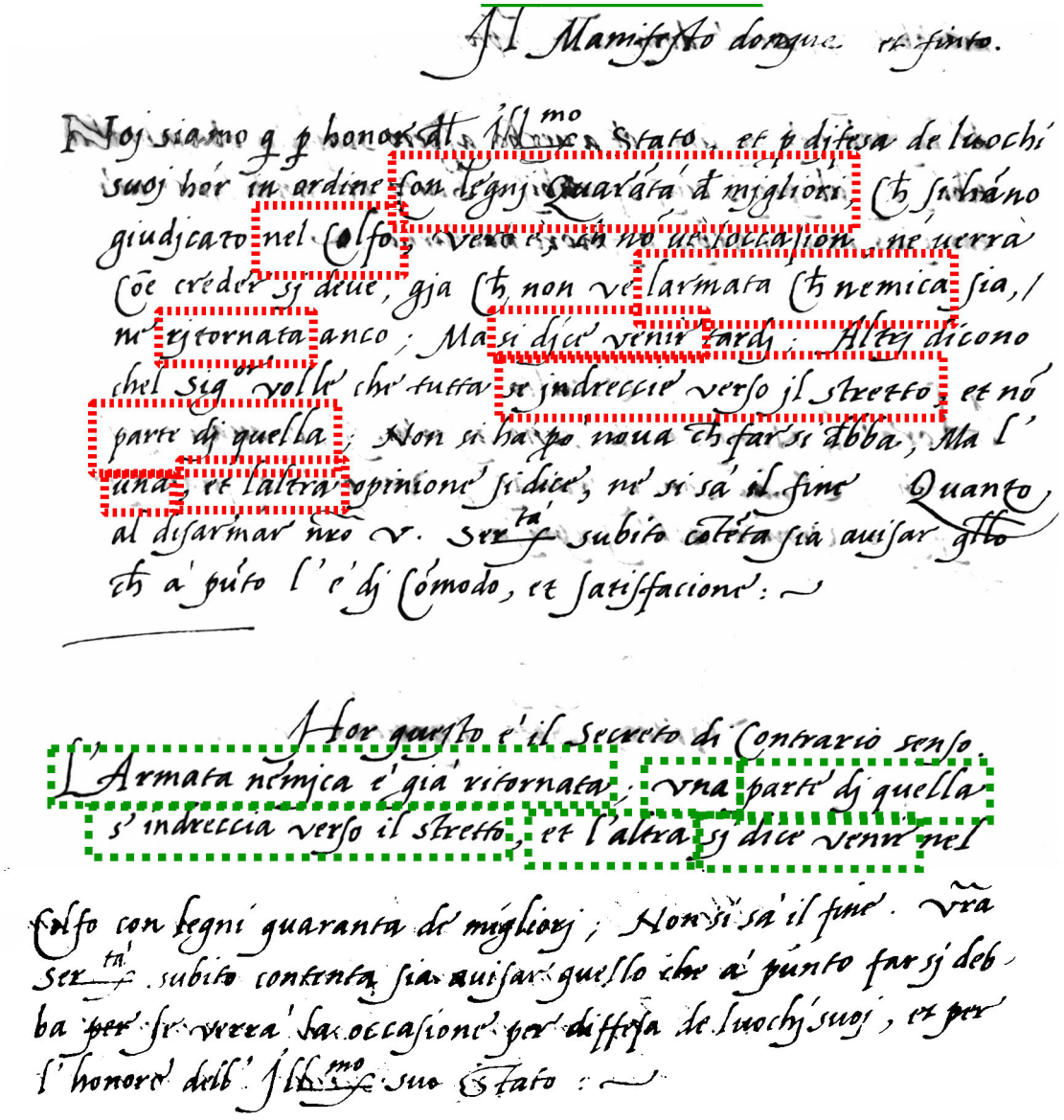


Figure 4: Example n.1. The red-border boxes show possible windows, but how was their position found? Note: Images are in HR and can be enlarged by zooming in with Acrobat or other pdf readers. *ASVe CX Cifre, chiavi e scontri di cifra b.6 misc.*

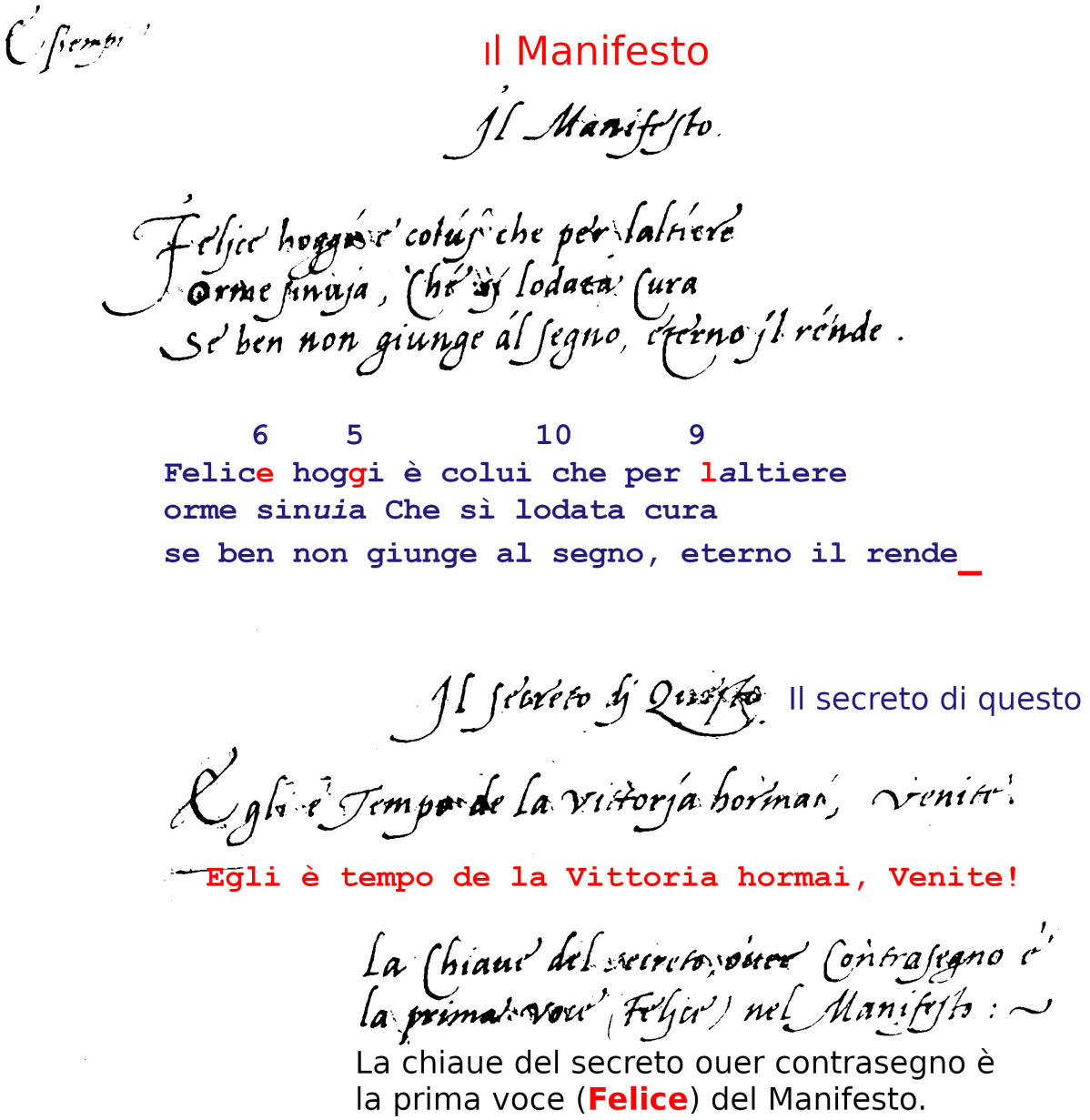


Figure 5: Example 2: The possible solution where letters of the key are interpreted as numbers of step forward. *Ibidem*

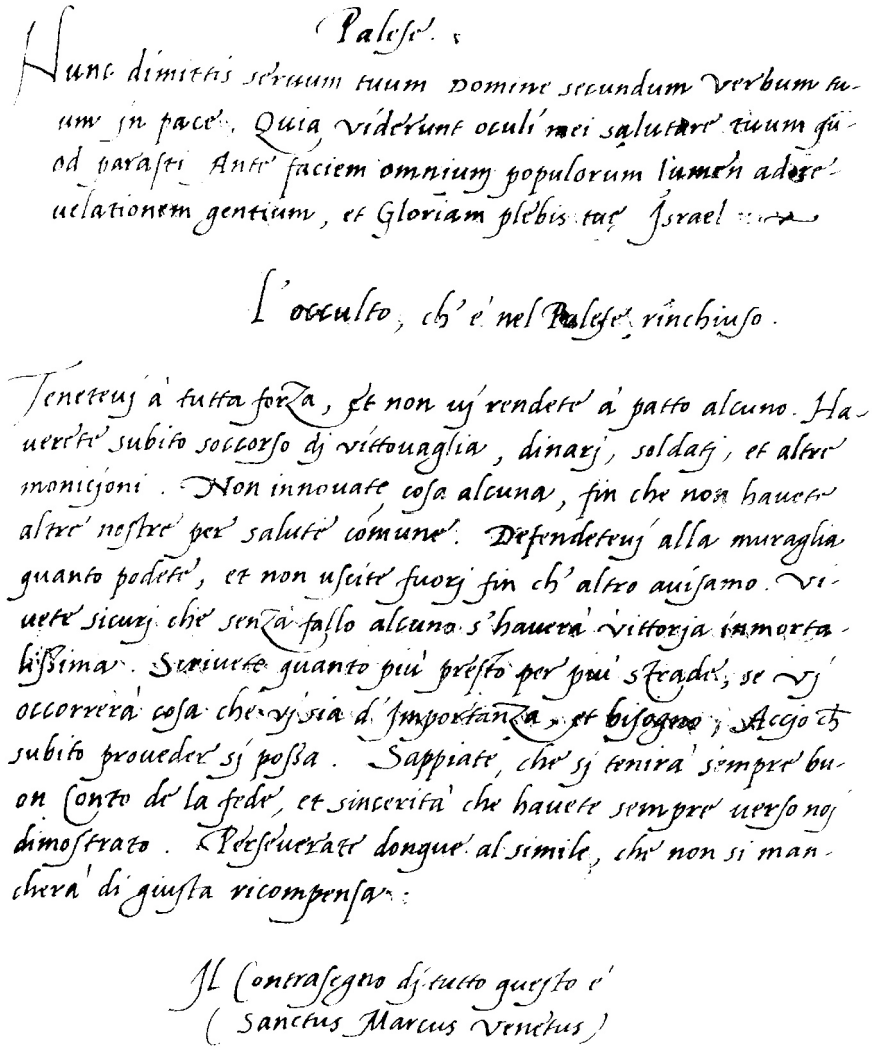


Figure 6: Example n.3, top, the *palese* fake text in Latin; middle: the true text in Italian; bottom: the keyword (*contrasegno*). *Ibidem*

Demystifying La Buse's Cryptogram and the Fiery Cross of Goa

Carola Dahlke
Deutsches Museum
Munich / Germany
c.dahlke@deutsches-museum.de

Abstract

The field of cryptology alone offers a multitude of exciting exhibits and stories for a museum of science and technology. But when secret ciphers meet seaborne piracy and rich treasures, it sounds like a perfect mix for a successful storytelling. However, an extensive study of an eyewitness account and contemporary reports on the legend of la Buse led to sobering contexts related to colonialism, inquisition and, apart from a small kernel of truth, to a large spool full of seaman's yarn.

1 Storytelling between Legend and Truth

A museum cannot assume that all visitors will be equally enthusiastic when it comes to complicated topics, i.e. in the field of cryptology. So in order to offer an attractive exhibition as well to non-experts, or to groups and families with different interests and prior knowledge, an additional range of simpler and more playful ways of presenting information is needed. Good storytelling brings an exhibition to life, and cryptology naturally offers wonderful opportunities here. The legend of the pirate la Buse's cryptogram combines everything that a lot of visitors' hearts desire: the golden age of piracy, rich treasure, and a mystic cryptogram. Nevertheless, as a scientific museum, curators want to make a clear distinction between truth and legend before exhibiting a story like this. For this purpose, this small study was carried out to sufficiently substantiate la Buse's tale with original documents - as much as possible.

2 The Legend of the Pirate La Buse

This is how the story is passed on: In the waters of the Indian Ocean, a ruthless French pirate known as la Bouche, the Mouth or la Buse, the buzzard, attacked a rich Portuguese cargo ship in 1721. He robbed the ship and the entire cargo, consisting of diamonds, jewelry, gold and silver bars as well as pearls, fine fabrics, spices, furniture and precious stones, estimated by historians to be worth up to 5 billion euros today. Religious articles from Goa Cathedral, located in India, had been on board, including the Golden Cross of Goa, which is said to have weighed more than a hundred kilograms, so that three men were needed to reload it. As well, it was said that he took the Viceroy of Goa hostage, and released him for ransom. After this successful raid, the pirate went into hiding for several years and was only discovered and executed in 1730. Shortly before his death, according to legend, he threw a cryptogram into the crowd - supposedly with the description of where to find his share of the pirate treasure. For centuries, the whereabouts of the cryptogram and the treasure were unknown.

2.1 The Mysterious Buccaneers of a Hidden Treasure

In 1934, the honourable French historian Charles de la Roncière (1870-1941), who worked at the Bibliothèque Nationale in Paris, wrote a short paperback novel.[1] In the beginning of his novel, de la Roncière explained why he wrote the story: One day, a customer from a distant African country had come to the National Library and had asked for a book on the Keys of Solomon, because her neighbour from a small island in the Indian Ocean had made strange discoveries on her property: stone sculptures, rock carvings, and finally she had also unearthed three bodies that were identified as slaughtered buccaneers. Following de la Roncière's novel (p. 5), the notary on the island then handed over a cryptogram to the lady with the words: "It is your property, where the pirate's treasure must lie. Here is everything you need to

[1] It was rather cheaply published at Le Masque in Paris, a publisher for small inexpensive fiction, i.e. a sort of literature known as dime novel in US-American context.

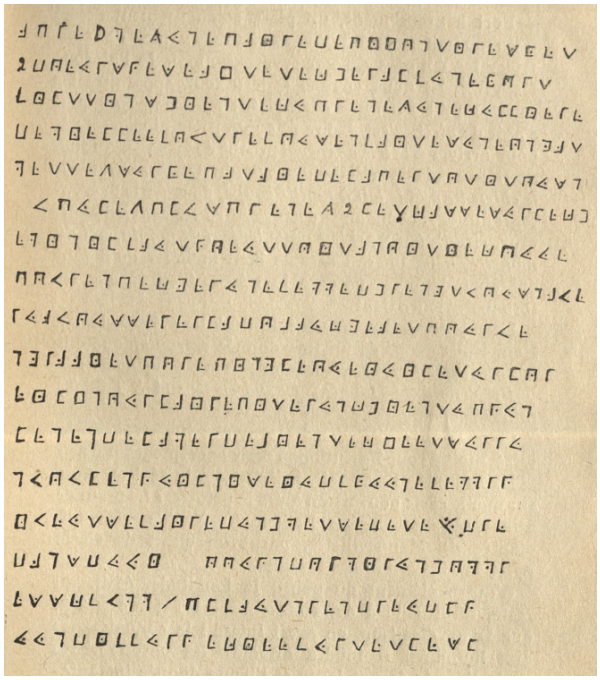


Figure 1: The alleged cryptogram of la Buse, taken from de la Roncière, p. 9 (1934)

find it." De la Roncière, upon the lady's request for assistance, deciphered the cryptogram for her, by promising not to give away any details on the whereabouts of the alleged treasure.

In his novel, de la Roncière published the cryptogram, the key he himself had found out, and the deciphered cleartext (see the Figures 1-3). He was of the opinion that some of the expressions used in the cryptogram could have originated from a Frenchman who must have come from the North-East of France, because of an obvious hard pronunciation of consonants. He cited the following words as examples of this: kort for corde, piter instead of bitter.[2] As well, he hinted that he already had exact knowledge as where to find the treasure. After the short cryptologic work in the beginning of the novel, a couple of chapters on East Indian piracy followed mainly based on the historical works of Johnson (1726), Hamilton (1727), and Grandidier and Grandidier (1907). In the end of the novel, de la Roncière related the supposed treasure to the great pirate's raid of 1721. He suggested that the cryptogram might have been written by the French pirate la Buse because la Buse had been part of the raid, and he had been from Calais (North-East of France). In the very end of the novel, de la Roncière delivered clear hints in the form of names printed in italic font that referred to locations on the Seychelle island Mahé.[3]

Concerning la Buse's execution, there is archival material in the National Archives d'Outre-mer that document his execution. Several accounts have already analysed the relevant sources, e.g. the books by Guët (1886, p. 219f), de Kerdéland (1961, p. 162), and Briseul (2019), who devotes a whole chapter on the death of la Buse. De la Roncière also based his story on the relevant documents, and quoted from a letter from December 1730 that Gouverneur of Bourbon island Pierre-Benoît Dumas wrote to the Count de Maurepas: "In 1728, the *La Méduse*, commanded by d'Hermitte and sent to ensure navigation between Bourbon and Madagascar, caught la Buse in the vicinity of Fort-Dauphin, where he had made his retreat, and brought him in chains to Bourbon. Although la Buse claimed amnesty, a ruling on 17 July 1730 proved that since he had continued his life as a pirate, he was excluded, and he was immediately hanged on

[2] See de la Roncière (1934, p. 8):"Si l'on peut émettre une hypothèse sur l'auteur de ces lignes bizarres, c'est que certaines prononciations dures semblent révéler un homme du Nord-Est de la France: *k'unne, kort* (pour *corde*), *piter* (pour *bitter*, qui consiste à enrouler un cordage autour de la bitte d'un navire)" (English translation: "If we can speculate on the author of these bizarre lines, it's that certain harsh pronunciations seem to reveal a man from the North-East of France: k'unne, kort (for rope), piter (for bitter, which means winding a rope around the bitt of a ship)"

[3] See de la Roncière (1934), p. 112f: "...Vous êtes *curieuse*, *cousine Thérèse*, et vous, *Marianne*, et vous, *Félicité*, de connaître la *silhouette* de l'archipel où gît le trésor? Mystère et discrétion. Pour que vous ne preniez pas *la mouche*, je vous convie à une promenade dans la *belle ombre* d'un *beau vallon*, comme dans un *port de consolation*. ...Peut-être y trouverez-vous, amis lecteurs, le mot de l'énigme. Tous ces noms en italique, *cousine, curieuse, silhouette, mouche, Thérèse, Félicité*, n'ont rien d'imaginaire. Ce sont des vocables gèographiques. Tous appartiennent à l'archipel que je veux pas autrement nommer. S'il est doté d'une anse la Blague, il a aussi, - et les mots seuls suffisent à attester la véracité de cette histoire, - une anse des Forbans. La découverte de leur trésor me servira, un jour, d'épilogue. Derrière le glacis de falaises qui masquent son gîte, les fouilles ont commencé..." (English translation: "...Are you *curious, cousin Thérèse*, and you, *Marianne*, and you, *Félicité*, to know the *silhouette* of the archipelago where the treasure lies? Mystery and discretion. So that you don't catch *the fly*, I invite you to take a walk in the *beautiful shade* of a *beautiful valley*, as if in a *harbour of consolation*. ...Perhaps you, dear readers, will find the answer to the riddle. All these names in italics - *cousin, curious, silhouette, fly, Thérèse, Félicité* - are not imaginary. They are geographical terms. They all belong to the archipelago that I don't want to reveal. If it has a Joke cove, it also has - and the words alone are enough to attest to the veracity of this story - a Forbans' cove. The discovery of their treasure will one day serve as my epilogue. Excavations have begun behind the brink of cliffs that conceal its hideaway..."

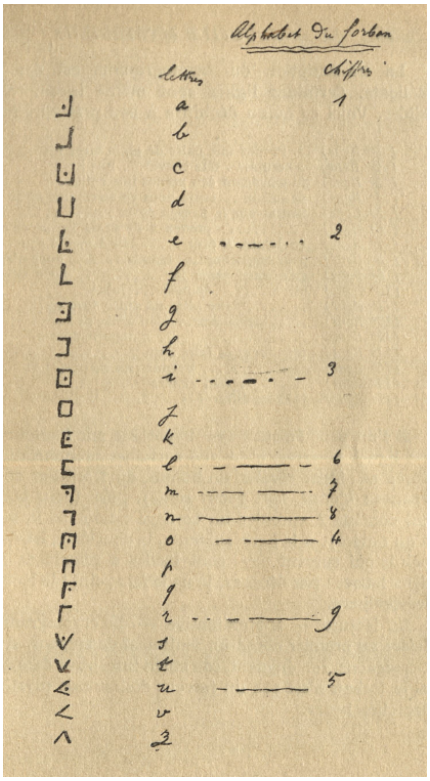


Figure 2: The masonic key, deciphered by de la Roncière, p. 7 (1934)

the beach at Saint-Denis to the applause of the populace"[4]. At the end of the letter, Dumas also mentioned that the rhubarb was starting to grow again on Bourbon island. De la Roncière referred to this rather unspectacular remark and added a personal note, declaring that he would have preferred another ending of the story: "Tradition has it that the forban, who hid his treasure, would hand the crowd his cryptogram before marching to his death, and that his last words would be like a testament: 'For whoever discovers this.'"[5]

Shortly after its publication, a detailed summary of the novel appeared in the daily French newspaper Le Temps on the 5th of May 1934 by G. Lenotre alias Louis Léon Théodore Gosselin, and an American version was published in July 1934 in the Milwaukee Journal (R.S. Fendrick, 1934).

1. Prenez une paire de pijon virez les
2 cœurs... tête cheval... une kort
fil winshient écu prenez une cullière
de mielle .. outre vous en faites une ongat
mettez sur le pasage de la .....
..... prenez 2 liv cassé sur le ch-
(ch)emin, il faut ..... toit à moitié couvé
pour empecher une femme ..... vous n'avé
qu'à vous serer la ..... pour ve-
nir ..... épingle ... juilet .....
..... faire piter un chien turq un
..... de la mer.. bien sèche et sur ..
..... k'unne femme q-
(qu)i veut se faire d'un .....
dans ..... dormir un homm(e) r
..... faut rendre ... q
(q)u'un diffur (?) .....

Figure 3: The deciphered cryptogram of la Buse, from de la Roncière, p. 8 (1934)

In the years that followed, no document could be found in which de la Roncière provided more detailed information, particularly regarding the origin and whereabouts of the cryptogram. However, de la Roncière described a small continuation of his story in a booklet that was printed in 1940 for the French army for reading and distraction in wartime. On pages 26-27, he retold a summary of his first novel on the pirate's treasure, but then added another episode to it: By accident, a version of the newspaper from the 5th of May 1934 written by G. Lenôtre (1934) had ended up in Cameroon, and had been read by an islander from the Seychelles. She sent her mother Mme D. to visit de la Roncière in Paris. De la Roncière did not reveal more than the first letters of the last names, but hinted that Mme D. owned not only a property on Coëtivy island, but as well complete Silhouette Island - which refers to the family name Dauban - the family that owned Silhouette island until 1960. Mme D. discussed with de la Roncière whether the treasure could have been on her properties. She had seen a ship anchored overnight, and on the next day a large hole in the ground would have indicated two buried urns that may have been dug up (de la Roncière, 1940, p. 27). De la Roncière then emphasized that there were usually several pirate hideouts and that more could be found. His last sentence was directed to Mme S.[6], who could now rely on the help of the radiesthesist Abbott Mermet

[4] See Dumas (1730): "Le sieur d'Hermitte, dans le dernier voyage qu'il a fait à Madagascar, ayant arrêté et amené ici le nommé Oliver Le Vasseur, dit la Buze, fameux capitaine forban, son procès lui a esté fait à la requête du Procureur Général, et il a été pendu par arrest du Conseil..."

[5] See de la Roncière (1934, p. 108f)"La rhubarbe commence a multiplier ... J'aurais préféré une autre phrase finale. La tradition veut que le forban, qui cacha son trésor, tendît à la foule son cryptogramme avant de marcher au supplice et que ses dernières paroles fussent comme un testament: 'Pour celui qui le decouvrira.'"

[6] Author's remark: Mme S. is most probably the islander and mentioned daughter of Mme D.

to find the treasures.[7]

Although unlikely, de la Roncière's story triggered a veritable treasure hunt that continues in earnest to this day. Numerous books and filmic interpretations reproduced this story at second and third hand, and a large community is busy with deciphering and interpreting the supposed cryptogram again, and keeps hunting the pirate's treasure, describing the main artefact, the Fiery Cross of Goa, as a famous piece of art.

2.2 La Buse's alleged Cryptogram

Since its publication in 1934, the authenticity of the cryptogram was regarded with great doubt. As far as we know today, de la Roncière's paperback novel is the only source that exists. To this day, not any other publication on original sources, or other statements of de la Roncière himself on the whereabouts of the original cryptogram is known to the author. Allegedly, de la Roncière claimed that the cryptogram was kept in the Bibliotheque Nationale. However, the author was unable to find a source for this. Apparently, the newspaper article in the Milwaukee journal by R.S. Fendrick (1934) has often been cited to be an interview with de la Roncière, with deeper insights on the cryptogram's origins, but it is not. According to information circulating today, the cryptogram was attributed to the heirs of a pirate called Nageon de l'Estang. It seems that this further information comes from a book by the journalist Robert Charroux, alias Robert Grugeau, written in 1962. The author of this article has tried to retrieve any citable source Charroux based his story on, but so far without success.

Written in a simple masonic cipher, i.e. easy decipherable. It had therefore been less of a concern for scientists than for adventurers and treasure hunters. De la Roncière delivered not only the cryptogram, but as well the masonic key and the deciphered text ready for interpretation. But the content of the cryptogram sounds more like a cooking recipe than a clue to a treasure. Accordingly, there are virtually no citable peer-review publications on the subject. Some profound books on piracy can be found, see e.g. Rennie (2013) and Fox (2014), and as well a handful of critical websites. But even though it was assumed by the scientific community that the cryptogram was a hoax, there was still the big question of where the famous pirate treasure was hidden. At least, both the French pirate la Buse and the described Portuguese cargo ship, with the Viceroy of Goa on board, had existed.

But were these facts enough to include the legend of la Buse into a science museum's gallery on cryptology? After all, myths and hoaxes were also part of the history of cryptology and were therefore given a small stage in the exhibition - of course, presented with a corresponding wink of an eye. The author of this study agreed, but for a scientific presentation, all details of the legend, and of the famous Cross of Goa, were to be researched and documented as thoroughly as possible.

3 The Sources

In order to get to the true story, contemporary sources had to be sought whose authors had been involved as closely as possible in the raid on the Portuguese ship. The name of the gallon was *Nossa Senhora do Cabo e São Pedro de Alcantara*[8], and since it had the Viceroy of Goa on board, it was indeed easy to find not only mentions in several books of that time, but as well one hand-written eyewitness account of the famous raid in the Indian Ocean. First of all, the books of Grandidier and Grandidier (1907) and Grey (1933) provided meticulously researched and sourced overviews. They enabled to find several very informative historical resources, of which the following were particularly important:

- The eyewitness account of the British seaman Richard Lazinby (1722), who worked as second mate on the ship *Cassandra* when it was taken by the pirates Edward England and John Taylor in 1719, and from then on sailed as their pirate ship. Lazinby was forced to serve the pirates as a prisoner, and was only released together with the Viceroy of Goa when this man was exchanged for ransom. In 1722, Lazinby described his personal experiences, which can be viewed as handwritten manuscript in the British Library.

- The report of the Viceroy of Goa, published in the Parisian newspaper Le Mercure (1722), authored anonymous. By and large

[7] "Puisse Madame S. ... déceler le gîte, grâce à la radiesthésie de l'abbé Mermet!" (May Madame S. ... find the place, thanks to Abbé Mermet's dowsing!)

[8] Sources claim see e.g. de Bucquoy (1771, p. 66) that it was the former Dutch Gallon *Gelderland*

it is noticeable that the article emphasises the Portuguese-French collaboration.

- The eyewitness account of the Dutch seaman and gunner Clement Downing (1738), who worked for the East India Company, and later wrote his memoirs about his days in the Indian Ocean.
- The contemporary descriptions of Captain Charles Johnson (1726), who apparently published under a pseudonym [9].
- The travel reports of the surveyor and cartographer in the service of the East India Company Isidore de Bucquoy (1771), who met in person John Taylor and La Buse on his travels in 1722. Although he was held prisoner for some time, he had been treated friendly and also had intensive conversations with Taylor. In his book, he quotes Taylor's account of the raid of the *Nossa Senhora*, and of the years after the raid.

4 Discussing the presumably Real Story

From the sources mentioned above, a presumable real story emerged. Of course it must be mentioned here that the sources did not depict one completely clear and uniform story, and definitely there are especially administrative archivals that still need to be investigated. But from the knowledge that could be gathered so far, especially surprising was the fact that the different accounts did not correspond concerning the pirates in command. According to almost every report, the British pirate John Taylor was the captain and boss of the entire pirate gang. And while Taylor appeared in every report, la Buse was only sometimes mentioned. However, it can be assumed that he was involved in the raid on the Portuguese ship because in addition to the sources mentioned above, documents from the Bourbon authorities (see the letter of govenor Dumas (1730)), condemned la Buse alias Oliver Levasseur to death, partly for his parting in the raid.[10]

On the 8th of April 1721, two pirate ships, one of them for sure the *Victory* and the other most certainly the *Cassandra* with eyewitness Lazinby on board, arrived in the North of Bourbon island (Réunion). Lazinby (1722) counted about 300 pirates on bord of the two ships, Le Mercure wrote of 480 pirates. At Saint-Denis, the pirates found a large Portuguese cargo ship lying at anchor that had lost all of her masts and two thirds of her canons (Lazinby, 1722). According to the sources, there was almost no resistance when they took the ship, because the crew of the Portuguese ship thought at first sight that both pirate ships were British Company (see e.g. Downing (1738), p. 66).[11]

On board of the Portuguese ship was the Viceroy of Goa, namely the comte d'Ericeira Luís Carlos Inácio Xavier de Meneses, first Marquis of Louriçal. After some discussion among the pirates, they took 2000 Risdales respectively Dollars[12] for ransom and set the Viceroy and all the other prisoners, among them eyewitness Richard Lazinby free.[13] The pirates then made off to St. Mary with all the ships and the valuable cargo of the *Nossa Senhora do Cabo e São Pedro*.

4.1 The Cargo

In St. Mary it is said according to Johnson (1726, p. 139) that the pirates turned the booty into money, and that the rich booty was shared among 280 surviving pirates. This meant that each pirate received about 42 small diamonds, and a reasonable amount of cash (about 4000 Pounds according to Johnson). The rich cargo of the Portuguese Ship consisted on the one hand of plunder and diamonds that were estimated by the Viceroy of Goa (who told Lazinby) of three to four million dollars. Later, Taylor in person told de Bucquoy (1771, p. 64) that the values were about 30 Mio Gulden[14].

[9] There is no evidence that a certain Captain Charles Johnson existed, but his reports show that he must have been nearby or even an eyewitness to the events described. For many years, his works were attributed to Daniel Defoe, but today this thesis is disputed.

[10] In his book *Sailing East*, Baylus C. Brooks explores why la Buse was captured and ultimately executed.

[11] The *Victory* was said to be the former *Petersborough Galley*, and the *Cassandra* was a captured British ship of the East India Company.

[12] Johnson cited Risdales, and Lazinby cited Dollars

[13] There circulated a legend about the price negotiation for the Viceroy's ransom, which Bernardin de Saint-Pierre (1800, p. 218) heard about 50 years later during his travels on Bourbon, but it did not match at all the sources examined: the governor of Bourbon island, M. Desfourges and the Viceroy were supposedly sitting at dinner when the *Nossa Senhora* was attacked. Suddenly, la Buse appeared between them and said that the Viceroy was now his prisoner. When the governor asked how much la Buse wanted for him, la Buse said, 1000 piasters. That was too little for such a worthy gentleman, said the governor, go and ask for more or nothing. La Buse then said he wanted nothing and the viceroy was free.

[14] "Man hat mich mehr als einmal versichert, daß der Werth laut der Factur und des Geständnisses der Privatinteressenten, über dreyssig Millionen Gulden betragen habe; welches mit

But as well, the intermediate deck of the *Nossa Senhora do Cabo* was filled with more than 200 slaves that had been taken on board at Mozambique, and had been caged in the Portugese ship to be sold at Madagascar (Lazinby, 1722; Johnson, 1726). About 60 of them had already died in the short battle when the *Nossa Senhora* was taken. The report of the Portuguese Viceroy also mentioned the slaves in the belly of the ship (Le Mercure, 1722). It must be concluded from the sources that the valuable cargo of the ship *Nossa Senhora* included slave trade under cruel conditions. Unsurprisingly, at this time there was a slave trade around Madagascar, from which both the East India Company and the pirates profited. What is surprising, however, is the fact that this has so far been completely omitted in most renditions of the legend of the pirate La Buse.

4.2 The Whereabout of the *Nossa Senhora do Cabo*

In St. Mary, the pirates replaced the broken *Victory* with the *Nossa Senhora do Cabo*, which was repaired[15], and from then on sailed under the name *Victory* (see e.g. Downing (1738, p. 67)). In the first few months after the raid, Taylor was captain of the *Victory* and la Buse commanded the ship *Cassandra* (de Bucquoy, 1771, p. 66f).[16]

However, during summer 1722, a major dispute arose between Taylor and La Buse and they parted ways. Once again the ships were divided. From then on, John Taylor was said to have commanded the *Cassandra* until his surrender in 1723 (Grey, 1933, p. 327f), (Johnson, 1726, p. 140f), (de Bucquoy, 1771, p. 68). Taylor and his crew surrendered their ship and most of their accumulated booty to Porto Bello[17], and were subsequently granted amnesty. They were only allowed to take what they had on them - according to de Bucquoy, each of the pirates only kept some gold and the diamonds for a new start in their pockets (de Bucquoy, 1771, p. 71). John Taylor had his share of diamonds with him, and spent it living a settled life with wife and children. But his wealth did not last long. In 1744 Taylor was said to be a poor and miserable fisherman (de Bucquoy, 1771, p. 71f).

La Buse got the *Victory*, and only little is reported about him from then on. According to Guët (1886) and Grey (1933, p. 327f), he kept on with his pirating life, but burnt the *Victory*, probably along the coast of St. Mary.[18]

4.3 The Golden Cross of Goa

What was not mentioned anywhere, however, was the legendary Cross of Goa. It could not be found in any source that was investigated. Nor does de la Roncière himself mention the cross in his book. It seems, most surprising, that the much-described Golden Cross of Goa had never existed, and only entered the legend after the paperback novels of de la Roncière in 1934 and 1940. So far, the authors could not define the point when this has happened. Nevertheless, the authors found mentions of golden or wooden crosses related to the events in 1721 or earlier:

- The Viceroy of Goa possessed a golden cross of the order, most probably hanging on a neclace, which was mentioned in his account. During the raid, the pirates had taken it from him, but then handed it back together with his sword made of gold (Le Mercure, 1722, 64)[19] It is possible though to interpret a sword of gold as some kind of golden cross.
- In the accounts of de Bucquoy (1771, 37), he cited from a not specified second-hand source a description on the cargo that had been accumulated in the belly of the *Nossa Senhora*,

Recht ein Schatz genennet werden kann."

[15] According to Downing (1738, p. 66-67), they forced the *Cassandra*'s assistant carpenter to repair and refit the *Nossa Senhora*

[16] According to the sources, it is quite possible that the *Cassandra* had been renamed by the pirates into *Fancy* (as well called *la Fantasie* (Le Mercure, 1722), or *La Défense* (Grandidier and Grandidier, 1907, p. 65) and de Bucquoy (1771, p. 68)) and was therefore called both names in the accounts. Formerly, a ship called *Fancy* had been in the ownership of the pirates, but had been given to the earlier commander of the *Cassandra*, Captain Macrae, who fled 1720 with the first *Fancy* that was damaged (see e.g. the accounts of Grey (1933, p. 308f).)

[17] De Bucquoy speaks of 121 tons of gold

[18] According to de la Roncière (1940, p. 27), la Buse tried to surrender in 1724, by returning some of the sacred vases from Goa to the gouverneur of Bourbon island, but his request was dismissed. Unfortunately, no source other than de la Roncière was found to prove this event. However, there is another piece of information on this, supported by archive data cited by Guët (1886, p. 219): The amnesty granted by a decision of the Bourbon Supreme Council on 20 January 1724 included la Buse, but only on condition that no further acts of piracy would be committed. La Buse, who was suspicious, preferred not to profit from this and continued his successful piracy business.

[19] "Ils lui rendirent même son épée, quoique d'Or, and sa Croix de l'Ordre de Christ". Please note that a famous oil painting from Pompeo Batoni shows Luís Carlos Inácio Xavier de Menezes with a small golden cross hanging around his neck. The painting is in the possession of the Museu Condes de Castro Guimarães.

where among raw diamonds, neclaces and silver coins as well golden crosses were mentioned.[20]

- Lazinby (1722) described the erection of a large, wooden cross in November 1721 that should attract cargo ships to take him, and the Viceroy of Goa, and more stranded crew members home to Europe.[21]

- There was a legend from the year 1619 about the Flaming Cross of Goa (i.e. the Holy Cross of Boa Vista[22]) that burned on a hill outside of a church in the night and allegedly healed sick people. This miracle was said to appear in the hardest times of the Goa Inquisition, when about 16,000 native Hindus were massacred in the name of Christianity.

5 Winding up Seaman's Yarn

Putting everything together what we know today, we must assume that the well-known and respected historian Charles de la Roncière wrote a breath-taking paperback novel, where he mixed real facts on piracy with his own ideas. After all, as a historian and librarian he had very good access to many books and archival material. The following list is intended to provide an overview of which aspects of the story came from which sources:

- The Portuguese cargo ship *Nossa Senhora* existed, and it had the viceroy of Goa on board during the pirates' raid in 1721.

- The French pirate La Buse existed, and according to the sources, he most probably was taking part in the raid on the *Nossa Senhora*.

- Concerning the treasure that is still object for treasure hunters today, contemporary sources did point out the valuable amount of diamonds and plunder, but as well the taken ships were part of the so-called treasure, and about 200 slaves that the *Nossa Senhora* contained. After making the cargo to cash the amount had been divided among some hundred pirates. When the pirates parted in 1722, from Taylor's crew it was said that they had handed over almost all of their share of the booty to Porto Bello in exchange of amnestie. Following the historical sources, a potential treasure would then only come from the share of la Buse's crew.

- La Buse throwing a cryptogram into the watching crowd shortly before his execution, was not found in any contemporary literature. For the first time, this part of the legend occurred in the novel by de la Roncière in 1934. The assumption is very likely that de la Roncière invented this part of la Buse's legend, and he even admitted to be disappointed by the - rather objective - govenor's version of the execution on Burbon island, and his interest in rhubarb (see the original text in footnote No. 5).

- The cryptogram itself showed up in the novel by de la Roncière the first, and only time. All later publications referred to him. To this day, the cryptogram itself has nowhere been found as a stand-alone manuscript, which leads to the conclusion that de la Roncière made up the cryptogram and its origin. It should also not be ignored that the Parisian publishing house Le Masque, where Roncière published, was well known for light inexpensive fiction, and humorous crime novels (e.g. like the novels by Agatha Christie) - i.e. for literature that is in general not fully credible. De la Roncière was surely fully aware of that. Had he intended to write a scientific and reliable publication, he most probably had chosen another publisher. In his novel from 1934, and in his continuation from 1940, he used a very relaxed and entertaining writing style. As well, hints can be found that he himself did

[20] "In dem Portugiesischen Kriegsschiffe wurde so viel Beute gefunden, daß die alten Seeräuber müde waren, dieses Handwerk weiter fortzusehen. Dieser Schatz machte nebst ihren zuvor geraubten Reichthümern eine ansehnliche Summe aus. Ihr Raub bestund vornehmlich in rohen Diamanten, goldenen Creutzen, und Ketten und gemünzten Silber: Nesseltuch, seidene Zeuge und andere dergleichen Dinge wurden für Lappen und Lumpen gehalten." (So much booty was found in the Portuguese warship that the old pirates were tired of continuing this trade. This treasure, in addition to their previously stolen riches, totalled a considerable sum. Their plunder consisted chiefly of rough diamonds, golden crosses, and chains and coined silver: nettle and silk clothes, and other such things were taken for rags and tatters).

[21] "On the first of November last arrived the Triton French Ship from Mocha. Last from Island Mauritius where had stayed 40 days during which time had taken possession of the said Island; by erecting a large Cross and leaving a French flag flying, the Governour of this place had some time before been in expectation of ships from France for the purpose, but none coming."

[22] See the article by Evelyn Siqueira in the Gomantak Times, published on 22 Feb 2022, https://www.gomantaktimes.com/my-goa/art-culture/the-fascinating-story-of-the-cruz-dos-milagres-in-old-goa

not take the story very seriously (e.g. he explained the possibility of a Forbans' cove, because a Joke cove already existed on Mahé island (see de la Roncière (1934, p. 113) and footnote No. 3), and he referred to Abbott Alexis Mermet to be helpful in finding the real treasure by dowsing (see de la Roncière (1940, p. 27) and footnote No. 7).

- We cannot exclude that someone else brought a cryptogram to de la Roncière, and he then related it to la Buse. But then, the cryptogram would exist as a real artefact - which has not been found or its existence confirmed. If there are other sources for this, the author would be pleased to receive information. In his book about Treasures of the World, Robert Charroux (1962) recited in 1962 de la Roncières story, but concerning the origin of the cryptogram, he introduced a certain Madame Savy, heir of a pirate called Nageon de L'Estang, apparently for the first time. However, there is no source or indication of truth and there also seems to be no official confirmation of this story from the Savy family. So far, it remains unclear where Charroux got this information from. It has to be said that Robert Charroux's publications are very controversial in the scientific community. In many circles, his work is regarded as pseudoscience or pseudohistory, which does not necessarily support the credibility of his sources.

- The Golden Cross of Goa was nowhere to be found, neither in contemporary sources nor in any publication from de la Roncière. Therefore the author assumes that it entered the legend after 1940. However, after de la Roncières publication, the legend, and the size of the buried treasure developed into an exciting-sounding sailor's yarn and moved further and further away from the facts of the contemporary sources. If the famous Cross of Goa, which plays such an important role in nowadays interpretation of the legend of la Buse, is derived from the legend of the inquistition in Goa in 1619, as described above, it seems almost cynical that treasure hunters today are looking for this Flaming or Fiery Cross.

What can be deduced from this study - should this story be included in a scientifically based exhibition? Still, the author is of the opinion that myths and hoaxes in cryptology can certainly be shown as easy accessible exhibits. And generally speaking, at a time when information is more accessible than ever before and is disseminated in abundance, it might be a useful message to visitors to stay critical with all kinds of statements and stories. Therefore, a two-pronged solution was developed for the exhibition: A short audio play provides the legend, discusses the cryptogram, and addresses the topic of the countless treasure hunters. In view of the clear evidence of slave trade, and a probable link to Christian inquisition, the obvious colonial context must also be included in this case. For this reason, an in-depths station next to the audio play is planned to include the recent findings from this study, i.e. to describe the historical sources, and to provide an opportunity to familiarise with contemporary documents and historical research.

Acknowledgments

Firstly, I would like to thank my colleagues Katja Rasch and Luise Allendorf-Höfer for their critical review. I would also like to thank three anonymous reviewers for their valuable input, which enabled me to uncover further important sources and integrate them into the work. I would also like to thank the library of the Deutsches Museum, in particular Florian Preiss, for his help in obtaining historical sources in several complicated matters.

References

Jacques-Henri Bernardin de Saint-Pierre. 1800. *A Voyage to the Isle of France, the Isle of Bourbon, and the Cape of Good Hope; with Observations upon Reflections, Nature and Mankind*. J. Cundee, Ivy Lane, London.

Charles-Mézence Briseul. 2019. *La Buse: De Calais à l'île Bourbon: un destin pirate*. Feuille Songe.

Robert Charroux. 1962. *Trésors du Monde: Enterrés, Emmurés, Engloutis*. Fayard.

Jacobs de Bucquoy. 1771. *Sechzehenjährige Reise nach Indien*. Christian Gottlob Hilscher, Leipzig.

Jean de Kerdéland. 1961. *La Nouvelle Course aux Trésors*. Robert Laffont, Paris.

Charles de la Roncière. 1934. *Le Flibustier mystérieux. Histoire d'un trésor chaché*. Le Masque, Paris.

Charles de la Roncière. 1940. *Explorateurs et Pionniers Français. Lecture et Délassement aux Armées*. Librairie Larousse, Paris.

Clement Downing. 1738. *Die neuesten Unruhen auf der Ost-Indischen Küste von Clement Downing, lebendigen Zeugen der meisten Sachen anjetzo aus dem Holländischen ins Teutsche übersetzt*. Johann Friedrich Rüdiger, Nürnberg.

Pierre-Benoît Dumas. 1730. *Letter de M Dumas, December 1730, Correspondance générale de Bourbon, 1727-1731*. FR-ANOM COL C3/5/002 ff. 120-121, Centre des Archives d'Outre-Mer, Aix en Provence.

Ed T. Fox. 2014. *Pirates in Their Own Words*. Raleigh NC: Lulu.com.

G. Lenôtre. 1934. Le Trésor de la Buse. *Le Temps, May 5th*, page 3.

Alfred Grandidier and Guillaume Grandidier. 1907. *Collection des Ouvrages anciens concernant Madagascar. Tome V: Ouvrages ou Extraits d'ouvrages anglais, hollandais, portugais, espagnols, suédois et russes (1718-1800)*. Comité de Madagascar, Paris.

Charles Grey. 1933. *Pirates of the eastern seas (1618-1723): a lurid page of history*. S. Low, Marston and Co., ltd, London.

Isidore Guët. 1886. *Les Origines de file Bourbon et de la Colonisation Française à Madagascar*. Charles Bayle, Paris.

Alexandre Hamilton. 1727. A New account of the East Indies, being the observations and remarks of Captain Alexander Hamilton, who spent his time there from the year 1688 to 1723.

Charles Johnson. 1726. *Histoire des Pirates Anglois, Depuis leur Etablissement dans l'Ile de la Providence, jusqu'à present*. Etienne Ganeau et Guillaume Cavelier, Paris. Seconde Edition corrigée.

Richard Lazinby. 1722. *Richard Lazinby's account of the proceedings of the pirates who boarded the Cassandra in 1719, Letters 97-99*. IOR/E/1/13 ff. 165-178v, British Library, London.

Le Mercure. 1722. Relation du Voyage de son Excellence M. le Comte d'Ericeira; Grand de Portugal, cy-devant Viceroy and Capitaine General des Indes Orientales pour Sa Majesté Portugaise. *Le Mercure, May*, pages 54–68. Chez Guillaume Cavelier, André Cailleau et Noel Pissot à Paris.

Neil Rennie. 2013. *Treasure Neverland: Real and Imaginary Pirates*. OUP, Oxford.

R.S. Fendrick. 1934. Cryptogram clue to Pirate Hoard. *The Milwaukee Journal, July 15th*, page 4.

An early French digit cipher: deciphering a letter from the King of France to the Duke of Nevers (1592)

Camille Desenclos
University of Picardie Jules-Verne
camille.desenclos@u-picardie.fr

George Lasry
The DECRYPT Project
george.lasry@gmail.com

Abstract

We deciphered a single letter written in 1592 by Henry IV, King of France, to Louis de Gonzague, Duke of Nevers, held in the Bibliothèque Nationale de France (BnF). The ciphertext mostly consists of contiguous digits, and demonstrates an early use of digit ciphers in 16th-century France. In this letter, Henri IV exposes some parts of his current military strategy against the Catholic League. After deciphering the letter, we were able to locate the original cipher table in another BnF manuscript, illustrating how codebreaking may assist historical research both to reconstruct the content of encrypted letters and to identify anonymous cipher tables.

1 Introduction

Louis de Gonzague, Duke of Nevers played an important role in the French early modern history, particularly because of his political and diplomatic activity during the French Wars of Religion (Boltanski, 2006). Moreover, he was one of the main players in cryptographic practice in the last two decades of the 16th century: dozens of manuscripts (now preserved at the BnF) contain encrypted letters addressed to or written by the Duke of Nevers, and one of them (BnF, fr. 3995) even contains 68 cipher tables which were used by him.

While the letters of the Duke of Nevers are well known to historians (Boltanski, 2006; Le Person, 2002; Le Roux, 2000; Wolfe, 1988), and the majority of these letters has been deciphered as soon as they were received by the recipient (see for instance the interlinear decipherment in the letter from Henri IV to the Duke of Nevers (19 April 1591) at BnF, fr. 3615, fol. 52), one letter written by Henri IV to the Duke of Nevers stands out in this corpus. Not only hasn't been this letter deciphered (or at least its original decipherment hasn't been preserved along with the letter) but, above all, the cipher did not match the one used in the other encrypted letters from the French King to the Duke.

It would certainly have been possible to directly compare this letter with the cipher tables that are preserved in the BnF collections. However, the use of digits makes comparisons more complex, or at least more time-consuming than symbols which are easier to spot in cipher tables. Another approach has therefore been chosen: deciphering the letter with cryptanalytic methods (and thus contributing to the DECRYPT project), and then comparing the reconstructed key with the preserved tables. This approach also has the advantage of questioning the use of cipher keys: for instance, are all the characters well represented in the table? As shown in Pierrot et al. (2023), the decryption of letters and, more broadly, the collaboration between cryptographers and historians not only gives access to the content of a letter, but also provides a better understanding of how a cipher works and was used in practice and contributes to the consolidation of the development and application of modern codebreaking techniques. This paper describes each stage of this collaborative effort. In Section 2, we present the letter, its writer and recipient and their correspondence. In Section 3, we describe how we deciphered the letter, and in Section 4, we compare the reconstructed key with the original cipher table which we later found in the BnF. In Section 5, we present the deciphered letter and the analysis of its contents. We summarize our findings in Section 6.

2 A sensitive correspondence at the end of the French Religion Wars

Since August 1589 and Henri IV's accession to the throne, many noblemen rallied the new King, while the Catholic League, with the support of Spain, kept fighting and denying any acceptance of Henri IV as the legitimate King of France. In September 1592, Henri IV whom many cities including Paris still refused to surrender, was thus campaigning against the Catholic League's armies. In 1591, after reconquering Noyon,

among other cities, Henri IV tried from November 1591 to April 1592 to besiege Rouen, without any success, due to the help of Spanish armies led by the Duke of Parma, governor of the Spanish Netherlands. However, the situation started improving for the new King as the royal armies gained some successes both East and South of the Kingdom. The King was then in Noyon and, notably, tried to prevent any Spanish assistance to the League's armies, in Picardy (led by the Duke of Aumale) and in Brittany (led by the Duke of Mercœur).

To contribute to this reconquest of his own Kingdom, Henri IV could rely, among others, on the Duke of Nevers whose catholicity however made him long hesitant between his faith (and rallying the League) and his loyalty to the French Monarchy. Despite some acquaintances with the League in the late 1580's, Nevers didn't join the League, but he didn't rally immediately, at least in an official way, to the new King either (Boltanski, 2006). In May 1590, Nevers finally rallied to Henri IV and acted for him in three different ways: fighting against the Catholic League in his lands (duchy of Nevers) and the lands of his son (government of Champagne), encouraging his clients to rally to the King, and helping the negotiations with some moderate representative of the League.

Many letters from Henry IV to the Duke of Nevers from this specific period have been preserved in the BnF. In particular, the "Manuscrit français 3620" (fr. 3620) especially contains 73 letters from the King to the Duke from May 1590 to February 1593. However, only the letter written on 1592, September 12 (folios 70-71) is encrypted (see Appendix 1). It presents itself as a two-page text (folio 70 recto and verso) and an address (folio 71 verso). According to the handwriting, the letter has been written by one or several clerks (from the Secretary of State for War) and then signed both by the King and by his Secretary of State for War and "Maison du roi", Martin Ruzé de Beaulieu. A later marginal addition at the top of the letter (probably added when receiving the letter or soon after[1]), as well as a mention within the letter, allows us to qualify it as a partial duplicate: the same letter has already been sent to the Duke of Nevers through another way. However, when sending it again, a postscript has been added to it, making our letter both a duplicate and a unique letter. As of today, the initial version of this duplicate hasn't been found. On the contrary, we have been able to locate two copies of the first letter (which doesn't contain the postscript) in BnF, fr. 3615, fol. 89 and in BnF, fr. 4003, fol. 10-11. The first copy is only identified as a "12 September" letter and is, moreover, preserved along letters from 1591. In addition, a copy, not of the whole duplicate, but only of the postscript, has been found in BnF, fr. 4003, fol. 17[2].

The first page of the letter was almost fully encrypted, without any decipherment, whereas the second page was surprisingly in cleartext. If the use of ciphers for the first page makes perfectly sense as detailed above (neither the sickness of the French King nor the military strategy can be broadly known by the Catholic League), ciphers could have been used of protect military information in the second page. Some hypotheses can be formulated to explain the sudden end of ciphertext with a new page. Unlike the information on the first page, the one on the second page could have been already more broadly spread within the French Kingdom. It may also have been written in cleartext to threaten any opponent in case of interception and demonstrate the French King's military preparation. It could finally be a way to preserve information by splitting the writing process: a confident clerk who knows the cipher and copies it by writing the first page; another clerk who writes the second page. The apparent difference between the handwritings (the cleartext within the ciphertext vs the cleartext in page 2) seems to confirm this hypothesis but without excluding the other ones.

Finally, the cipher in use in this letter mostly consists of continuous segments of digits, making cryptanalysis more challenging. There are occasional short segments of cleartext. A few of the digits have a o-looking diacritic on top.

[1] Additional mentions are often written by one of the recipient's clerks on margins or on the back of letters. They identify the sender, the date (the sending and sometimes the receiving date), and the nature of the document (duplicate, copy). Sometimes a short abstract of the letter is written above. Theses mentions are mostly written to help with mail management (reading, classifying).

[2] These copies have been compared to our transcription after decipherment. There is a perfect match.

3 Deciphering the letter

Homophonic ciphers in 16th-century France usually consisted of a set of symbols to represent letters of the alphabet – more than one per letter, as well as a nomenclature with symbols representing common words, persons, places, punctuation signs (Desenclos, 2021). Those ciphers often included nulls, which have no meaning and should be ignored when deciphering an encrypted message. In the 1590s, French diplomatic letters (Monts de Savasse, 2004) and, even more, other letters from Henri IV to the Duke of Nevers[3] are still encrypted in that way. None of these letters, however, used the same cipher as the 1592 letter; it remains thus the only use we have found so far of this cipher table.

Despite the use of digits (Latin and/or Greek characters and/or symbols are the most commonly used at that time in France), we assumed that the 1592 letter from Henri IV to the Duke of Nevers might have been encrypted with a homophonic cipher, with or without a nomenclature. The first challenge was to decompose the contiguous segments of digits into groups of digits, each group representing a letter of the alphabet, a combination of letters, a word, a person, a place, or a null. This process is trivial if it is known, upfront, that all those groups of digits have the same number of digits, e.g., two digits. In such a case, the length of any segment of digits between two segments of cleartext should be even (divisible by two). We determined that this was not the case with this ciphertext.[4]

Next, we assumed that one of the digits (0 to 9) might be a null symbol. After removing the digit presumed to be a null, we would expect the length of the digits segments between the cleartext parts to be even.[5] As this experiment did not yield any result, it became apparent that the groups may have a variable number of digits, maybe some groups having two digits (10, 11, etc.) and some having 3 digits (100, 101, etc.). The task of decomposing segments of continuous digits into groups of digits in such a case is highly challenging, and to date, no method exists to do that automatically.[6]

Lasry et al. (2021) had shown that in several cases of papal ciphers, groups of digits representing nomenclature elements (e.g., words, places, or people) would always start with the same digit and would be longer than the homophones (groups of digits representing letters of the alphabets), e.g., three digits vs. two digits. Accordingly, we then assumed that nomenclature elements in the current cipher were encoded with three digits and always started with the digit 1, while homophones consisted of two digits, and would not start with the digit 1.[7]

This assumption, which later turned out to be only partially correct, was quite useful. Not only could we consistently decompose segments between cleartext parts, but also single lines of digits, or segments between the beginning of a line and a cleartext segment on that line.[8] We then transcribed the resulting groups of digits (742 groups in total), and fed them into an algorithm which recovers the key of a homophonic cipher from ciphertext only (Kopal, 2019), obtaining meaningful fragments of Middle French text. With additional manual processing, we were able to recover the full key, shown in Appendix 2.

Our assumption about groups of three digits which start with 1 representing nomenclature elements (words, places, people) or nulls, while useful, turned out to be mostly wrong. While 121, 124, and 130 indeed are nulls, the other groups with three digits are homophones, e.g., 100, 101, and 102 representing P, or 106, 107, and 08 representing R. The symbols with a “o” on top were found to represent doubled consonants, e.g. LL or SS. Also, we found that the letter E was represented by ten homophones, the letter A by eight homophones, I and V by seven each. Curiously, the writer used only two homophones for O, but we suspected that the original table would have included ten

[3] 110 letters from Henri IV to the Duke of Nevers (1589-1594), including 9 encrypted letters, can also be found in BnF, fr. 3615 and fr. 3626.

[4] Neither was the length of those segments divisible by three, assuming that all groups of digits contain three digits.

[5] Or divisible by three.

[6] For more details on such variable-length homophonic ciphers, see (Lasry et al., 2021).

[7] At this stage, we ignored the digits with a “o” on top.

[8] If the clerk enciphering the letter would not split the digits of a single homophone into two lines. This assumption turned out to be correct.

homophones for each vowel.[9] One group, 51 with a dot on top, was assumed to be a place, which could be interpreted ("Picardye") only after finding the original table.

4 An original digit cipher

Taking advantage of the reconstructed key, we were able to local the original cipher table in the BnF, fr. 3995, fol. 141, shown in Appendix 3 and now referred to as the 1592-cipher. Although this letter could have been deciphered by comparing it with all the cipher tables held at the BnF or by looking for copies of the letter, the current work had two benefits. Firstly, the table on fol. 141 was not identified as a "cipher between the King of France and the Duke of Nevers" but as a "chiffre commun entre messieurs les secrétaires d'Estat et messieurs du conseil. Pour bailler à monsieur de Nevers" [transl.: common cipher between the Secretaries of State and the members of the Council. To be given to the Duke of Nevers]. If this mention doesn't date the cipher, its creation date can be put between May 1590 (death of the previous cardinal de Bourbon who wouldn't have been listed in the nomenclature along with Henri IV's supports) and June 1592 (death of François de Bourbon, Duke of Montpensier). According to Nevers' position (he is one of the members of the Council), this sharing makes perfect sense; sharing a similar cipher between several people certainly weakens the security granted by the encryption but it eases and shortens the writing and encryption process. It questions mostly about the possible use of such a cipher: was it intended to be used for global correspondence (the same letter written and sent to every member of the Council) and/or for specific correspondence (any correspondence with any member of the Council)? This letter would indicate the later use but doesn't explain why the 1592-cipher was used when a specific cipher between the King and the Duke already existed (the 1591-cipher, see above).

In fact, having deciphered a letter encrypted with the 1592-cipher not only demonstrates how the cipher was used in practice but also helps to understand how and why the 1592-cipher has been used for this letter. The reconstructed key is essentially correct, but incomplete, as would have been expected given that it was recovered from a short letter with only 742 groups. As expected, in the original table there are 10 homophones for each vowel (A, E, I/J, O, U/V). There are three homophones per consonant, except for X, Y, and Z which have only two. The original table also lists additional groups for nulls and for doubled consonants.

The 1592-cipher also includes symbols for names of people ("Noms"), cities ("Villes"), and provinces ("Provinces"), but used only once ("Picardye") in the letter, although there are several names and places spelled out in cipher. However, this isn't a misusage of the cipher table but the result of one of the limitations of any cipher table. Cipher tables are indeed conceived at one specific time and had to consider both the time situation (names and places that are the most subject to be frequently used) and usability. This final criterion is even more important in a war context when letters must be encrypted and deciphered quickly and when names of places are as important as names of persons but cannot always be anticipated. It is thus understandable not a have a symbol for Sancy's name (Nicolas de Harlay, sieur de Sancy has been one of the King's representatives towards European protestant princes and helps as well in negotiations within the Kingdom). On the contrary, the absence of any symbol for Charles de Lorraine, Duke of Aumale, governor of Picardy and one of the Catholic League's leaders, is more surprising. As Aumale's name is present in the nomenclature of the 1591-cipher used between the King and the Duke, the use of the 1592-cipher raises some questions.

The nature of our letter as a duplicate could be a reason. To secure the contents as much as possible, using another cipher can level up the difficulty of cryptanalysis and preventing the understanding of the letter as a duplicate of a previous one. As both the King and the Duke already possessed the 1592-cipher, it was easier to use it again without putting at jeopardy a new cipher table by sending it via unsecure roads. In fact, the reason for this change of cipher doesn't seem much to be linked to the military context but to the writing context. All the encrypted letters from the King to the Duke, from 1591 (BnF, fr. 3615, fol. 52) to 1594 (BnF, fr. 3626, fol. 53) have been countersigned by Louis Potier de Gesvres and thus encrypted and written by his clerks. In contrast, the 1592 letter has been

[9] For example, A is represented by 90 to 93 as well as by 96 to 99. We were expecting that A should be represented by the full range of 90 to 99, including the missing 94 and 95.

countersigned by Martin Ruzé de Beaulieu. We may thus assume that Ruzé didn't have access to Potier's cipher table (1591-cipher) and needed then to use the only table he already possessed and could use with the Duke of Nevers, in this case the common table between Secretaries of State and members of the Council.

As for the 1592 letter, we have been able to find back the cipher table (1591-cipher) used for the encrypted letters from the King to the Duke that have been countersigned by Potier and encrypted by his clerks; it can be found in BnF, fr. 3995, fol. 67 and it is shown in Appendix 4. According to a mention on the previous folio, this cipher was used between the King and the Duke of Nevers and had been created in January 1591. Unlike the 1592-cipher but like the vast majority of ciphers that have been created at the same time for King's correspondence (Monts de Savasse, 2004), the 1591-cipher consists mostly of graphical symbols, except for double letters and common words which employ groups of digits. Vowels have five or six homophones (vs. 10 in the digit cipher), and three homophones per consonant. Unlike the 1592-cipher, this one proposes encryption options for common words (here entitled "monosylabes"), for double letters and for the intitulatios. These later encryption options were often used, especially during French Wars of Religion but they were rarely mentioned on cipher tables. In fact, there is the only example in BnF, fr. 3995 (68 cipher tables), and only two provide encryption options for the intitulatios but on the attached guidelines rather than directly on the cipher tables. Finally, the most interesting feature of the 1591-cipher is the use of false plaintext words to represent names and words, such as *les*[10] to represent the city Metz, or *ny*[11] for word *estrangers*[12]. False plaintext words like *tout*[13] or *ce*[14] are also employed as nulls, resulting in a mix between traditional ciphers and jargons. But unlike common jargons, rather than metaphorical words (plants, animals, names from the Bible, etc.), common words are used, that can be easily confused with other parts of the letter, especially if there is only a single word on the middle of ciphertext. However, the global security of the 1591-cipher isn't higher than the 1592-cipher. The size of the nomenclature is quite the same: 59 names for each and only a dozen of common words in supplement in the 1591-cipher. In fact, although the 1591-cipher seems more secure as it relies on more different features, the 1592-cipher provides a higher degree of cryptographic security, as the digits were written continuously without any visual clue as to how to separate them into groups.

Moreover, the 1592-cipher and its use, even if limited, confirm the expansion, in France, of an encryption experiment - digit ciphers - at the turn of the 1590's. Although ciphers still use letters and symbols (Monts de Savasse, 2004), they quickly rationalize themselves by using digits (Desenclos, 2021). In fact, several exclusively digit ciphers have been created and/or used and some of them have been created slightly earlier than we initially thought. At least 23 encrypted letters (using only digits) are sent to the Duke of Nevers in 1589-1591 (BnF, fr. 3422, fr. 3614, fr. 3616, fr. 3623, fr. 3633, fr. 3634, fr. 3646, fr. 3976, fr. 3977) and one of them was written to François de Montholon, keeper of the seals of Henri III.[15] However, in the BnF collections, we have found only two instances of the use of digit ciphers in France in the 1580's/1590's: some of Nevers' correspondence, and another correspondence from Spain.[16] In these years, the Duke of Nevers uses digit ciphers with the French Monarchy (like Montholon or La Vieuville[17] but the latter belongs as well to Nevers' patronage), with Italians such as the Duke of Mantua[18] and Camillo Volta, his agent in Rome[19], or with his wife.[20] Moreover, several cipher tables, using only digits, have been identified in the BnF, fr. 3995, and all belong to the Duke of Nevers. The 1592-cipher is thus, for now, the only documented example of a cipher

[10] Plural of *the*.
[11] Middle French spelling for *ni* (*neither*).
[12] Strangers.
[13] All.
[14] This.

[15] BnF, fr. 3614, fol. 89, 20 April 1589.
[16] BnF, fr. 3977, fol. 130, letter from Bernardino de Mendoza to the duke of Parma, 17 April 1589.
[17] BnF, fr. 3977, fol. 191, letter from La Vieuville to the duke of Nevers, 13 August 1589.
[18] BnF, fr. 3646, fol. 93, 10 November 1589.
[19] The French manuscript 4689 held for example 16 letters from Camillo Volta to the Duke of Nevers from 1585 to 1589. They are written in Italian and use only digits in their encryption.
[20] BnF, fr. 3634, fol. 63. The letter may be earlier than 1589 but we can only estimate its date between 1587 and 1592. Two cipher tables (digits only) for the correspondence between the duke and its wife can be found in BnF, fr. 3995, fol. 2 and 10.

which has certainly been used by the Duke of Nevers but has been created neither by him nor for its sole use as it was intended for the members of the King's Council. Moreover, three letters written by Armand de Gontaut, Duke of Biron, to Henri IV in 28 September and 7 October 1591 (BnF, fr. 3645, fol. 15 and 33) and by de Jean de Chaumont, sieur de Guitry, Jean de Gontaut-Biron, baron de Salagnac and Jean de Vivonne, marquis de Pisani to Henri IV in 9 November 1591 (BnF, fr. 3645, fol. 92) confirm the shared use of this 1592-cipher. Not only is this cipher used beyond the sole entourage of the Duke of Nevers but it is one of the rare examples of a shared use between several people of the same cipher.

However, we don't know yet if the use of digit ciphers had been introduced by the Duke of Nevers, the Monarchy, or by the Catholic League with whom the Duke of Nevers held an encrypted correspondence in the late 1580's.[21] Another cipher table, probably produced in 1589-1590 and using only digits except for 3 symbols (BnF, fr. 3413, fol. 109), weakens this last hypothesis as names of both Protestants and Catholics are mixed in the nomenclator. The presence of the names of Italian princes (Mantua, Parma, etc.), still in the nomenclator, suggests a link with the Duke of Nevers but this table equally fulfills the needs of the French monarchy and rationalizing the cipher by using digits might be a suggestion from François Viète, Henri IV's main cryptographer whose work made him aware of the advantages of digit ciphers. However, without excluding the two other hypotheses, Nevers' letters, as well as a 1562 letter from the Duchess of Mantua to the Duke of Nevers (her son)[22] provides evidence at least of his significant role in the broader diffusion of digit ciphers in France and a strong Italian influence[23]. Nevertheless, relying on the analysis of Nevers' ciphers and his use of digit cipher requires some caution: Nevers's letters and cipher tables form the main part of the letters and cipher tables that we have found and identified yet for the 1580's and 1590's. As the research project is still ongoing, the Duke of Nevers cannot be considered as the sole or main user and creator of these digit ciphers even though he is one of the main actors of the French cryptographic practices.

At this point, we can only affirm that using digit ciphers remains until the mid-1590's, without overshadowing the common use of Latin/Greek letters and symbols, or combining them with digits (e.g., mostly digits but some Latin letters as in BnF, fr. 3625). From this period, we can only find some very intermittent uses of digits until a new experiment in the 1610's and their systematic use in the mid-17th century.

5 The deciphered letter

As often while deciphering early modern letters, the encrypted parts don't hide some state secrets and this letter is no exception. Moreover, this one doesn't differ from the regular letters that the King was sending, especially at war, and that consisted mainly of status reports from the King. As a member of the Council, Nevers needed indeed to be kept informed, so that he could still fight against the League's armies and reinsure some noblemen about the King's intention.

The letter starts thus with some information about the King's health. If this is quite common, it is here particularly significant as a fever has delayed the King's departure towards the North of his Kingdom. The main part of the encrypted text, however, reports about the military strategy of both the King and the Duke of Nevers, according to the information the latter provided to the King and conversely. As extra protection, in addition to the encryption, no specific information (name, location) is given except for the location of the rebel armies. As they are the ones that would like to intercept the letter, it won't harm the King's interests if this information was discovered.

The second part of the letter (in cleartext) keeps going with the status report, now elaborating on the League's military situation and making clear the King's effort to oppose it. In fact, the information in cleartext does not portray the League in an advantageous light (which explains the absence of encryption): their armies in Picardy would be diminished, the King's officer prepared to face them. Moreover, the Spanish army wouldn't be able to help them as Coeverden has been taken back by the Dutch armies as part of the Eighty Years War. The

[21] See for example BnF, fr. 3976.

[22] BnF, fr. 4687, fol. 5, 10 January 1562.

[23] If the recent publications and current research haven't revealed an important use of digit cipher in France, their use in Italy, moreover much earlier, has been demonstrated. See Meister (1906) and Lasry and al. (2021).

letter, whose full transcription is given above, ends with two pieces of information: the King's recovery and the late arrival of the Duke of Bouillon to the King's camp. As the King isn't sick anymore and Bouillon would have arrived before the letter could be intercepted, this isn't confidential and is thus mentioned in clear: none of this could be used by the Catholic League against the King or the Duke.

Transcription[24]

[Mention marginale en haut de la lettre]
Dupplicata horsmis ce qui est adjousté à la fin

Mon cousin,

Je pensois monter à cheval ce matin *pour* aller à Chauny[25] et retourner dès ce soir en ceste ville pour en partir demain et me rendre lundy à Crecy[26] mais *une* grosse fievre m'a pris et me tient il y a quatorze heures sans apparence de diminution ni que les medecins sache[n]t dire quelle elle sera à ce soir ou demain matin *je vous en mandray des nouvelles.*

Cependant je trouve tres bon vostre advés[27] de passer mon armee où vous m'avez escrit pour les raisons que vous me mandez *et* pour oster tout umbrage à celuy duquel vous m'avez envoyé le double *de* la lettre qu'il m'escrit au subjet de laquelle je vous asseure n'avoir jamais pensé *comme* je le luy fais entendre par ma response et mande encores au sieur de Sancy *qui est* de ses amis de le luy dire de ma part. Vous pouvez executer vostre desseing sans peril *car* le duc d'Aumalle s'est retiré avec toutes les forces d[e] Picardye en telle diligence qu'il n'a osé sejourner en aucun lieu que une heure à La Fere; encores a il laissé de ses plumes à la garnison de Chauny.

Le reste de leur armée est fort diminué et s'est logé le long de la riviere d'Esne[28] *tirant de Soissons à Retheil. Toutesfois de peur qu'elle n'entrepreigne quelque chose ou qu'elle donne allarme à ceux de mes villes de Challons, d'Esparnay*[29] *et autres lieux de mon pais de Champagne, nous voyans esloignez, j'escris aux sieurs de Thommasin*[30]*, de Vignolles*[31]*, à ma court de parlement*[32] *et de madite ville de Challons qu'ilz soyent diligens à se garder de surprise et qu'ilz s'asseurent de mon brief retour comme je le leur ay promis. Les nouvelles du Pais Bas continuent la prise de Couverden*[33] *et la deffaicte des trois regimens de lansquenetz et de la cavallerie qui estoit envoyée pour le secourir s'il est ainsy le duc de Parme*[34] *ne peult entreprendre de venir en mon royaume de plus de trois moys ; j'ay adverty mon cousin le cardinal de Bourbon*[35]*, les sieurs de mon conseil et do*[36] *de se trouver aujourd'huy ou demain à Senlis afin de les y prendre et de les mener avec moy à Meleun*[37] *où il me tarde que je ne sois desja arrivé.*
Priant, sur ce, Nostre Seigneur, qu'il vous ayt, mon cousin en sa sainte et digne garde. Escrit à Noyon le XII[e] *jour de septembre 1592.*

[Signé] *Henry*

Mon cousin, depuis la presente escritte et fermée qui est le dupplicata de celle que je vous ay envoyée par une aultre voye et que j'ay gardée jusques à ce jourd'huy pour la vous faire tenir par ceste commodité, ma fievre, graces à Dieu, m'a quicté du tout et espere qu'elle ne me reprendra plus de façon que je partiray demain sans faillir pour me rendre en mon armée suivant ce que je vous ay mandé. Encores fusse je party dès aujourd'huy si mon cousin le duc de Bouillon[38] *fust arrivé hier comme je l'attendois pour donner ordre à noz garnison mais s'estant voulu purger il ne sera icy que tantost.*

[Signé] *Henry*
[Signé] *Ruzé*

[24] The parts in *italics* are in clear and the text have been edited according to the rules in Barbiche 1980.
[25] Chauny, Aisne, France. It's 20 km away from Noyon where the King is staying.
[26] Crécy-sur-Serre, Aisne, France. It is 50km away from Noyon and mostly close to La Fère (which is mentioned later in the letter).
[27] To be understood as "advis".
[28] To be understood as "Aisne".
[29] Epernay, Marne, France.
[30] Philippe de Thomassin, governor of Châlons.
[31] Bertrand de Vignoles, governor of Épernay.
[32] Since 1589, the magistrates who have stayed faithful to the King (and not to the Catholic League) have settled in Tours for one part and in Châlons for this other part.
[33] Coeverden, Drenthe, Netherlands. It doesn't concern the French Wars of Religion but the Eighty Year's War that opposed Spain to the rebel Norther Provinces of the Netherlands. In early September 1592, Coeverden has been taken back by the rebel armies.
[34] Alexander Farnese, Duke of Parma and governor of the Spanish Netherlands.
[35] Charles II of Bourbon, cardinal de Bourbon since 1590.
[36] It is either a writing mistake or a symbol that hasn't been written in the nomenclature.
[37] Melun, Seine-et-Marne, France.
[38] Henri de La Tour d'Auvergne, Duke of Bouillon.

Translation

[Marginal note above the letter] *Duplicate except for what is added at the end.*

My cousin,

I thought of riding this morning to Chauny and going back from this city already from this evening, and leaving Crecy on Monday, but I have caught a high fever which started 14 hours ago, without any sign of abating, and the physicians not knowing whether it will continue until this evening or tomorrow morning. I will update you on this.

Nevertheless, I found highly appropriate your advice to move my army to the place you wrote to me, for the reasons you mentioned and to avoid any suspicion from the person from which you have sent me a duplicate of a letter, about the topic which I assure you of never thinking of, as I asserted in my response, and having asked Sieur of Sancy, who is one of his friends, to convey to him on my behalf. You will be able to carry your designs without any risk as the Duke of Aumalle has retired with all his forces from Picardy, so diligently that he has not dared to stay in any place rather that one hour in La Fere and he has lost some men while facing the garrison of Chauny.

The remainder of their army is significantly diminished and has taken position along the Aisne rive, from Soissons to Rethel. However, out of concern that it [this army] main underact some action or that it may cause my people in the cities of Chalons, Epernay, and other places in my Champagne country to be alarmed, seeing us afar, I am writing to the sieurs de Thommasin, de Vignolles, to the court of my parlement, and of the said city of Challons that they must be diligently avoid being surprised, and that the should be confident of my return soon, as I have promised to them. The news from the Low Countries follows the capture of Coeverden and the defeat of the three regiments of Lansquenets and of the cavalry sent to save it, so that the Duke of Parma will not be able to enter my realm the next three months. I have instructed my cousin the Cardinal de Bourbon, the members of my council, and 'do' to be today or tomorrow in Senlis to take them and bring them to me to Melun where I soon arrive soon.

With this, praying Our Lord to keep you, my cousin, in his holy and worthy guard. Written at Noyon, the twelfth day of September 1592.

[Signed] *Henry*

My cousin, since writing and sealing the current letter which is a duplicate of the letter I had sent you via another channel and that I had kept until today in order to hand it over to you with this commodity, my fever, thanks God, has gone, and I hope it will not return so that I can without fault leave tomorrow to join my army as I had informed you. Also, I would have departed today if my cousin the Duke of Bouillon had arrived yesterday as I was expecting to assume the command of my garrison, but having wanted to purge himself, this will not happen soon.

[Signed] *Henry*

[Signed] *Ruzé*

6 Conclusion

With a systematic survey of encrypted documents as well as cipher tables in archives and libraries, it is sometimes possible to match an encrypted letter with its original cipher table, even if cipher tables were not systematically preserved by their users nor passed to archives and libraries. If deciphering can be performed without the original cipher table, it helps accurately decipher the document and sometimes better identify its origins.[39] For this letter, it allowed us to understand the reason for this uncommon use of a digit cipher in the correspondence between the King and the Duke as well as the dissemination of digit ciphers in France at the end of 16th century and to document an effective shared use of a same cipher between, at least, 6 persons.

In terms of cryptanalysis, digit ciphers, especially when documents consist of contiguous segments of digits, present more challenges, as opposed to homophonic ciphers employing graphical symbols, which can be compared more easily. In this work, finding the original cipher table was much easier after deciphering the letter using cryptanalytic means.

Finally, the main challenge in cryptanalysis was the fact that the homophones have variable lengths, and a breakthrough was possible only

[39] Especially if the encrypted document does not mention the sender or recipient in cleartext, and those can be identified only after the document is deciphered.

via lucky guesses and trial-and-errors. A generic algorithm to decompose contiguous sequences of digits would be helpful in solving other similar challenging cases.[40]

As for the historical part, while it hasn't led to a breakthrough in the understanding of the French Wars of Religion, it has enabled us to lay the foundations for a better understanding of some cryptographic issues in times of civil war.

Funding

The work of one of the authors has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Bernard Barbiche, 1980. *Conseils pour l'édition des documents français de l'époque moderne* [http://theleme.enc.sorbonne.fr/cours/edition_epoque_moderne/edition_des_textes]

Arianne Boltanski, 2006. *Les ducs de Nevers et l'État royal : genèse d'un compromis (ca 1550-ca 1600).* Genève, Droz.

Camille Desenclos, 2021. Écrire le secret quoditien. Pratiques de la cryptographie au sein de la diplomatie française (XVI^{e} siècle – premier XVII^{e} siècle), *in* Guido Braun and Susanne Lachenicht (ed.), *Spies, espionnage and secret diplomacy in the early modern period.* Stuttgart, Kohlhammer, 85-103.

Nils Kopal, 2019. Cryptanalysis of homophonic substitution ciphers using simulated annealing with fixed temperature, *in Proceedings of the 2nd International Conference on Historical Cryptology,* 107-116.

George Lasry, Béata Megyesi and Nils Kopal, 2021. Deciphering papal ciphers from the 16th to the 18th Century, *Cryptologia*, 45:6, 479-540.

Aloys Meister. 1906. *Die Geheimschrift im Dienste der Päpstlichen Kurie von Ihren Anfängen bis zum Ende des XVI. Jahrhunderts*, volume 11. F. Schöningh, Paderborn, 1906.

Xavier Le Person, 2002. *« Practiques » et « practiqueurs » : la vie politique à la fin du règne de Henri III (1584-1589).* Genève, Droz.

Cécile Pierrot, Camille Desenclos, Pierrick Gaudry and Paul Zimmermann, 2023. Deciphering Charles Quint (A diplomatic letter from 1547), *in Proceedings of the 6th International Conference on Historical Cryptology,* 148-1459.

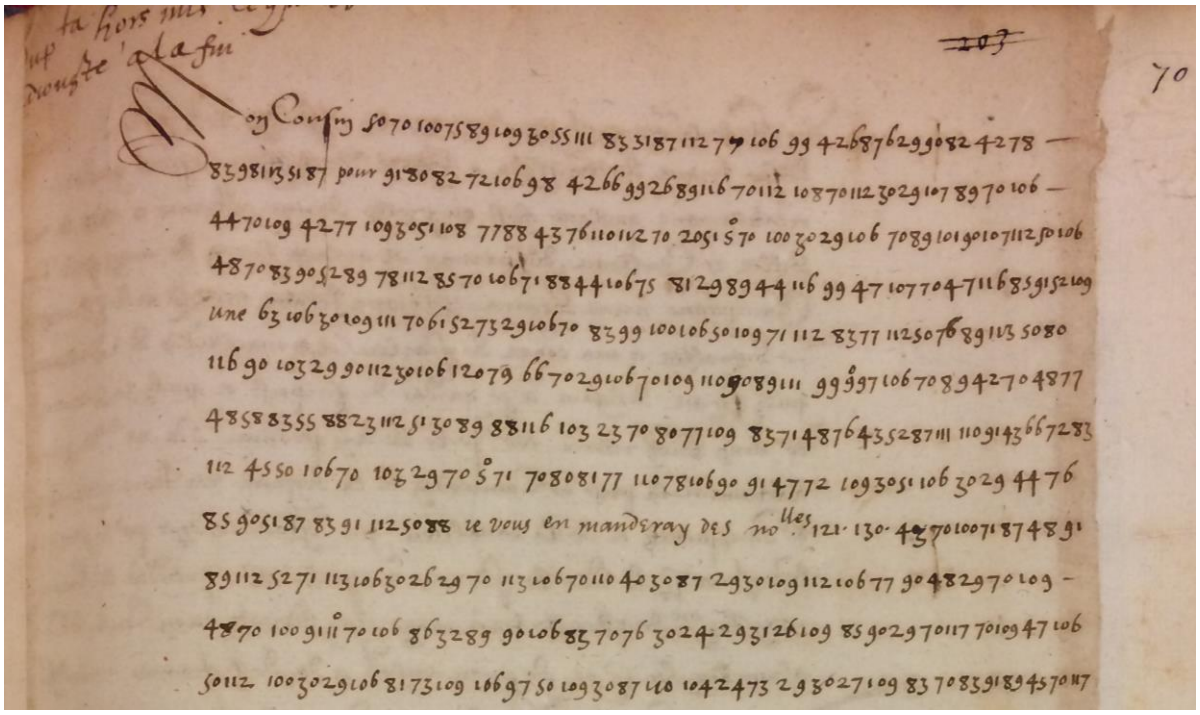
Nicolas Le Roux, 2000. *La faveur du roi : mignons et courtisans au temps des derniers Valois (vers 1547-vers 1589).* Seyssel, Éditions Champ-Vallon.

Jacques de Monts de Savasse (ed.), 2004. *L'Europe d'Henri IV : la correspondance diplomatique du secrétaire d'État Louis de Revol, 1588-1593.* Grenoble, Presses universitaires de Grenoble.

Michael Wolfe, 1988. "Piety and political allegiance: the duc de Nevers and the protestant Henri IV, 1589-93", *French history*, 2:1, 1-21.

[40] For additional examples, see Lasry & al, 2021.

Appendix 1 – Letter from Henri IV to the Duke of Nevers, 12 September 1592 (beginning of the first page)



Source: BnF, fr. 3620, fol. 70

Appendix 2 – The reconstructed key of the 1592 cipher

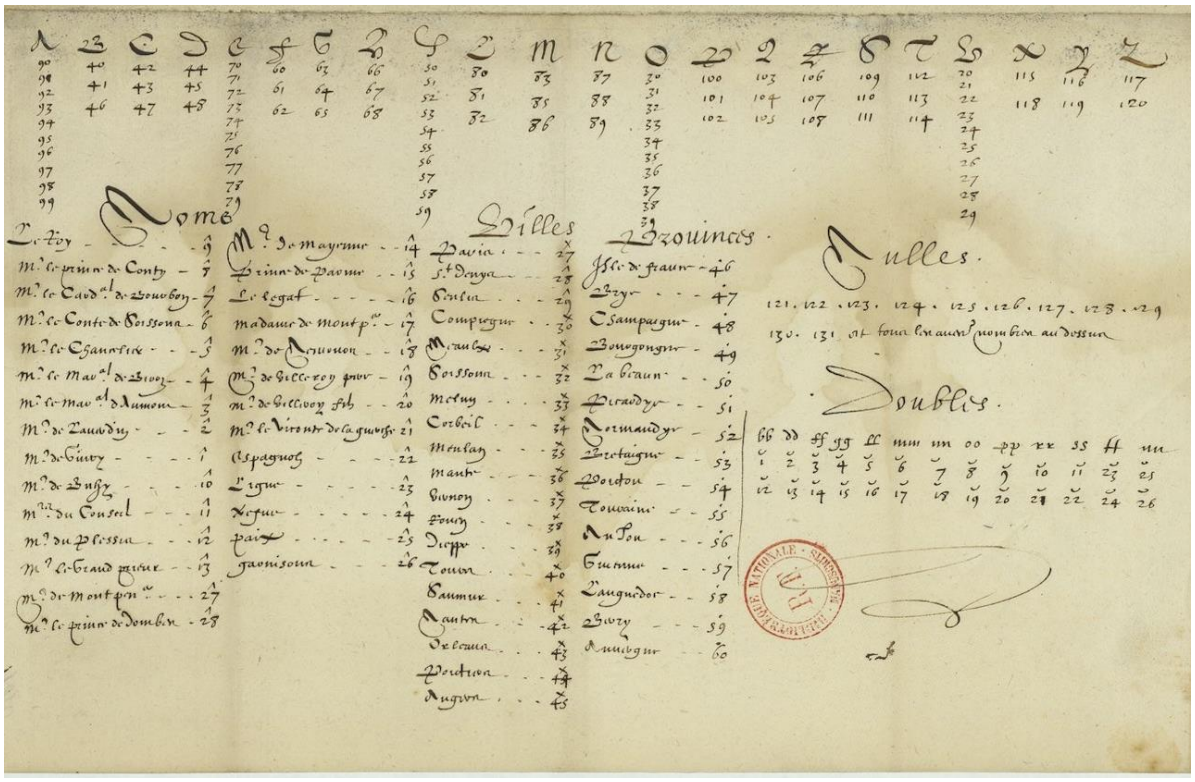
A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T	V	X	Y	Z
90	40	42	44	70	60	63	66	50	80	83	87	30	100	103	106	109	112	20	115	116	117
91		43	45	71	61	65	68	51	81	85	88	31	101	104	107	110	113	23		119	120
92		47	48	72	62			52	82	86	89		102		108	111		24			
93				73				53										26			
96				74				55										27			
97				75				58										28			
98				76														29			
99				77																	
				78																	
				79																	

LL	PP	SS	TT
5	9	11	23

Space or null		
121	124	130

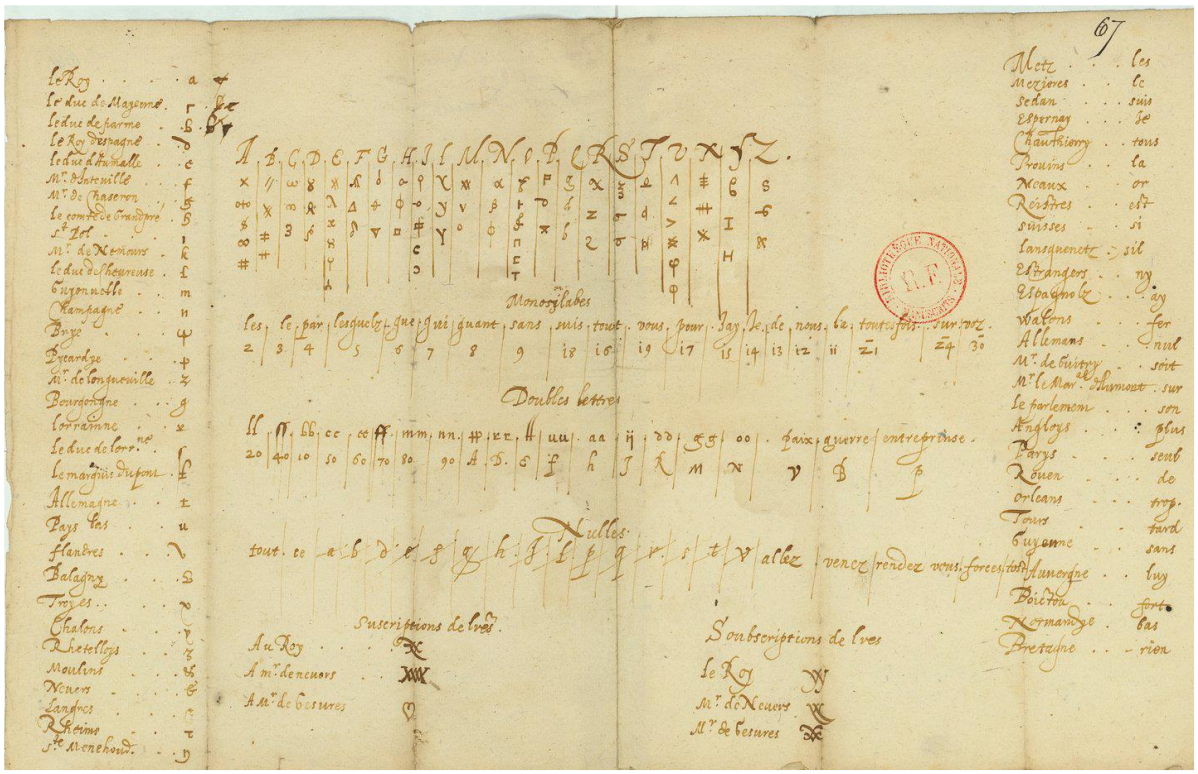
Source: The authors

Appendix 3 – The 1592 cipher - “Chiffre commun entre messieurs les secrétaires d’Estat et messieurs du conseil. Pour bailler à monsieur de Nevers” (ca. 1590-1592)



Source: BnF, fr. 3995, fol. 141 (©Gallica)

Appendix 4 – The 1591 cipher used between Henri IV and the Duke of Nevers, January 1591



Source: BnF, fr. 3995, fol. 67 (©Gallica)

Send Someone to Finish Fredenburgh's Works. A Dutch Ciphertext (1689) from Suriname

Jörgen Dinnissen
Historian,
The Netherlands
dinnissen.jorgen@gmail.com

Nils Kopal
University of Siegen,
Germany
nils.kopal@uni-siegen.de

Abstract

A ciphertext without its corresponding key was found in the archives from the Dutch colony Suriname, in the National Archives at The Hague. We were able to decrypt it through cryptanalysis and with the use of CrypTool 2. The revealed plaintext contains a letter with military sensitive information and the name of Fredenburgh, who served as Governor ad interim from 1688 to 1689. It was sent in May 1689 by Governor Van Scharphuijsen to his directors in Amsterdam. Since 1689, the Society of Suriname (SvS) used ciphers for its militarily sensitive information. Ciphertext U3 was sent during the Nine Years' War (1688-1697) when the Dutch were at war with the French. The letter was encrypted as a precaution against possible interception by the (French) enemy.

1 Introduction

In the 17th century the Dutch divided planet Earth beyond Gibraltar into two chartered joint-stock companies. The Dutch East India Company (VOC) was born in 1602 and was empowered to trade, settle, conquer, administer, and defend their monopoly in the East, this is Asia. The Dutch West India Company (WIC) was born in 1621 and had their monopoly in the West, this is the Atlantic. The colony Suriname was subject to the charter-territory of the WIC.

After losing Dutch Brazil in 1654, the Dutch Republic's States General shouldered almost all the Atlantic military operation costs, as the WIC was financially depleted. The leadership in the Dutch Republic had little choice if they did not want to destroy the Dutch economic position in the Atlantic completely. The weakness of the WIC and the increasing competition of England and France in the Atlantic forced her to make this choice.[1]

From 1630 to 1667 Suriname was an English colony. In 1667, during the Second Anglo-Dutch War (1665-1667), Dutch from the States of Zeeland captured this colony.[2] The States of Zeeland considered Suriname as their property although neither the States General of the Dutch Republic nor the WIC had issued a patent for this. The States of Zeeland sold Suriname in 1682 to the WIC. The WIC, not wishing to undertake colonization alone, collaborated in 1683 to establish a Dutch private company, the Sociëteit van Suriname (English: Society of Suriname). This company had three participants, each holding equal shares in the society's responsibilities and profits: (1) the WIC, (2) the city of Amsterdam, and (3) the family Van Aerssen van Sommelsdijck.

The main crops for the Dutch in Suriname were sugar, cotton, and indigo. In the 18th century the single most important export product became coffee. The economy of this "transplantation

[1] Knaap (2015) and Schwartz (2014).

[2] 'Zeeland' was one of the five chambers of the WIC. It is also a province, situated in the southwest of the Netherlands, near the border with Belgium, consisting of several islands. Hence its name, meaning 'zee-land' (English: sea-land). Its capital is Middelburg. Notice, the conquest of the colony Suriname in 1667 was paid by the government of Zeeland (States) and not by the company (WIC chamber of Zeeland).

colony"[3] depended on enslaved people working on its plantations.[4]

The ciphertext U3 (1689a) is a one-page manuscript found in the Sociëteit van Suriname (SvS) archive at the Dutch National Archives, in The Hague, the Netherlands. See *Appendix 1*. In 2022, Mark Ponte noticed the ciphertext U3 and passed it on to the first author of this paper for further cryptanalysis due to the absence of a corresponding key. The name 'U3' is based on the first two codes of the ciphertext (see Figure 1). The ciphertext probably dates from 1689 because it was found physically after an archive document from April 21st 1689, in the folder with the inscription 'Number 8. Letters and papers from Suriname to the Society starting with October 22, 1688 and ending on [illegible month] 28, 1689'.[5]

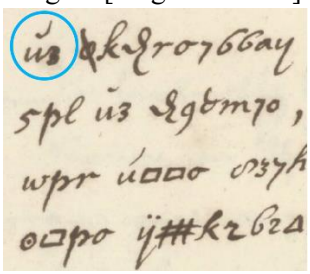


Figure 1: The first two codes of U3 (1689a), in the blue circle, gave the ciphertext its name.

This paper is organized as follows. In *Section 2*, we show the various cryptanalysis attempts that ultimately led to a breakthrough. In *Section 3*, we show the revealed plaintext and the corresponding reconstructed key. We also analyze the key and compare it with two other keys from 1739. Also, we discuss the quality of the encryption, performed by the encryptor. *Section 4* examines the automatic cryptanalysis component in CrypTool 2 (CT2). We answer questions, which are of interest of future "solvers" using CT2: Why didn't the cryptanalysis work properly in the first attempt? What tips and tricks can we take into account next time when we use it for other cryptanalysis tasks? *Section 5* discusses the use of ciphers in Suriname and describes to what extent the content of the revealed ciphertext confirms that it indeed dates from 1689. Finally, *Section 6* concludes this paper.

2 Cryptanalysis

In this section we show the various attempts that ultimately led to the breakthrough and with which attempt we were finally able to decipher the ciphertext.

2.1 First attempt: automatic cryptanalysis in CrypTool 2

Because there was no corresponding key, we used the automatic cryptanalysis component "Homophonic Substitution Analyzer" in CT2 to automatically cryptanalyze the ciphertext. Unfortunately, the first attempt yielded no readable words (see Figure 2).

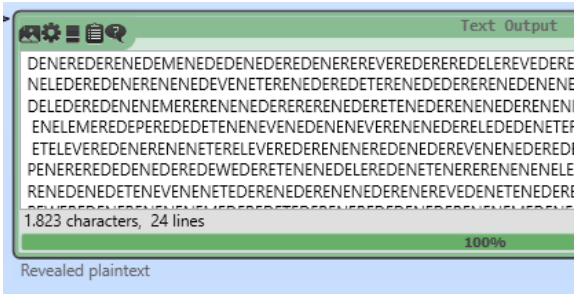


Figure 2: First attempt automatic cryptanalysis in CT2.

2.2 Second attempt: using the old and new key-1739 in CrypTool 2

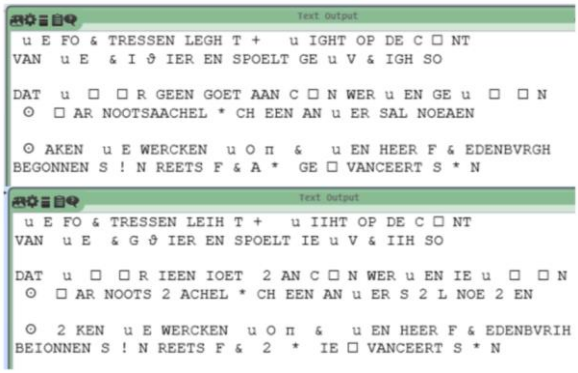


Figure 3: Second attempt with using the old (top) and the new key-1739 (bottom) in CT2 for decryption.

[3] A "transplantation colony" is a territory established by moving a population from one region to another, often by a colonizing power, to exploit new lands and spread their influence.

[4] Fatah-Black (2019).

[5] Original in Dutch: 'No 8. Brieven & Papieren van Suriname aan de Societijt beginnende met den 22 octob[er] 1688 ende eijndigende met den 28 [onleesbare maand] 1689'. The most recent date on a document in this folder is from August 1689. So the month must be August or later.

Karl de Leeuw (2000) writes that the WIC had one cipher, introduced in 1739, that was used over and again during the Fourth Anglo-Dutch War (1780-1784). It was specially designed for the encryption of military messages. In De Leeuw (1997) the new key-1739 is presented. If we look at the original source in the archives of the SvS we see that it is a complete and original (not reconstructed) key (New Key, 1739). On the previous page of this source we find the "old key", this is old key-1739 (Old Key, 1739). We used both keys for our second attempt in CT2. The substitution component and a handmade plain- and ciphertext alphabet, yielded better results than the automatic attempt. We were able to read Dutch words when we used the old and new key-1739: E.g. ‘spoelt’, ‘wercken’, ‘heer’, and ‘reeds’ (see Figure 3).

We observed that the old and new key-1739 do not decipher the ciphertext U3 (1689a) completely but the five Dutch words found above are basically enough to put the remaining puzzle together for the missing parts of the unknown key by hand. But before we started solving this puzzle we looked more closely at the pages before and after the ciphertext in the physical archive.

2.3 Third attempt: pin to ‘fortressen’ in CrypTool 2: our breakthrough with automatic cryptanalysis!

We noticed that the ciphertext page (see *Appendix 1*) and the page after it (see *Appendix 2*) contain about the same amount of text. On both pages the first word has two characters, the second word ten with double ‘6’ and double ‘s’ on position 7 and 8 (see Figure 4). At first, we thought that those two pages were not related because the size and the color of the paper and handwriting differed. But what happens if we pin the second word in the automatic cryptanalysis mode in CT2 to ‘fortressen’ (English: fortresses)?

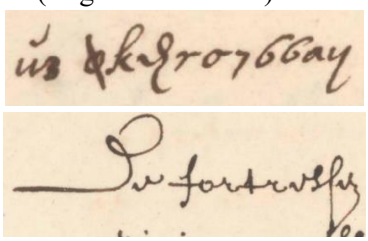


Figure 4: First two deciphered words of ciphertext U3 (1689a) and plaintext U3 (1689b).

That was the breakthrough! Immediately the first readable words appeared automatically in Dutch (see Figure 5). The first five words of the revealed plaintext are correct 17th century Dutch: *de fortressen legh te dicht* (English: the forts are too close). This means that the plaintext U3 (1689b) is the revealed plaintext of ciphertext U3 (1689a) and that we were able to relate them. So we were also able to decipher by hand (we never had to follow this path to the end) and automatic means.

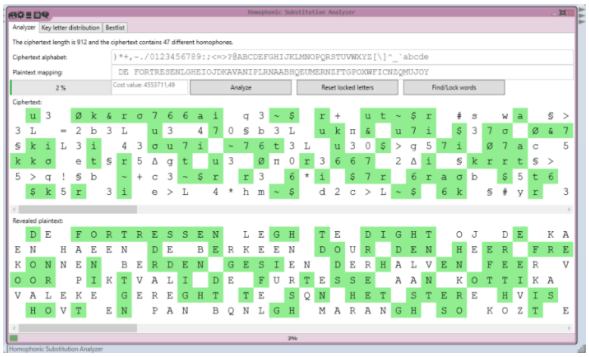


Figure 5: Breakthrough in CT2 for U3 (1689a): Pin the second word to ‘fortressen’.

3 Revealed plaintext and technical analysis

In this section we show the revealed plaintext and the reconstructed key of U3. After that, we analyze the key technically and compare it to two other keys from 1739. Finally, we discuss the quality of the encryption done by the encryptor.

3.1 Reconstructed key and revealed plaintext

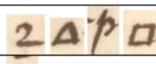
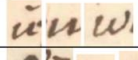
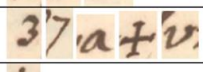
Latin letter	U3 (1689) reconstructed
A	
B	
C	
D	
E	
F	

Table 1: A part of the reconstructed key of U3 (1689a).

With some effort the complete key – this is the key that produces the most readable plaintext with the fewest errors – could be reconstructed (see Table

1). Also, see *Appendix 3* for the complete reconstructed key.

The revealed plaintext is in Dutch and reads:

DE FORTRESSEN LEGH TE DIGHT OP DE CANT VAN DE RIVIER, EN SPOELT GEDURIGH WEGH SOO DAT DAAR GEEN GOET AAN CAN WERDEN GEDAAN, MAAR NOOTSAACKELYCKE EEN ANDER SAL MOETEN MAKEN, DE WERCKEN DOOR DEN HEER FREDENBURGH BEGONNEN, SYN REETS VREY GEAVANCEERT, SYN GEDETACHEERT EN VAN AGHTEREN OPEN, GELYCK BY DE KAART SAL KONNEN WERDEN GESIEN, DERHALVEN VEEL VOLK VAN NODEN HET SELVE TE DEFENDEREN, ONS MAGASYN HEEFT MAAR VOOR DRIE WEECKEN KOST MEER, DAAROM SYNDE DE BERCK NA DE BARBADOS VOOR VICTUALY. DE FORTRESSE AEN COTTICA IS NOGH WYNIGH AEN GEVORDERT, EN DERVE DAAR MEEDE NIET VOORT VAAREN EER DEN INGENIEUR COMT ALSO EEN YEGELYCK OORDEELT DAT SEER QUALYCK GELEGT TE SYN, HET STERCK HUYS IN PARA VAN SLEGTE STOFFE OPGEMAACKT IS SEER VERVALLEN, SO DAT GROTE REPARATIEN DAAR AAN SAL MOETEN WERDEN GEDAAN, DAT IN SARANMICA VAN HOUT EN VAN WYNIGH BALANGH, SOO KOMT EEN ENGELS SCHIP MET PAARDEN EN WYNIGH VICTUALIE SO DAT DE BARCK NIET NA BARBADOS GAAT

The complete plaintext translated into English:

The forts are too close to the side of the river, (and the water) continually washes away so that it is not possible to moor there properly. But necessarily someone else will have to finish the works that sir Fredenburgh started. This (work) is already quite advanced and is detached and open at the rear. As can be seen on the map. Therefore many people are needed to defend themselves. Our warehouse only has food for three weeks. That's why a ship is to Barbados for provisions. The fortress on Cottica is still little advanced. And therefore, partly for this reason, do not sail any further before the engineer arrives. The strong house in Para is actually very poorly laid, made of bad material. It is very dilapidated and needs major repairs. The (fort) in Saranmica is wooden and of little importance. An English ship comes with horses and little provisions, so the ship does not go to Barbados.

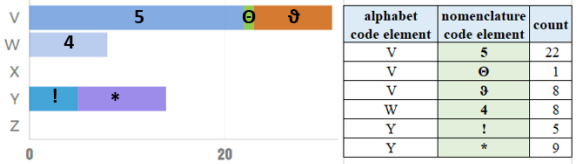
See *Appendix 5* for the complete transcription, description, and translation into English of U3 (1689a and 1689b).

3.2 Technical analysis of the three keys

In ciphertext U3 (1689a) 45 unique codes were used: twenty-one Latin letters, eight numbers (from 2 to 9), seven Greek letters, one astrological symbol (sun), and eight other symbols. Notice, that there are also four codes for punctuation. When we look at the used alphabet code elements of cipher U3 and cipher old-1739 we see that they are very similar. Cipher new-1739, on the other hand, deviates more from cipher U3 because it uses nine words in nomenclature elements. See Table 2.

Sorting nomenclature code elements		U3-1689 Code	Old-1739 Code	New-1739 Code
syllable	astrological	1	1	1
	Greek	7	9	9
	Latin	21	22	22
	number	8	9	8
	symbol	8	4	5
word	symbol			6
	Greek			3
Sub Total		**45**	**45**	**54**
punctuation		4		
Grand Total		**49**	**45**	**54**

Table 2: Three keys, sorting of their cipher codes.



alphabet code element	nomenclature code element	count
V	5	22
V	Θ	1
V	ϑ	8
W	4	8
Y	!	5
Y	*	9

Figure 6: U3 (1689) alphabet code elements over nomenclature code elements (count) in color.

In the revealed plaintext, only Latin letters were used. Twenty-two of the twenty-six Latin letters were used. The letters J, U, X, and Z were not used, which is typical for ciphers and texts of that time. For example, letters U and V share the same ciphertext symbol. Each alphabet code element has 1 to 5 code elements (homophones). See Figure 6 for more details. There are no nomenclature elements with words (only with syllables). Thus, we can conclude that the cipher for U3 is a **homophonic substitution cipher**. The same is true for cipher old-1739. Cipher new-1739 on the other hand, is a homophonic substitution cipher but with nine nomenclature elements with

words. Figure 6 shows the alphabet code element frequencies of V, W, and Y. See *Appendix 4* for the complete list.

We define an assignment as a direct mapping between a plaintext element (for example the letter E) and a code element (for example the 7). If two keys encrypt the E the same way, we say they share an assignment.

If we compare the plaintext elements, code elements and their assignments we see that cipher U3 shares 71% of its assignments with cipher old-1739. For example, code element Δ is in both keys assigned to plaintext element A. The shared assignment between U3 and new-1739 is 54%. See Table 3 and 4. Also see *Appendix 6* for the details of the comparison between the three keys.

Indicator	Key U3	Key Old-1739	Key New-1739
Plaintext elements. Count	22	24	33
Code elements. Count	45	45	54

Table 3: Three keys, counting of their cipher elements.

Element	Indicator	Key U3 *A* compared Key Old-1739 *B*	Key U3 *A* compared Key New-1739 *B*
plaintext	Shared plaintext elements. Count	22	22
	Shared plaintext elements. In %	**91.7%**	**66.7%**
	Not shared plaintext elements from key *A*. Count	0	0
	Not shared plaintext elements from key *B*. Count	2	11
code	Shared code elements. Count	33	32
	Shared code elements. In %	**73.3%**	**59.3%**
	Not shared code elements from key *A*. Count	12	13
	Not shared code elements from key *B*. Count	12	22
assignment	Shared assignments. Count	32	29
	Shared assignments. In %	**71.1%**	**53.7%**
	Not shared assignments from key *A*. Count	13	16
	Not shared assignments from key *B*. Count	13	25

Table 4: Three keys, shared and not shared elements.

What is the relationship between the three keys? 1) Key old-1739 is probably derived, in an unknown year, from key U3, 73% (count 33) of their code elements are shared, 12 codes get replaced by another code. 71% of their code elements get the same assignment.

2) Key new-1739 is directly derived on December 2nd, 1739 from key old-1739. The manuscript literally reads: "old" and "new secret alphabet" (Dutch: oud en nieuw secreet alphabet). In *Appendix 6* we see that they share 82% of their code elements and 78% of their assignments. In key new-1739 nine nomenclature elements with words are added. For example, code element 'ThreeHouses' for plaintext element 'buscruyt' (English: gunpowder). It makes cryptanalysis harder.

As a rule of thumb one can say that with a shared assignment of 50% onwards a plaintext is sufficiently readable to be able to reconstruct the original key. Therefore, plaintext U3 is already quite readable if we use key old-1739 and key new-1739 for decrypting on ciphertext U3 (see Figure 3). Clearly, this also depends on the shared plaintext letters' frequencies, e.g. if the letter E is shared.

The person(s) who designed key old and new-1739 apparently had little knowledge of and hands-on experience with cryptanalysis because key U3 and old and new-1739 overlap in such a way that the plaintext is sufficiently readable after decryption to reconstruct its complete key.

3.3 Quality of encryption?

How good was the employee of the company who did the encryption of U3 (1689a)? Did they make any mistakes? The encryptor addressed nine wrong code elements to a plaintext element. See Table 5 for the exact mistakes.

Details of Mistakes
Code ϕ (count 23) with plaintext H should read plaintext **K** (count 2)
Code ϕ (count 23) with plaintext H should read plaintext **K** (count 2)
Code i (count 41) with plaintext N should read plaintext **M** (count 2)
Code i (count 41) with plaintext N should read plaintext **M** (count 2)
Code 2 (count 27) with plaintext A should read plaintext **T** (count 1)
Code 2 (count 27) with plaintext A should read plaintext **E** (count 1)
Code Ø (count 8) with plaintext F should read plaintext **V** (count 1)
Code 5 (count 22) with plaintext V should read plaintext **F** (count 1)
Code 7 (count 41) with plaintext E should read plaintext **Y** (count 1)

Table 5: U3 (1689a) nine mistakes made in the encryption.

Line	Word in ciphertext (a) not in Dutch (b)	Word in Dutch (b) not in ciphertext (a)	Translation into English. Missing word in bold
2		wegh	continually washes **away**
6	reets		work is already **quite** advanced
22		schip	English **ship** comes with horses

Table 6: Missing words compared.

Is ciphertext U3 (1689a) based on the Dutch plaintext U3 (1689b)? Or is the Dutch plaintext the decrypted plaintext from the ciphertext? When we look in detail at the missing words (Table 6), then it is most likely that the Dutch plaintext U3 (1689b) is the decrypted plaintext from the ciphertext U3 (1689a). The decryptor added an insert symbol '^' and 'schip' (English: ship) after decryption, to make a contextually correct sentence out of it (Figure 8).

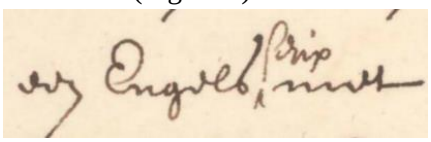


Figure 7: Line 22 in Dutch plaintext with insertion of word 'schip'.

4 Lessons learned

In this section, we look in detail at the automatic cryptanalysis component in CT2. Why didn't it work properly in the first attempt? What tips and tricks can we take into account next time when we use it for cryptanalysis?

4.1 Fourth attempt: automatic cryptanalysis in CrypTool 2

The first attempt to cryptanalyze the ciphertext with automatic cryptanalysis in CT2 failed. Now that we know what output to expect, we can take a closer look at how the automatic cryptanalysis for homophone substitution in CT2 works exactly. Was the transcription afterwards not as CT2 expected? Were the settings wrong? Another constraint?

In the fourth attempt, CT2 was able to decipher U3 (1689a) using the Homophonic Substitution Analyzer (see Figure 8).

4.2 Tips and tricks for automatic cryptanalysis

There are a few things that have to be taken into account when working with the Homophonic Substitution Analyzer of CT2 to allow it to actually decipher a ciphertext:

1) The transcription can be done and provided in different ways, but the analyzer has to know how it was done: (a) each symbol of the ciphertext was transcribed with a single transcription letter or (b) each symbol was transcribed using a word and spaces were used to separate these words. If the analyzer is not set to the correct transcription mode ("**single letter**" or "**symbol separated**"; here with the space symbol), it fails in tokenizing the ciphertext into individual code elements for the cryptanalysis.

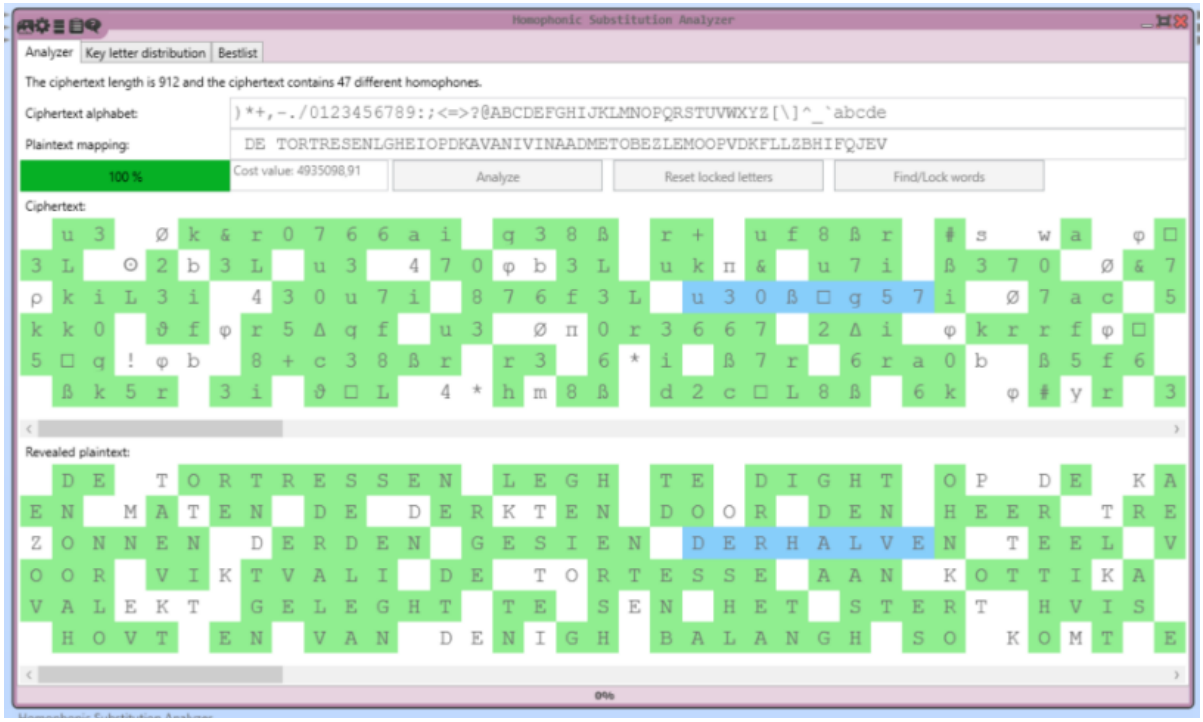


Figure 8: Fourth attempt: automatic cryptanalysis in CT2 using the Homophonic Substitution Analyzer for U3 (1689a).

2) The initial key for a first trial decryption of the automatic algorithm is created using different models of **randomness**: (a) it is equally distributed or (b) based on the assumed language of the plaintext. Depending on the cipher's type and difficulty, it makes more sense to set the key generation to equally distributed letters (e.g. to 1 for a monoalphabetic substitution cipher or e.g. to 2 for a simple homophonic substitution cipher). Here, some trial and error is needed but when the correct initial key generation is found, the automatic algorithm performs much better.

3) Set the correct **language** to be used at both, the Homophonic Substitution Analyzer as well as the used Dictionary component. The analyzer uses a language model for the cryptanalysis to improve the key and it uses the provided dictionary to find already deciphered words in the decrypted plaintext. If the languages are not set to the same (and correct) language, here in our case Dutch, the automatic cryptanalysis performs badly or fails.

4) Work **iteratively**: don't rely only on the automatic cryptanalysis. Use it as a starting point: it will find small parts of the plaintext and key. Then, correct these and lock these (green marked letters in Figure 8) in the analyzer's GUI. After that, restart the automatic process to let it improve what you found so far. By doing so, you improve the decryption (and the key) each time a little more.

After following all these tips and tricks, the automatic analysis on ciphertext U3 performed much better. Still, manual work was needed to decipher the texts completely, but the Homophonic Substitution Analyzer is a helpful tool to support the cryptanalysis process. Also, being in possession of a (different) key allowed us to enter they key and decrypt the ciphertext using another component, namely the Substitution cipher component of CT2.

5 The bigger picture

In Section 5 we discuss the use of ciphers in communication with the Dutch plantation colony Suriname and see to what extent the content of the revealed ciphertext confirms that it indeed dates from 1689.

5.1 Use of ciphers in the colony Suriname

What do we know about the use of ciphers in the colony Suriname? In 1710, in a reply to a letter from February 21st of the same year, the Gentlemen X of the WIC gave orders to the Council in Guyana, instructing them that 'encrypted letters' (Dutch: cijfer letteren) should use the old form instead of the new, advising to be used with great care to avoid mistakes. Otherwise, it could not be properly 'decrypted' (Dutch: ontcijffert werden).

In 1782, almost a hundred years later than ciphertext U3 (1689a), a couple of maps from the Suriname River with ciphertext written on them were sent to the Netherlands. De Leeuw writes about the used cipher (1997, 171): "Wollant used a secret code that was specially designed for the encryption of military messages. (...) quite simple in design and it had not changed since the outbreak of the War of Jenkins' Ear in 1739." Some years later he wrote (De Leeuw, 2000, note CLXV): "The West Indian Company, for instance, only had one cipher that was introduced in 1739 and was used over and over again during the 4th Anglo-Dutch War."

5.2 Dutch at war with England and France

In the 17th and 18th centuries the Dutch fought the following wars against England and France.

Four wars against the English:

- 1652-1654 First Anglo-Dutch War
- 1665-1667 Second Anglo-Dutch War
- 1672-1674 Third Anglo-Dutch War
- 1780-1784 Fourth Anglo-Dutch War

Five wars against the French:

- 1672-1678 Franco-Dutch War
- 1688-1697 Nine Years' War
- 1701-1714 War of the Spanish Succession
- 1740-1748 War of the Austrian Succession
- 1792-1797 War of the First Coalition

5.3 Dating U3 based on its content

Based on the physical place in the SvS archive between October 22, 1688 and [illegible month] 28, 1689 we assume that ciphertext U3 (1689a) dates from 1689 and was sent from Suriname.

What can we say about U3's dating based on its content?

- "Finish the works that Sir Fredenburgh started" (line 4-6). Abraham van Vredenburgh, commandeur and Governor ad interim between July 1688 and March 1689 after the murder of Governor Van Sommelsdijck and commandeur Verboom on July 19, 1688 (Verboom died of his injuries nine days later).[6]
- "This (work) is already quite advanced and is detached and open at the rear" (line 6-7). Referring to fort Zeelandia.
- "As can be seen on the map" (line 8). Referring to an unknown map.
- "Do not sail any further before the engineer arrives" (line 15). This seems to indicate that it is a report between an unknown reporter and the Governor in fort Zeelandia, Suriname.
- "Our warehouse only has food for three weeks" (line 10-11). Not specific enough.
- "A ship is (going) to Barbados for provisions" (line 12).
- "The fortress on Cottica is still little advanced" (line 13-14). The Fortress on Cottica is fort Sommelsdijk. Situated where the Commewijne River and the Cottica River meet. Construction on this fort started in 1684.[7]
- "The strong house in Para needs major repairs" (line 17-20). The strong house in Para is fort Houttuyn. At the mouth of the Para River in the Suriname river.
- "The (fort) in Saranmica is wooden and of little importance" (line 20-21). Unknown fortress or strong house on the Saramacca (also written as: Sarameca) or the Carameca River. On March 21st, 1689 a list was drawn up of the ammunition and other supplies present in Suriname. The following forts and posts are mentioned: fort Zeelandia, fort Sommeldijck, post Para, and post in Surammaca (Ammunition, 1689). In July 1689 Van Scharphuijsen (1689c) writes that he intends to leave the post in Surammeka and accommodate the soldiers elsewhere because this post does not provide any service to the Colony and everyone became very ill. Apparently this was a post that only existed for a few years.
- "An English ship comes with horses and little provisions" (line 22-23). This indicates that it was a time when they were not at war with England.

Based on the content of U3 we can draw the following conclusions: It's about the Dutch plantation colony Suriname because fort Cottica and strong house Para are mentioned and the plaintext is written in Dutch. When this letter was written, the Dutch were not at war with England, otherwise the latter would not have supplied provisions. The strongest affirmation that this letter was indeed written in 1689 is the fact that the Governor ad interim of Suriname is mentioned: Abraham Vredenburgh (written as: Fredenburgh).

5.4 Sender and receiver

Who was the sender of this message and who was supposed to receive it? The letter was written in May 18, 1689 and sent by Governor Van Scharphuijsen around that date from Paramaribo, Suriname to the Directors of the SvS in Amsterdam (number 4 "a small letter in code" in Letterbook, 1689; folio 315, 316 "secret letter from Scharphuijsen" in Register, 1689).[8] On August 4, 1689 the Directors read the letters from Van Scharphuijsen and don't mention the secret letter in code (Dutch: secrete missive of briefje in caracters) by name (Directie SvS, 1689a and

[6] Oudschans Dentz (1942).

[7] Meiden (1987).

[8] No date is mentioned in this source. All letters from the letterbook are listed consecutively. Folio 315, 316 states in Register (1689): "secret letter from the aforementioned Governor Scharphuijsen at this date" (Dutch: in dato dezen). And points back to Van Scharpenhuijsen's letter of May 18, 1689.

1689b). In their letter to Van Scharphuijsen from September 10, 1689 they reply to the sent letters from May. In it they write (Directie SvS, 1689c):

- Pleased to read that the French attack on Suriname had been repulsed and that the construction of our forts is progressing.
- We sent three ships in July to Suriname before. All three were taken – two of them were sunk – by the French enemy and the shipped goods have been lost.
- With these two ships we send to you once again: victuals, materials, ammunition of war, cannons, money and 50 good and experienced soldiers.

Was this what they were waiting for in Suriname? In October 1688 they had already started building fort Zeelandia and fort Van Sommelsdijck, advancing in defense of the colony against possible enemies (Vredenburgh, 1688). On March 12, 1689 the new Governor Van Scharphuijsen was installed in the 'Raad van Justitie en Politie' and Van Vredenburg is appointed as second in command (Raad notulen, 1689). On May 4, 1689 Van Scharphuijsen writes that they are working on the fortifications, they are awaiting an engineer and requests a few masons and carpenters. He also indicates that there is a lack of victuals (Van Scharphijsen, 1689a). The next day the French enemy arrive.

On May 5, 1689 Dutch ship Waapen van Amsterdam is taken at the mouth of the river Suriname by the French. Between May 8 and 12, the French shelled Fort Zeelandia from the river with their 10 ships, but did not come ashore. The Dutch fire back. Many deaths occur, especially on the French side. On May 14 the Dutch know for sure that the French sailed away into the sea (Scharphuijsen, 1689b and Vredenburgh, 1689a and 1689b).

In his letter from May 18, 1689 Van Scharphuijsen (1689b) writes that they are preparing for a second battle with the French and that is why even more people have been put to work to complete the forts. Request for ammunition of war, victuals, good officers, and warships.

Ciphertext U3 (1689a) in a secret letter with more details about the state of defense in May 1689 in Suriname. A caption to a map. It was considered so militarily sensitive that it had to be encrypted to be unreadable in case the map and its caption was intercepted by the (French) enemy.

6 Conclusion

Our investigation revealed that the content of the revealed plaintext letter U3 (1689) contains information about the progress of the construction of various forts in the Dutch plantation colony Suriname and the amount of food in the warehouses. Both are of military strategic importance.

Dutch plaintext U3 (1689b) is the revealed plaintext of ciphertext U3 (1689a). We have been able to relate these two texts.

Since 1689, the SvS used ciphers for communicating about militarily sensitive information. Ciphertext U3 was sent by Governor Van Scharphuijsen, in May 1689, from Suriname to his directors in Amsterdam, during the Nine Years' War (1688-1697) when the Dutch (together with others) were at war with the French. The letter – a caption to a map – was encrypted as a precaution against possible interception by the (French) enemy. This message was apparently never intercepted because both the ciphertext and its revealed plaintext were found in the archive of the SvS in the Netherlands.

Acknowledgements

We wish to express our gratitude to Mark Ponte for transferring ciphertext U3 to our possession.
We would like to thank Tiemen Bosma, George Lasry, Hugo Araújo, and the three anonymous reviewers for their assistance in the preparation of this article.
This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT - Decryption of Historical Manuscripts.

Link to the appendices

At the end of this paper *Appendix 3* is added. *Appendices 1*, *2*, *4*, *5*, and *6* can be found in DECODE record number 7841 via the following link: https://de-crypt.org/decrypt-web/RecordsView/7841.

References

Ammunition, 1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 302r-309r (scan 0322-0329).

Directie SvS, 1689a. Register van resoluties, Directors SvS from 4-8-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 3, unfoiled (scan 220-222).

Directie SvS, 1689b. Minuutresoluties, Directors SvS from 4-8-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 19, 176v-179r (scan 183-185).

Directie SvS, 1689c. Letter Directors SvS from 10-9-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 91, 115r-126v (scan 122-134).

Fred. Oudschans Dentz, 1942, Eenige bladzijden uit het leven der commandeurs van Suriname 1680 – 1804. *De West-Indische Gids* jaargang 24, nummer 1 (1942), 161-166.

Gentlemen X, 1710. 1710-9-5. NL-HaNA, WIC, 1.05.01.02, inventarisnummer 2, 1708 nov. 5 - 1710 okt. 4: scan 11 (unfoiled), f176v.

Karl de Leeuw, 1997. 'Geheimschrift op enkele kaarten en plattegronden van de verdedigingswerken rond de Surinamerivier, 1782', *Tijdschrift voor Zeegeschiedenis* (1997 nummer 2), 160-178.

Karl de Leeuw, 2000. *Cryptology and statecraft in the Dutch Republic*.

Karwan Fatah-Black, 2019. *Sociëteit van Suriname 1683-1795: het bestuur van de kolonie in de achttiende eeuw*.

Letterbook, 1689. Letters from Suriname to Amsterdam. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 91, 286r (scan 306).

Gerrit Knaap, Henk den Heijer, and Michiel de Jong, 2015. *Oorlogen Overzee. Militair optreden door compagnie en staat buiten Europa, 1595-1814*, 296-323.

Meiden, G. van der, 1987. *Betwist Bestuur. Een eeuw strijd om de macht in Suriname 1651-1753*, 41-69.

Old Key, 1739. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 86: unfoiled, scan 0002.

New Key, 1739. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 86: unfoiled, scan 0003.

Raad notulen, 1689. Minutes Raad van Justitie en Politie from 12-3-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 352r-353v (scan 372-374).

Register, 1689. Sent letters from Suriname to Directors SvS in Amsterdam. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, unfoiled (scan 001-011).

Scharphuijsen, 1689a. Letter Governor from 4-5-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 288r-291v (scan 308-312).

Scharphuijsen, 1689b. Letter Governor from 18-5-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 293r-292v (scan 313-315).

Scharphuijsen, 1689c. Letter Governor from 11-7-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, unfoiled (scan 545-546).

Stuart Schwartz, 2014. 'Looking for a New Brazil: Crisis and Rebirth in the Atlantic World after the Fall of Pernambuco', in: Michiel van Groesen (editor), *The Legacy of Dutch Brazil*.

U3, 1689a. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 315r (scan 0335).

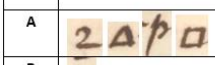
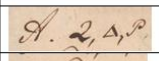
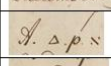
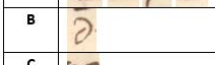
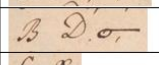
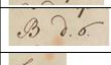
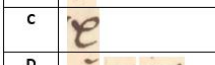
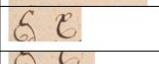
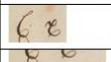
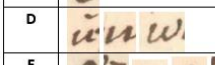
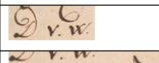
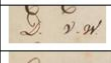
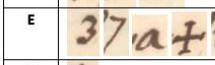
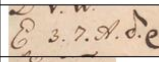
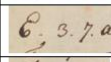
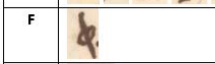
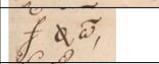
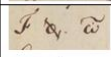
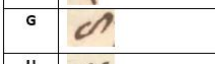
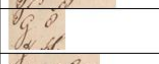
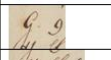
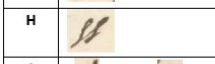
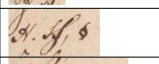
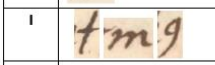
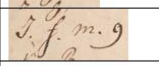
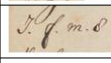
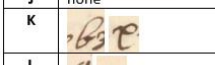
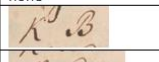
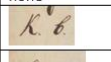
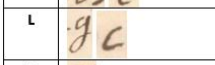
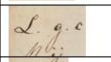
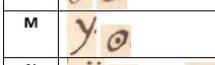
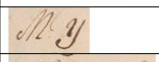
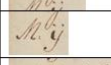
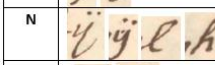
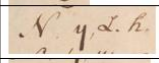
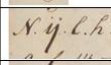
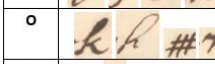
U3, 1689b. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 316r (scan 0336).

Vredenburgh, 1688. Letter from 23-10-1688. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 78r-82r (scan 094-098).

Vredenburgh, 1689a. Letter from 16-5-1689. NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 368r-372r (scan 388-392).

Vredenburgh, 1689b. Memorie without date (probably 16-5-16890). NL-HaNA, Sociëteit van Suriname (SvS), 1.05.03, inv.nr. 219, 374r-376v (scan 394-397).

APPENDIX 3. Reconstructed Key U3 (1689), Old Key (1739) and New Key (1739)

Latin letter	U3 (1689) reconstructed	Old key (1739)	New key (1739)
A			
B			
C			
D			
E			
F			
G			
H			
I			
J	none	none	none
K			
L			
M			
N			
O			
P			
Q			
R			
S			
T			
U	none	none	none
V			
W			
X	none		
Y			
Z	none		
NC	no nc	no nc	9 nomenclature elements (nc)

Nomenclature (NC) elements and corresponding nomenclature code elements in the new key (1639):

French encrypted newspaper advertisements in the 19th century

Elonka Dunin
Codebreaking-guide.com
elonka@gmail.com

Didier Müller
Ars Cryptographica
madimu2@gmail.com

Klaus Schmeh
Codebreaking-guide.com
klaus@schmeh.org

Abstract

We present a ciphertext database created by one of the authors. It contains over 3700 encrypted newspaper advertisements published in the French newspaper *Le Figaro* between 1875 and 1897. The collection includes over 2300 solved messages, which have been encrypted in almost 90 different crypto systems, as well as over 1400 unsolved cryptograms. We introduce some of the most interesting solved and unsolved advertisements, including messages based on ciphers, codes, and steganography. It will become clear that in addition to the messages contained in the database so far, thousands more encrypted advertisements from French newspapers remain to be catalogued and deciphered.

1 Introduction

Encrypted advertisements in British newspapers have been covered in several publications. Already in 1881, the book *The Agony Column in The Times, 1800-1870* by Alice Clay listed over 200 encrypted and some thousand unencrypted advertisements (Clay 1881). 125 years later, a different book with a similar title, *The Agony Column Codes & Ciphers* (Palmer 2005) was published by Tony Gaffney, using the pen name Jean Palmer. This work had a focus on encrypted advertisements, such as the story of English polar explorer Richard Collinson, who used encrypted advertisements in the London-based newspaper *The Times* to receive messages from his family during a five-year worldwide journey. Collinson's story is also covered in (Rabson 1992) and (Gutoskey 2022). The book *Codebreaking: A Practical Guide* by two of the authors of this paper covers a number of encrypted ads in British newspapers, too (Dunin 2023).

Contrary to British encrypted newspaper advertisements, French ads of this kind have never received much coverage in the literature. In his classic work *The Codebreakers*, David Kahn reports about the French cryptologist Étienne Bazeries (1846-1931), who discovered his interest in encryption technology through enciphered newspaper advertisements (Kahn 1996). However, Kahn provides no information about the advertisements themselves. Hervé Lehning introduces several French encrypted newspaper ads in a short article titled "Les messages chiffrés du Figaro en 1890" (Lehning 2018), but his collection is small. We are not aware of a comprehensive publication about French encrypted newspaper advertisements.

To shed more light on this subject, one of the authors of this work (Didier Müller) in 2022 started to systematically scan the French newspaper *Le Figaro* in order to find encrypted advertisements. Didier has authored a well-known French crypto-book (Müller 2018) and is a leading French crypto history expert. As he found out, *Le Figaro* launched a new column, *La Petite Correspondance* (Small Correspondence), in its Sunday supplement on January 3, 1875. This column was dedicated to messages exchanged between private individuals. At a time when the telephone was still in its infancy, many readers of the daily newspapers took advantage of this means of communication. Some of the advertisements in this column were encrypted.

A rapid success, *La Petite Correspondance* was also published in *Le Figaro* on Thursdays, starting in May 1875, before finally becoming a

January 3, 1875

PETITES ANNONCES DU FIGARO

La ligne composée de 34 lettres, 1 fr. 50. — La double ligne dans la Petite Correspondance, 70 lettres : 3 francs.
S'adresser de neuf heures à six heures aux bureaux du *Figaro*.

ON PAIE LES ANNONCES COMPTANT

Les Petites annonces sont reçues jusqu'au vendredi soir, six heures, pour dernier délai.

August 17, 1879

PETITE CORRESPONDANCE

PIERRETTE.— 1° Le prix de la Petite Correspondance est de 6 fr. la ligne de 40 lettres ; 2° on peut envoyer timbres-poste avec l'annonce ; 3° toute copie reçue la veille est insérée le lendemain.
(Administration.)

Figure 1. At the beginning in 1875, an advertisement in the *Petite Correspondance* cost 1.50 francs (which would be around €15 or $16 today) per line. In 1879 the price was increased to 6 francs.

daily feature. Its title changed to *Renseignements* (Information), then *Arrivées – Départs – Renseignements* (Arrivals – Departures – Information) on July 15, 1878, before being revived on February 25, 1879. It was later renamed *Correspondances personnelles* (Personal correspondence, May 6, 1884), then *Correspondances particulières* (Special correspondence, April 12, 1886), then *Correspondances personnelles* again (July 2, 1886).

At the beginning, an advertisement in the *Petite Correspondance* cost 1.50 francs per line with a line consisting of 34 letters (Figure 1). According to the website historicalstatistics.org, this would be around 15 Euros (16 US-Dollars) today in 2024, based on the gold price. In 1879, an advertisement became much more expensive, at a price of 6 francs for a line with 40 letters. This rate didn't change until at least 1896.

In the *Petite Correspondance* column and its successors, Didier found many encrypted advertisements. He set up a database with his findings, which he made available online at www.apprendre-en-ligne.net/crypto/Figaro (see Figure 2). As of April 2024, Didier's collection contains 3743 cipher messages, 2314 (62 %) of which have been broken by himself or Julien Cavillon, Maximilian Bärtl, Mark van de Beek, or Pierre-Yves Chansigaud.

So far, Didier has scanned the content of *La Petite Correspondance* and its successor columns through 1896, which means that his research project is still ongoing. Nevertheless, he has collected enough material for this first publication.

2 The ciphers used

Here are some facts about the nearly two thousand deciphered advertisements in Didier's database:

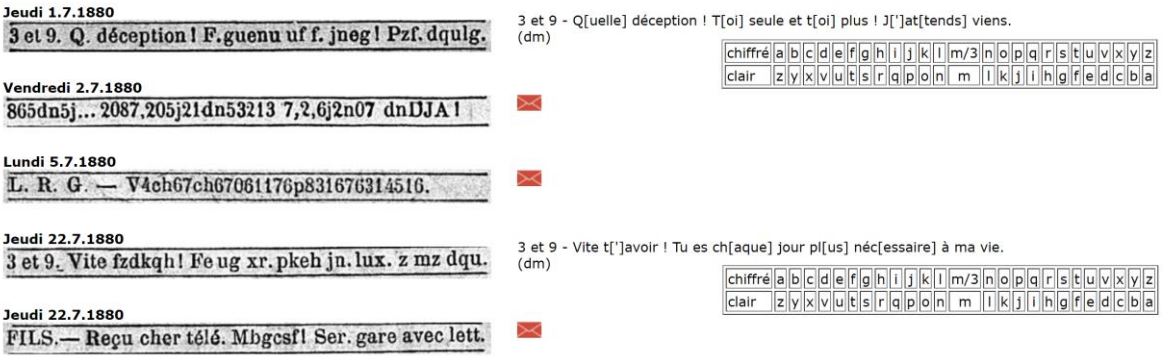


Figure 2. This excerpt from Didier's database shows five encrypted newspaper advertisements from *Le Figaro* published in July 1880. Two of these, which used simple substitution, have been solved. The other three remain unsolved.

So far, 86 different encryption systems have been identified.

- Most encryption systems used are simple substitutions. The Caesar cipher is very common (938 appearances), with the letter "w" (570 out of 938), which is very rare in the French language, sometimes being omitted (368 out of 938). The most frequent key (offset) used for the Caesar cipher is 1 (511 out of 938). This means that over 13% of the ciphertexts Didier found in *Le Figaro* are based on one of the simplest ciphers. Why do so many encryption systems remove the letter "w"? At the end of the 19th century, "w" was not really considered a letter in French. It wasn't until 1964 that the Robert dictionary first declared "w" to be the 23rd letter of the French alphabet.
- Another very frequent system is the Atbash cipher (175 appearances). The Atbash cipher is a substitution code that maps the letters A to Z to the letters Z to A.
- Other types of simple substitutions appear 72 times. Many of these ciphers are based on an alphabet that includes numbers. In some cases, numbers replace vowels, while the rest stays in the clear.
- There are a few polyalphabetic ciphers in the collection (215 appearances). The most sophisticated one is the one that can be referred to as TRn. It uses three shifts of the Caesar cipher and nulls, which indicate that the words that follow are written backwards.

Didier has encountered only a few homophonic ciphers (6 appearances). Usually, the letter "e" has several homophones, which means that it can be replaced by several symbols.

- There are also only a few transposition ciphers (53 appearances) in the database. Most of these consist of writing from right to left.
- So far, Didier has only been able to identify and decipher one code based on a codebook, namely *Benoît Brunswik's Dictionnaire pour la correspondance télégraphique secrète* (Dictionary for secret telegraph correspondence), which was used 76 times (Brunswik 1869). Considering that codebooks were very popular in the 19th century, it seems likely that others were used as well (one candidate is provided below).
- There are also more original ciphers, including combinations of encrypted and plain language, mixed alphabets, and the use of null characters to confuse a potential solver.

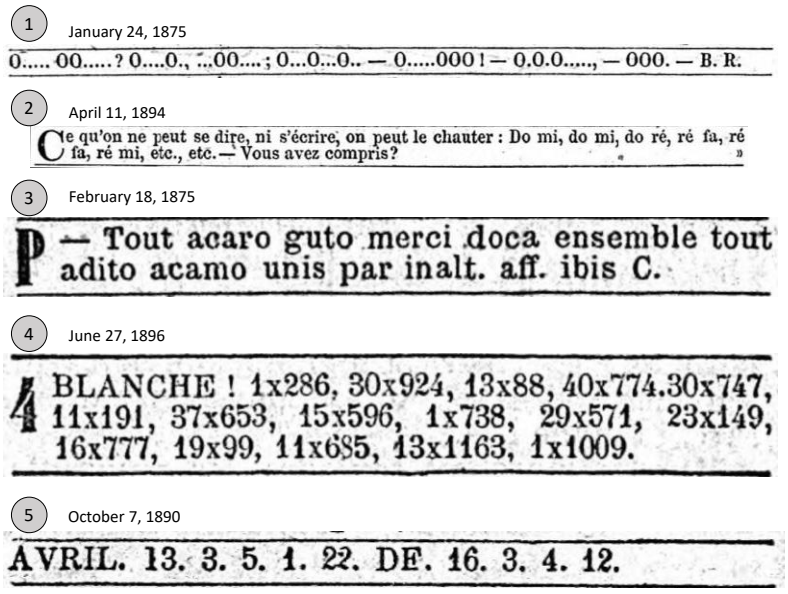


Figure 3. There are over 1300 unsolved messages in the collection. Five of them are listed here. Many of the ads are short and therefore provide only little material for analysis.

Although many of the encryption systems used in *Le Figaro* are fairly simple to solve, there are a few difficulties that a cryptanalyst might encounter:

- Most advertisements were certainly written by amateurs. As a result, there are numerous errors, including misspellings and incorrectly shifting some letters.
- To save money, many words were abbreviated. Sometimes, the abbreviation is so short that it is hard to reconstruct the original word.
- There are also typographical errors: "c"'s that are "e"'s, "h"'s that are "b"'s, and so on.

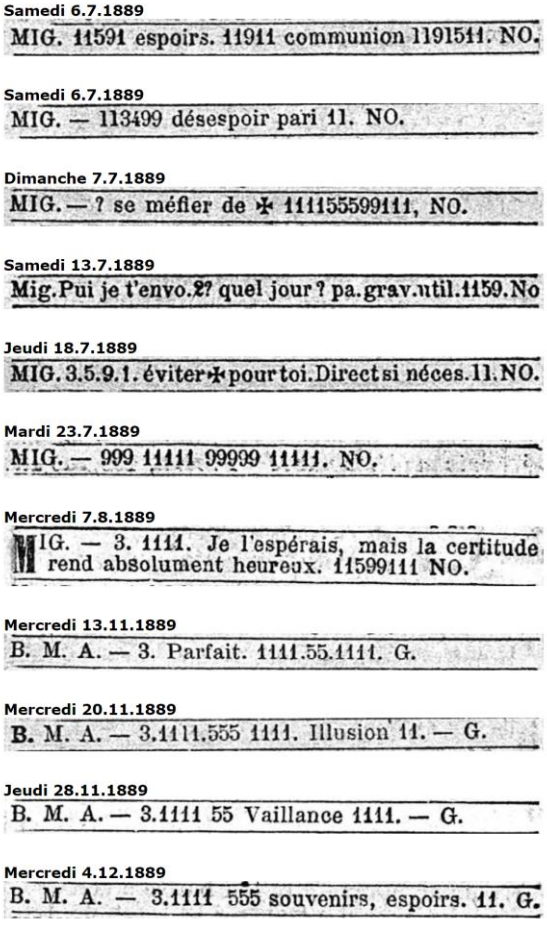


Figure 4. At least 147 ads start with the letters MIG or B. M. A. As can be seen, many of these messages contain digit groups such as 1111, 555, or 9999. The meaning of these ads is unknown.

3 Content of the ads

The vast majority of the deciphered advertisements in Didier's database represent messages that were exchanged between lovers. A few others are about the health conditions of the sender. Messages with a commercial background are very rare. To our regret, we haven't encountered any outstanding background stories so far. We still hope to find something that is as spectacular as the aforementioned story of the English polar explorer Richard Collinson, who in the 1850s established a worldwide communication system based on encrypted ads in *The Times*.

Most of the ads were apparently written by men for women, because in those days women stayed at home and it was dangerous for them to receive mail from lovers. The women, on the other hand, probably replied with letters. As they weren't financially independent, they often couldn't afford to pay the price of an advertisement. For this reason, we usually have only half of the dialogue.

4 A selection of unsolved messages

As mentioned, many of the encrypted advertisements in Didier's collection are still unsolved. Figure 3 shows a few examples, which are covered in more detail below.

4.1 A binary message

The message in Figure 3.1 mainly consists of words encoded with the characters "0" and ".". This suggests that a binary code was used, similar to the Bacon cipher, the Morse alphabet or the ASCII code (which was only invented in the 1960s). So far, we have not been able to make sense of this message.

4.2 A musical message?

The advertisement in Figure 3.2 is in French plaintext, and can be translated to English as follows:

What can be neither said nor written can be sung: Do mi, do mi, do re, re fa, re fa, re mi. Do you understand?

The terms “do”, “re”, “mi” and “fa” probably correspond to the musical notes c, d, e and f. Unfortunately, no note lengths are given. A possible explanation is that the notes in question form the beginning of a piece of music. The notes actually sound like a melody, which is, as we know, not always the case with a message disguised as sheet music. However, we are not aware of any tune that starts this way. Another suggestion is that the notes represent the beginning of a musical choir warm-up, so may refer to such an event.

Another potential explanation is that the Solresol artificial language was used. Solresol is a music-based artificial language developed by the Frenchman François Sudre starting in 1817 (Wikipedia 2023). The idea behind it is similar to that of the much better-known Esperanto, although the implementation differs considerably.

Figure 5: This message series was probably encrypted with a code based on a codebook.

The vocabulary of Solresol is based on the tonal syllables do, re, mi, fa, sol, la, si (or ti). Frequently used words consist of one, two or three tonal syllables. Thus, si stands for “yes” and do for “no”. Doredo means “time”, dorela “year” and doresi “century”. More specific terms are formed from four or five syllables. One can communicate with Solresol not only by speaking, but also by singing, whistling, or using symbols.

Is “Do mi, do mi, do re, fa re, fa re, fa mi” a Solresol message? We have consulted experts, such as via the Solresol online community on Discord, but to our regret we have only received negative responses.

4.3 A non-existing language

The message in Figure 3.3 appears to be mainly made of French words, which, however, don't form a meaningful sentence. Perhaps, they come from a codebook. It seems also possible that each word stands for a letter or number, though such a system would have led to a costly waste of bandwith.

4.4 A list of alleged multiplications

Figure 3.4 starts with the word *BLANCHE!* (White) and then shows a list of multiplications:

1×286, 30×924, 13×88, 40×774, 30×747, 11×191, 37×653, 15×596, 1×738, 29×571, 23×149, 16×777, 19×99, 11×685, 13×1163, 1×1009

One possible explanation is that this message was created with a book cipher. If so, “30×924” might stand for the 30th letter on page 924 or the 924th letter on page 30 or something similar. As determining the 924th letter in a text can be laborious and error-prone, it seems possible that a book mainly displaying tables was used. Or perhaps, an additional cipher was used, which mapped the true position of the letter to a larger number.

On the other hand, it seems possible that the alleged multiplications refer to entries of a codebook, perhaps with some super-encryption. As the message is short and no other message of this kind is known, it might be difficult to find the solution.

4.5 A list of numbers

The message in Figure 3.5 can be transcribed as follows:

AVRIL. 13. 3. 5. 1. 22. DE. 16. 3. 4. 12

As all the numbers are between 1 and 26, it seems likely that they need to be substituted with letters (A=1, B=2, C=3, …). Such a substitution renders the following result:

APRIL. M. C. E. A. V. DE. P. C. D. L

We don't know what this means.

5 An unsolved ad series

The collection described in this work includes many series of advertisements, which were apparently published by the same person. One of the longest and most interesting ones consists of at least 147 ads, which were published between May 29, 1887 and May 7, 1890. Figure 4 shows a few of these cryptograms. Until August 7, 1889, these messages start with MIG, then with BMA. As can be seen, many of the messages contain digit groups such as 1111, 555, or 9999. To our regret, we have no idea how such a cryptogram can be deciphered.

6 Another unsolved ad series

Figure 5 shows another series of encrypted advertisements from *Le Figaro*. All eight messages were published between October and December 1880. No other ads of this kind appear in Didier's collection. As can be seen, some of the messages contain cleartext passages, some don't. The ciphertext parts mainly consist of five-digit groups, which is typical for a code based on a codebook. The best way to break such an encryption is usually to find the codebook that was used. Anyone interested in solving this mystery should probably start by checking French codebooks that were available in 1880.

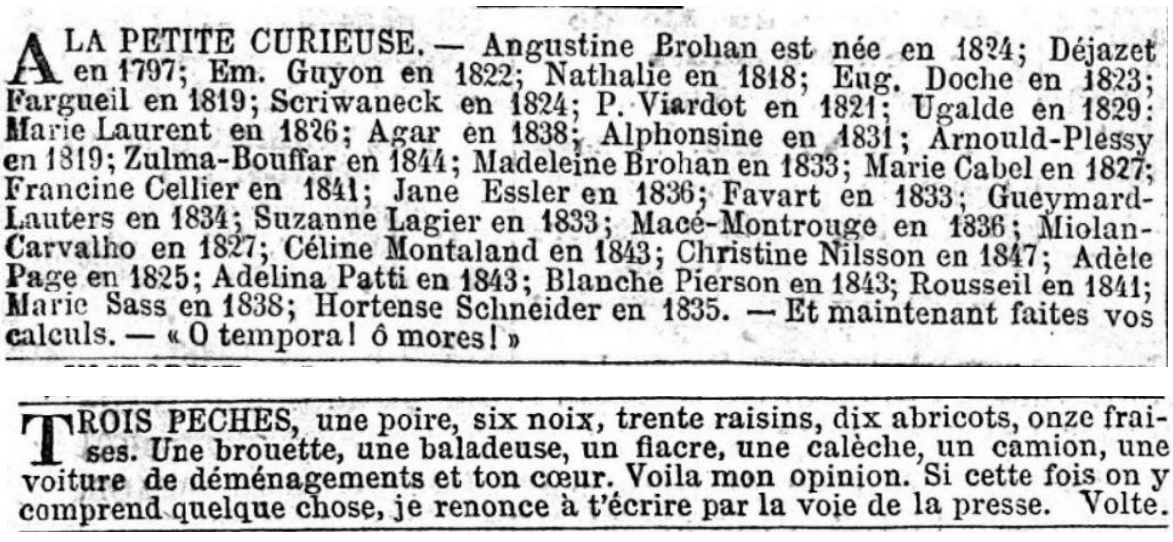
7 Steganography used?

Figure 6 shows two advertisements that might represent steganographic messages.

7.1 A LA PETITE CURIEUSE

The title of this message translates to "To the curious little girl". It lists the names of about 30 persons along with their birth years. It finishes with the words "and now draw your conclusions, O tempora, O mores". The latter is a famous Cicero quote, which can be translated as "Oh the times! Oh the customs!"

It seems unlikely that somebody spent the money for a newspaper advertisement just to publish names, birth years, and some trivial statements. We therefore believe that this



A LA PETITE CURIEUSE. — Angustine Brohan est née en 1824; Déjazet en 1797; Em. Guyon en 1822; Nathalie en 1818; Eug. Doche en 1823; Fargueil en 1819; Scriwaneck en 1824; P. Viardot en 1821; Ugalde en 1829; Marie Laurent en 1826; Agar en 1838; Alphonsine en 1831; Arnould-Plessy en 1819; Zulma-Bouffar en 1844; Madeleine Brohan en 1833; Marie Cabel en 1827; Francine Cellier en 1841; Jane Essler en 1836; Favart en 1833; Gueymard-Lauters en 1834; Suzanne Lagier en 1833; Macé-Montrouge en 1836; Miolan-Carvalho en 1827; Céline Montaland en 1843; Christine Nilsson en 1847; Adèle Page en 1825; Adelina Patti en 1843; Blanche Pierson en 1843; Rousseil en 1841; Marie Sass en 1838; Hortense Schneider en 1835. — Et maintenant faites vos calculs. — « O tempora! ô mores! »

TROIS PECHES, une poire, six noix, trente raisins, dix abricots, onze fraises. Une brouette, une baladeuse, un fiacre, une calèche, un camion, une voiture de déménagements et ton cœur. Voila mon opinion. Si cette fois on y comprend quelque chose, je renonce à t'écrire par la voie de la presse. Volte.

Figure 6. These two messages contain lists of persons and objects. They might represent steganographic messages.

message has a hidden meaning. As with other of the messages, we have no idea about what the true content of this advertisement might be.

7.2 TROIS PECHES

This message translates to:

Three peaches, one pear, six nuts, thirty raisins, ten apricots, eleven strawberries. One wheel-barrow, one lamp, one cab, one carriage, one truck, one van, and your heart. Here is my opinion. If this time you understand something of it, I will stop writing to you via the press. Please.

Again, it is unlikely that somebody would invest a considerable amount of money just to publish a list of fruits and vehicles. We therefore believe that there is a hidden meaning behind these lines, but we don't know any details.

8 Conclusion

Didier's database is a treasure trove for anybody interested in the history of encrypted communication. In this work, we have given an overview on the first 3165 finds and provided a number of examples of solved and unsolved messages from this corpus.

The current collection is still far from complete, as advertisements published in *Le Figaro* after 1896 and in other French newspapers need to be found and reviewed. Nevertheless, our work shows that several thousand encrypted advertisements were published in France in the 19th century. This is considerably more than the thousand ads that are mentioned in the aforementioned book of Tony Gaffney, which covers more newspapers (8) and a longer time span (1804-1909). This comparison suggests that considerably more encrypted advertisements were published in France than in Great Britain at the time.

8.1 Open questions

It is obvious that additional research work is necessary in order to get a good understanding of French encrypted advertisements in the 19th century:

- The collection includes over a thousand unsolved cryptograms. It will be an interesting task to break these.
- It should be checked whether other French newspapers, especially *Le Monde*, contained encrypted advertisements as well.
- So far, Didier has reviewed the content of *La Petite Correspondance* and its successors until 1897. Later issues of *Le Figaro* need to be scanned, as well.
- If it is indeed true that there were more encrypted messages in French newspapers than in British, it would be an interesting question to figure out why.

8.2 Call for help

We want to use this publication to launch a call for help. To further extend the database, Didier would appreciate the help of other crypto enthusiasts. Perhaps, one or more readers are interested. Check here for details: https://www.apprendre-en-ligne.net/crypto/Figaro/contribuer.html.

Acknowledgments

The following people have assisted in extending the database by deciphering advertisements contained in it: Julien Cavillon, Maximilian Bärtl, Mark van de Beek, and Pierre-Yves Chansigaud. The authors want to thank these persons.

References

Alice Clay. 1881. *The Agony Column in The Times, 1800-1870*. Chatto and Windus, Piccadilly

Benoît Brunswik. 1869. *Dictionnaire pour la correspondance télégraphique secrète*. Veuve Berger-Levrault & Fils, Paris

Elonka Dunin and Klaus Schmeh. 2023. *Codebreaking: A Practical Guide. Expanded edition.* No Starch Press, San Francisco

Ellen Gutoskey. 2022. *How Victorian Explorers and Pining Lovers Used Coded Newspaper Ads to Communicate*. https://www.mentalfloss.com/posts/victorian-coded-messages-franklin-expedition

David Kahn. 1996. *The Codebreakers*. Scribner, New York: 244

Hervé Lehning. 2018. *Les messages chiffrés du figaro en 1890*. https://blogs.futura-sciences.-com/lehning/2018/03/10/messages-chiffres-figaro-1890/

Didier Müller. 2018. *Les codes secrets décryptés (3rd edition)*. Nymphomath Editions, without location

Jean Palmer. 2005. *The Agony Column Codes & Ciphers*. Authors OnLine, Gamlingay

John Rabson. 1992. *All are well at Boldon a Mid-Victorian Code System*. Cryptologia Volume 16, 1992 (2):127-135

Wikipedia. 2023. *Solresol*. https://en.wikipedia.org/w/index.php?title=Solresol&oldid=1217986188

The TICOM DF-114 Cryptanalytic Device - A Theory of Operation and Computer Simulation

Magnus Ekhall
magnus.ekhall@gmail.com

Abstract

The M-209 cipher machine was used extensively by the U.S.A. during World War II. It is known that German cryptanalysts under certain circumstances were able to decipher M-209 enciphered messages using pen-and-paper techniques. A German wartime document found by the allies' Target Intelligence Committee (TICOM) in 1947 describes a electromechanical machine that supposedly could be used as an aid when breaking M-209 enciphered messages. The document, designated DF-114 by TICOM, is quite technical but does not describe how the device would work.

This paper suggests a theory of how the device could have been used, and by creating a computer simulation of the device described in DF-114 explores the viability of the theory.

1 The M-209 Cipher Machine

The M-209 is a mechanical cipher machine used by the U.S. military during World War II. Developed by the Swedish company AB Cryptoteknik as C-38, it was licensed to the U.S. military under the designation "Converter M-209". In 1942 national production of M-209 started in the U.S.A. by L. C. Smith & Corona Typewriters Inc. In total more than 140,000 M-209 units were manufactured (Kahn, 1996).

The M-209 has six pin wheels of different sizes: 17, 19, 21, 23, 25 and 26 pins. The pins can be individually set to either an inactive or an active state. For each letter to be enciphered all wheels steps forward one step. Since the sizes of the six wheels are relatively prime, this gives a cycle size of $17 \times 19 \times 21 \times 23 \times 25 \times 26 = 101,405,850$ letters.

Each wheel has a number of letters printed on it corresponding to the number of pins of that wheel. Since the wheels differs in size, the available letters also differs:

Wheel 26:	A-Z
Wheel 25:	A-Z except W
Wheel 23:	A-X except W
Wheel 21:	A-U
Wheel 19:	A-S
Wheel 17:	A-Q

At any time, the six wheels present one pin each to a drum consisting of 27 bars. On each bar two metal lugs can be configured to either a neutral position or to take a position corresponding to one of the six wheels. When the M-209 is enciphering a letter, the drum spins one revolution and all 27 bars are in turn interacting with the pins currently presented by the wheels.

If a lug on a bar meets an active pin, that bar is then activated. Depending on the configuration of the pins and lugs, any number of bars from 0 to 27 could be activated at the end of the revolution. This number of active bars decides what alphabet is used to encipher the letter (Lasry et al., 2016).

A Beaufort reciprocal alphabet is used as a basis for encryption in the M-209:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ZYXWVUTSRQPONMLKJIHGFEDCBA
```

With this alphabet letter A would be enciphered to Z, B to Y and so on.

The number of active bars when enciphering a letter changes the alphabet used for enciphering that specific letter: The encryption alphabet is rotated that many steps to the right.

For example, if there are three active bars on the drum, the alphabet used for that letter would look like this:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
CBAZYXWVUTSRQPONMLKJIHGFED
```

Note that it is only possible to create 26 different Beaufort alphabets this way.

In this case letter A would be enciphered to C, B to B and so on. With DF-114 terminology the M-209 in this case is said to have introduced a "skip" of three spaces in the reciprocal alphabet (TICOM, 1948b).

Since the alphabet used is reciprocal, deciphering takes place using the same alphabet as when enciphering and the cipher is thus symmetrical.

Another way to describe the encryption process carried out by the M-209 is that it switches between up to $2^6 = 64$ indexes, depending on the active pins and the lugs on the drum. Each index points to one of the 26 possible Beaufort alphabets. Several of the indexes therefore must point to the same Beaufort alphabet.

1.1 Keys

The cryptographic key of the M-209 consists of two parts: the state of the pins on the wheels and the position of the lugs of the drum. The key was changed at regular intervals, typically daily. The key settings were distributed in advance to all parties that should be able to communicate with each other. Typically each U.S. army division and corps had their own set of keys (Friedman et al., 1950).

When a message was enciphered a procedure was used to set the starting position of the six wheels in a secure manner. The procedure allowed the sender to inform the receiver of the starting position, and also make sure that the starting position is different for each new message. If this procedure is followed, it makes it difficult for an eavesdropper who does not have the correct key to break the cipher.

Different procedures might have been in use, but one that is known to have been used worked as follows (Pokorn, 1945):

1. The operator configures the lugs and pin wheels according to the daily key. This is typically done once per day (or possibly other time period as agreed).

2. The operator **randomly** sets the six wheels and writes down the corresponding letters of the wheel.

3. The operator randomly selects a letter of the alphabet.

4. This letter is encrypted repeatedly with the M-209.

5. The enciphered letters that result from the aforementioned encryption are used to set the six wheels to the real starting position for the encryption of the cleartext message. If a letter is generated that does not exist on that specific wheel, the next encrypted letter is used. This is repeated until viable starting positions for all wheels have been found.

6. The real message is enciphered.

The six letters from point 2 and the random letter from point 3 is part of the message indicator which is transmitted, unencrypted, as part of the message. On the receiving end the operator performs the same operation with the data from the message indicator and will thus decipher the message with the identically configured M-209 as the sender.

By using this procedure every message is enciphered with the six wheels set to a random position. Furthermore it is not easy to obtain the correct starting position for the six wheels from the information in the message indicator alone (Pokorn, 1945).

2 German cryptanalysis of the M-209

A considerable amount of M-209 traffic was broken and read by the Germans during World War II. About 5-10% is an estimate of the amount of traffic read purely by cryptanalytic means. In practice the Germans depended on messages in depth[1] or on M-209 operator errors in order to be able to read the messages (Friedman et al., 1950; TICOM, 1948c). This shows that the procedure outlined in section 1.1 was not strictly followed at all times.

From a broken message it is sometimes possible to deduce the M-209 lug settings, and the sequence of active and inactive pins on the M-209 wheels (Pokorn, 1945). This is called the *relative setting*.

To be able to decipher further messages the *absolute setting* is needed. The absolute setting consists of the additional knowledge of where in the sequence of active and inactive pins the letter "A" is on the circumference of each wheel.

[1] Two or more messages enciphered using the same M-209 settings.

With this information, it is possible to use the message indicator as described in section 1.1, and all messages enciphered with the same key would then be readable. Typically that would be all traffic for one network for one day.

Obtaining the relative setting was estimated to take two to three hours. Obtaining the absolute setting was considerably more difficult, estimated to take between twelve hours and four days (Friedman et al., 1950) and for that, different pen-and-paper techniques were developed (TICOM, 1948a).

3 The TICOM DF-114 document

In 1947 a German document was found dug down in the ground at a camp at Glasenbach just outside Salzburg, Austria. The document was translated into English and got the designation DF-114, document number 2785 by the Target Intelligence Committee (TICOM, 1948b). The title of the document is "Technical Note on machine treatment of AM-1 compromised texts in depth of 5", and it is clear already from the title that the document is related to the M-209 converter which was called "AM-1" by the Germans. Even though the document is translated into English there are some words here and there that are left in German, or sometimes written both in English and in German. The original German title of the document is given as *Techniches Erlaueterung zur maschinellen Bearbeitung von AM-1 Kompromisstextlösung auf 5er Texttiefe*. It is not clear who the original author is or who the target audience for this document is.

The document consists of 13 pages and a TICOM cover page. Of the 13 pages, four are a textual description of an electromechanical apparatus and nine pages are technical drawings of various kinds: electrical schematics, mechanical drawings and an overview.

The text describes what the device consisted of and how the various parts were interconnected. It fails to describe how the device was used: what was the input and what was the output? There are frequent references to the M-209 in the document, so it is clear also from the contents of the document that this design targeted the M-209.

On the cover page, TICOM writes: "From this paper the precise purpose of the device is not clear, neither is the manner in which it is supposed to function."

The document was declassified in 2010 (TICOM, 1948b).

3.1 Theory

The theory that will be investigated in this paper is whether the device can be used to test for possible cribs given an enciphered message and given that you know the pin and lug settings of the day. That is, the relative setting of the M-209 is assumed to be known, but the message that is to be investigated does not need to be in depth.

As mentioned in section 2, the Germans mostly relied on messages in depth, or messages close to being in depth in order to read M-209 traffic.

When having both the cleartext and the ciphertext of a message there were techniques for reconstructing the pin and lug settings: the relative setting (Pokorn, 1945). One remaining problem is that unless there is knowledge of the absolute setting, the rest of the traffic from that day can not be read since it is not possible to use the message indicator as described in section 1.1.

The theory will be described in detail in section 5 but first an analysis of the device described in the DF-114 document is needed.

4 The DF-114 device

This section describes the device as a direct and objective interpretation of the contents of the DF-114 document.

The device described in DF-114 is stated to consist of three main parts and a number of auxiliary parts. The main parts are:

1. Skip box
2. Distributor
3. Switching device

Figure 1 shows the main parts of the DF-114 device and how they are connected.

4.1 Skip box

The skip box (German: *Sprungkasten*) gets its name from the German nomenclature of how the M-209 works. As described in section 1 above, at each position in the message the M-209 introduces a 0 to 27 shift of the reciprocal alphabet used to encipher or decipher the current letter. This shift is called "skip" (German: *Sprung*) by the author of the DF-114 document (TICOM, 1948b).

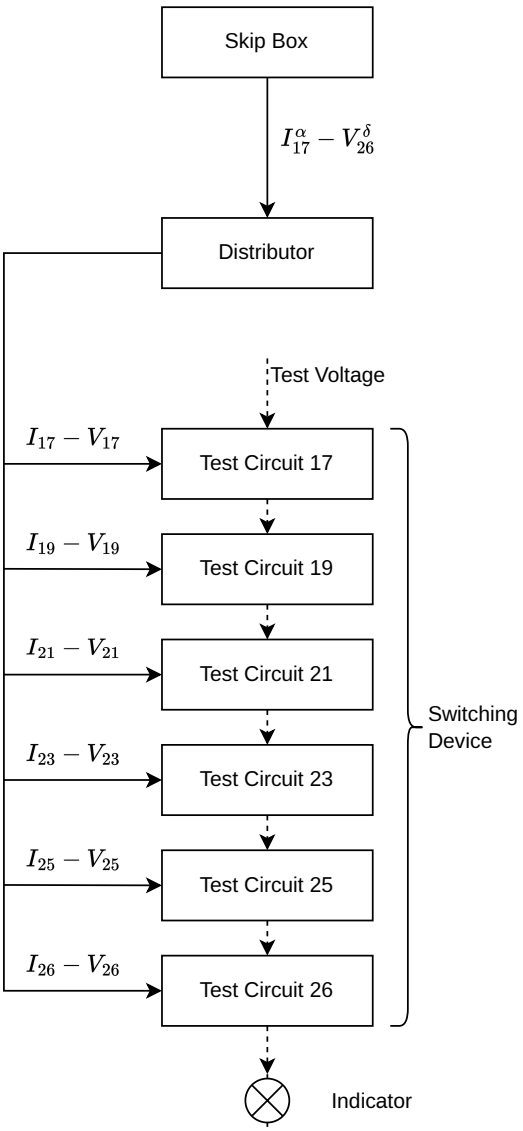


Figure 1: The DF-114 device is described as having three main parts: a skip box, a distributor and a switching device. The switching device consists of six test circuits.

The skip box contains five cylinders. Each cylinder has 26 slots, presumably one slot per letter in the alphabet. Each slot is divided into four sections. One sheet metal plate can be inserted into each section, giving four possible plates in each of the 26 slots of a cylinder. Each metal plate has six teeth that each represent an active or inactive state, depending on whether a tooth is present or has been removed. Each metal plate measures 60 millimeter in width by 15 millimeter in height. In the DF-114 document the teeth are described as "corresponding to the six wheels of the AM-1 or M-209" (TICOM, 1948b).

The five cylinders can be manually rotated independently from each other using a knob on the cylinders. Directly below the cylinders in the skip box are spring loaded electrical switches. Only one of the 26 cylinder slots, the cylinder slot that is facing directly downwards, affects these switches. There is one switch per tooth of a metal plate. So, for one cylinder there are 4×6 electrical switches, which makes $4 \times 6 \times 5 = 120$ in total for all five cylinders.

In DF-114 the five cylinders are denoted with roman numerals from I to V. The four metal plates in a cylinder slot are labeled with Greek letters α to δ, and the six teeth of a metal plate are denoted 17, 19, 21, 23, 25 and 26 which equal the number of pins of the six wheels of the M-209. So for example IV_{19}^{γ} denotes cylinder 4, metal plate 3 and tooth 2 on that metal plate. At all times, only the cylinder slot that is facing downwards and thus affects the electrical switches is considered.

Figure 2 shows a technical drawing from the DF-114 document depicting a cylinder, seen from the side. Figure 3 illustrates how a cylinder is slotted.

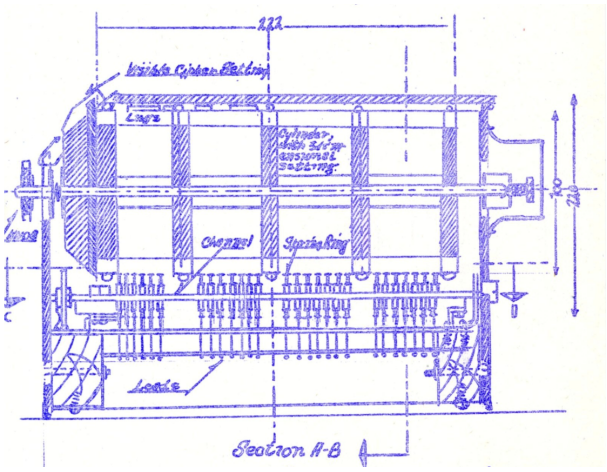


Figure 2: A cross-section of one of the cylinders of the skip box. The spring loaded switches are seen below the cylinder. Source: DF-114.

4.2 Distributor

The distributor (German: *Verteiler*) is connected to the skip box with 120 cables, one for each electrical switch in the skip box. The distributor consists of five wheels, one for each cylinder of the skip box. All cables from skip box cylinder I is connected to wheel I of the distributor and so on. In terms of signal names that means that $I^{\alpha-\delta}$ is connected between cylinder I of the skip box and

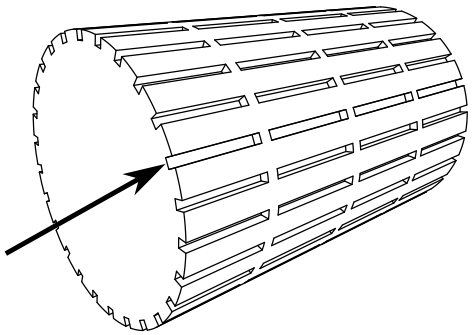


Figure 3: Modern rendition of a DF-114 cylinder. The arrow points to one of the 26 slots which in turn is divided into four sections.

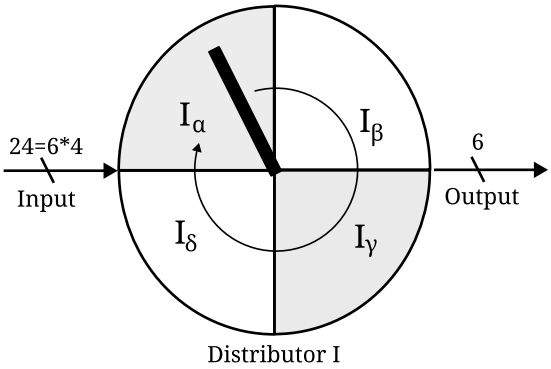


Figure 4: Illustration of the first distributor. The input consists of four groups of six signals from the skip box, denoted α to δ. The output is one of the four groups, but which one changes for every step of the arm.

wheel I of the distributor.

Each distributor wheel has an arm (German: *Zeiger*) that connects the incoming α, β, γ or δ signals to the output of the distributor. That is, each wheel has 4×6 input signals and outputs 6 signals at any time. The wheel or arm can step forward, making the arm connect the next group of signals, (α to δ). In DF-114, wheel I is labeled "low speed", and wheel V is labeled "high speed". This suggests that the five wheels of the distributor moves much like the odometer of a car: once wheel V has iterated through its four inputs, wheel IV will move one step forward. This leads to the distributor, in sequence, outputting all $4^5 = 1024$ different combinations of $I^{\alpha-\delta}$ through $V^{\alpha-\delta}$.

In DF-114 the distributor is described as having "a stepped advance of the five wheels corresponding to the five cylinders" and notes that "the purpose of the distributor is to scan all [4^5] combinations" (TICOM, 1948b).

The total number signals output from the distributor at any time is $5 \times 6 = 30$.

4.3 Switching device

The switching device (German: *Schaltapparat*) consists of six test circuits (German: *Diskussionskreise*). Each test circuit correspond to one of the six teeth of a metal plate and therefore also one of the rotors of the M-209: 17, 19, 21, 23, 25 or 26.

The six test circuits are connected in series, and only if all six circuits produce a positive result is the objective of the whole device is met and the machine stops.

As input, a test circuit has five of the signals coming from the distributor. For example, test circuit 17 is given input from I_{17}, II_{17}, III_{17}, IV_{17} and V_{17}. In addition to this input, a test circuit has a test voltage which comes from the previous test circuit in the series connection or in the case of the first test circuit, a 90 Volt signal. The test voltage is passed through a binary tree implemented with electro-mechanical relays. The relays implement what is nowadays called a demultiplexer, with the output selected by the five control signals from the distributor. A demultiplexer connects one input to one out of several outputs, in our case 32. Since there are five control signals, there are $2^5 = 32$ possible outputs, and only one of these 32 outputs will be connected to the incoming test voltage.

Each of the 32 outputs of the relay circuit is connected to a lamp socket. In a lamp socket there can either be a lamp or not, this is one of the parameters that the operator controls depending on the current job the machine is working on.

After the 32 lamp sockets, the signals are merged into one, and connected to the next test circuit or, in the case of the last test circuit, to a circuit that will halt the machine. When halted the operator can see exactly which six light-bulbs that are shining and then resume the run with the press of a button. Figure 5 illustrates one of the test circuits.

All in all, the switching device has a 90 Volt test voltage as input, $5 \times 6 = 30$ signals coming from the distributor, $32 \times 6 = 192$ lamp sockets and one output for the test voltage.

5 Detailed theory of operation

The previous section is a description and interpretation of what is actually shown and written in the

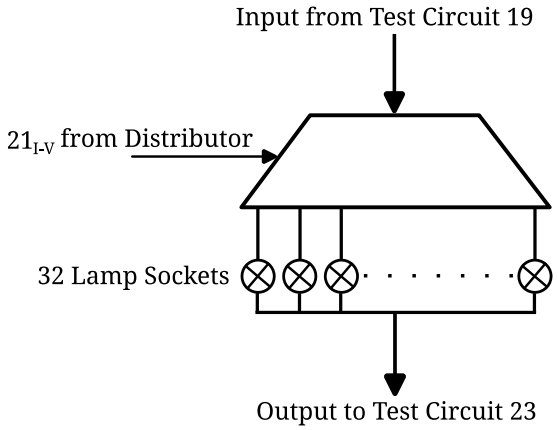


Figure 5: Illustration of one of the six test circuits of the switching device: in this case test cirucit 21 is shown. The five signals 21_{I-V} selects to which of the 32 lamp sockets the input shall be connected. The presence or absence of a lamp in that socket determines if the test signal is passed on to the next test circuit or not.

DF-114 document. What is *not* described in the document is with what data the device should be configured, and what the output of the device is. The following is a theory of one way the DF-114 device could have been used.

As mentioned in section 3.1, the theory investigated here is whether the DF-114 device would be useful as a tool to speed up the process of finding cribs.

The theory is implemented by the following steps:

1. Populate the five cylinders with teethed metal plates according to the known pin and lug setting of the M-209. See section 5.1 below for a detailed description.

2. Put lamps in lamp some of the sockets of the switching device according to the pin settings of the M-209. See section 5.2 below for a detailed description.

3. Assume a five letter cleartext somewhere in the message.

4. Calculate the number of skips the M-209 would have to produce in order for the cleartext letter to be enciphered to the ciphertext letter for each of the five letters.

5. Set the five cylinders so that for each cylinder the right skip is active, according to the calculation done with the crib.

6. Start the device.

When the device is running, the distributor will test all 1024 different combinations of teethed metal plates for the given five skips. In essence, this creates the up to 1024 ways of creating the given ciphertext from the given cleartext.

The switching device will for each of the 1024 combinations test whether the resulting pin sequences are present on all six wheels of the M-209. If that is the case, the machine stops and the operator should investigate this specific setting further. The setting to be investigated can be deduced from the six light-bulbs that are shining (one light bulb from each of the six test circuits of the switching device).

5.1 Cylinder teethed metal plates

The six engaging pins of the M-209 wheels can assume $2^6 = 64$ different combinations. Each of the 64 combinations gets converted to a number of skips in the range $0,\ldots,27$, depending on the lug settings of the M-209. Note that a skip of 26 is equivalent with a skip of 0, and 27 is equivalent with 1, so in practice there are 26 distinct skips: one for each letter of the alphabet. This leads to the fact that some of the 64 pin combinations must result in the same number of skips, since there are only 26 different results. So, for some of the pin combinations the M-209 would encrypt a given cleartext letter to the same ciphertext letter (Pokorn, 1945).

This is the reason why it is possible to have four teethed metal plates in each cylinder slot of the DF-114 device. Each of the 26 slots in the cylinder is populated with the metal plates that generate the same number of skips. If there are more than four metal plates that would generate the same number of skips, the DF-114 document mentions that there is an option to add a fifth skip possibility using "the lead to the switch device". Exactly how this fifth skip possibility would work is not very well described.

Of course, if there is a case where more than five combinations produce the same number of skips this will then potentially lead to missed solutions.

5.2 Lamps of the switching device

As mentioned in section 4.3 each of the six test circuits of the switching device gets as input five signals from the distributor. The theory is that these five signals correspond to a presumed sequence of

Wheel size	Distinct sequences
17	13.67
19	14.82
21	15.90
23	16.90
25	17.86
26	18.29

Table 1: The average number of distinct pin sequences of length five with different wheel sizes. Result of a computer simulation of 1,000,000 samples per wheel size.

five adjacent pins on one of the wheels of the M-209. Each of the $2^5 = 32$ lamp sockets that are part of the test circuit corresponds to one specific five pin sequence. Let - denote an inactive pin and + denote an active pin: then the pin sequences of the lamp sockets range from `-----`, `----+`, `---+-`, `---++`, and so on, up to `++++-`, `+++++`.

On a M-209 wheel where the pins have been randomly selected there is an expected value of the number of distinct pin sequences of length five over the whole of the wheel. The average number of distinct sequences depends on the length of the wheel and has been determined by computer simulation. The results are visible in table 1.

Given the information in table 1 and the fact that the six test circuits of the switching device are connected in series, it is possible to calculate the probability that a totally random pin sequence would pass through all six test circuits.

$$P = \frac{13.67}{32} \times \frac{14.82}{32} \times \frac{15.90}{32} \times \frac{16.90}{32} \times \frac{17.86}{32} \times \frac{18.29}{32}$$

$$P = 0.0166$$

The probability is thus on average 1.66%.

6 Computer simulation

To test the feasibility of the theory described in section 5 a computer simulation was made of the DF-114 device. The simulation of the DF-114 device was written as a Python program which takes as input the known pin and lug setting (the relative setting), a five letter cleartext word to be tested and the corresponding five letters of ciphertext.

The pin settings for each wheel are used to calculate all distinct pin sequences of length five. This corresponds to which sockets on the test circuits an operator would put light-bulbs into. This data is stored per wheel in an associative list.

Furthermore, the pin and lug settings are used to calculate the number of skips that each of the $2^6 = 64$ wheel pin combinations result in. This corresponds to putting the teethed metal plates in the correct place in the cylinders. Note that while the DF-114 device cylinders are described as having four places for metal plates ($\alpha - \delta$) per skip, the simulation does not have this limitation. For an M-209 key there is a possibility that there will be more than four pin combinations that result in the same number of skips. If this happens then the DF-114 device might not find a solution, but the computer simulation will.

The five letter suggested cleartext letters and the corresponding ciphertext is used to calculate how many skips that must be carried out by the M-209 for each letter for the theory to hold true. This corresponds to what the five cylinders of the skip box would be set to prior to starting the DF-114 device; which of the 26 slots of the cylinders that would be engaging with the electrical switches of the skip box.

This concludes the setting-up part of the simulation and the simulation of the DF-114 device can now be started.

The simulation of a running DF-114 device is done by iterating through the set of skips resulting from the Cartesian product of the active skips from the five cylinders. With each iteration the six test circuits are simulated by looking up whether the produced pin sequences are part of the six corresponding wheels. If all six tests are passed the simulation prints which skips were used to produce the results.

6.1 Example simulation

Assume that there is an M-209 enciphered message where the ciphertext `XEWIR` is believed to correspond to the cleartext `THEZM`[2].

The pin and lug settings are known, and are shown for reference in table 2 and table 3.

The number of skips needed to encipher each of the letters are shown in table 4.

Let S_x denote the set of M-209 wheel pin combinations that produce a skip of length x. With the

[2]The letter Z was used in place of a space between words since the M-209 did not have a specific space character. The plain text message would thus have been “THE M” in this case.

Wheel 17	++-+---+--+--++-+
Wheel 19	-+-+++-++---++-+--+
Wheel 21	--+-++-++---++-+--+++
Wheel 23	++----++-+-+++---++++-+
Wheel 25	+--++-+--+++--+--++-+-+--
Wheel 26	++-+---++-+-++----++-++---

Table 2: Pin settings. A dash denotes an inactive pin, a plus denotes an active pin.

3-0	0-6	1-6	1-5
4-5	0-4	0-4	0-4
0-4	2-0	2-0	2-0
2-0	2-0	1-0	2-0
2-0	2-0	2-0	2-5
2-5	0-5	0-5	0-5
0-5	0-5	0-5	

Table 3: Lug settings on the 27 drum bars of the M-209.

T → X: 17 skips
H → E: 12 skips
E → W: 1 skip
Z → I: 8 skips
M → R: 4 skips

Table 4: Number of skips needed for each letter of the suggested cleartext of the example message.

key used in the example, the sets of interest are shown in table 5.

S_{17}:	+-++-,+--++,-+++--,--++++
S_{12}:	+---+-,-++---,----++
S_1:	++++++,--+---
S_8:	+--+--,--++-+
S_4:	+-+---,+----+

Table 5: The different wheel pin combinations that produce a certain number of skips.

In this case the number of skip combinations that will be tested is the Cartesian product of the set sizes:

$$|S_{17}| \times |S_{12}| \times |S_1| \times |S_8| \times |S_4| = 96$$

Each skip combination can be seen as a matrix of 6×5 elements. As an example, consider the skip combination consisting of the first member of each of the sets in table 5. This is shown in table 6.

The rows of the matrix in table 6 are M-209 wheel pin combinations that produce the required

+-++-+
+---+-
++++++
+--+--
+-+---

Table 6: One of the 96 skip combinations that are tested.

number of skips. Each column can then be seen as a sequence of active and inactive pins that must be present somewhere on that specific wheel. For example, for the first wheel (size 17), there must be a pin sequence of +++++ somewhere in its pin sequence.

This is looked up by the simulator in the associative per-wheel-list that was calculated earlier. If all six columns exists on their respective wheel the result is positive and the skip combinations are printed. It means that it is possible to encipher the cleartext to the ciphertext with the M-209 key settings given as input.

As seen in table 2 there is no sequence of +++++ anywhere on wheel 17. This particular combination is thus not possible and will not be printed.

7 Results

In order to test the feasibility of this theory of operation the computer simulation was run with different sets of input data: 94 different M-209 keys and three different plaintexts were used for a total of 282 simulations.

All simulations finds the correct solution but also a number of false solutions where the pin patterns happen to exist on the six wheels but it is not the correct solution. In this case the ciphertext would decipher to the correct plaintext, but the rest of the message would not decipher correctly.

From a feasibility perspective it is of interest to see how many false solutions that is expected since each solution had to be investigated further, perhaps on a real M-209 converter. In a perfect situation either the correct solution should be found, or none at all. In practice, the fewer false solutions the better.

Figure 6 shows a histogram of the number of solutions for the simulations. The median number of solutions is 6, and the mean is 8.6.

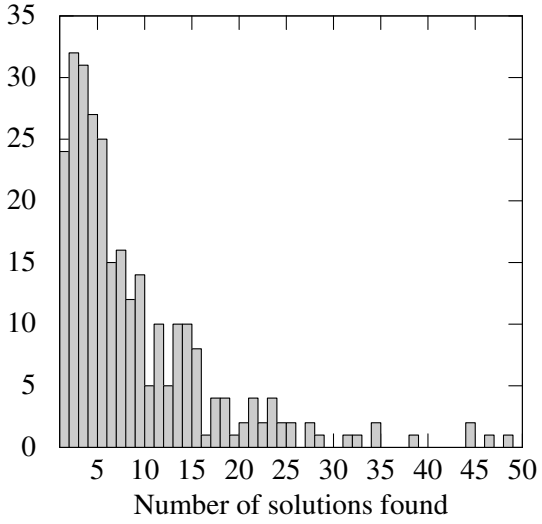


Figure 6: Histogram showing the number of solutions for the 282 simulations.

8 Conclusion

Not much is known of the DF-114 device and neither the device itself nor parts of it have never been found. There exists some evidence which suggests that it has existed and that the use was to speed up the process of finding the absolute setting given the relative setting (Ekhall and Schmeh, 2023). It is not impossible that the DF-114 device is capable of doing so, but it is still not known how the device would be operated in that case.

The theory presented in section 5 is different: it suggests that the DF-114 device can be used to test for cribs given the relative setting.

While the simulation result shows that it is possible to use the DF-114 device in this manner, it also shows that it would sometimes lead to a large number of false solutions which would need to be investigated further. The median number of solutions is six, which perhaps is not terribly bad, but the maximum number of solutions found is 49 and that would lead to considerable manual work. The possibility of a large number of false solutions can be interpreted in two ways:

1. The theory presented in this paper is correct and the DF-114 device would generate a large number of false solutions. The device would thus not have been easy to use and this may explain why there is not a lot of historical records mentioning it. The historical footprint of the DF-114 device is very small, basically limited to the DF-114 document itself and an interview referenced in (Ekhall and Schmeh, 2023).

2. The theory is wrong and there is another, more feasible, way which the DF-114 device would have been used. This is clearly a possible situation and is an area for further research.

Acknowledgments

The author would like to thank the reviewers for their invaluable feedback, which significantly improved the quality of this paper.

References

Magnus Ekhall and Klaus Schmeh. 2023. A WW2 device for breaking the M-209 encryption machine. In *Proceedings of the 6th International Conference on Historical Cryptology HistoCrypt 2023*, HistoCrypt 2023. Linköping University Electronic Press.

W Friedman, B Miller, K Perrin, A Sinkov, F Austin, W Pettengill, and M Lane. 1950. Special conference on M-209 security, minutes of meeting. https://www.nsa.gov/Portals/75/documents/news-features/declassified-documents/friedman-documents/patent-equipment/FOLDER_371/41755249079440.pdf. Ref ID: A66657, Folder 371, NSA, William F. Friedman Collection of Official Papers.

David Kahn. 1996. *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*. Scribner.

George Lasry, Nils Kopal, and Arno Wacker. 2016. Ciphertext-only cryptanalysis of Hagelin M-209 pins and lugs. *Cryptologia*, 40(2):141–176.

Alfred Pokorn. 1945. Report by Alfred Pokorn, of OKH/CHI, on M. 209. https://catalog.archives.gov/id/23890261. TICOM/I-175. NARA, College Park, NAID: 23890261.

TICOM. 1948a. Determination of the absolute setting of the AM 1 (M-209) by using two messages with different indicators. https://catalog.archives.gov/id/26466553. Document T-2795, TICOM/DF-105. NARA, College Park, NAID: 26466553.

TICOM. 1948b. German cryptanalytic device for solution of M-209 traffic. https://catalog.archives.gov/id/23889821. Document 2785, TICOM/DF-114. NARA, College Park, NAID: 23889821.

TICOM. 1948c. Report on the solution of messages in depth of the American chipher device

M-209. https://catalog.archives.gov/id/23889823. Document 2794, TICOM/DF-120. NARA, College Park, NAID: 23889823.

Artificial neural network for hoax cryptogram identification

Floe Foxon
University of Leeds
United Kingdom
floefoxon@protonmail.com

Abstract

Numerous putative cryptograms remain unsolved. Some, including the Dorabella cryptogram, have been suggested as hoaxes, i.e., some sort of gibberish with no meaningful underlying plaintext. The statistical properties of a putative cryptogram may be modelled to determine whether the cryptogram groups more closely with real or with randomly generated plaintext. Ten thousand plaintexts from an English-language corpus, and ten thousand (pseudo-)randomly generated English-alphabet gibberish texts were studied through their statistical properties, including the alphabet length; the frequency, separation, and entropy of n-grams; the index of coincidence; Zipf's law, and mean associated contact counts. An artificial neural network (deep learning) model was fitted to these data, with a cross-validated mean accuracy of 99.8% (standard deviation: 0.1%). This model correctly predicted that arbitrary, out-of-sample simple substitution ciphers represented meaningful English plaintext (as opposed to gibberish) with probabilities close to 1; correctly predicted that arbitrary, out-of-sample gibberish texts were gibberish (as opposed to simple substitution ciphers) with probabilities close to 1; and assigned a probability of meaningful English plaintext of 0.9996 to the Dorabella cryptogram.

1 Introduction

Lists of dozens of putative unsolved cryptograms have been published, such as the 'Top 50 Unsolved Encrypted Messages' by Klaus Schmeh (2023) and 'Famous Unsolved Codes and Ciphers' by Elonka Dunin (2023), and even these are not exhaustive. Some of these putative cryptograms have remained unsolved for many decades, such as the first and third Beale ciphers, published in 1885; the Voynich manuscript, which drew modern attention in 1912; and the Dorabella cryptogram, published in 1937.

It seems probable (though the author will not prove) that the longer a putative cryptogram goes undeciphered, the more likely it is to be identified as a hoax; that there is no cipher to be solved at all, and that the 'cryptogram' was only a fake, perhaps designed to attract media attention or sell merchandise. Indeed, the three putative cryptograms mentioned above have been described as 'bamboozlement' (Kruh, 1982), 'gibberish' (Gaskell and Bowern, 2022), and a 'full-fat hoax' (Pelling, 2019).

However, proving a negative is challenging (and often times practically impossible). Cryptanalysts must be wary of the example of Z340, a cryptogram that went unsolved for 51 years and was considered to be a possible pseudo-cipher (Juzek, 2019) before being solved completely in 2020 by Blake et al. (2021).

Thus, there is a need in classical cryptology to develop more sophisticated ways of distinguishing between real but unsolved cryptograms and actual hoaxes. In data science and statistics, a popular and effective way of categorizing data is with machine learning classification algorithms. In essence, this involves taking records (which may represent anything from animals to cryptograms) and applying an algorithm to these data which identifies, for each record, which category the record most likely belongs to. For example, the properties of vocalisations made by particular species of bird may be used to identify the species of an unknown bird (Qian et al., 2015).

These methods may be extended to cryptology by studing the linguistic and other statistical prop-

erties of putative cryptograms. With an appropriately trained model, a putative cryptograms of unknown status (real or hoax) may be identified rigorously and accurately (at least in theory).

The Dorabella cryptogram is of particular interest to the application of machine learning methods in this study due to its simplicity (having only a short alphabet) and brevity (having only 87 characters). The Dorabella cryptogram has been comprehensively described in other works (e.g. Bauer (2017)). Briefly, it is apparently a simple or monoalphabetic substitution cipher (MASC) prepared by Edward Elgar in an 1897 letter to an acquaintance named Dora Penny. It is assumed to be a MASC because the same symbols appear in definite MASCs in other cryptologic writings by Elgar. Despite its apparently simple encryption method, no solution has been generally accepted by the cryptologic community, hence its possible identification as a hoax (Elgar was allegedly known to be cruel), or as a different type of cipher. Cryptanalytic methods designed for MASCs have yet to yield a solution to the Dorabella cryptogram, but have identified interesting properties such as possible vowels (Schmeh, 2018).

The aim of the present study is to build an accurate machine learning model using statistical properties of cryptograms designed specifically for the Dorabella cryptogram. The model developed is used to determine whether the Dorabella cryptogram is statistically more likely to be real, or more likely to represent a hoax.

2 Methods

2.1 Data

To create a training and testing set of real (i.e., meaningful) English-language plaintexts, a collection of Wikipedia articles totalling 1.8 million English words were used as an English-language corpus (Davies, 2015). Ten thousand successive blocks of text were taken from the corpus, each 87 characters in length (the same length as the Dorabella cryptogram; all lowercase). This provided ten thousand real English plaintexts. An example of one of the 10,000 87-character texts from the English-language corpus used in the study is as follows:
turnwasinvestedwiththeduchyforhimselfand-
hisheirsalbertsruleinprussiawasfairlyprosperous
(from the Wikipedia article for Albert, Duke of Prussia).

To create a training and testing set of fake (i.e., gibberish) English-alphabet plaintexts, characters from the English alphabet (all lowercase) were pseudo-randomly selected (with replacement) to create a string of gibberish. Ten thousand such gibberish texts were generated, each 87 characters in length as above. An example of one of the 10,000 87-character gibberish texts used in this study is as follows:
idqbnjbnalbclvdfqypkzbwhddivepqjobbfrliplhusg-
onwshzdktdmbrtowispplvymrbsqzvhkramedbtdgk.

Since there were ten thousand texts in each of the real and fake sets, the data were evenly balanced, enabling a fair learning phase in the model.

The values of linguistic and other statistical properties were calculated for each text. These properties were as follows.

- Alphabet length: The unigram alphabet length is the number of unique single characters (unigrams) in the text. E.g., the text ‘aabc’ has a unigram alphabet length of 3 (the alphabet is the set { ‘a’, ‘b’, ‘c’ }). The bigram alphabet length is the same as above but for unique pairs of characters (bigrams). E.g., the text ‘aabc’ has a bigram alphabet length of 3 ({ ‘aa’, ‘ab’, ‘bc’ }). Finally, the trigram alphabet length is the same as above but for unique trios of characters (trigrams). E.g., the text ‘aabc’ has a trigram alphabet length of 2 ({ ‘aab’, ‘abc’ }).

- Average frequency: The average unigram frequency is the average number of occurrences of each unigram in the text. E.g., the text ‘aabc’ has an average unigram frequency of $1.\dot{3}$ from $\frac{2+1+1}{3}$. The average bigram/trigram frequency is the same as above but for bigrams/trigrams.

- Average distance: The average unigram distance is the average number of ‘steps’ between repeated occurrences of unigrams in the text. E.g., the text ‘abba’ has an average unigram distance of 2 (from $\frac{3+1}{2}$). The average bigram/trigram distance is the same as above but for bigrams/trigrams.

- Entropy: the first-order Shannon character entropy or unigram entropy is given by $H_1 = -\sum_{i=1}^{n} p_i \log_2 p_i$, where p_i is the probability of occurrence of each unigram (i.e.,

the number of occurrences of the unigram divided by the total number of occurrences of all unigrams). The bigram/trigram entropies are the same as above but for bigrams/trigrams.

- Index of coincidence: The index of coincidence (IC) measures the evenness of the distribution of characters in the text (greater IC means greater unevenness) and is given by $IC = \frac{n}{N(N-1)} \sum_{i=1}^{n} n_i(n_i - 1)$, where N is the text length, n is the unigram alphabet length, and n_i is the number of occurrences of the i^{th} character in the unigram alphabet. E.g., the text 'abbba' has $IC = 0.8$ from $\frac{2}{5(5-1)}[2(2-1)+3(3-1)]$.
- Zipf's exponent: The exponent α in the equation $f = \frac{K}{r^{\alpha}}$ is obtained by regressing the unigrams' frequencies of occurrence f on their ranks r, where the most frequently-occuring unigram has rank $r = 1$, the second most frequently-occuring unigram has $r = 2$, etc. Zipf's exponent measures the gradient or slope of $\log(f)$ against $\log(r)$. Natural languages have $\alpha \approx 1$.
- Average mean associated contact counts: Burleson (1989) defines the variety of contact count (VCC) for a given unigram as the number of unique unigrams adjacent to (i.e., immediately next to or contacting) the root unigram. E.g., the unigram 'a' in the text 'cab' has $VCC = 2$, while 'c' and 'b' both have $VCC = 1$. Burleson (1989) then defines the mean associated contact count (MACC) for a given unigram as the sum of the VCC values for each adjacent unigram divided by the VCC of the root unigram. E.g., the unigram 'a' in the text 'cab' has $MACC = 1$ from $\frac{1+1}{2}$, while 'c' and 'b' both have $MACC = 2$. To provide a single statistic for the entire text, the author defines the average MACC (AMACC) as the sum of the MACC values for each unigram in the alphabet of the text divided by the alphabet length. E.g., the text 'cab' has $AMACC = 1.\dot{6}$ from $\frac{1+2+2}{3}$.

2.2 Model

To classify texts, an artifical neural network was implemented in Python with the TensorFlow machine learning software library and Keras deep learning API. In this model, the target variable was the binary category of text (real or fake); the fit data were the statistical properties of text (i.e., unigram, bigram, and trigram alphabet lengths, average frequencies, average distances, and entropies; as well as the IC, Zipf's exponent, and AMACC). Briefly, a simple two-layer sequential model was implemented. The input layer contained 15 nodes (one for each input feature) and used the rectified linear unit activation function. The output layer used the sigmoid activation function. Binary cross entropy was used as the loss function for binary (real/fake) classification. The Adam algorithm was used as the optimizer for efficiency.

$N = 20,000$ Dorabella-like texts ($10,000$ real and $10,000$ fake, as described above) were used in training and testing the model. 5-fold cross-validation was used to evaluate the model.

No general, formal equation exists to estimate the sample size required for accurate and precise classification with neural networks. In the context of quantitative linguistics, Kubáček (1994) suggests that a sample size in the thousands may be necessary for a representative count of linguistic entities. In this study, the sample size (number of texts) was in the tens of thousands; three orders of magnitude greater than the number of fit variables. Thus, sample size is unlikely to present an issue (classification models may still be accurate with few data as long as the data are high quality).

All analyses were conducted in Python version 3.8.16 with the packages Numpy version 1.21.5, Pandas version 1.5.2, Scipy version 1.7.3, Uncertainties version 3.1.6, Natural Language Toolkit (NLTK) version 3.7, Scikit-learn version 1.0.2, and TensorFlow version 2.10.0.

3 Results

Linguistic and other statistics for the real texts, fake texts, and Dorabella cryptogram are shown in Table 1. Values for the Dorabella cryptogram are closer to the real texts except in the trigram statistics and AMACC.

A satisfactory model was obtained. Across the five folds, the average model accuracy and standard deviation were 99.8% (0.1%).

To further test the out-of-sample performance of the model, a real 87-character MASC was created from the plaintext of Ellie's essay in the theatrical play *The Whale* and the JavaScript 'Simple Substitution Cipher' generator available from Practical

Cryptography (Lyons, 2023). The model correctly predicted that this real MASC represented real English text with a probability of 0.99999 (giving a probability of fake text of just 0.00001). Likewise, the model correctly predicted that an out-of-sample 87-character random string generated with the website `random.org` was random text with a probability of 0.9999998 (giving a probability of real text of just 0.0000002).

Finally, the model was applied to the Dorabella cryptogram. The model classified the Dorabella cryptogram as real English text with a probability of 0.9996.

Statistic	Corpus	DB	Random
Alphabet length			
Unigram	19.7 (1.4)	20	25.1 (0.9)
Bigram	64.9 (5.5)	69	80.9 (2.1)
Trigram	78.5 (5.6)	83	84.8 (0.5)
Avg. frequency			
Unigram	4.4 (0.3)	4.4	3.5 (0.1)
Bigram	1.3 (0.1)	1.2	1.1 (0.0)
Trigram	1.1 (0.1)	1.0	1.0 (0.0)
Avg. distance			
Unigram	14.4 (2.9)	14.8	17.8 (2.6)
Bigram	7.1 (2.7)	4.7	1.8 (1.0)
Trigram	2.5 (2.6)	0.4	0.1 (0.2)
Entropy			
Unigram	4.0 (0.1)	4.0	4.5 (0.1)
Bigram	5.9 (0.2)	6.0	6.3 (0.1)
Trigram	6.2 (0.1)	6.4	6.4 (0.0)
IC	1.3 (0.1)	1.2	1.0 (0.1)
Zipf's exp.	0.6 (0.1)	0.5	0.4 (0.1)
AMACC	7.7 (0.7)	6.9	7.0 (0.4)

Table 1: Statistics for 10,000 87-character samples from an English-language corpus (Corpus), the Dorabella cryptogram (DB), and 10,000 randomly-generated 87-character English strings (Random). For the Corpus and Random results, means are presented with standard deviations.

4 Discussion

This study demonstrates the use of deep learning methods to classify putative cryptograms probabilistically. The results of this study preliminarily suggest that the Dorabella cryptogram is perhaps more likely to represent real underlying English plaintext than it is to represent purely random gibberish. This does not mean that the Dorabella cryptogram is necessarily a real cryptogram, only that it is less likely to be random gibberish. Of course, other possibilities exist, including the Dorabella cryptogram as non-random gibberish (i.e., gibberish that is designed to look more like real text than purely random text), as a non-MASC cipher, or as encrypted shorthand. Future studies can explore these possibilities by expanding the binary classification model of the present study to multi-class models with data representing other possible classifications.

The present study is not the first to apply classification or clustering methods to classical cryptography. For example, the Neural Cipher Identifier (NCID) by Leierzopf et al. (2021) uses an ensemble neural network classifier for 55 standardized classical cipher types. More relevantly, Juzek (2019) clustered true ciphers and pseudo-ciphers using support-vector machines on entropy. The present study expands upon the NCID by including random or gibberish text as a possible classification (which the NCID does not), and expands upon the Juzek analysis by including other statistical properties besides entropy. Yet more statistical properties may be included in future works, as well as other classification types.

The results of the present study comport with other recent analyses of the Dorabella cryptogram. Schmeh (2018) concluded that "frequency count and the contact counts of the Dorabella Cryptogram are consistent with the English language," and Hauer et al. (2021) reported "evidence for English as the language of the cipher" from n-gram language models, and that "the occurrence of several pairs of mirrored symbols is unlikely to be due to chance, suggesting that Dorabella is not a hoax." Still, no convincing MASC solution to the Dorabella cryptogram has yet been found, even using modern, state-of-the-art algorithms with high decipherment rates (Wase, 2023).

It is not certain whether these findings are accurate, and the author emphasises the need for larger multi-class models to better understand the nature of the Dorabella cryptogram and other putative cryptograms. One limitation in applying deep learning methods for classification is the necessity for the training dataset to reflect the underlying plaintext of the putative cryptogram; if the plaintext is written in French or Old English but the training dataset consists of Modern English text, accurate classification is not possible. Thus, the corpus must be appropriate for the use case.

References

Craig Bauer. 2017. *Unsolved! The history and mystery of the world's greatest ciphers from Ancient Egypt to online secret societies*. Princeton University Press.

Sam Blake. 2021. The solution of the zodiac killer's 340-character cipher. https://blog.wolfram.com/2021/03/24/the-solution-of-the-zodiac-killers-340-character-cipher/.

Donald R. Burleson. 1989. The "macc"–a statistical technique for vowel isolation. *The Cryptogram*, March–April:4–6.

Mark Davies. 2015. The wikipedia corpus. https://www.english-corpora.org/wiki/.

Elonka Dunin. 2023. Famous unsolved codes and ciphers. https://web.archive.org/web/20231025073911/https://elonka.com/UnsolvedCodes.html.

Daniel E Gaskell and Claire L Bowern. 2022. Gibberish after all? voynichese is statistically similar to human-produced samples of meaningless text. In *Proceedings of the 1st International Conference on the Voynich Manuscript*.

Bradley Hauer, Colin Choi, Anirudh Sundar, Abram Hindle, Scott Smallwood, and Grzegorz Kondrak. 2021. Experimental analysis of the dorabella cipher with statistical language models. *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt 2021, June 20-22, 2022, University of Amsterdam, The Netherlands*, pages 70–79.

Tom S. Juzek. 2019. Using the entropy of n-grams to evaluate the authenticity of substitution ciphers and z340 in particular. *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt 2019, June 23-26, 2019, Mons, Belgium*, 158:117–125.

Louis Kruh. 1982. A basic probe of the beale cipher as a bamboozlement. *Cryptologia*, 6(4):378–382.

Lubomír Kubáček. 1994. Confidence limits for proportions of linguistic entities. *Journal of Quantitative Linguistics*, 1(1):56–61.

E. Leierzopf, N. Kopal, B. Esslinger, H. Lampesberger, and E. Hermann. 2021. A massive machine-learning approach for classical cipher type detection using feature engineering. *Proceedings of the 4th International Conference on Historical Cryptology HistoCrypt 2021*.

James Lyons. 2023. Simple substitution cipher. http://practicalcryptography.com/ciphers/simple-substitution-cipher/.

Nick Pelling. 2019. Dorabella cipher: timeline, texts, and keith massey... https://web.archive.org/web/20230528045150/https://ciphermysteries.com/2019/11/15/dorabella-cipher-timeline-texts-and-keith-massey.

Kun Qian, Zixing Zhang, Fabien Ringeval, and Björn Schuller. 2015. Bird sounds classification by large scale acoustic features and extreme learning machine. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1317–1321. IEEE.

Klaus Schmeh. 2018. Examining the dorabella cipher with three lesser-known cryptanalysis methods. In *Proceedings of the 1st International Conference on Historical Cryptology, HistoCrypt*, pages 145–152.

Klaus Schmeh. 2023. The top 50 unsolved encrypted messages. https://web.archive.org/web/20230601011324/https://scienceblogs.de/klausis-krypto-kolumne/the-top-50-unsolved-encrypted-messages/.

Viktor Wase. 2023. Dorabella unmasced – the dorabella cipher is not an english or latin monoalphabetical substitution cipher. *Cryptologia*, pages 1–10.

How the Machines Were Assisted by Women

Elizabeth S.D. Fricker
History of Science
129 Dickinson Hall
Princeton University
efricker@princeton.edu

Abstract

The following paper analyzes nature of skill in the cryptographic work performed by female employees of the World War II British cryptography program at Bletchley Park. Using anecdotes from women who worked to decrypt the Lorenz cipher, I show the hidden skills which were integral to their work and to the war effort. This paper aims to reconstruct their experiences and deepen our understanding of what it took to operate decryption machines, especially focusing on the Colossus computer.

1 Introduction

Even at the height of World War II, a new recruit to Station X, the Government Code and Cypher School (GC&CS)'s highly classified site dedicated to breaking Axis military codes located in the Bletchley Park estate, would be forgiven for assuming that the Bletchley train station was the wrong stop.[1] Arrival was typically inauspicious. Recruits were told to take the train on a designated date and call a designated number on a telephone at the Bletchley Station kiosk on arrival. A rare few were told to show whoever was in the small hut by the front of the station an envelope they had been given, but by no means were they to hand the envelope over. From there, they would be given further directions (and further individuals to show their mysterious envelope to) before they entered Bletchley Park itself (Hill, 2004: 24-25). Almost none knew what they would be doing; some had not even worked out where they were.

One of the best-kept secrets of the Second World War, the cryptography program that ran out of Bletchley Park reportedly shortened the war "by not less that [sic] two years and probably by four years" (Hinsley, 1993). This number may be exaggeration—it is certainly impossible to verify—but its meaning is clear. Bletchley Park was a critical part of Great Britain's war effort. Bletchley Park was more successful in cracking German codes than Japanese, but the information they gained proved critical for major Allied operations such as the 1944 Normandy landings.A more thorough account of this can be found in Kenyon (2019). The popular narrative around Bletchley Park focuses on Alan Turing and the breaking of the German Enigma codes. In comparison, the breaking of the German Lorenz codes and use of the Colossus machines has received less attention.

The work performed by Colossus operators, all of whom were women, was often described in caretaking, maternalistic terms. This is not a coincidence. Seth Koven and Sonya Michel (1990) have demonstrated how in the realm of politics and welfare, women utilized their authority as mothers to gain leverage and respect, and Paula Baker (1984) has shown how women's activism, particularly regarding welfare, allowed for a conceptualization of women as political actors that did not violate feminine norms or the Victorian doctrine of "separate spheres." So, caretaking language rendered the operator's work legible within existing gender constructs, and provided them authority regarding the machines, while positioning them as less skilled than and subservient to the men they worked under. Indeed, managers and labor officials leveraged stereotypes about women's natural inclinations and skills to argue that they were particularly well suited for early computing work. For an example of this phenomenon which intersects with race and culture, see Nakamura (2014). Margot Lee Shetterly (2016) demonstrates the significance of race, which is reinforced by Elsa Barkley Brown's (1992) argument that "women" inher-

[1] The location was also called Special Duties X (SDX) or HMS Pembroke V, depending on which organization was sending the recruit. For example, HMS Pembroke V was exclusively used by the Navy.

ently cannot be a universalized category, blind to race and class.

The Colossus operators' statuses as predominantly white, middle-class, educated women was important, as it allowed them access to particular professional networks that working class women and women of color could not participate in. For example, Mundy (2017) shows how women were selected for personality traits and demographic qualities as well as their technical skills. However, in an attempt to understand their work beyond the lens of maternalism and femininity I have found it useful to conceptualize these women as technicians, as described by Steven Shapin (1989). Additionally, Jennifer S. Light (1999) shows how female computer operators' lack of authority in both social and technical settings led to their erasure from the historical record by those who did have power. Shapin (1989) argues that scientific technicians are invisible when the apparatus is working and that the technicians' work are always subject to the scientist's decisions because the scientist had authority, not the technician. Similarly, even as Colossus operators accrued technical and tacit knowledge about the Colossus, they were still described as "mindless" workers who followed the instructions of cryptographers. Historians of science and technology have developed a robust history of management practices that examines the importance of social networks in the production of knowledge and technology as well as the manipulation of power structures in society to reify the authority of an "elite" group. For further reference, see Ensmenger (2010), Jones (2016), Schaffer (1988), Kasson (1977), Nelson (1995), and Roediger and Esch (2012). Feminist historians have further explored management practices, showing how the physical separation of men and women in the workplace served to reinforce the gendered division of labor in white-collar offices. For more on this, see Strom (1989) and Fine (1990). Taking Bletchley Park as a case study provides an example where managers were more concerned with completing tasks than they were with social management. For instance, administrators typically cited office romances as a reason to keep men and women separated at work—heterosexual friendships, courtships, marriages, and affairs all happened at the Park without much comment, as long as the work continued effectively. Additionally, Bletchley Park was a military operation, so Bletchley Park workers had to contend with a set of management practices and cultural norms which often ran in direct contrast to scientific and educational norms which they were familiarized with from universities.

In recent years there has been a growing focus on the history of "hidden labor(ers)" within the history of computing, with scholars examining the work of "human computers," early programmers (who were often women), and institutional support workers who helped make the personal computer industry viable. These works encourage us to expand our understanding of what it actually means to work with computers, as well as challenge our presumptive vision of what a "computer programmer" looks like. For further reference, see Irani (2015a, 2015b). Nooney (2020), Roberts (2019), and Rankin (2018). In addition, Anne Secord (1994) provides an early model for considering alternate locations of scientific practice by examining a working class, "artisanal" science. The women at the center of this paper would not have considered themselves "programmers," as the term did not yet exist, but it is important to note that the work they performed required a number of skills central to modern computing. In addition, the postwar masculinization of computing as a profession has been well documented, most notably by Janet Abbate (2012) and Mar Hicks (2017, 2021). Similarly, while not specific to computing, Oldenziel (1999) describes the construction of engineering as a distinctly masculine enterprise—this provides critical context for Abbate's description of the pivot to "software engineering." As a result, we are less aware of what women in early computing actually *did*. Both Hicks and Abbate discuss the Colossus operators, but Hicks focuses more on the challenges of their postwar work, especially the pay inequality and bans against employing married women. Abbate, on the other hand, describes the process by which women's work was undervalued at Bletchley Park. My aim in this paper is to examine the work of the Colossus operators in detail, in order to understand the tacit knowledge and skill necessary to perform codebreaking with machines.This relates to Ruth Schwartz Cowan's (1983) argument that, contrary to popular narratives, the introduction of household technologies made domestic labor more difficult through changing standards of cleanliness and increased technical skill required from housewives. Similarly,

while the use of machines like the Bombe, Heath Robinson, and Colossus certainly shortened the time it took to crack ciphers, their use also mandated the mastering of an entirely new set of skills and new expectations regarding the pace of work. It is thus necessary to address the gendered dimensions of labor and the ways in which these influence the perceptions of women's work, but my goal is to look past these narratives in order to establish a sense of what it meant to operate a Colossus computer.

2 The Lorenz Cipher

In the summer of 1942 Helen Pollard (later Helen Currie) sits at a trestle table, head bent over her papers. She is accompanied by another young woman and a handful of "army boys," or young men of a lower rank. Occasionally, a man (either military or civilian) will deposit sheets of squared paper in front of her, with lines of dots and crosses drawn across them. She picks up her pencil and begins carefully adding the markings together. She has a chart to consult if she needs one, but by now, she can complete the calculations in her head. A few months later, Pollard and her companions moved from their room at the back of the Bletchley Park manor house to a hut on the property. Here, she meets the British Tunny machine for the first time.

Helen Pollard joined the Auxiliary Territorial Service (ATS) during the 1938 Munich crisis, which saw the nation of Czechoslovakia forced to cede territory to Nazi Germany through military pressure from Germany, Poland, and Hungary and diplomatic pressure from the United Kingdom and France. In 1942, she became a signals operator, or someone who operated teleprinters, switchboards, and cipher operators, earning a promotion to lance corporal for her efforts during her time at the signals school in Trowbridge.[2] That summer, she had her interview to work at Bletchley Park. Pollard only remembers two of the questions she was asked: "Would I like to work in the country? Could I keep a secret?" With her answers of "yes" and "I think so," respectively, her experience as a typist, training with military intelligence, and rank, she was assigned to the "Testery," the team headed by Major Ralph Tester. Solely dedicated to breaking the German Lorenz codes, the work of the Testery has received far less historical attention than Alan Turing and his Hut 8. However, the Lorenz ciphers demand attention, as it was reserved for communications amongst the highest levels of German High Command, including Adolf Hitler himself by early 1944.

Once assigned to operating the Tunny machines, Helen Pollard's day began by pressing the key numbers given to her by the cryptographers on her team to set the machine. Then, the rest of her shift was given over to typing. She would enter coded messages into the machine, and if her settings and typing was correct, out came a string of perfect German. In her own words, "It seemed like magic at first" (Currie, 2010: 266). Pollard could not read the German she produced, but she learned to tell when the machine was returning gibberish. More often than not this was an error in the interception of the message: reception was typically bad, meaning that there could be gaps in the message given to Pollard. She rarely made mistakes of her own. Incoherent messages required trial and error to get right. She would push the buttons again and start typing until she saw clear German. Finished messages went into an out-basket, off to who knows where. Pollard had only been told that her job was to type.

Despite assertions from Testery cryptographers that the work done by "backroom staff" like Helen Pollard was "very important, but often very boring and using very slow manual processes," Pollard enjoyed her work (Roberts, 2017: 83). As she describes it, "The work never got tedious. There was something about the atmosphere at Bletchley Park that generated an all-pervading excitement" (Currie, 2010: 266). She even describes the early stages of her time at Bletchley Park as idyllic, though this did not last. The war caught up with Helen Pollard in August 1943, when the airman she had married only ten weeks prior was killed. Sympathies abounded, but the only allowances she was given for her bereavement came from mathematician and fellow Testery employee Peter Hilton, who gave her "the best and clearest message he could find, which gave me no trouble to type. It was his way of bringing me comfort" (Currie, 2010: 266). Pollard does not identify this herself, but the episode reveals the fact that her work required skill. If it was truly mindless, then her grief would not have had any impact on her

[2] Women were drafted from volunteer organizations like the ATS to be signals operators in WWII to combat manpower shortages.

ability to type the messages and would not have needed accommodation.

Perhaps by the level of command of the men who used it, Lorenz was drastically more complicated than the Enigma codes. In terms of pure machinery, the Lorenz encrypting device had four times as many wheels as the Enigma machine, meaning that there were more possible encryption keys. As British Captain Raymond C. "Jerry" Roberts describes it, Lorenz "was the 'secret writer' of which Hitler dreamed. The Lorenz SZ40/42 was a wonderful machine: it should never, ever have been broken" (Roberts, 2017: 58). However, perhaps surprisingly, the Lorenz machines were easier to operate than Enigma machines. Enigma devices required someone to prepare the message, someone to convert it into Morse code, and someone to transmit it. Lorenz machines only needed one person, by comparison. An operator only needed to type the message on the machine's keyboard, and the machine did the rest (Roberts, 2017: 65). From this, it may seem counterintuitive that the Germans kept using Enigma once they had developed Lorenz, but every message sent was a chance for the Allies to intercept and break the code.

The Bombe machines, developed to crack Enigma codes, did not work on Lorenz. The British had no ways of even seeing a Lorenz SZ40/42, let alone acquiring one to take apart and analyze. How could they ever hope to break the code? It seems callous to attribute it to luck, but in truth, the Germans were careless during the relatively infrequent instances that they used Lorenz. Early on in the war, the wheel settings of the machines (which dictate the encryption) were not changed as frequently. It took until January 1944 for them to be changed daily, despite the fact that it was not tremendously difficult to change the wheel settings. By then, however, the British had already spent years attacking and solving Lorenz. On August 30, 1941, a Lorenz-encoded message of around 4,000 characters was sent from Athens to Vienna.[3] Then, a miracle: the message was not received correctly, and the receiving operator sent a plain-language request for it to be retransmitted. This piqued the interest of the British. They tuned in, to find that the message had been retransmitted using the same encryption settings. On its own, this would have meant nothing, but critically, the sender made a number of small alterations to the text. These edits, mainly switches to shorthand, allowed cryptanalyst Brigadier John Tiltman to decipher the two messages and determine their keystream.[4] To his dismay, though, Tiltman was unable to determine how the keystream was generated.

The problem was handed to young mathematician Bill Tutte. Captain Jerry Roberts (2017: 73) describes Tutte's process of breaking Lorenz as follows:

> Working with a sample of intercepted messages, Tutte, by intellect and intuition alone, deduced that the Lorenz code was generated by a machine with a series of rotors. He noticed that the first of these had forty-one teeth and then went on to work out all twelve rotors correctly. The Lorenz machine was broken – how the twelve wheels worked, their lengths and their functions – everything worked out.

And so, the British could crack Lorenz. The process was not streamlined in any way. Throughout the war, cryptographers were required to break Lorenz ciphers by hand. Machinery helped but could not solve the codes alone. Helen Pollard was right in the middle of this. Her new team, the Testery was the only team to work on Lorenz by hand. She joined surprisingly early, as well. In 1941, the team had eight or nine people, which jumped to a staff of 118 by the end of the war, "including twenty-four ATS girls, nine cryptanalysts and all kinds of support staff, all organized in three shifts working round the clock" (Roberts, 2017: 76).

The mismatch between Pollard's testimony of her work and Roberts' perception follows the pattern of mechanizing clerical work identified by historian Sharon Hartman Strom (1986: 64), namely that as machines were utilized more and more for bookkeeping and secretarial work, women were increasingly given those jobs as they were deemed less likely to protest the "deskilling" of that labor. This is not a perfect analogy, as women were overwhelmingly needed to occupy home front jobs in Britain during the Second World War, as most of the male labor force had

[3] Note that the average Enigma message ran around 300 characters.

[4] The string of characters used to encode the message.

been drafted. For one exceptional window, women were called to work jobs that would not previously have been available to them, travel, alone, into unknown spaces for the first times in their lives, and do everything they could to support the war effort. This is not to say that pre-war social mores vanished entirely. In fact, stereotypical views of women were used to justify giving them early computer work. The perception was that "Programming requires lots of patience, persistence and a capacity detail and those are traits that many girls have"(Gürer, 2002: 176). Indeed, the women of Bletchley Park were trusted to work eight-hour shifts in which the slightest inaccurate detail could have major ramifications, all with minimal breaks. Whether or not they held these traits innately (if at all), were trained into them, or simply rose to the occasion, the women did perform as needed, reinforcing beliefs about them. Moreover, "the expectations and ideals of peacetime society—particularly in regard to a patriarchal, heteronormative family—continued to influence how and where women were deployed and who was called up" (Hicks, 2017: 26). For instance, the majority of the women at Bletchley Park were young and unmarried. Some were married during their time at the Park, though in another moment of wartime exceptionalism, this did not inherently spell the end of their employment (Strom, 1986: 64). In general, as Mar Hicks (2017: 27) describes, "the government recognized and tried to mitigate the 'double shift' effect that accrued for married . . . in order to keep wartime factories running efficiently and keep factory floor accidents down," something it had not done during peacetime.

3 "Deskilled"

However, the perception that these forms of labor were "deskilled" did not necessarily translate to reality. Strom (1986: 64) describes the real-time decisions clerical workers had to make both to keep their machines running and to ensure that they were being used in the most effective manner. Helen Pollard certainly needed to apply forms of judgement to her work. She had to evaluate the messages that the machines returned and determine whether or not they were meaningful, even though she could not read German. In addition, the female Bletchley Park employees employed in traditionally "clerical" roles, such as registrars, were similarly required to make swift, critical decisions. They needed to keep track of all of the assignments and schedule them in such a way that prioritized more important codes and made the most effective use of the machines, which is a form of mathematical optimization (Hicks, 2017: 37). In a workforce where everyone's knowledge was limited for security purposes, these women had to juggle several moving parts, often with imperfect information at hand. For a similar analysis of the importance of administration in a mechanized office, see Strom (1986: 66). For example, after Wren and Colossus operator Dorothy Du Boisson's team expanded, she was reassigned to be a registrar, where she was responsible for assigning the paper tapes that both the Colossus and Heath Robinson machines used, and which held encoded messages and recording where each message was. She remembers that "We kept a register in which we recorded the date and identity of each tape. . . We knew exactly where every tape was, and how much machine time had been spent on it" (Copeland, et al., 2010: 163). So, Du Boisson needed to evaluate the priority of each code despite not knowing what they actually said, assign them in such a way that she maximized the information gained, and keep meticulous track of where everything was, even when the tapes had to pass through multiple people and stations.

Even the seemingly mundane act of mending the paper tapes was an essential job and high-skill job. Eleanor Ireland remembers using Bostik, a liquid adhesive, to put broken tapes back together. Even stretched tapes posed problems, though, as that could skew results. Operators had to keep a watchful eye on them, even as they spun at high speeds, in order to prevent errors. This was another hidden skill, and it came with the same pressure as every other task at Bletchley Park. Incorrect information could cost lives, and "Given that these young women had family and friends on the line of fire, the idiosyncracies of their machines were as frightening as they were frustrating" (Hicks, 2017: 33-34). Their ability to maintain composure and keep the machines running smoothly far exceeded the "patience" and "persistence" that supposedly made women such capable programmers.

The Tunny machines that Helen Pollard worked on were not the only technologies involved in deciphering the Lorenz codes. Built by engineer

Thomas "Tommy" Flowers, Member of the Most Excellent Order of the British Empire, and first made functional in December 1943, the Colossus machine was the world's first electronic computer.The narrative of ENIAC as the first electronic computer was promoted by Hungarian-American mathematician John von Neumann, who was not aware of the existence of Colossus. See B. Jack Copeland (2010: 101). A drastic improvement on its predecessors, the "Heath Robinson"[5] machines, Colossus quickened the statistical analysis necessary to break Lorenz (*Colossus*, 2012). In particular, the Colossus determined the patterns of the "cams," or the ridges on the wheels of the Lorenz encryption device which could be set to "active" or "inactive," and which would dictate the resulting code. The Colossus was also capable of determining the starting positions of the wheels of the Lorenz device. Those two pieces of information were essential for decrypting Lorenz encoded messages.

While developing the Colossus computer, Flowers faced numerous concerns over the glass valves that made up his digital switching techniques, as his doubters knew that the valves were prone to breaking. In response, Flowers posited that the valves were less likely to fail if they were kept working all the time (*Colossus*, 2012). His assertion was correct, and the valves proved sufficient for keeping up with the workload. In fact, labor at Bletchley Park ran twenty-four hours a day, seven days a week anyways, with workers taking one week with a nine a.m. to four p.m. shift, switching to a four p.m. to midnight shift the week after, then switching again to a midnight to nine a.m. shift. Working weeks were six days long, but many worked twelve days straight so that they could have a full weekend to visit home (Hill, 2004: 51). The Park was always active in some capacity, so there was no point in considering turning the Colossus off. In practice it was always running and always needed.

Much like the British Tunny machines, the Colossus was entirely operated by women. A full list of the Colossus operators can be found at "Computing Herstory - Computing History." Colossus computers were primarily operated by Max Newman's "Newmanry" team, which worked hand-in-hand with the Testery team on Lorenz. At the start of the war, the Newmanry team was comprised of one (male) cryptographer, two (male) engineers, and sixteen "Wrens" (the nickname for the Women's Royal Navy Service, or WRNS). By 1945, their numbers jumped to twenty-six cryptographers, twenty-eight engineers, and 273 Wrens (Copeland et al., 2010: 158). It should be noted that the cryptographers and engineers were still all men; despite the fact that some women received training to become engineers, they would have been classified as Wrens at Bletchley Park. So, the Newmanry and the Testery saw similar increases in staff, though the breakdowns of their respective personnel differed. The Testery team did not feature engineers and used ATS girls, not the more prestigious Wrens. Additionally, the majority of the Newmanry team was women, which was never true of the Testery.

Even if the majority of the Newmanry team were women, they still played second string to the men. Even the subset of the Wrens with university educations were never promoted to senior positions or made cryptographers (Copeland et al., 2010: 159). Yet, the women were incredibly capable. In a report on the Tunny machines, their "cheerful common sense" was noted, along with the fact that "'several' of the Newmanry's 273 Wrens 'showed ability in cryptographic work' and that 'several others were trained by the engineers to undertake routine testing of the machines'" (Copeland et al., 2010: 159). Moreover, Dorothy Du Boisson reports that at first, the Wrens operated the Colossi under the direction of the cryptographer. However, she adds, "After a while, a formula was developed so that the operator could [set the machines] herself" (Copeland et al., 2010: 163). She states that this freed the cryptographer for "more important work," which directly conflicts with the utility of the Colossus (Copeland et al., 2010: 163). The Colossus hastened codebreaking and allowed for forms of "brute force" cryptanalysis that a human being could not feasibly perform. Setting the machines properly was important, and the Wrens had learned how. Nevertheless, the Newmanry Wrens were not always utilized to their full ability, as they faced restrictions on what titles they were allowed to hold due to their sex.

Working on a team like the Newmanry proved an odd mix of ingrained hierarchy and social im-

[5]These machines were nicknamed "Heath Robinson" by the Wrens because they were reminiscent of the complicated machinery drawn by British cartoonist William Heath Robinson.

pudence. Women were barred from positions of authority, but typical boundaries of etiquette were frequently broken. For example, Du Boisson recalls the heat emitted by the machine's numerous glass valves, stating that "If we got too hot or sleepy we went out for a splash from the static water tank (for use in the event of enemy action or fire). Someone suggested that we [go] topless, but we did not take up the offer" (Copeland et al., 2010: 162-163). In another instance, engineer Ken Myers recalls a Wren propping her pocket mirror up on one of the Tunny machines' transformer panels to reapply her lipstick. The mirror fell, evaporating through contact with the extremely hot machinery, and the lipstick shot across the girl's throat, giving the effect that it had been cut (Copeland et al., 2010: 170).

4 Operating the Colossus

For Joanna Stradling, meeting the Colossus was love at first sight. Mere moments after she had signed the Official Secrets Act, she "saw this astonishing machine the size of a room. It was ticking away, and the tapes were going around and all the valves, and I thought, what an amazing machine. There were valves and transistors and flippy-flappy things. Like magic and science combined!" (Dunlop, 2015: 125-126). Barred from entering university by her distant and yet overbearing father, Stradling had settled for a domestic science college. For her, service to her country equally presented a personal opportunity. In 1944, at age 19, she walked into a WRNS recruiting office in Gloucester, a decision she called "the first time I had taken control of my own life" (Dunlop, 2015: 89). She chose to be trained in "light electrical work" before she was sent to Bletchley Park, because of this, she had a better sense of what she would be doing than most of her fellow recruits.

At first, Stradling's work in Bletchley Park revolved around the paper tapes that span through the Tunny, Heath Robinson, and Colossus machines. She delivered them, mended them when they tore, and loaded them into the machines. In the orbit of the Colossus which had so thoroughly captured her attention but kept from working directly with the machine, she was left unsatisfied. Her least favorite task was repairing the tapes, leading to a particular vendetta towards the Heath Robinson machines, since on that particular device "the tapes had to be put on wheels, and the wheels weren't sprung so you couldn't move them around and regulate the tension of the tape. It was always ripping or exploding" (Dunlop, 2015: 126). Throughout everything, she retained her desire to work on the Colossus.

It almost did not happen. The Colossus was named as such because of its size—only tall girls could operate it, because they could reach the top of the machines (Dunlop, 2015: 125; Slimming, 2021: 121). But luckily for Joanna, she had long arms and had arrived just after the D-Day landings. There were two Colossi at Bletchley Park by then, and neither could be turned off for fear of the vacuum tubes breaking. Their utility was becoming increasingly apparent, as well: "in the Newmanry some twenty-five Fish keys were broken in August, far more than any previous month. Computer power had come into a league of its own and Colossi were hatching at the rate of one a month" (Dunlop, 2015: 209). Women with Joanna's abilities were needed to staff the machines, and she was quickly promoted.

By the end of the war, there were ten Colossus computers, with an eleventh having been commissioned. They were not identical devices, though. Each iteration improved on the last, and they developed at such a rate that even seven decades later Joanna Stradling can remember that she worked specifically on numbers five, seven, and nine (Dunlop, 2015: 209). Whether truly necessary or not, she treated the machines tenderly, recalling that "You had to make sure it's little eye was clean or it wouldn't be able to see…I was interested in how it did what it did, so I was determined to look after it and make life easy for it" (Dunlop, 2015: 210). Stradling personified the Colossi, perhaps because her job led to her feeling as though she was communicating with the machines. Unlike Helen Pollard, who learned how to read the small signs that she was getting nonsense from the Tunny machines, Stradling knew "if [she] got a positive result; [the Colossus] would tell you. It would go click click click whizz, the tapes would all run down and you knew you had a fit" (Dunlop, 2015: 210). Much like Helen Pollard, Stradling never knew what the decryptions said, and she never asked.

Joanna Stradling was spared from dismantling the Colossi at the end of the war, which was a blessing, considering how much she loved them. She was allowed to keep a small part of one to

remember the machines and the Park as a whole, as she was still sworn to secrecy and had no expectations of that changing. She "got a switch and kept it in my Wren belt-pocket for years" (Dunlop, 2015: 258).

There was some recognition that operating machinery required skill, but that skill was deprioritized in comparison to more theoretical abilities. In his 1944 article titled "Careers for Girls," published in the popular British journal for mathematics and math education, L. J. Comrie (1944: 92) writes that "No employer would engage an untrained girl to take dictation and type, but if his research department gets a new calculating machine he is, in general, forced by circumstances to engage a girl who has never seen a machine before." Comrie assumes that this translates to the girl not knowing any theory, only the mechanical process she must perform, but as women like Helen Pollard and Joanna Stradling demonstrate, what it actually meant was that the women worked out the underlying principles on their own. The entirety of "Careers for Girls" is rife with sexist and derogatory ideas. Comrie (1944: 90-91) writes that computer operation "holds a peculiar fascination for those temperamentally suited for it, and that it is well within the capacity of properly trained. . . girls." He does not see this kind of work as particularly rewarding or engaging and is only promoting it for its economic utility. In addition, Comrie (1944: 94) does not see the women as having lasting intrinsic value. In his words, "they can be made proficient, give good service in the years before they (or many of them) graduate to married life, and become experts with the housekeeping accounts!" Comrie advocated training the women because they made his business easier, not because he thought they deserved socioeconomic mobility, work they could be proud of, or the right to skilled labor. The work was deemed valuable, but the women were not. Even marriage itself had economic utility: "Turnover through marriage was supposed to ensure women didn't tire of the work or require promotion to better—and better paid—work, as that would throw his system out of alignment," though this did not always work in practice (Hicks, 2017: 22).

Computer operation required genuine expertise, and in the case of Bletchley, this expertise had to be gained under high-stress and nearly inhospitable conditions. The Colossus' predecessors had warmed the spaces that they were in. The Colossus was warmer. They were built with more than a thousand vacuum tubes each and were never switched off, leading to an at times oppressive heat. In addition, Heath Robinson and Tunny machines were noisy, due to their use of paper tapes, and due to the nature of the codes being deciphered their surroundings were always chaotic and intense. The rooms they were housed in were also improperly ventilated as they were kept in hastily-erected huts on the grounds of the Bletchley Park manor house, built solely for utility and out of necessity when the decryption program outgrew the bounds of the house. These problems, too, increased after the Colossus was installed (Hicks, 2017: 35-36). Work shifts were grueling and led to exhaustion.

The women faced difficulties at "home," as well. Bletchley was a small place, and many of them needed to billet with locals. Some were sent to the equally hastily built (and therefore somewhat unsound) lodgings of the nearby army camp (Hill, 2004*colon* 111). Conditions varied, but many women faced a combination of "contaminated water, poor heating, and noisy living quarters" (Hicks, 2017: 36). At the very least, those in billets could not escape intertwinement with the personal lives of their host families, for better or for worse. The Wrens faced their own particular challenges, as for a short time their supervisors, "unaware of the skilled, nonphysical nature" of their work called for mandatory drills in the morning, "which resulted in exhaustion and illness for many workers until their supervising officers were eventually told to stop the drills" (Hicks, 2017: 36).

5 ENIAC

To further situate the women of Bletchley Park in the history of computing, it is helpful to take a short diversion into one of the Colossus' closest analogs: the Electronic Numerical Integrator and Computer, or ENIAC, built at the University of Pennsylvania in 1945 and primarily used for the United States Army's Ballistic Research Laboratory. Unlike the Colossus, ENIAC was not operational until December 1945 and therefore could not be used during World War II, but like the Colossus its first operators were women. Specifically, ENIAC was first programmed by six female mathematicians who had to learn how to commu-

nicate with the machine before ever laying eyes on it. The literature on ENIAC is far more robust than the literature on the Colossus, mainly due to how heavily classified the existence of the Colossi was. However, the discourse around the Colossi since their declassification has mirrored early conversations around ENIAC. For one thing, as Jennifer S. Light (1999: 473) identifies, "while the War Department urged women into military and civil service and fed the media uplifting stories about women's achievements during the war, its press releases about a critical project like the ENIAC do not mention the women who helped make the machine run." The result of this, Light argues, is that even though the ENIAC was classified as "intelligent" for its ability to perform calculations, women computers were never given the same esteem. Moreover, the female programmers were almost entirely written out of ENIAC's story.

Another aspect of ENIAC was that it "made a fundamental distinction between hardware and software: designing hardware was a man's job; programming was a woman's job" (Light, 1999: 469). This separation never took place at Bletchley Park, as women like Margaret Boulton were tasked with constructing Colossi. Regardless, early programming required a detailed understanding of the hardware. ENIAC's original operators could not practice on the machine, so "They learned how ENIAC worked by talking with the original design engineers, studying their logic diagrams, and sharing ideas with the other programmers" (Fritz, 1996: 27). By the time they were in front of ENIAC, they knew how it worked, and because of that they could make it perform as they needed it to. Rather than a monotonous plugging in of data, programming mechanical computers was dynamic, creative, and challenging (Strober and Arnold, 1986: 139). Colossus and ENIAC operators had no precedents to rely on and the information they were given about their respective machines was limited, sometimes to the point of insufficiency. They could ask mathematicians and engineers in their programs for help, but they alone were responsible for making the machines run properly. The Bletchley Park women would have been especially disincentivized from asking for support unless absolutely necessary, as doing so would have meant not only that they were not working, but that someone else on their team was pulled away from another task. Given how frequently they state that other members of their teams were doing "more important" work, they would have tried to fix everything that they could themselves.

6 New Skills

Cora Pounds, who was only seventeen when she began working at Bletchley Park in 1944, represents a trend of women found themselves capable of things that they had previously been told they lacked the skills for, as identified by historian Liza Mundy (2017). As Mundy (2017: 184) writes, "As they went about their assignments, women found competencies within themselves they had not known existed. Jane Case—told, growing up, that she was bad at math—turned out to have a ready facility with numbers." Cora had not been a good student. She had a poor memory and did not pass her School Certificate exam on her first try (Dunlop, 2015: 35). Cora's interests had always lay in fashion (influencing her choice to join the Wrens, who notoriously had the most stylish uniforms), not academics. Despite this, she performed to the high standards of Bletchley Park. Regrettably, her perception of her work aligned with the stereotype of meaningless drudgery, as she recalls "'There were so many messages being sent in, when one didn't come good you were sent another one.' She was checking for a match, but she didn't know why or what it meant. 'Checking checking. It was very boring. There were no written messages" (Dunlop, 2015: 124). Cora never really learned what she did, which meant that she never really understood her own worth (Dunlop, 2015: 125). Furthermore, when asked about Max Newman, the head of her team, she does not remember meeting him. Historian Tessa Dunlop interprets this as Newman believing that the Wrens who worked for him did not want to have their work explained to them, and so he made no efforts to. The men who worked for him enjoyed a much closer relationship, as he encouraged collaboration and informal brainstorming sessions. Dunlop's (2015: 125) interpretation of this is that "Professor Max Newman was a model employer—unless you were a low-status girl, that is." It could be the case that Newman disregarded the women under him due to their gender, but it seems more likely that it was due to compartmentalization. As Captain Roberts noted, even when the Colossus was put in use, cryptographers still worked on break-

ing Lorenz codes by hand. In problem-solving, more heads are often better, so Newman's encouragement of group thinking makes strategic sense. On the other hand, the female Colossus operators were working on a different part of the decryption process. Information was heavily compartmentalized in Bletchley Park, and it was not unusual for members of the same team to not know what the other was doing. Time was also a factor. Eleanor Ireland remembers being trained to operate the Colossus by another Wren, and she remembers that "when I was given a new Wren to instruct, I was worried about leaving her alone for very long, and would hurry back from meal-break to make sure nothing awful had happened" (Copeland et al., 2010: 164). Mistakes and breaks cost time, and as Ireland continues, "We knew we were working against the clock and that people's lives depended on what we were doing" (Copeland et al., 2010: 164). The Wrens were far more familiar with the Colossus than Max Newman (who likely never operated one himself), and so in the name of efficiency it was better for them to teach each other than to turn to their team leader. It is true that female Bletchley Park employees were limited in ways that male employees were not, but in the case of Max Newman and Cora Pounds, pure sexism is an insufficient, even misleading, explanation for what Pounds experienced.

7 Conclusion

It is easy to forget the network of people necessary to maintain the Lorenz decryption project. The fact that a great deal of decryption was relegated to a select group of men, as well as the fact that these men have traditionally written the accounts of what went on at Bletchley Park, has resulted in a misunderstanding of what this labor actually looked like. Women are often excluded from the narrative, and when they are mentioned (often in passing) it is generally assumed that what they were doing was dull and tedious. In truth, while not every woman who worked at Bletchley Park enjoyed her time there, there were many who found their work fulfilling and meaningful. They found excitement in learning to operate technology like the Tunny machines and the Colossus computers. They kept themselves motivated with an understanding that what they were doing was of critical importance, even if they were not given enough information to know where their efforts fit in or even what they were doing. They formed attachments to the environment, the people, the devices, and the work that characterized so much of their wartime experiences. Once an individual was assigned to Bletchley, the chances were slim to none that they could be transferred. The stakes of secrecy were too high.

All of the Bletchley Park employees understood the importance of silence. They hid the true nature of their work from their friends, families, and even each other, both during and after the war. Even when the Bletchley Park program was declassified, though, more accounts came from men than women. The work done by women was seen as less glamorous, dynamic, and compelling than that of their male counterparts, but it was equally important and meaningful. The decryption of the Lorenz ciphers was a critical part of Allied success in the Second World War, and speed was essential. Even the Colossus computers have been recognized for their ability to expedite the code-breaking process, but they needed the women to function.

Acknowledgments

My thanks to Professors Laura Edwards and Matthew L. Jones for their insights and guidance.

References

Algis Valiunas. "Turing and the Uncomputable." *The New Atlantis*, no. 61 (2020): 44–75.

Anne Secord. "Science in the Pub: Artisan Botanists in Early Nineteenth-Century Lancashire." *History of Science* 32, no. 3 (December 1, 1994): 269–315.

B. Jack Copeland. "Colossus and the Rise of the Modern Computer." In *Colossus*: *The Secrets of Bletchley Park's Code-Breaking Computers*, edited by B. Jack Copeland, Reprint edition. Oxford: Oxford University Press, 2010.

B. Jack Copeland, Catherine Caughey, Dorothy Du Boisson, Eleanor Ireland, Ken Myers, and Norman Thurlow. "Mr. Newman's Section." edited by B. Jack Copeland, Reprint edition. Oxford: Oxford University Press, 2010.

Brian J. Winkel. *The German Enigma Cipher Machine*: *Beginnings, Success, and Ultimate Failure*. Artech House, 2005.

Chris Christensen. "Polish Mathematicians Finding Patterns in Enigma Messages." *Mathematics Magazine* 80, no. 4 (2007): 247–73.

Colossus: *Creating a Giant*, 2012. https://www.youtube.com/watch?v=knXWMjIA59c.

"Computing Herstory - Computing History." Accessed April 21, 2023. https://www.computinghistory.org.uk/pages/58327/Computing-Herstory/.

Daniel Nelson. *Managers and Workers*: *Origins of the Twentieth-Century Factory System in the United States, 1880-1920.* The University of Wisconsin Press, 1995. https://hdl.handle.net/2027/heb08761.0001.001.

David Kahn. "Codebreaking in World Wars I and II: The Major Successes and Failures, Their Causes and Their Effects." *The Historical Journal* 23, no. 3 (1980): 617–39.

David Kenyon. *Bletchley Park and D-Day*: *The Untold Story of How the Battle for Normandy Was Won*. Yale University Press, 2019. https://doi.org/10.2307/j.ctvpbnpmf.

David R. Roediger and Elizabeth D. Esch. *The Production of Difference*: *Race and the Management of Labor in U. S. History*. Oxford, UNITED STATES: Oxford University Press, Incorporated, 2012. http://ebookcentral.proquest.com/lib/princeton/detail.action?docID=916040.

Denise Gürer. "Pioneering Women In Computer Science" 34, no. 2 (2002).

Dermot Turing. *Prof*: *Alan Turing Decoded* : *A Biography.* Stroud, Gloucestershire: The History Press, 2015.

———. *X, Y & Z* : *The Real Story of How Enigma Was Broken*. Gloustershire [England]: History Press, 2018.

Elsa Barkley Brown. "'What Has Happened Here': The Politics of Difference in Women's History and Feminist Politics." *Feminist Studies* 18, no. 2 (1992): 295–312. https://doi.org/10.2307/3178230.

F. W. Winterbotham. *The Ultra Secret*. 1st U.S. ed. New York: Harper & Row, 1974.

Gary M. Bateman. "THE ENIGMA CIPHER Machine." *American Intelligence Journal* 5, no. 2 (1983): 6–11.

Google Arts & Culture. "The Women of Bletchley Park." Accessed April 21, 2023. https://artsandculture.google.com/story/the-women-of-bletchley-park/qgVxQIAxvdB-JA.

Gwen Watkins. *Cracking the Luftwaffe Codes*: *The Secrets of Bletchley Park*. Reprint edition. Frontline Books, 2013.

Sir Harry Hinsley. "The Influence of ULTRA in the Second World War." 1993. https://www.cix.co.uk/~klockstone/hinsley.htm.

Helen Currie. "An ATS Girl in the Testery." In *Colossus*: *The Secrets of Bletchley Park's Code-Breaking Computers*, edited by B. Jack Copeland, Reprint edition. Oxford: Oxford University Press, 2010.

Jan Slimming. *Codebreaker Girls*: *A Secret Life at Bletchley Park*. Barnsley: Pen & Sword Military, 2021.

Janet Abbate. *Recoding Gender*: *Women's Changing Participation in Computing*. The MIT Press, 2012. https://www.jstor.org/stable/j.ctt5vjp2p.

Jennifer S. Light. "When Computers Were Women." *Technology and Culture* 40, no. 3 (1999): 455–83.

Jerry Roberts. *Lorenz*: *Breaking Hitler's Top Secret Code at Bletchley Park*. New edition. The History Press, 2017.

John F. Kasson. *Civilizing the Machine*: *Technology and Republican Values in America, 1776-1900*. New York: Penguin Books, 1977.

Joy Lisi Rankin. *A People's History Of Computing In The United States*. Cambridge, Massachusetts: Harvard University Press, 2018.

Kathy Kleiman, *Proving Ground*: *The Untold Story of the Six Women Who Programmed the World's First Modern Computer* New York; Boston: Grand Central Publishing, 2022.

L. J. Comrie. "Careers for Girls." *The Mathematical Gazette* 28, no. 280 (1944): 90–95. https://doi.org/10.2307/3606392.

Laine Nooney. "The Uncredited: Work, Women, and the Making of the U.S. Computer Game Industry." *Feminist Media Histories* 6, no. 1 (January 1, 2020): 119–46. https://doi.org/10.1525/fmh.2020.6.1.119.

Lilly Irani. "Justice for 'Data Janitors.'" *Public Books* (blog), January 15, 2015a. https://www.publicbooks.org/justice-for-data-janitors/.

———. "The Cultural Work of Microwork." *New Media & Society* 17, no. 5 (May 1, 2015b): 720–39. https://doi.org/10.1177/1461444813511926.

Lisa M. Fine, *The Souls of the Skyscraper*: *Female Clerical Workers in Chicago, 1870-1930* (Philadelphia: Temple University Press, 1990), https://www.amazon.com/Souls-Skyscraper-Clerical-Workers-1870-1930/dp/0877226741.

Lisa Nakamura. "Indigenous Circuits: Navajo Women and the Racialization of Early Electronic Manufacture." *American Quarterly* 66, no. 4 (2014): 919–41.

Liza Mundy. *Code Girls*: *The Untold Story of the American Women Code Breakers of World War II*. First edition. Boston: Hachette Books, 2017.

Mar Hicks. *Programmed Inequality*: *How Britain Discarded Women Technologists and Lost Its Edge in Computing*. MIT Press, 2017.

———. "Sexism Is a Feature, Not a Bug," March 9, 2021. https://doi.org/10.7551/mitpress/10993.003.0011.

Margot Lee Shetterly. *Hidden Figures*: *The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race*. HarperCollins, 2016.

Marion Hill. *Bletchley Park People*: *Churchill's "Geese That Never Cackled."* Stroud, Gloucestershire: Sutton, 2004.

Matthew L. Jones. *Reckoning with Matter*: *Calculating Machines, Innovation, and Thinking about Thinking from Pascal to Babbage*. Chicago, IL: University of Chicago Press, 2016. https://press.uchicago.edu/ucp/books/book/chicago/R/bo24836963.html.

Michael Smith. *Station X*: *The Codebreakers of Bletchley Park*. London: Channel 4 Books, 1998.

Myra H. Strober and Carolyn L. Arnold. "Integrated Circuits/Segregated Labor: Women in Computer-Related Occupations and High-Tech Industries." In *Computer Chips and Paper Clips*: *Technology and Women's Employment*, by National Research Council (U.S.)., Vol. 1. Washington, D.C.: National Academy Press, 1986.

Nathan L. Ensmenger. *The Computer Boys Take Over*. The MIT Press, 2010. https://mitpress.mit.edu/9780262517966/the-computer-boys-take-over/.

Paula Baker. "The Domestication of Politics: Women and American Political Society, 1780-1920." *The American Historical Review* 89, no. 3 (1984): 620–47. https://doi.org/10.2307/1856119.

Peter Calvocoressi. *Top Secret Ultra*. M & M Baldwin, 2001.

R. K. Shyamasundar. "The Computing Legacy of Alan M. Turing (1912–1954)." *Current Science* 106, no. 12 (2014): 1669–80.

Richard James Aldrich. *GCHQ*: *The Uncensored Story of Britain's Most Secret Intelligence Agency*. HarperPress, 2010.

Ruth Oldenziel. *Making Technology Masculine*: *Men, Women, and Modern Machines in America, 1870-1945*. Amsterdam University Press, 1999. https://www.jstor.org/stable/j.ctt46mtdk.

Ruth Schwartz Cowan. *More Work for Mother*: *The Ironies of Household Technology from the Open Hearth to the Microwave*. New York: Basic Books, 1983.

Sarah T. Roberts. *Behind the Screen*: *Content Moderation in the Shadows of Social Media*. New Haven: Yale University Press, 2019.

Seth Koven and Sonya Michel. "Womanly Duties: Maternalist Politics and the Origins of Welfare States in France, Germany, Great Britain, and the United States, 1880-1920." *The American Historical Review* 95, no. 4 (1990): 1076–1108. https://doi.org/10.2307/2163479.

Sharon Hartman Strom. "'Machines Instead of Clerks': Technology and the Feminization of Bookkeeping, 1910-1950." In *Computer Chips and Paper Clips*: *Technology and Women's Employment*, by National Research Council (U.S.)., Vol. 1. Washington, D.C.: National Academy Press, 1986.

Simon Schaffer. "Astronomers Mark Time: Discipline and the Personal Equation." *Science in Context* 2, no. 1 (1988): 115–45. https://doi.org/10.1017/s026988970000051x.

Steven Shapin. "The Invisible Technician." *American Scientist* 77, no. 6 (1989): 554–63.

Tessa Dunlop. *The Bletchley Girls*: *War, Secrecy, Love and Loss*: *The Women of Bletchley Park Tell Their Story*. London: Hodder & Stoughton, 2015.

W.B. Fritz. "The Women of ENIAC." *IEEE Annals of the History of Computing* 18, no. 3 (1996): 13–28. https://doi.org/10.1109/85.511940.

Exploring the Alignment of Transcriptions to Images of Encrypted Manuscripts

Goio Garcia, Pau Torras, Alicia Fornés
Computer Vision Center, Computer Science Dept.
Universitat Autònoma de Barcelona, Spain
goio.garcia@autonoma.cat
{ptorras,afornes}@cvc.uab.cat

Beáta Megyesi
Department of Linguistics
Stockholm University
Sweden
beata.megyesi@ling.su.se

Abstract

The automatic transcription of encrypted manuscripts is a challenge due to the different handwriting styles and the often invented symbol alphabets. Many transcription methods require annotated sources, including symbol locations. However, most existing transcriptions are provided at line or page level, making it necessary to find the bounding boxes of the transcribed symbols in the image, a process referred to as alignment. So, in this work, we develop several alignment methods, and discuss their performance on encrypted documents with various symbol sets.

1 Introduction

Historical encrypted manuscripts often employ invented alphabets for encryption. These alphabets are usually composed of digits, Latin or Greek letters, Zodiac or alchemical signs, invented symbols and diacritics. Not surprisingly, the automatic generation of transcription of such documents is a challenge due to difficulties in dealing with many different symbols and handwriting styles presented in a few lines or pages. Despite the advances in the image processing research field, current Handwritten Text Recognition (HTR) techniques are based on deep learning architectures, which require a considerable amount of labelled data to train. Unluckily, in the case of encrypted documents, the amount of available transcriptions is insufficient and the few existing transcriptions are provided at line or page level, without any labelling of the area where the symbol is located. Normally, in HTR, the symbols in a document are typically labelled with bounding boxes with respect to the coordinates of its top left and bottom right points. The consequence of the lack of information about the symbols' location is that only end-to-end segmentation-free transcription methods can be employed, restricting the use of other high-performing types of recognition methods, such as symbol spotting and symbol classification models.For this reason, it is desirable to devote efforts in aligning the existing transcriptions represented as text sequences to the actual cipher symbols in the image, that is, leverage the information granted by the transcription, such as the presence and order of symbols, and obtain accurate bounding boxes, so that all type of recognition methods can be applied.

In this work, we explore different methods for aligning the transcriptions with symbol images, a task referred as text-to-image alignment.

2 Previous Work

The alignment of graphical sources to their transcriptions has been explored through various means in the literature. The most common is the idea of employing a text recognition model and then finding the mapping between the ground truth and the recognised text. The earliest instances of this idea can be found in (Tomai et al., 2002; Kornfield et al., 2004), in which the authors segment the page into words and then align the ground truth using dynamic programming algorithms. Further work combine Hidden Markov Models (HMMs) with the Viterbi algorithm for the word image-to-text alignment phase (Rothfeder et al., 2006; Feng and Manmatha, 2006; Toselli et al., 2007; Fischer et al., 2011), with the latter also dealing with writing-transcript inconsistencies. Good word segmentation is usually difficult to obtain, so this family of techniques is usually combined with some form of correction algorithm that prevents catastrophic failure cases (De Gregorio et al., 2022; De Gregorio et al., 2023).

In (Torras et al., 2021) an alignment method is proposed where an attention-based Sequence-to-Sequence model is trained to recognise symbol lines from the Copiale (Knight et al., 2011) and

Borg (Aldarrab et al., 2018) ciphers. The highest attention activations are used to locate the correspondence between each symbol and its display in the image. To account for possible transcription mistakes, the Levenshtein algorithm is used to find the minimum edit path between the ground truth and the prediction of the model. The same authors propose the use of Recurrent and Convolutional Neural Network-based recognition architectures trained using the Connectionist Temporal Classification (CTC) loss (Torras et al., 2023). The alignment is found by forcing the known ground truth sequence when decoding the output of the model – for every slice of the input image, the corresponding transcription class is generated. Their advantage over Sequence to Sequence models is that they require much less data and compute time to train and can recover from mistakes much more reliably.

3 Connectionist Temporal Classification

In this work, we build from the foundation of CTC-based alignment methods and attempt to find ways of preventing its common failure cases. In this section, we briefly introduce the CTC loss to describe the improvements we propose.

CTC models (Graves et al., 2006) are a popular paradigm for sequence-learning. They have the capability to process arbitrary length input and output sequences, as long as the input sequence is longer than the target. In addition, they do not need alignment information during training. Given an input sequence $X = (x_i)_{i=1}^{T}$ and a target sequence $L = (l_i)_{i=1}^{U}$, where each element of the target sequence belongs to an alphabet A of length a, the CTC model predicts an output sequence $Y = (y^{(i)})_{i=1}^{T}$, where $y^{(i)} \in \mathbb{R}^{a+1}$ is a vector containing the probability for every symbol in alphabet $A' = A \cup \emptyset$ at time-step i. Symbol $\emptyset$ is known as the "blank" character. By selecting a symbol for each time-step, one obtains a path π. Since input sequences may differ (e.g. in speech or handwriting there are varying speeds or character widths), the CTC output should represent the same target sequence despite different input sequences.

The CTC defines a mapping from paths to target sequences $\mathcal{B}$. For some path π, $\mathcal{B}$ removes consecutive repeated symbols and then deletes every blank symbol. We denote as $\mathcal{B}^{-1}(L)$ the set of feasible paths, all paths that map to target sequence L. To obtain the probability of a target sequence L, we must marginalize over all feasible paths

$$p(L|X) = \sum_{\pi \in \mathcal{B}^{-1}(L)} p(\pi|X) \tag{1}$$

where the probability of path π is simply the product of probabilities for the selected symbol at every time-step. The loss function is taken as the negative log-probability of the target path

$$L_{ctc} = -\log p(L|X). \tag{2}$$

4 Methods

For text-to-image alignment of symbols, we test three different methods, based on the CTC model.

Method 1: Maximum entropy regularization

The CTC loss function has been modified with the goal of discouraging peaky distributions, as proposed by Liu et al. (2018) and using the implementation given in the same paper. A new regularization term is introduced, which encourages exploration when decoding by maximizing the entropy of the set of feasible paths and as a result distributes the probabilities over multiple alignments, avoiding the convergence into a single feasible path. This way the CTC decoding process is improved, as it requires for the probabilities of multiple feasible paths to be summed. The parameter β controls the importance of this term.

Given a target path L and an input sequence X, the new loss function reads

$$L_{entctc} = L_{ctc} - \beta H(p(\pi|L, X)) \tag{3}$$

where $H(p(\pi|L, X))$ is the entropy of the set of feasible paths. The entropy of feasible paths is given by the Shannon entropy

$$H(p(\pi|L, X)) = -\sum_{\pi \in \mathcal{B}^{-1}(L)} p(\pi|L, X) \log p(\pi|L, X) \tag{4}$$

Method 2: Define anchors based on single model confidence

By leveraging the reduced overconfidence in incorrect outputs of Method 1, it is possible to identify high confidence alignments directly through the CTC output probabilities. These high confidence alignments are referred to as anchors, and are used to further subdivide the image and repeat predictions on the low confidence segments while keeping the anchored alignment intact. An example of the process is illustrated in Figure 1.

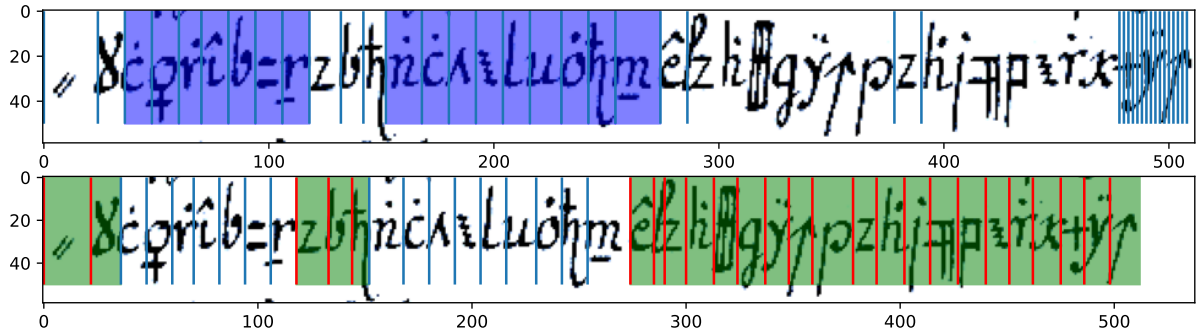


Figure 1: Top image: initial alignment of a Copiale string (blue vertical bars) outputted by the $\beta = 0.1$ regularized model, and selected anchors based on single model-confidence (blue regions). Bottom image: final alignment after reprediction of non-anchored alignments (green regions). The changed alignments are marked in red.

The determination of the threshold for what constitutes a high confidence alignment is performed automatically through Otsu's method (Otsu, 1979), which identifies the optimal cutoff point between two classes by minimizing the intraclass variance. Then, for the high confidence class, the anchor threshold is taken by substracting 1.5 standard deviations from the median.

A final filter is performed on the alignments that have been classified as high confidence, only leaving as anchors those for which their direct neighbours are also high confidence. The reason for this is that the confidence outputted by the CTC is still high for the first incorrect alignment, but falls off for the following ones.

Method 3: Define anchors based on multiple model agreement

By training two separate models and comparing their decoded outputs, it is possible to select anchors based on similarity of both alignments. Alignments were selected as anchors when their Intersection over Union (IoU) was over a 0.95 threshold. The separately trained models included the original unregularized model, the regularized model presented in Method 1 and a version of the regularized model trained on horizontally inverted images (H.Flipped model).

5 Experiments

Given the three methods described above, we carry out experiments with various symbols sets.

5.1 Datasets and Metrics

The models have been evaluated in two different datasets. Annotations include the 1-dimensional bounding boxes of every symbol present in the image as ground-truth. a) Copiale: Training dataset of 253747 images of line fragments from the Copiale cipher, extracted from a dataset of 648 lines of text. Validation dataset of 126 images of full lines, and test dataset containing 139 images of full lines. b) Borg: Training dataset of 2868 images of line fragments from the Borg cipher, extracted from 178 lines of text. Validation dataset composed of 22 images of full lines, and test dataset containing 49 images of full lines. Line fragments used for training are a form of data augmentation and may include overlap between several images.

To evaluate the models' performance, we use the Average Intersection over Union (AIoU) metric and the Hits@X metric. AIoU quantifies the average overlap of the predicted bounding boxes with the ground truth, with 1 being ideal and 0 being worst. The Hits@X metric indicates the percentage of bounding boxes that have achieved an IoU greater than X%.

5.2 Results of Method 1

The first goal was to determine the optimal value for the regularization parameter β. Values 0.05, 0.1 and 0.2 were tested, and the results for the best value is shown in Table 1. The best performing value depends on the dataset: 0.1 for Copiale, and 0.05 for Borg.

While the overall performance of the model has been reduced with the exception of the Borg dataset when $\beta = 0.05$, by looking at the distribution of output confidences we have observed that a reduction in high confidence alignments has been achieved, as desired.

Table 1: Results for all tested methods. Regularized columns always show results for the best β parameter on each dataset, 0.1 for Copiale and 0.05 for Borg.

			Method 1	Method 2		Method 3	
		Unreg.	Regularized	Unreg.	Reg.	Unreg. + Reg.	Reg. + H.Flip.
Copiale	AioU	0.807	0.763	**0.826**	0.794	0.813	0.780
	Hits@25	0.972	0.935	0.994	0.976	0.976	0.953
	Hits@50	0.952	0.911	0.973	0.951	0.954	0.925
	Hits@75	0.763	0.695	0.785	0.722	0.780	0.720
Borg	AioU	0.545	0.559	0.645	**0.654**	0.552	0.559
	Hits@25	0.786	0.748	0.960	0.893	0.780	0.755
	Hits@50	0.685	0.681	0.788	0.797	0.683	0.678
	Hits@75	0.284	0.394	0.340	0.429	0.305	0.389

5.3 Results of Method 2

The described method has been applied to both datasets, yielding results shown in Table 1. There is a significant performance uplift when compared to Method 1 for the Borg dataset, while in the Copiale dataset it achieves similar performance. The usage of the regularized model over the unregularized one does not have a clear impact. For instance, in the case of the Borg dataset it increases the rate of accurate alignments at the same time it reduces the number of rough alignments, causing the average IoU to stay within a 1% range of each other. Usually, it is desirable to maximize the number of perfect alignments even at the cost of reducing the quality of other alignments. This is because in case the model predictions are manually corrected, perfect alignments do not need intervention and so it results in a workload reduction. In the Copiale dataset, however, the overall performance is better with the unregularized model. This seems to indicate that the regularized model is better for smaller datasets with few labelled data.

5.4 Results of Method 3

Method 3 outperforms the baseline and Method 1 in the Copiale dataset when using the agreement between the unregularized and regularized models, as shown in Table 1. However, Method 2 is still the best-performing one in both datasets. Using the model trained on horizontally flipped images does not seem to provide a significant advantage.

One of the main challenges that have been observed when using the agreement between the regularized and unregularized models, is the fact that both models tend to produce the same mistakes, marking the incorrect predictions as anchors. As they have been trained using the same data and have a very similar architecture, this situation is not uncommon. In the case of the agreement between the regularized and the H.Flipped model, it is hard to obtain any anchor at all, since the regularized model tends to produce mistakes at the end of the image, while the H.Flipped model produces them at the start, leaving no region where both models agree. The performance of this method might be improved if a different architecture can be used in the aggregation process.

6 Discussion and Conclusions

In this paper, we have explored different methods to improve the accuracy of handwritten text-to-image alignment of symbols to known transcriptions using CTC models. We presented a regularized model based on maximum entropy of feasible paths, and two methods making use of the concept of anchors: high confidence alignments that remain fixed during subsequent predictions. The effectiveness of each method depends on the dataset, but Method 2 achieves the best results in general, with similar performance as the original model for the Copiale dataset, and a significantly improved alignment for the Borg dataset. Method 1 accomplishes its goal of reducing high confidence outputs, but does not improve performance by itself in the Copiale dataset. However, the regularized model can be a useful tool when used in conjunction with other methods. Although Method 3 is an improvement over single-model predictions, the similarity in the predictions of the aggregated models leads to a lack of capacity to detect anchors and correct errors. Further work will include experiments on encrypted sources with various alphabets.

Acknowledgments

This work has been partially supported by the Swedish Research Council (grant 2018-06074, DECRYPT), the Spanish projects PID2021-126808OB-I00 (GRAIL) and CNS2022-135947 (DOLORES) and the Spanish FPU Grant FPU22/00207. The authors acknowledge the support of the Generalitat de Catalunya CERCA Program to CVC's general activities.

References

Nada Aldarrab, Kevin Knight, and Beáta Megyesi. 2018. The Borg.lat.898 Cipher. https://www.su.se/english/research/research-projects/decipherment-of-historical-manuscripts/the-borg-cipher-1.688283.

Giuseppe De Gregorio, Ilaria Citro, and Angelo Marcelli. 2022. Transcript Alignment for Historical Handwritten Documents: The MiM Algorithm. In Cristina Carmona-Duarte, Moises Diaz, Miguel A. Ferrer, and Aythami Morales, editors, *Intertwining Graphonomics with Human Movements*, Lecture Notes in Computer Science, pages 45–60, Cham. Springer International Publishing.

Giuseppe De Gregorio, Giuliana Capriolo, and Angelo Marcelli. 2023. End-to-End Transcript Alignment of 17th Century Manuscripts: The Case of Moccia Code. *Journal of Imaging*, 9(1):17.

Shaolei Feng and R. Manmatha. 2006. A hierarchical, HMM-based automatic evaluation of OCR accuracy for a digital library of books. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '06, pages 109–118, New York, NY, USA. Association for Computing Machinery.

Andreas Fischer, Volkmar Frinken, Alicia Fornés, and Horst Bunke. 2011. Transcription alignment of Latin manuscripts using hidden Markov models. In *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, HIP '11, pages 29–36, New York, NY, USA. Association for Computing Machinery.

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376.

Kevin Knight, Beáta Megyesi, and Christiane Schaefer. 2011. The Copiale Cipher. In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web*, pages 2–9, Portland, Oregon. Association for Computational Linguistics.

E.M. Kornfield, R. Manmatha, and J. Allan. 2004. Text alignment with handwritten documents. In *First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings.*, pages 195–209.

Hu Liu, Sheng Jin, and Changshui Zhang. 2018. Connectionist temporal classification with maximum entropy regularization. In *Advances in Neural Information Processing Systems*, pages 837–847.

Nobuyuki Otsu. 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

Jamie Rothfeder, R. Manmatha, and Toni M. Rath. 2006. Aligning Transcripts to Automatically Segmented Handwritten Manuscripts. In Horst Bunke and A. Lawrence Spitz, editors, *Document Analysis Systems VII*, Lecture Notes in Computer Science, pages 84–95, Berlin, Heidelberg. Springer.

C.I. Tomai, Bin Zhang, and V. Govindaraju. 2002. Transcript mapping for historic handwritten document images. In *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, pages 413–418.

Pau Torras, Mohamed Ali Souibgui, Jialuo Chen, and Alicia Fornés. 2021. A Transcription Is All You Need: Learning to Align Through Attention. In Elisa H. Barney Smith and Umapada Pal, editors, *Document Analysis and Recognition – ICDAR 2021 Workshops*, volume 12916, pages 141–146. Springer International Publishing, Cham.

Pau Torras, Mohamed Ali Souibgui, Jialuo Chen, Sanket Biswas, and Alicia Fornés. 2023. Segmentation-Free Alignment of Arbitrary Symbol Transcripts to Images. In Mickael Coustaty and Alicia Fornés, editors, *Document Analysis and Recognition – ICDAR 2023 Workshops*, Lecture Notes in Computer Science, pages 83–93, Cham. Springer Nature Switzerland.

Alejandro H. Toselli, Verónica Romero, and Enrique Vidal. 2007. Viterbi Based Alignment between Text Images and their Transcripts. In Caroline Sporleder, Antal van den Bosch, and Claire Grover, editors, *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007).*, pages 9–16, Prague, Czech Republic, June. Association for Computational Linguistics.

On the tracks of Félix-Marie Delastelle

Rémi Géraud-Steward
École normale supérieure,
PSL Research University,
Paris, France
remi.geraud@ens.fr

David Naccache
École normale supérieure,
PSL Research University,
Paris, France
david.naccache@ens.fr

Abstract

"Can not find any info on Delastelle — Nothing on record in this country." (William Friedman, 18 Jan 1955, NSA Archives A63734) Following these words, the then-director of the US National Security Agency hailed contacts in Europe, hoping that someone would fill in this missing information. The initial inquiry was sent to Friedman by amateur American cryptographer William Maxwell Bowers; in 1963, Bowers would publish under a pseudonym all that he could find on the matter (The Cryptogram 1963, preserved under reference VF 54-30 at the US National Cryptologic Museum). Since this document, which had a very limited audience, almost no new information on Delastelle was published, and indeed most of the information available widely today on Delastelle is at best fragmentary.

In this paper we reopen that case, reviewing information about the life and work of Félix-Marie Delastelle, establishing data overlooked by earlier historians, correcting several oft-repeated errors and bringing novel documents to public awareness.

Félix-Marie Delastelle is briefly mentioned in Kahn's monograph (Kahn, 1996, Chap. 8) as one of the great French cryptographers of the late 19th century — along the likes of Kerckhoffs, de Viaris, Valerio, and Bazeries — and "the only major writer on cryptology of the time who was not in the military". A single paragraph sketches the biography of that unusual individual. Kahn drew the (partially incorrect) information verbatim from Bowers' *Cryptogram* article[1] (Bowers, 1963), which is based "on documents in the Mairie of Saint-Malo and on recollections of Delastelle's niece (Kahn, 1996, Note 242)."

As part of his personal interest in the matter, Bowers contacted NSA's director William Friedman, in the hope that he would know additional details — he did not. This triggered an investigation from the NSA[2,3] (Figure 1) that seems to have culminated in a different, but similarly laconic summary[4]. While both sources point out that Delastelle was a complete outsider in the field of cryptography, and highlight his detailed knowledge of the techniques of his time, neither manage to peer into any details of his life or works besides his famous "bifid" and "trifid" ciphers. Delastelle died in 1902 (more on that later).

Our investigation followed the tracks that these sleuths had initiated. We obtained documents from the Tobacco Administration where Delastelle worked, got access to his civil status registry, found published and unpublished work of his, followed his formative years and early career, tracked down his family properties and finances, and scoured various archives to collect evidence and crumbs of context. This information allows us to draw a much more precise, though still incomplete, portrait of this evasive and famous amateur cryptographer. At the very least, it makes it possible to correct mistakes in earlier accounts that are (at the time of writing) uncritically repeated by multiple sources, including Wikipedia[5].

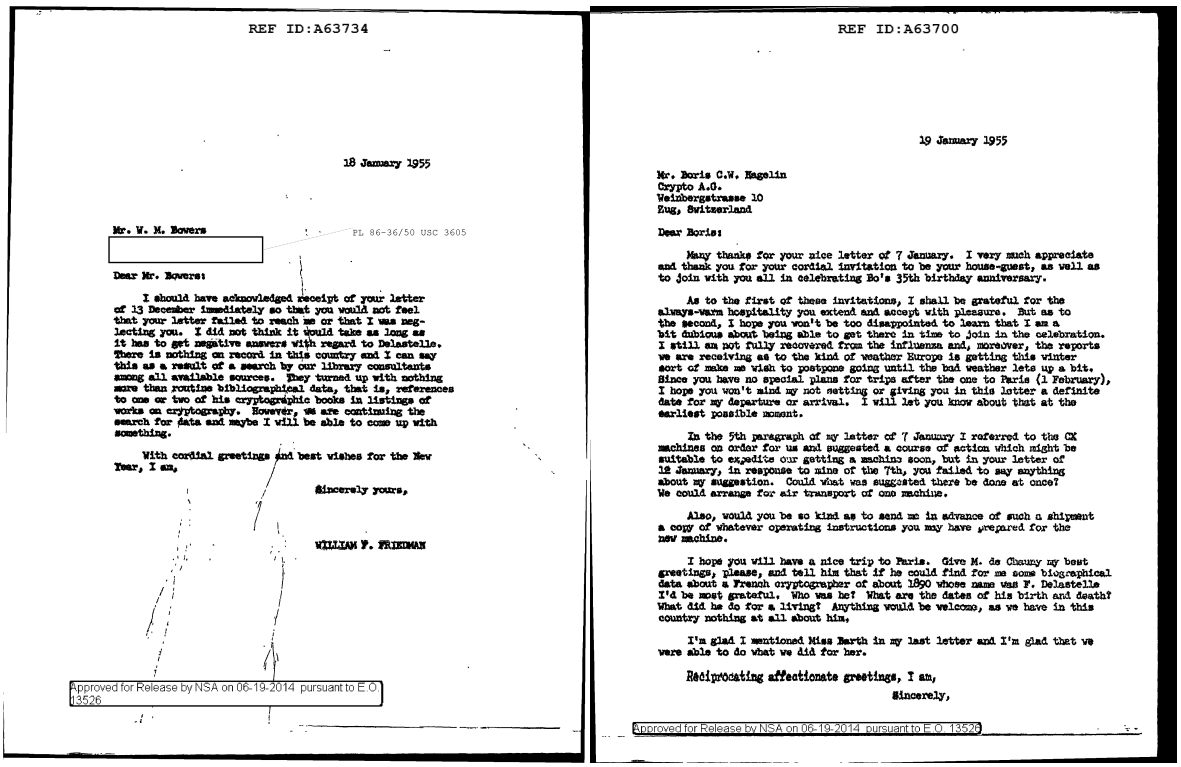
[1] Bowers also wrote a series of earlier leaflets, *Practical cryptanalysis*, edited by the American Cryptogram Association. Volume 2 (1960) and 3 (1961) discuss Delastelle's ciphers. His later *Cryptogram* article added some information and focuses on biographical aspects that appear in Kahn's account.

[2] Letter from William F. Friedman to William Bowers, 18 January 1955, Archives of the NSA, ref A63734.

[3] Letter from William F. Friedman to Boris Hagelin, 19 January 1955, Archives of the NSA, ref A63700.

[4] Letter DK 54-40, 5 May 1969, National Cryptologic Museum.

[5] At the time of writing, the Google search engine when queried for an image of "Félix-Marie Delastelle" only returns



REF ID:A63734

18 January 1955

Mr. W. M. Bowers

[redacted]

PL 86-36/50 USC 3605

Dear Mr. Bowers:

I should have acknowledged receipt of your letter of 13 December immediately so that you would not feel that your letter failed to reach me or that I was neglecting you. I did not think it would take as long as it has to get negative answers with regard to Delastelle. There is nothing on record in this country and I can say this as a result of a search by our library consultants among all available sources. They turned up with nothing more than routine bibliographical data, that is, references to one or two of his cryptographic books in listings of works on cryptography. However, we are continuing the search for data and maybe I will be able to come up with something.

With cordial greetings and best wishes for the New Year, I am,

Sincerely yours,

WILLIAM F. FRIEDMAN

Approved for Release by NSA on 06-19-2014 pursuant to E.O. 13526

REF ID:A63700

19 January 1955

Mr. Boris C.W. Hagelin
Crypto A.G.
Weinbergstrasse 10
Zug, Switzerland

Dear Boris:

Many thanks for your nice letter of 7 January. I very much appreciate and thank you for your cordial invitation to be your house-guest, as well as to join with you all in celebrating Bo's 35th birthday anniversary.

As to the first of these invitations, I shall be grateful for the always-warm hospitality you extend and accept with pleasure. But as to the second, I hope you won't be too disappointed to learn that I am a bit dubious about being able to get there in time to join in the celebration. I still am not fully recovered from the influenza and, moreover, the reports we are receiving as to the kind of weather Europe is getting this winter sort of make me wish to postpone going until the bad weather lets up a bit. Since you have no special plans for trips after the one to Paris (1 February), I hope you won't mind my not setting or giving you in this letter a definite date for my departure or arrival. I will let you know about that at the earliest possible moment.

In the 5th paragraph of my letter of 7 January I referred to the CX machines on order for us and suggested a course of action which might be suitable to expedite our getting a machine soon, but in your letter of 12 January, in response to mine of the 7th, you failed to say anything about my suggestion. Could what was suggested there be done at once? We could arrange for air transport of one machine.

Also, would you be so kind as to send me in advance of such a shipment a copy of whatever operating instructions you may have prepared for the new machine.

I hope you will have a nice trip to Paris. Give M. de Chauny my best greetings, please, and tell him that if he could find for me some biographical data about a French cryptographer of about 1890 whose name was F. Delastelle I'd be most grateful. Who was he? What are the dates of his birth and death? What did he do for a living? Anything would be welcome, as we have in this country nothing at all about him.

I'm glad I mentioned Miss Barth in my last letter and I'm glad that we were able to do what we did for her.

Reciprocating affectionate greetings, I am,

Sincerely,

Approved for Release by NSA on 06-19-2014 pursuant to E.O. 13526

Figure 1: Friedman's letters to Bowers and Hagelin, Archives of the NSA.

It was indeed difficult to find accurate and verifiable information about the man. Delastelle moved a lot across France, he often changed the way he himself spelled his name (seemingly on a whim), and he lived a very discreet life. While digitised databases and declassified documents are tremendous help and easily accessible, we cannot undo the damage of time and wars — archives were lost or destroyed, documents and witnesses did not survive. Bower's main source of information, Delastelle's niece, died in 1970.

One key result of this research[6] is that contrary to the established narrative, his interest and contributions to the field didn't come out of nowhere out of the boredom of retirement following a nondescript career: he was a lifelong inventor, tinkerer, and mathematician.

Félix-Marie Delastelle was born January 5, 1840 in the French city of Saint-Malo, at 8:00 AM, the son of 37-years old Captain Charles Delastelle and 27-years old Louise Cécile Delastelle (née Moison). The Delastelle family — sometimes written "de Lastelle"[7] — was a rather wealthy family of seafarers and merchants long associated with the city of Saint-Malo[8], and later with Madagascar (Micouin and Harel, 2008; Fontoynont, 1935). Félix-Marie was preceded by three children (see fig. 2): Charles Hyacinthe (named after his paternal grandfather, born 21 November 1833 and deceased 31 October 1834, before the age of one); Charles Louis (born in 1835); Auguste Michel (born in 1837), who shall play an important role model. Félix-Marie was followed by a younger sister Louise Marie Joséphine (born in 1843).

The father Charles is presumed lost at sea, with an official date of death of 16 November 1842[9,10]. It is unclear how the family dealt with this loss. Presumably, it put some financial pressure on them[11];

pictures of the completely unrelated French mathematician Camille Jordan!

[6] Which is extended in the full version of this article to include all new documents and additional biographic details.

[7] The name is absent from contemporary French nobiliaries (Dayre de Mailhol, 1843 1898).

[8] The name Delastelle is one of those to whom the famous French poet Chateaubriand, writing about his future sepulture in Saint-Malo, expresses gratitude in a 1831 letter to the then-Mayor of the city, Louis-François Hovius (de Chateaubriand, 1997 original publication 1849, Appendix 1).

[9] Civil registry for the birth of Louise Josephine, Saint-Malo Archives.

[10] Civil registry for the marriage of Louise Joséphine with Édouard Béhier, Saint-Malo Archives.

[11] Louise Cécile owned appartments that she later rented as barracks to the local soldiers dispatch; the other owner was Mayor Hovius. *Rapports et délibérations du Conseil général*

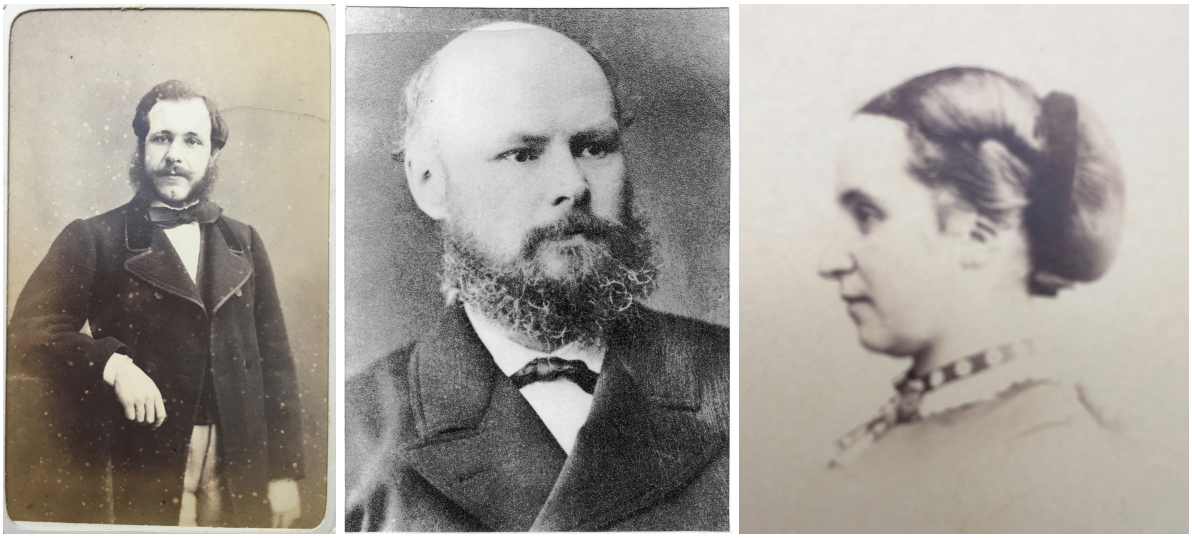


Figure 2: Left: Auguste Michel Delastelle. Middle: Félix-Marie Delastelle. Right: Louise Marie Joséphine. Dates unknown, family archives.

Louise Marie would be sent at age 10 to a girl-only boarding school (*Pensionnat de Mlle A. Egault*, Saint-Malo), where she stayed until at least 1858.

1 School years and first inventions (1852–1861)

Félix-Marie Delastelle entered the local *collège de Saint-Malo* in 1852 (aged 11), receiving multiple certificates of excellence (*Témoignage de complète satisfaction*). He applied to the *Lycée Impérial de Rennes* — the first *lycée* in Brittany, and the most prestigious at that time — to prepare for the national and very selective *baccalauréat* exam.

On 17 October 1858, as a 18 years old *lycée* student, Félix-Marie Delastelle submits an invention to the French Academy of Sciences (temporarily renamed *Institut Impérial de France*) — his first publication[12]. Spelling his name *Félix de Lastelle*, he describes the *chrono-barométrographe*: a precision weather instrument dedicated to recording variations of atmospheric pressure during a day. He hints at further similar inventions, such as a device to record temperature variations, which he hopes to describe in a further letter once some technical details are figured out, but there is no evidence that he went through with this promise.

Delastelle sent a second letter of invention to the French Academy of Sciences in December 1860, on the matter of chemistry, more specifically an algebraic notation appropriate for complex syntheses: *Note sur l'Annotation chimique*[13,14]. Here too, the letter concludes with a promise to pursue research in that field, a promise not followed through.

Delastelle would have attempted the baccalauréat in 1861, although we have no evidence that he did. That year, 2872 students obtained this most coveted diploma, including the first ever woman, Julie-Victoire Daubié. Delastelle was not one of them. An academic career was impossible without this diploma. Thus, contrary to an oft-reprinted myth, Félix-Marie Delastelle *did not* attend university, and in particular did not attend any Grande École.

2 Early Carreer at the Tobacco Administration (1861–1889)

Félix-Marie Delastelle joined the French Tobacco Administration on 1 December 1861 as unpaid intern (*surnuméraire des tabacs*) — two years after his brother Auguste did the same. Such a position was not easy to attain as it was required that the applicant be financially independent (hence a deposit of 4000 francs) and required that the applicant successfully passes multiple rounds of tests and checks

d'Ille-et-Vilaine, session held on 14 December 1878. Bibliothèque nationale de France, département Droit, économie, politique, 8-LK16-103.

[12] Archives of the French National Academy of Sciences, Letter 331, No 464, Session of 18 October 1858.

[13] Archives of the French National Academy of Sciences, Letter 341, No 449, Session of 2 December 1860.

[14] This letter is also signed *Félix de Lastelle* and indicates the same address as the 1858 communication, confirming that they were written by the same person.

(Block, 1877 1885)[15]. It is noted in his application that Delastelle could speak English and Italian.

After these two years of unpaid internship, in 1863, Félix-Marie Delastelle was accepted as a paid employee (*commis*), then promoted to *vérificateur* that same year. In theory, to reach higher positions (as he did), it was necessary to go through yet another series of exams, before the age of 35[16].

While his siblings start having families[17,18], Félix-Marie remains isolated, single and reclusive, applying himself to tasks unrelated to his professional obligations: on 22 September 1875 he applies for a French patent on candle manufacture and lighting[19] also submitted to the London patent office[20] and also attempted to obtain an Italian patent that same year[21]. He applied again, perhaps with a variation of his original idea, for a British patent in 1877 or 1878[22].

In 1879, Félix-Marie is promoted to *contrôleur de culture* and sent away from his hometown, in the city of Morlaix (Lot-et-Garonne department) working for Mr. Delestre[23]; he is then sent to Chambéry (Savoie department) in 1880. There, he applied for a British patent on lightweight ambulance beds on 7 December 1885[24]. Notably, he did not apply for a similar patent in France.

Promoted to *contrôleur principal* in 1886, he moved to Béthune (Pas-de-Calais), then to Dieppe (Normandy) in 1889, where he applied for a French patent on a new telescope model[25].

3 The Mathematical Turn (1890–1900)

Up to that point, it seems difficult to identify an underlying theme to Delastelle's various inventions, and they seem disconnected from his daily activity of overseeing tobacco transactions. But starting around 1889, his interest in mathematics and especially algebra solidified itself.

He sent a memoir to the French Academy of Sciences titled *Contributions à la théorie des équations algébriques* (contributions to the theory of algebraic equations). This memoir was read on 15 September 1890 by Hermite and Darboux[26,27]. He presents this work as motivated by some ongoing research of his where extant approaches due to Lagrange and Sturm are deemed too laborious.

The 32-page essay concerns itself with polynomials, extending a theorem of Budan and Fourier (Akritas, 1981). Its main result establishes a condition for a degree-m polynomial to have all its roots in $\mathbb{R}$, and locating them, together with an *algorithm* (called "algorithme des différences") to ease manual computations of the relevant invariant, which Delastelle claims is also helpful in obtaining continuous fraction approximation of roots.

The methods used in this work are not new, relying on a method of differences (à la Budan) and root symmetry considerations[28]. Perhaps for this reason, the Academy did not reply and made no public comment.

Delastelle would send a final unsolicited mathematical note to the Academy of Sciences in 1892, on Cryptography this time[29]. The note is prefaced by a letter insisting that the extreme simplicity of this method makes it available to "anyone who can read or write" in a matter of minutes. This is the

[15] *Programme des conditions d'admission à l'administration des tabacs, Moniteur du 28 fevrier 1865, Ministère des Finances.*

[16] *Examen des postulants de première série*, Moniteur, 28 February 1865; Le Temps, 14 October 1875, p. 3 ; Journal de la gendarmerie de France, 11 March 1877, p. 105.

[17] On 24 February 1872 his sister Louise Joséphine marries in Saint-Malo tax inspector Édouard Pierre Béhier; Auguste and Félix-Marie are among the witnesses. Three days after, 27-29 February 1872, his older brother Auguste Michel gets married too, in Montreuil (Pas-de-Calais) to Georgina Francès Amam Reid Dalton, the daughter of a retired British General in the Royal Artillery. Félix-Marie is again a witness.

[18] Interestingly, Félix-Marie and Louise write their name "de Lastelle", including on the invitations — all other family members write "Delastelle".

[19] *Amélioration dans la fabrication des bougies*, Number 109753, under the name "Delastelle". Cited in The Commissioner of Patents' Journal, "Grants of provisional protection for six months", Entry 2641, p. 2460, 27 oct 1876; "Notice to proceed" Oct 31 1976, p. 2479; cited in The Engineer, Volume 42, p 316, Morgan-Grampian (Publishers), 1876.

[20] *Improvements in the manufacture of candles and improved apparatus connected therewith*, 14 June 1876, under the name "de Lastelle". Also mentioned in *The London Gazette*, issue 24377, p. 5806, 31 October 1876.

[21] Cite in: Subject matter index of patents for inventions (attestati di privative industriali) granted in Italy from 1848 to January 1, 1886. Reference 26 I). 1876. II, 7, 612 (18, 65), pi. 156.

[22] *Arranging wicks in candles for increasing light*, Applications to the Great Britain Patent Office, Reference 2361.

[23] *Almanach national*, 1879, p. 348.

[24] *Movable structure for ambulances &c.*, application number 15,017.

[25] *Catalogue complet des brevets français délivrés au 3 janvier 1891*, publication du cabinet Émile Barrault, S. (divers) 297,796 dépôt du 26 août 1890, Téléscopes.

[26] *Comptes rendus de l'Académie des Sciences*, tome 2, 1890, p. 412.

[27] Delastelle's letter is dated 12th September 1890.

[28] Delastelle himself considers them well-known, and refers to Joseph Serret (Serret, 1866), but it is unclear which edition. The first edition dates from 1849, and contains one of the first systematic expositions of Galois theory.

[29] *Comptes rendus de l'Académie des Sciences*, 1892, p. 344. Session of 16 August 1892.

first known description of the *bifid cipher*, which is traditionally dated to a decade later. Once again the Academy was uninterested, and Delastelle would stop attempting to catch their attention any further.

At about the same time, Delastelle becomes visibly active in the French mathematical community. He contributes questions, problems, solutions and articles to *Les Tablettes du chercheur*[30] and *L'Intermédiaire des mathématiciens* (edited by Gauthier-Villars), a form of early precursors to Internet forums. Both were hugely popular amongst mathematicians, and included many contributions by famous names (including e.g., Henri Poincaré, Camille Jordan, Jacques Hadamard, and Adolf Hurwitz).

Delastelle's contributions to *L'Intermédiaire* (see Appendix A for the original French) are varied. From a modern cryptographer's outlook, some questions suggest that he may have investigated the mathematics of exponentiation modulo primes, with an eye to statistical and combinatorial properties. However to the best of our knowledge this was never fleshed out and he likely dropped these ideas to favour simpler notions.

Delastelle also sent a series of longer, self-contained articles with title *Cryptographie nouvelle*. These articles span over two years and are scattered across multiple issues (1892 p. 308–310, 1893 p. 24, 37, 54, 69, 81, 130, 140, 147). His last contribution is found in volume 13.

4 Retirement, final years, and death (1900-1902)

In 1893, Félix-Marie lives in Dieppe, 12 rue des tribunaux[31], under the name "Delastel"[32]. He gets promoted to *entreposeur* and leaves for Marseille (Bouches-du-Rhône). That same year he publishes his first book on cryptography (Delastelle, 1893), building from his collection of earlier published articles. The retail price is 3 Francs. He then moves to Marmande (Lot-et-Garonne).

In 1896, his very successful brother Auguste — who reached, in record time, the position of *directeur de la culture* — is knighted in the Légion d'Honneur. Félix-Marie's highest position in the administration, reached in 1898, is merely *inspecteur*. While employed at the Tobacco Administration, Félix-Marie Delastelle was regularly evaluated by superiors: although they unanimously portray an educated and competent man, many point out with regret that his career was unavoidably limited by his being "*original*" — an expression that appears verbatim in multiple reports.

Félix-Marie Delastelle retired in 1900 and returned to his hometown of Saint-Malo, more specifically the nearby town of Paramé. There he would complete his final book (Delastelle, 1902) — finished on 15 May 1901.

He was awarded a pension of 2332 Francs, for 37 years, 1 month and 26 days of service — with no military service and no compensation for any widow or children[33].

The narrative of Delastelle's death requires some corrections. On 1st April 1902, Félix-Marie Delastelle died age 62 in Saint-Ideuc, in the *Ker-Cadoc* villa. Remarkably, Bowers (and Kahn after him, and the NSA after them) misread the situation and thought that Félix-Marie died upon learning of his brother's death — the reality is much less poetic.

Indeed, the cause of death of Félix-Marie, at 5 PM, is unknown[34]. The next day, Auguste took his wife and the train to attend to his brother's funerary service. The train arrived at Le Mans at 11 PM, another train left for Saint-Malo at 12 PM — rushing to make this connection on time, 65 years old Auguste died of a heart attack[35].

Félix-Marie Delastelle was buried in the Cimetière Rocabey (Section 3 S–40) in Saint-Malo, besides other family members. His book, *Traité élémentaire de cryptographie*, was published posthumously by Gauthier-Villars, in 1902.

Acknowledgements

The authors would like to thank Tiphaine Colas and Marc Jean (Archives municipales de Saint-Malo), Isabelle Maurin-Joffre (Service des Archives et du Patrimoine historique de l'Académie des Sciences), Amélie Noiré (Musée des Transmissions), Patrick Hébrard, Vanessa Gratzer, Patrick Hebrard, Éric Latour (Coutot-Roehrig).

[30] *Les Tablettes du chercheur : journal des jeux d'esprit et de combinaisons*, editor B. Decolombe, owner and publisher P. Dubreuil.

[31] The street is now named rue Victor Hugo.

[32] *Fonds Ancien et Local*, Médiathèque Jean Renoir, Dieppe.

[33] *Bulletin des lois de la République française*, décret N°59.972, p. 1332–1333, 12 October 1900. Bibliothèque nationale de France.

[34] Civil registry for the death of Félix-Marie Delastelle (Saint-Ideuc, Ille-et-Vilaine).

[35] Civil registry for the death of Auguste Delastelle (Le Mans, Sarthe); also cited in *L'Avenir de la Dordogne*, 10 April 1902; also cited in *La Croix de la Charente*, 13 April 1902.

References

Alkiviadis G Akritas. 1981. On the Budan–Fourier controversy. *ACM SIGSAM Bulletin*, 15(1):8–10.

Maurice Block. 1877–1885. *Dictionnaire de l'administration française*, volume 2.

William Bowers. 1963. All that you ever wanted to know about Felix Delastelle, French cryptologist. *The Cryptogram*, XXX:79–82, 85, 101, 106–109.

Camille Philippe Dayre de Mailhol. 1843–1898. *Dictionnaire historique et héraldique de la noblesse française, rédigé dans l'ordre patronymique, d'après les archives des anciens Parlements, les manuscrits de d'Hozier et les travaux des auteurs, contenant un vocabulaire du blason, la notice des familles nobles existant actuellement en France, avec la description et le dessin de leurs armes*.

François-René de Chateaubriand. 1997 (original publication 1849). *Mémoires d'Outre-tombe*. Gallimard.

Félix-Marie Delastelle. 1893. *Cryptographie nouvelle, assurant l'inviolabilité absolue des correspondances chiffrée*. P. Dubreuil (18 bis rue des Martyrs, Paris).

Félix-Marie Delastelle. 1902. *Mathématiques appliquées. Traité élémentaire de cryptographie*. Gauthier Villars.

Dr Fontoynont. 1935. Napoléon de Lastelle (1802–1856). *La Revue de Madagascar*, 11:91–107.

David Kahn. 1996. *The Codebreakers: The comprehensive history of secret communication from ancient times to the Internet*. Simon and Schuster.

Françoise Micouin and Patrick Harel. 2008. *De Saint-Malo à l'Isle de France*, volume I.

Joseph Alfred Serret. 1866. *Cours d'algèbre supérieure*, volume 1. Gauthier-Villars.

A Delastelle's contributions to *L'Intermédiaire*, original French

- [Vol 1 p. 285 (ref 624, Iia)] Étant donné l'énoncé de la question 330, il serait fort intéressant, notamment du point de vue de la décimation cryptographique, de pouvoir calculer l'ordre probable de sortie des boules, principalement pour $n = 25$ et $n = 27$. Ce calcul est-il possible ?
  - [For reference, question 330, by Émile Lemoine, is the following] n boules $A_1, A_2, \ldots, A_n$ sont rangées en cercle et se suivent dans l'ordre $A_1, A_2, \ldots, A_n$. On part de A_1 sur la circonférence, en marchant dans le sens $A_1, A_2, \ldots, A_n$; on compte 1 sur A_1, 2 sur A_2, 3 sur A_3, etc. On convient d'enlever la boule sur laquelle on compte p. Cela fait, on continue à marcher dans le même sens en recommençant à compter $1, 2, 3, \ldots, p$ sur chaque boule que l'on rencontre et en enlevant celle sur laquelle on compte p; on continue ainsi jusqu'à ce qu'il ne reste plus qu'une seule boule. Peut-on savoir d'avance quel numéro portera cette boule? (...) Y a-t-il des valeurs de n et de p pour lesquelles on puisse donner une solution si l'on ne peut trouver la solution générale?
  - A response to this question is given in Vol. 3, p. 109, by Adr. Akar.
- [Vol 1 p. 314 (ref 639, 125b)] L'étude de la question n°400 m'a conduit à la notion, nouvelle pour moi, des nombres CIRCULAIRES. Les multiples de ces nombres jouissent de la propriété de reproduire toujours les *mêmes chiffres* disposés dans le *même ordre*. Quand le nombre des chiffres du produit surpasse celui du multiplicande, il suffit d'écrire le produit en *hélice* pour en trouver les chiffres primitifs. Certains nombres sont MULTICIRCULAIRES. Cette propriété est indépendante de la base du système de numération. Les nombres *circulaires* ont-ils déjà été étudiés ?
  - [For reference, question 400, by Charles-Ange Laisant, is the following] Si l'on multiplie le nombre $123456789 = N$, écrit dans le système décimal, par $2, 4, 5, 7$ ou 8, les différents produits

obtenus s'écrivent en permutant simplement les chiffres de N. Cette propriété, que j'ai constatée par hasard, m'a conduit à la proposition *empirique* suivante, que j'ai vérifiée sur un grand nombre d'exemples, et que je crois vraie, mais dont je n'ai cependant aucune démonstration : soit $123\ldots n = N$ un nombre écrit dans le système de numération de base $n+1$. Si l'on forme le produit de ce nombre par un mutiplicateur inférieur à n et premier avec n, ce produit s'écrira au moyen des n chiffres $1,2,3,\ldots,n$, pris chacun une seule fois et convenablement permutés. Quelque correspondant pourrait-il me donner une démonstration de ce théorème d'Arithmétique, ou en établir l'inexactitude?

  – [On this question, in Vol. 3, p. 100, Palmström says] J'ait fait en 1893 sur ces nombres que M. Delastelle nomme *circulaires* une Conférénce à la Société polytechnique de Bergen; mais je n'ai rien publié à ce sujet; ma question dans l'*Intermédiaire* avait pour but de m'informer auparavant de ce qui avait été fait sur le sujet.

- [Vol. 3, p. 128 (ref 846, A3d)] Je désirerais savoir si la proposition suivante est connue: soient $f(x) = 0$ une équation de degré m, ℓ et L les limites entre lesquelles les racines réelles sont toutes comprises, et h une quantité moindre que la différence des deux racines inégales les plus rapprochées. Substituons à x, dans le polynôme $f(x)$, une suite de nombres en progression arithmétique: $\ell, \ell+h, \ell+2h, \ldots, \ell+(n-1)h, \ell+nh$, dont le premier est *au plus* égal à la limite inférieure des racines et le dernier *au moins* égal à la limite supérieure de ces mêmes racines. Écrivons le résultat de ces substitutions en une colonne verticale, que nous désignerons par u; formons dans les colonnes suivantes le tableau des différences des divers ordres $\Delta_1, \Delta_2, \ldots, \Delta_n, \ldots, \Delta_m$; puis indiquons par V_n le nombre de variations de signes que renferme chaque colonne Δ_n du tableau ainsi formé.

  Ceci posé, si chaque colonne contient une variation de plus que la colonne suivante, on a pour chacune d'elles la relation $V_n = m - n$ et les racine sde la proposées sont toutes réelles et inégales. Si, pour un ecolonne quelconque, on trouve $V_n < m-n$, c'est-à-dire $V_n + d = m - n$, on aura, pour toutes les colonnes $u, \Delta_1, \ldots, \Delta_{n-1}$ qui précèdent Δ_n la relation $V_{n'} + d = m - n'$ et la proposée possèdera d racines imaginaires.

  Si, indépendamment des variations d perdues par une colonne quelconque Δ_n, une colonne de différences d'ordre inférieur Δ_ν, $\nu < n$, perd un nouveau nombre d' de variations, les racines imaginaires de la proposée s'élèveront à $d + d'$, et ainsi de suite.

  Si enfin la colonne Δ_1 présente plus de variations que la colonne u, la proposée a pour chaque signe (ou couple de variations) perdu deux racines doubles ou deux racines imaginaires, ce que l'on sait distinguer.

- [Vol. 5, p. 57] Après les multiples réponses à cette question, je crois qu'il n'est pas sans intérêt d'exposer brièvement la théorie des *nombres circulaires* [...] Les nombres circulaires me paraissent appelés à rendre d'utiles services dans beaucoup de calculs, notamment dans la recherche de *racines primitives*[36]. Leur détermination facile est encore simplifiée par la propriété dont ils jouissent d'avoir la seconde moitié de leurs chiffres formée du complément à 9 (ou, plus généralement, à $a-1$) de chacun des chiffres de la première moitié.

- [Vol. 5, p. 77] Soient p un nombre premier absolu, et $a, b, c, \ldots$ ses racines primitives. En écrivant la suite $a, a^2, a^3, a^4, \ldots, a^{p-1}$ et en retranchant de chaque terme tous les multiples de p qu'il peut contenir, on obtient la suite naturelle des nombres de 1 à $p-1$ disposés dans un ordre quelconque. Les autres racines $b, c, d, \ldots$ donneront cette même suite dans des ordres différents. Ceci posé : (1) La somme des puissances n-ièmes de $a, b, c, \ldots$ est congrue avec zéro, relativement au module p, lorsque n est diviseur ou multiple d'un diviseur de $p-1$:

$$a^n + b^n + c^n + \cdots = 0 \bmod p$$

  (2) La somme des n racines de a^n jouit de la même propriété, ainsi que toutes les puis-

[36] This can refer to two different notions in modern parlance: either the *multiplicative units* of $\mathbb{Z}/p\mathbb{Z}$, or the *discrete logarithm* of a number modulo p.

sances de ces racines, à l'exclusion de celles qui sont égales à n ou multiple de ce nombre. Ce théorème peut encore s'énoncer: dans toute suite calculée comme il est dit plus haut, la somme des nombres équidistants est toujours congrue à zéro pour le module p. Peut-on donner une démonstration élémentaire de ce théorème?

- [Vol. 5, p. 78] Lorsque le module est puissance d'un nombre premier p, la suite, qui contient $p^{\nu-1}(p-1)$ termes, jouit de propriétés analogues à celles de la question précédente; mais il faut alors que n ait un facteur commun avec $p^{\nu-1}(p-1)$; lorsque ce facteur est égal à $p-1$, $p(p-1)$, $p^2(p-1),\ldots$, la somme des $p^{\nu-1}, p^{\nu-2},\ldots$ termes, congrue avec zéeo pour le module p, ne l'est pas pour le module p^{ν}; elle a pour valeur les multiples $p, 2p, \ldots, p(p^{\nu-1})$ ou $p^2, 2p^2, \ldots, p^3, 2p^3, \ldots$ chacun augmenté du plus petit des termes trouvés. Peut-on démontrer ce théorème?

- [Vol. 5, p. 93] Delastelle solves a diophantine problem posed by Escott in the 1896 issue, ref 153.

- [Vol. 5, p. 111] Delastelle solves an interpolation problem posed by Cyp. Stephanos in the 1898 issue.

- [Vol. 5, p. 112] Delastelle solves a problem posed by A. Goulard in the 1897 issue.

- [Vol. 5, p. 148] L'étude des grands nombres, ou si l'on veut, des fractions décimales illimitées, paraît bien délaissée. Il me semble cependant qu'elle pourrait fournir des résultats intéressants. Pour en donner un aperçu, soit proposé de trouver la fraction génératrice du nombre illimité écrit dans le système B,

$$N = \frac{a}{B^{\alpha}} + \frac{b}{B^{2\alpha}} + \frac{c}{B^{3\alpha}} + \frac{d}{B^{4\alpha}} + \cdots$$

Si les quantités $a, b, c, d, \ldots$ sont liées par les relations

$$\begin{aligned} b &= a + \delta_1, \\ c &= a + 2\delta_1 + \delta_2, \\ d &= a + 3\delta_1 + 3\delta_2 + \delta_3, \\ e &= a + 4\delta_1 + 6\delta_2 + 4\delta_3 + \delta_4, \\ &\ldots \end{aligned}$$

on aura

$$N = \frac{a}{B^2-1} + \frac{\delta_1}{(B^2-1)^2} + \frac{\delta_2}{(B^2-1)^3} + \cdots + \frac{\delta_{n-1}}{(B^2-1)^n}$$

les différences $\delta_n, \delta_{n+1}, \delta_{n+2}, \ldots$ étant égales à zéro. Si

$$b = aq, \quad c = aq^2, \quad d = aq^3, \quad \cdots$$

on aurra $N = a/(B^2-q)$. Ces formules ont-elles été publiées?
La solution de l'équation

$$N = \frac{x}{B^y-1} + \frac{\delta_1}{(B^y-1)^2} + \frac{\delta_2}{(B^y-1)^3} + \cdots + \frac{\delta_{n-1}}{(B^y-1)^n}$$

où $x, y, \delta_1, \delta_2, \ldots, \delta_{n-1}$ sont inconnues ne semble pas offrir de sérieuses difficultés. Il n'en est pas de même de l'équation

$$N = \frac{x}{B^y - z}$$

et je serai reconnaissant à qui pourra m'en donner la solution (...). [Le problème a été résolu par Brocard, p. 11 du volume 13]

- [Volume 13, publié en 1906, p. 163 (ref 996 A31)] Je désirerais une solution *algébrique* de l'équation

$$mA^x - By^{f(x)} = C$$

que je sais seulement résoudre par l'arithmétique, lorsque les quantités $A, B, C < B$ et m étant remplacées par des nombres entiers, les inconnues x et y ont des valeurs entières et positives.

There are more contributions by Delastelle in *L'Intermédiaire*: in Vol. 8, p. 31 and Vol. 12, p. 218. Unfortunately we could not find these issues in archives.

The Philosophy of Secrecy: Towards a Historical Analysis of Cryptography, Privacy, and Information Organization

Harry Halpin
Nym Technologies
Place Numa-Droz 2
2000 Neuchâtel, Switzerland
harry@nymtech.net

Abstract

The philosophical definition of privacy is conflated with the secrecy of individual life as guaranteed by the nation-state. We trace the origin of this conception of the nation-state as the guarantor of liberal privacy, and in parallel investigate the claim (by Schmitt) that the historical origin of the modern nation-state is given by the keeping of secrets. From these contradictory claims, we show how the phenomenon of state secrecy and the surveillance of citizens is inherent in the historical development of sovereignty. Finally, we demonstrate the centrality of the history of cryptography to the philosophy of history.

1 Introduction

Within cryptography, the goal of preventing access to information has long lacked philosophical analysis and only recently been subjected to rigorous historical analysis in English (Kahn, 1967).[1] Within its long history, only in the last decades has cryptography been employed to enforce the fraught philosophical concept of privacy. This narrative is held by Whitfield Diffie, who stated that he co-invented public key cryptography to enforce "an individual's privacy as opposed to government secrecy" (Levy, 1994). The prevailing doctrine of privacy holds that it is a relatively new individual right which has only come to the forefront due to its ease of violation by technological developments such as photography and mass media (Westin, 1967). The right to privacy is historically enshrined in legal protections by the state itself rather than technical protections. Thus, due to failures of governments to enforce the right to privacy, cryptography can be used as an privacy-enhancing technology by individuals to enforce control over their own secrets, from any adversary including their own government.

This state of affairs, in which privacy is under threat by private (often corporate actors) and out-of-control governments, could be considered to be a mutation within capitalism (Zuboff, 2018). We would like to turn such a notion on its head. As shown by the history of cryptology, secrecy is constitutive of information organizations of the modern state. This inversion allows us to then consider the increase of government secrecy and the mass surveillance of their own populations to be a historical continuity rather than aberration of the history of the state. It also allows us to reconsider the spread of cryptography from the state to the individual as a shift in the historical landscape of sovereignty, rather than a merely defensive position against some legal lacuna regarding privacy rights and the increasingly digital individual self.

2 The Modern Philosophical Conception of Privacy

The modern concept of privacy is widely considered to have been historically inaugurated in the 19th century by Warren and Brandeis in their *Right to Privacy* (1890). They state that "the protection afforded to thoughts, sentiments, and emotions [...] so far as it consists in preventing publication, is merely an instance of the enforcement of the more general right of the individual to be let alone." Shortly thereafter, the United States assembled a patchwork of common law provisions to protect individual privacy and the United Nations put forward a right to privacy in Article 12 of the Universal Declaration of Human Rights, despite the lack of common constitutional legal precedent: "No one shall be subjected to arbitrary interference with his privacy, family, home or cor-

[1] It should be noted that pre-modern cryptography was systematized in a number of works by German historian Aloys Meister (1902).

respondence, nor to attacks upon his honour and reputation." Europe then followed with a constitutional approach to privacy, reaching its apogee with the Data Protection Regulation that applies principles of consent to an individual's personal data in terms of computer processing.

Despite these seemingly solid legal precedents, considerable philosophical trouble remains even in philosophical legal theory, as there is a long-standing argument over what 'privacy' means (for example, whether privacy constitutes a new right or can be reduced to property rights (Posner, 1977)), with prominent theorists such as Solove claiming that there is no unified philosophical definition (Solove, 2005). At the same time, there has been a revival of interest in using cryptography to preserve the secrecy of personal data and the individual right to privacy (Diffie and Landau, 2010). Although there is newfound philosophical interest in digital privacy in the age of digital data (Véliz, 2021) and the history of privacy as a concept (Vincent, 2016), an historical analysis of the development of privacy and secrecy is missing. This is sorely needed given the well-known history of cryptography in terms of government secrecy (de Leeuw and Bergstra, 2007). Yet this seeming opposition between government secrecy and individual privacy via cryptography may be more dialectical than it first appears.

3 Historical Origins of Privacy

The origin of privacy may very well lie in the murky past of evolution, as even animals such as birds seem to have some natural inclination to privacy in terms of withdrawing into distinct dwellings or territories (Klopfer and Rubenstein, 1977), and the majority of tribes preferred human mating in private dwellings rather than in public (Ford and Beach, 1951). The code of Hammurabi also explicitly creates a law against the intrusion of people into the private dwelling of others (Konvitz, 1966). This division between the public and private as distinct realms becomes explicitly formalized in Aristotle, where the *public* as it relates to decision-making and argument in the polis is given immense value, in contrast to the private realm of domestic dwelling (Reeve, 1998). Public life is given a positive valence with humans being defined as public beings by Aristotle (Reeve, 1998) and the abolition of private life by a communism of women and commodities being put forward as an ideal political organization in Plato's *Republic* (Reeve, 2004).

In contrast, the private is typically given a negative valence as it belongs to the *oikos*, the household and familial life of an individual life which pre-exists, and so Aristotle thought to be inferior to the public life of the polis. The *oikos* is where infamously *idiotes* are constrained, such as women and slaves that are excluded from public life (Reeve, 1998). In this regard, the term 'privacy' and 'deprivation' both descend from *privatus*, being apart from the state and so in a deficient state of being (McStay, 2014). This predominant understanding of privacy as inferior to public life continued in Europe, with the Church in the medieval era functioning as a public administration of spiritual matters, although a private relationship of the clergy and 'holy men' to God retained importance and even became predominant with the rise of Protestantism (Vincent, 2016).

The defense of privacy as a liberal individual virtue then comes from a minoritarian reading of the ancient Greeks, as Socrates admits that his private questioning of virtues would not be possible as part of a public political discussion (West, 1979). Likewise, Aristotle recognizes the virtues of private intellectual pursuits, such as scientific work, as not being easily accomplished as part of public politics (Reeve, 1998). However, this underground tradition of privacy as a realm of virtue was not defended by law in ancient Greece, as is shown by the death of Socrates. The ascendance of the concept of the state as a commonwealth to defend an individual's property rights appears in the 17th century in Locke's *Second Treatise of Government*; privacy is then surprisingly given a positive reading as an individual is guaranteed a private domain – constituted by their property – beyond the possible tyranny of public political life (1689).

Privacy is again described as a virtue in *Utilitarianism* by Mill, which argues that a public government should only interfere in private liberties in order to prevent harm to others, as otherwise the state should guarantee the secrecy of an individual's private life (Mill, 1859). In this vein, Warren and Brandeis' definition of privacy as a legal right to be left alone by (1890) makes sense as a virtue, with the life of ordinary citizens being kept secret by default in contrast to public life. Thus it also follows that the secrecy of private life

is not obtained by prisoners who have violated the implicit legal social contract of liberal societies, although a certain right to mental privacy is obtained even in Bentham's *Panopticon*, namely that "it is to make them not only suspect, but be assured, that whatever they do is known" while leaving "thoughts and fancies to their proper ordinary, the courts above" (Bentham, 1791). With the advent of the Snowden revelations, it appears that the digital panopticon has come to pass.

4 Secrecy as the Foundation of the State

A history of secrecy – including the uses of cryptology – shows the inverse of the liberal hypothesis that the individual's life is the locus of secrecy: secrecy is the province of the state. There are three theories of the birth of the modern nation-state as given by political philosopher Carl Schmitt (Caygill, 2015). In his most well-known theory, the concept of the modern nation-state descends from a secularization of inherently theological concepts (Schmitt, 2005). His second theory is that the emergence of the nation-state was necessary to quell the internecine civil wars of religion at the end of the medieval era (Schmitt, 2008). The last lesser-known theory is that the nation-state descends from the keeping of *arcanum*, or secrets (Schmitt, 1996).

The rise of the literate priestly class that ruled early city-states and empires was at least in part due to their control over information, originating in their knowledge of the valuable relation between agriculture and time (as given by astronomy). This control of information was later extended into records of the storage of food surplus, which later expanded into the storage and transmission of information about resources of all types (Innis, 2022). This information was originally secret, as witnessed by early secret fertility cults and their relationship in Sumer to the anointing of the first kings. Early writing, which was not widely known, could have been considered to be 'secret' knowledge to the vast majority of illiterate people of ancient Mesopotamia. As time progressed in some civilizations such as China, literacy remained the province of a minority in employ of the government, but in other civilizations such as Mesopotamia the spread of literacy caused the invention of what is called 'secrecy statements', where in a text it was explicitly forbidden by the author to share the knowledge in writing with anyone except those with explicit permission given by their position or caste. This literally restricted the information to "one who knows" (*mūdû*), creating a tradition of making a document 'classified' (Mohr, 2022). Thus, we find at the very origins of civilization that writing was an apparatus of information organization on the threshold between the theological and the political insofar as writing was used to maintain the secrets of the emerging state. As more and more people became literate and began to violate these secrecy statements, what appears to be early cryptographic substitution ciphers developed in the historical record of Mesopotamia, protecting both the secrets of religion and the state. As public literacy also increased during the Roman empire, we see the return of substitution ciphers like the Caesar cipher and possibly even more complex ciphers (Reinke, 1962).

In this regard, ancient Greece's notion of the state as an absolutely public space was just as much of a mutation as the development of philosophy as the function of public reason, as witnessed by the persistence of fertility cults based on secret knowledge in ancient Greece and the continued use of ciphers by Greek city-states (Reinke, 1962). The lack of emphasis on privacy as a virtue in individual life in the medieval era was not due to the persistence of the notion of public virtues and private vices from Plato (Reeve, 2004), but was in a sense proportional to the growth of the power of secret knowledge in the Church, which naturally dealt with secrets due to the lack of public literacy in Latin (Innis, 2022). With the decline of empires and the rise of the power of the church came the rise of cryptology being applied to esoteric biblical secrets (Ellison, 2016), and the quest to discover the 'true' (Adamic) names of beings, a tradition transmitted in part from the Arabic medieval world to the early alchemists (Al-Hassan, 2004). However, the relationship between the spread of literacy and the need for cryptography by empires returned to the historical scene with widespread literacy in Arabic in the Middle East and Africa. In parallel to the quest for Biblical knowledge in Europe, the religious impulse to discover hidden knowledge in the Quran re-ignited the field of cryptology, as shown by the work of Al-Farahidi on permutations and the use of frequency analysis of Al-Kindi to break substitution ciphers; this work was politically mobilized by the rise of the

Arabic caliphates, who then weaponized advanced cryptography – as shown by the work of Ibn Adlan on cryptanalysis – to make their own internal communications secret (Schwartz, 2014).

This cryptographic work was then translated and merged with the alchemical tradition, where an *arcanum* of secrets was both simultaneously mystical and practical knowledge (for example, guild secrets as *arcana artis*). This later branched, with the more mystical side of arcana becoming *occultum* (occult) while almost any practical knowledge could become a *secretum*, or secret (including industrial trade secrets such as that of silk production). The use of cryptography became widespread amongst early scientists and intellectuals to defend themselves from the church, as demonstrated by the use of cryptography by Galileo in his trial (Marcus and Findlen, 2019). After the rise of the classical era of Renaissance cryptology given by Cardano, de Vignere, and Della Porta, the art of encoding and deciphering secrets became a profession in and of itself, with professional cryptographers being employed in the diplomacy of Italian city-states in the 15th century (Strasser, 2007). As the civil wars in the rest of Europe came to an end, these techniques were then absorbed into the emerging order of sovereign nation-states of France by cryptographers such as Rossignol and Wallis in England, leading to a new diplomatic order of cryptographers and cryptanalytic 'black chambers' throughout Europe. Although there was some usage of cryptography to conceal scientific discoveries and for other personal purposes (Lochrie, 2011), skilled cryptographers like Wallis were generally put into state service. This increasing monopoly of the nation-state on cryptography became one of the defining aspects of Europe that continued into the era of the world wars and the invention of digital computing, and the invention of modern cryptography by Shannon was originally classified as well (Shannon, 1945). It is only with the invention of public-key cryptography that cryptography became a matter of individual knowledge to the public (Diffie and Hellman, 1976).

5 Contradictions of Secrecy and Privacy

Simmel defines secrecy as the control of information, either by individuals or organizations (ranging from secret societies to the states) (1906). This definition parallels Nissenbaum's definition of privacy as the appropriate flow of information in a social context (2020). A secret has been defined by Bok as the attempt "to block information about it or evidence of it from reaching that person, and to do so intentionally" (Bok, 2011). A secret is a social relation, where information may only be transmitted to its intended reserpine(s). Of course, there is a difference between privacy and secrecy; Bok continues to define privacy as "the condition of being protected from unwanted access" (Bok, 2011), such that secrecy is considered to be non-consensually enforced, while privacy is typically assumed to be consensual. As shown earlier, privacy is ultimately a concept that is indexed to the development of the concept of a sovereign and autonomous human individual whose right to privacy is protected by law and with consent. On the other hand, secrecy is a much wider concept, whose evolution – while grounded also in individuals and networks of trade in Mesopotamia, as per the infamous example of protecting a secret recipe for ceramic glaze (Pearce, 1982) – is ultimately tightly interwoven with the rise of the political theology of state and the formation of hierarchy. As bluntly illustrated by the frontpiece of *Leviathan* by Thomas Hobbes, the state can be considered to be composed of individuals, so conflict is ontologically nullified. Yet this naïve view of the state formation ignores the rise of a ruling class, from kings and scribes to modern-day bureaucrats and cryptographers who are separated from the general population due to their information organization in terms of secrets. Contra Bok, privacy could be simply the application of secrecy to the individual, as her definition does not include the active blocking of access to information, such as via cryptography. Cryptography is simply the technical application of secrecy to information in the presence of adversaries.

It is precisely in this contradiction between state secrecy and individual privacy that the importance of the history of cryptography to the broader philosophy of history is revealed. The state as a purveyor of secrets may guarantee the private lives of citizens from each other, but not from the state itself, which requires transparency from its constituents in direct contradiction to the liberal conception of privacy outlined earlier. Contra Habermas (1985), the state itself is an information organization for the control and transmission of secrets, rather than a publicly transparent space of

democratic decision-making and communication as imagined by the Greeks. This is reflected in how the ancient concept of Greek *polis* arose from the *stasis* – civil war – provoked by persistence of private ties between individuals given by the *oikos*, insofar as the civil war was only resolved as the private lives of individuals were subordinated to the transparent public life of the state (Agamben, 2015). This transparency was viewed as a voluntary virtue in Greece, but has now become an enforced prescription by the state under the rubric of preventing civil war, and is no mere (perhaps temporary) mutation within capitalism (Zuboff, 2018).

This is exemplified by the state simultaneously using cryptography to enforce its own secrets against both other nation-states and its own population. This explains how legislation to increase government transparency is opposed while legislation to increase surveillance powers is supported by the state, and the practice of mass surveillance grows more powerful regardless of the law. In this manner, individuals like Julian Assange who expose state secrets weaken the monopoly of secrecy of the state and so naturally become enemies of the state. On the other hand, the spread of cryptographic techniques such as public-key cryptography outside of the state creates an inevitable schism. Cryptography then both empowers non-state actors (such as multi-national corporations and their digital platforms) to build forms of sovereignty via secrecy at the expense of the nation-state, while individuals can also use cryptographic techniques to preserve their liberal right to privacy technically, rather than only as a legal right that a state may not consent to due to its own security concerns. The public availability of cryptography can then be studied as a historic shift the focus of sovereignty to individuals that wish to escape the transparency that the state enforces on their own population, creating new forms of sovereignty, which leads to the inexorable need for nation-states to break the cryptographic technqiues used by individuals (including those in their own populaton) with the same concern once reserved for competing nation-states.

6 Conclusion

There is perhaps one aspect of being human that serves as the foundation for secrecy: the ultimate interiority of the human mind. With the advent of ubiquitous digital technology, the increased exteriorization of what appeared to be inner cognitive functioning has accelerated, even for the individual (Clark and Chalmers, 1998). This in turn signals the increasing importance of cryptography in terms of maintaining human individual autonomy via secrecy. As the individual becomes more embedded in collective and social technical apparatuses, the concept of privacy may very well be replaced by a broader notion of autonomy. The importance of secrecy as enforced by cryptography will only increase; the information organizations of the future that may very well succeed the Westphalian nation-state will also be based on secret communications between networks of super-empowered individuals, as prefigured in corporations and other social networks. A philosophical analysis should also not only cover the content of communication, but also the structure of the communication network (metadata), as can be defended by technical means like mixnets (Chaum, 1981). A more thorough theoretical meta-analysis is needed with increased reliance on a diversity of archival sources. Note that this treatment has been relatively limited in scope to Europe, and further work should be done in terms of other civilizations such as China and India, with their own histories of privacy, cryptography, and sovereignty. A history of cryptography can be conceived as not only a history of the nation-state, but as the philosophical and political conceptions of sovereignty itself and so an *arcanum* of the philosophy of history.

References

Giorgio Agamben. 2015. *Stasis: civil war as a political paradigm*. Stanford University Press.

Ahmad Al-Hassan. 2004. The Arabic Original of Liber de Compositonae Alchemiae: The epistle of Maryānus, the hermit and philosopher, to Prince Khālid ibn Yazīd. *Arabic sciences and philosophy*, 14(2):213–231.

Jeremy Bentham. 1791. *The Panopticon Writings*. Verso Books (reprinted 2020).

Sissela Bok. 2011. *Secrets: On the ethics of concealment and revelation*. Vintage.

Howard Caygill. 2015. Arcanum: The secret life of state and civil society. *The Public Sphere from Outside the West*, pages 21–40.

David Chaum. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90.

Andy Clark and David Chalmers. 1998. The extended mind. *Analysis*, 58(1):7–19.

Karl de Leeuw and Jan Bergstra. 2007. *The History of Information Security: A comprehensive handbook*. Elsevier.

W Diffie and M Hellman. 1976. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654.

Whitfield Diffie and Susan Landau. 2010. *Privacy on the line: The politics of wiretapping and encryption*. The MIT Press.

Katherine Ellison. 2016. *A cultural history of early modern English cryptography manuals*. Routledge.

Clellan S Ford and Frank A Beach. 1951. *Patterns of sexual behavior*. Harper Press.

Jürgen Habermas. 1985. *The theory of communicative action: Volume 1: Reason and the rationalization of society*. Beacon Press.

Harold Innis. 2022. *Empire and Communications*. University of Toronto Press.

David Kahn. 1967. *The Codebreakers: The comprehensive history of secret communication from ancient times to the internet*. Simon and Schuster.

Peter Klopfer and Daniel Rubenstein. 1977. The concept privacy and its biological basis. *Journal of Social Issues*, 33(3):52–65.

Milton Konvitz. 1966. Privacy and the law: A philosophical prelude. *Law and Contemporary Problems*, 31(2):272–280.

Stephen Levy. 1994. Battle of the Clipper Chip. New York Times.

Karma Lochrie. 2011. *Covert operations: The medieval uses of secrecy*. University of Pennsylvania Press.

John Locke. 1689. *Second Treatise of Government*. Hackett Publishing (reprinted 1980).

Hannah Marcus and Paula Findlen. 2019. Deciphering Galileo: Communication and secrecy before and after the trial. *Renaissance Quarterly*, 72(3):953–995.

Andrew McStay. 2014. *Privacy and Philosophy: New media and affective protocol*. Peter Lang.

Aloys Meister. 1902. *Die Anfänge der modernen diplomatischen Geheimschrift*. Ferdinand Schöningh.

John Stuart Mill. 1859. *On liberty, utilitarianism, and other essays*. Oxford University Press (reprinted 2015).

Sara Mohr. 2022. *Secrecy, Protection, and the Foundations of Knowledge in Ancient Mesopotamia*. Ph.D. thesis, Brown University.

Helen Nissenbaum. 2020. *Privacy in Context: Technology, policy, and the integrity of social life*. Stanford University Press.

Laurie Pearce. 1982. *Cuneiform cryptography: numerical substitutions for syllabic and logographic signs*. Yale University.

Richard Posner. 1977. The right of privacy. *Georgia Law Review*, 12:393.

Charles Reeve. 1998. *Aristote's Politics*. Hackett Publishing.

Charles Reeve. 2004. *Plato: Republic*. Hackett Publishing.

Edgar C Reinke. 1962. Classical cryptography. *The Classical Journal*, 58(3):113–121.

Carl Schmitt. 1996. *Roman Catholicism and Political Form*. Greenwood Publishing Group.

Carl Schmitt. 2005. *Political Theology: Four chapters on the concept of sovereignty*. University of Chicago Press.

Carl Schmitt. 2008. *The Concept of the Political: Expanded edition*. University of Chicago Press.

Kathryn Schwartz. 2014. From Text to Technological Context: Medieval Arabic Cryptology's Relation to Paper, Numbers, and the Post. *Cryptologia*, 38(2):133–146.

Claude Shannon. 1945. A mathematical theory of cryptography. Technical report, Bell Labs.

Georg Simmel. 1906. The sociology of secrecy and of secret societies. *American Journal of Sociology*, 11(4):441–498.

Daniel Solove. 2005. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154:477.

Gerhard Strasser. 2007. The rise of cryptology in the European Renaissance. In *The History of Information Security*, pages 277–325. Elsevier.

Carissa Véliz. 2021. *Privacy is Power*. Melville House.

David Vincent. 2016. *Privacy: A short history*. John Wiley & Sons.

Samuel Warren and Louis Brandeis. 1890. The right to privacy. *Harvard Law Review*, 4(193).

Thomas West. 1979. *Plato's Apology of Socrates: an interpretation, with a new translation*. Cornell University Press.

Alan Westin. 1967. *Privacy and Freedom*. Athenum.

Shoshana Zuboff. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

OVERLOOKED, FORGOTTEN, MISUNDERSTOOD
The "Other" SIGINT in World War II

David A. Hatch
Center for Cryptologic History
alldavid@aol.com

Abstract

Fascination with ULTRA, the extraordinary World War II cryptologic intelligence, by historians and the public alike has eclipsed knowledge of and study of a second important cryptologic intelligence source. Whereas ULTRA provided senior Allied leaders with deep insight into enemy strategic thinking and plans, tactical cryptologic intelligence often gave combat commanders the vital information needed to win actual battles. Tactical cryptologic intelligence is less known and studied partly because it is in the shadow of ULTRA. In addition, fewer sources about the practice of tactical cryptologic intelligence have survived, and those readily available are fragmented and anecdotal. Tactical cryptologic intelligence merits further research and analysis if we are to have a fuller understanding of cryptology in practice and intelligence in the world war.

When the subject of cryptology in World War II is raised, most everybody's first thought is ULTRA, the Allied exploitation of the German ENIGMA machine and other machine-generated cryptography. The topic almost automatically brings to mind exciting stories relating how Allied analytic successes against Axis cryptographic systems enabled good strategic decisions and facilitated victory by the Allied powers in World War II.

Declassification of the "ULTRA secret" in the early 1970s revealed an important but previously-unknown aspect of World War II intelligence. The new revelations fascinated public and professional historians alike --- all now had to revise what they thought were concrete facts about decision-making in the world war. The professional and lay fascination with ULTRA has not dimmed in the decades since. (Winterbotham, 1974)

ULTRA is a prime example of signals intelligence (SIGINT), a secret but common practice of most governments and militaries. SIGINT is a process consisting of collection of target communications, i.e., eavesdropping, then solution of cryptographic systems, and translation. Since the discipline of cryptanalysis remains central to the process, in Great Britain and the United States, the agencies and processes involved are designated "cryptologic."

During World War II, ULTRA was a shared enterprise of Great Britain, the United States, Canada, Australia, and New Zealand. Decrypts of high-grade Axis cryptosystems were produced by specialized personnel at secret centers in each country, the most famous of which was Bletchley Park in the UK. The product was distributed to an extremely limited audience of the most senior civilian and military officials in the five countries. (Benson, 1980)

This aspect of World War II is fairly well known today, and most histories of the war published since the 1970s have included at least a mention of ULTRA, often much more. However, it is less well known that the allies had a second SIGINT/cryptologic effort, one that also had a significant effect on military operations, but it is little mentioned today. Tactical SIGINT.

From the earliest days of the war, the Allied nations supported their fighting forces with tactical SIGINT units that accompanied deployed troops wherever they might be dispatched. These self-contained units performed all functions in the SIGINT cycle: intercept, cryptanalysis and/or

traffic analysis, translation, and distribution to the local commander. Whereas ULTRA product was distributed to a small number of senior officers, plus their senior intelligence staffers, and Army level (sometimes Corps level), tactical SIGINT usually went to corps-level commanders or below.

The Bletchley Park historian notes that after the landings in Normandy in June 1944, ULTRA remained a valuable source for senior officers needing a broader view of enemy intentions, but often tactical SIGINT was more important for daily combat operations. (Kenyon, 2019)

ULTRA contained insights into the enemy's strategic thinking, sometimes including operational plans, information appropriate for Allied senior commanders. While senior leaders could not share ULTRA with their subordinate commanders for security reasons, insights gained from ULTRA were reflected in the orders sent to subordinates from the top.

In fact, commanders operating near the front really did not need ULTRA. However valuable the texts to senior leaders, the decrypts generally did not apply to lower-echelon officers with more limited areas of responsibility. Division commanders primarily needed accurate information on enemy deployments and actions just beyond their horizon. This is what tactical SIGINT provided them.

Whereas ULTRA producers targeted high-grade cryptographic systems, such as the German ENIGMA machine or the Japanese Navy's JN-25, tactical SIGINT units worked against lower-level systems. Communications at or near the front lines frequently was encrypted, but usually in less difficult systems, "paper and pencil" ciphers that did not involve cumbersome machines or codebooks and could be used quickly. Trained cryptanalysts in a tactical unit usually could solve these systems without the heavy machine support needed for ULTA-level intercept.

Before the war, both Great Britain and the United States trained units in tactical intercept of enemy communications, intending them to provide direct combat support should war break out. This training was good, but did not reflect the reality the SIGINT units later encountered in actual combat. The British experience in North Africa against German tactical communications honed the talents of their combat SIGINT personnel, and UK veterans from the early days of the desert war shared the lessons from this direct experience with their American colleagues, once the United States entered the war.

As it deployed worldwide after 1942, the US Army established cryptologic headquarters in each theater of war to serve as a coordinating organization for tactical units operating in the war zones. In practice, tactical units served two masters: both theater SIGINT headquarters, and the intelligence staffs of the combat units to which they were assigned.

For the United States military, tactical SIGINT units were deployed at company strength. The usual practice in the field, however, was to split the company into multiple detachments that could be deployed individually in several areas within the combat zone assigned to a particular combat unit.

Tactical units were intended to operate in fast-changing situations, and would work in large trucks on which the container area had been converted for office work. The team would include intercept operators, cryptanalysts, and translators. Where possible, as in the campaign in Western Europe, tactical SIGINT companies would commandeer abandoned buildings for both operational use and housing.

Take, for example the Ardennes Offensive, commonly known as the Battle of the Bulge, in 1944. German units assembled and attacked under conditions of radio silence, thus preserving the element of surprise; neither ULTRA nor tactical SIGINT provided any warning. Once the attack was blunted, forcing German units to

retreat, they became chatty on the radio. US and British tactical SIGINT teams exploited these communications, enabling Allied success in regaining lost ground. (SRH-112)

Tactical cryptologic teams often warned about impending air raids. Their analysis of enemy communications enabled them to identify specific enemy units facing theirs, detect unit replacements or movements, or hear activities that might portend offensive actions.

During World War II, US air forces were subordinate to the Army, and received their SIGINT support through Army channels. A number of tactical SIGINT units received specialized training in air communications, and were allocated to work against the other side's air forces; these specialized units had the designation "radio squadrons mobile."

The US Navy also deployed tactical SIGINT units during combat operations. The first such unit accompanied American aircraft carriers on a series of harassment raids against Japanese-held islands just days after the United States entered the war. The SIGINT units proved so valuable from this starting point, the Navy trained more personnel and assigned a unit to each task force, sometimes to individual components of naval task forces. As in the Army, Navy tactical SIGINT units were self-contained, including intercept operators, cryptanalysts, and translators, the full range of necessary skills in each team.

In the US Navy, the primary task for the tactical SIGINT units was force protection --- to alert the task force commander at the earliest possible moment after the Japanese had detected the US force's presence, and then to give warning about enemy aircraft dispatched against American ships during combat operations.

The Royal Navy also used tactical cryptologists for force protection. For example, teams of tactical cryptologists many times provided warnings that enabled the protection of supply convoys. (Kenyon, 2023)

Who were the tactical SIGINT personnel?

In the United States, Army and Navy recruitment personnel were instructed to identify those who did especially well on aptitude tests, particularly in mathematics, and shunt those recruits or volunteers to the cryptologic services. Recruitment personnel also did the same with those who claimed knowledge of foreign languages. The US Army trained soldiers in field cryptology at several Signal Corps military posts within the United States, and also gave specialized language training to the tactical personnel.

Recent publications have rightly emphasized the wartime role of women in SIGINT. However, the thousands of women SIGINT personnel worked only on ULTRA; there were no tactical "code girls." Tactical SIGINT units operated near the front; US and British policy in the war barred women from combat roles.

The five Allied countries that shared ULTRA also had by far the largest and most widely deployed tactical cryptologic units. But, they were not entirely alone in this endeavor.

More research remains to be done on German SIGINT in the world war, but it appears that the uncoordinated German SIGINT endeavor operated primarily from fixed sites. One exception was a company of the 56th Signals 4

Battalion, a highly-proficient tactical SIGINT unit that supported General Rommel in North Africa. The National Archives in the United States and Great Britain hold many German wartime SIGINT documents and interrogation reports of practitioners, primarily strategic cryptologists, but also including some at the strategic level; this is, in the author's opinion, an underutilized resource. (Kahn, 1978; Bennett, 1989)

Polish forces in exile fought alongside the British Army in Italy, and had a Polish tactical SIGINT unit in support. The unit had been raised, trained, and equipped by the British, but operated directly with their countrymen in combat. (Skillen, 1989)

Why are tactical cryptologic operations in the war so little-known?

In the first place, material on tactical operations has been overwhelmed by the rich, entertaining, and valuable ULTRA material. There seems to be an endless ULTRA treasury containing not only technical data on cryptography and cryptanalysis, but also attention-grabbing stories that provide new twists to familiar wartime episodes. Although intelligence from tactical SIGINT often was crucial in the military operations of the day, in retrospect the stories seem mundane and are no threat to the attention-getting ULTRA material.

Because tactical units operated in forward areas and were vulnerable to capture, they destroyed their operational records very soon after they were used. The reports were tactical in nature, so often were not preserved at the intelligence headquarters, either. Thus, little remains in terms of original documentation about the activities and successes (or failures) each unit had.

Much of the surviving information about tactical SIGINT is administrative. During the war itself, tactical SIGINT units sent monthly operational reports to the senior SIGINT office in their theater of operations, but these were mostly technical data, both about their personnel situation and statistics (about the amount of enemy traffic processed, etc.). At the end of the war, many of these units wrote histories of their activities, but this was generally make-work for personnel who did not have sufficient service time to be demobilized, and was often written from memory not documentation.

Both the monthly reports and wrap-up histories, amid the statistics, contain anecdotes about the work of the units, but they are fragmented and often lack context. These histories also often contain the pawkish and acerbic comments that famously characterized American GI humor in the war, along with plenty of complaints about the inequities of life at the front.

What are the sources about tactical SIGINT in World War II?

A good administrative history is George Howe's *American Signals Intelligence in North Africa and Western Europe*. Howe as a US Army historian wrote the Army's standard volume on World War II in North Africa, and later in his career joined the National Security Agency, and documented the SIGINT aspects of the story. (Howe, 1980)

In Britain, Hugh Skillen, a veteran of tactical cryptologic operations, for decades sought to document the wartime effort. He held conferences and published privately several anthologies of reminiscences from tactical SIGINT personnel (primarily British, but a few key Americans.) (Skillen, 1989)

The National Security Agency has made available many of the war's end summary reports from tactical cryptologic units. These are found within a larger collection of declassified documents known as SRHs (Special Research History).

Some participants in tactical SIGINT operations, such as Charles David, have written articles or portions of larger memoirs that shed light on day-to-day work in the field. Unfortunately, there is no central reference point for these writings. (David, 1996)

Many personal accounts of tactical SIGINT activities appear in a series of conference papers edited by Hugh Skillen, *The ENIGMA Symposium*, with editions for 1992 through 2003. Despite the title, the volumes deal with tactical SIGINT in all theaters of World War II.

The way forward in studying wartime tactical cryptology is, of course, to synthesize the official reports at the US and British national archives, incomplete though they may be. This material may be supplemented with individual accounts of practitioners and users: many universities and large local libraries have conducted oral history interviews with veterans, including cryptologists, and collected their private memoirs. An initial

survey of the US Army Heritage and Education shows that their holdings include papers and interviews from senior officers who worked in cryptologic endeavors early in their careers, and can provide perspective on the tactical efforts.

It has been clear since the 1970s that the Allied success in the cryptologic aspects in all theaters of war in World War II was a key factor in the ultimate military victory. While most attention has been focused on ULTRA, more attention needs to be paid to tactical SIGINT as well. ULTRA kept Allied senior leaders exceptionally well informed in terms of grand strategy, but tactical SIGINT was also a necessary ingredient for victory at the frontline level, where lives were at risk and the success or failure of strategy was at stake.

This is a gap in knowledge that can and should be filled.

References

Ralph Bennett, 1980. *ULTRA and Mediterranean Strategy.* Morrow, NY, NY.

Robert Louis Benson, 1980. *History of U.S. Communications Intelligence During World War II: Policy and Administration.* Center for Cryptologic History, Fort Meade, MD.

Charles David, "A World War II German Army Field Code and How We Broke It.," *Cryptologia*, January 1996: 5-76.

George Howe, 1980. *American Signals Intelligence in North Africa and Western Europe*. Center for Cryptologic History, Fort Meade, MD.

David Kahn, 1978. *Hitler's Spies.* MacMillan, NY, NY.

David Kenyon, 2023. *Arctic Convoys: Bletchley Park and the War for the Seas.* Yale University Press, New Haven, CT.

David Kenyon, 2019. *Bletchley Park and D-Day: The Untold Story of How the Battle for Normandy Was Won.* Yale University Press, New Haven, CT.

Hugh Skillen, 1989. *Spies of the Airwaves: A History of Y Sections during the Second World Wa*r.

Hugh Skillen. *The ENIGMA Symposium*, editions for 1992, 1994, 1995, 1997, 1998, 1999, 2000, 2001, 2002, 2003. 6

SRH-112. "Post Mortem Writings on Indications of the Ardennes Offensive." National Security Agency, National Cryptologic Museum, www.nsa.gov., accessed January 10, 2024.

Frederick W. Winterbotham, 1974. *The ULTRA Secret.* Harper & Row, NY, NY.

Supporting Historical Cryptology: The Decrypt Pipeline

Mihály Héder[1], Alicia Fornés[2], Nils Kopal[3], Ferenc Szigeti[1] and Beáta Megyesi[4]
[1]Budapest University of Technology and Economics, Hungary
[2]Autonomous University of Barcelona, Spain
[3]University of Siegen, Germany
[4]Stockholm University, Sweden

Abstract

We present a set of resources and tools to support research and development in the field of historical cryptology. The tools aim to support transcription and decipherment of ciphertexts, developed to work together in a pipeline. It encompasses cataloging these documents into the Decode database, which houses ciphers dating from the 14th century to 1965, transcription using both manual and AI-assisted methods, cryptanalysis, and subsequent historical and linguistic analysis to contextualize decrypted content. The project encounters challenges with the accuracy of automated transcription technologies and the necessity for significant user involvement in the transcription and analysis processes. These insights highlight the critical balance between technological innovation and the indispensable input of domain expertise in advancing the field of historical cryptology.

1 Introduction

The Decrypt project (de-crypt.org) introduces a comprehensive suite of tools designed for the transcription and analysis of historical ciphers and keys. This initiative combines newly developed tools with enhanced versions of existing technologies to facilitate the study of historical encrypted sources. In this paper, we will give an outline of the tools, some under development, some in already in production — from transcription to decipherment — to work individually as well as together in a pipeline for efficient processing of historical encrypted sources. Noteworthy is that the pipeline is made accessible to users with an account to the decode database.

2 Elements of the Decrypt Pipeline

Our work centers on historical ciphers discovered across a diverse array of sources including both handwritten and printed documents of various lengths, featuring many distinct hand-writing styles and symbol sets. These documents, originating from different time periods and geographic regions in Europe, were created for a wide range of purposes such as diplomacy, military correspondence and even unconventional mediums like postcards.

The pipeline is illustrated in Figure 1 and will be described in the subsequent sections.

2.1 Cataloging Documents: the Decode Database

Upon discovery, historical ciphers or keys must be converted to digital images, accompanied by detailed metadata. This metadata typically includes details about the document's origins, like the sender and receiver, its creation or dispatch date and location as well as information about the medium. Accurately describing these historical sources with metadata is crucial as the contents are often completely unknown. This information can be recorded and stored in the Decode database (Héder and Megyesi, 2022). To make the processing of the encrypted sources easy, we developed an interface to allow users to create their own projects from which it is possible to start to utilize the entire pipeline for the analysis of the sources. Figure 2 shows a screenshot of user-specific projects.

2.2 Transcription

Once we have digital images, the next step is to turn the image into a machine readable text format, i.e. to provide a transcription of the content of the image. This step produces a document that contains the symbols seen on the digital images in a format that allows further processing, most often

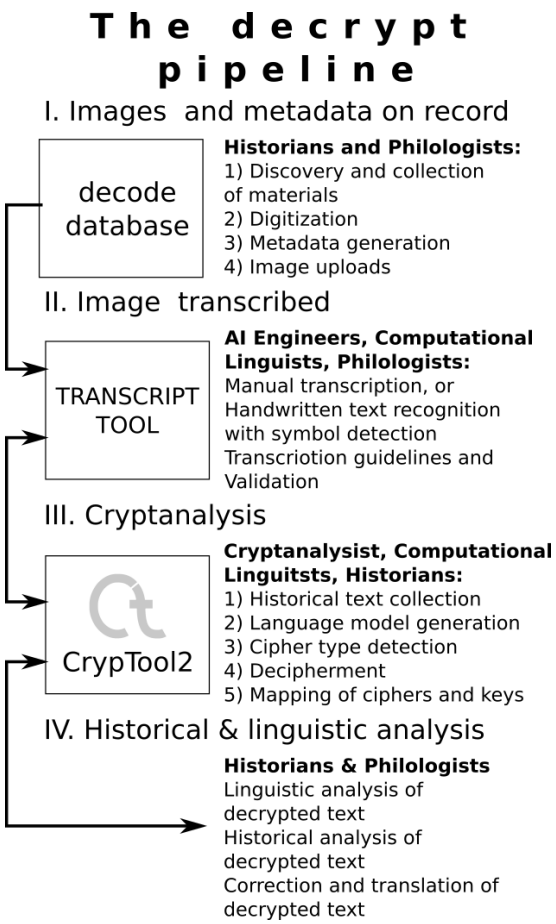


Figure 1: The Decrypt pipeline.

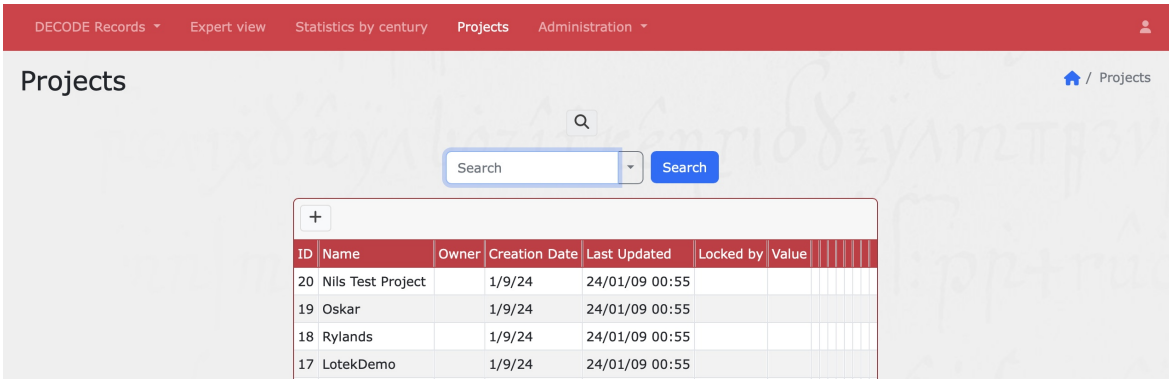


Figure 2: User-specific projects.

Unicode (UTF-8). The goal is to transcribe each symbol on the image as shown without leaving out any important information. Transcription can be carried out manually or (semi-)automatically.

George Lasry developed a tool for manual transcription called CrypTool Transcriber and Solver or shortly CTTS (Lasry et al., 2023), which was used, among others, for the transcription of the recently deciphered Mary Stuart letters. The tool is efficient but fully manual. CTTS allows the user to mark each symbol in the image and assign it into a cluster of similar symbols. Each cluster can then be assigned a label with which it can be transcribed in a unique way to result in a machine-readable text format. CTTS encourages a cyclic process of review and iteratively editing transcriptions and decryptions (Megyesi et al., 2024), which provides an excellent starting point to produce training data and manual transcriptions of shorter inscriptions that can be used to train better performing models. Figure 3 shows an image of CTTS currently being used to transcribe a ciphertext.

Transcription can be also done partly or entirely automatically using tools that are developed to specifically handle hand-written manuscripts. One of the most well-known transcription tools is Transkribus (Societas Cooperativa Europaea, 2024) which provides AI models trained on manually transcribed images for a wide range of European languages. However, these models are not directly applicable to encrypted sources, as these seldom contain character sequences known from plaintext languages. Therefore, the performance of these models are very low. We tested the three largest models of Transkribus on 20 different ciphertexts, all written as digits (0-9) and the models performed very poorly with a symbol error rate between 90% and 100%.

There is a great need of transcription tool(s) that can handle a wide range of symbol sets and handwriting styles. AI-based models can be an option, but they need large amounts of data, preferably transcribed ones to be used for training, which takes time to produce. Such AI systems requiring a lot of human interaction are usually referred to as human-in-the-loop AI. However, ciphertexts are usually short and training data is therefore often sparse. For this reason, some efforts have been put in developing tools that rely on efficient image processing algorithms and, at the same time, can learn from user feedback and corrections (Baró et al., 2019; Souibgui et al., 2021). Here, we turn our focus from a large set of manually produced training data to serve the AI to force the AI to learn from its own mistakes by correcting its output and fine-tune the model on the corrected data. We call our approach "AI-in-the-loop" (Beáta Megyesi, 2022) since the AI assists the user instead of the other way around.

The TranscriptTool (Szigeti and Héder, 2022) was developed by the Decrypt project to support the user for fast and efficient transcription of cipher images. This tool is an AI-in-the-loop system which includes models trained on large set of training data, but it is also able to learn from user input. The system takes the output of a pretrained image recognition algorithm and visualize it for the user for post-processing and error correction. The developed models are based on unsupervised clustering or few-shot learning, and they only require the manual correction of few lines. These models can be re-trained or fine-tuned with the examples provided by the user, which is of course a more resource-intensive process but leads to higher transcription performance. The TranscriptTool is illustrated in Figure 4 and described in more detail in (Megyesi et al., 2024).

CTTS and the TranscriptTool can be also used to mark the exact location of the symbols on the page (called "bounding boxes"). This way the digital image of the individual symbols can be cut out to separate files. These files can serve further machine learning projects as real, historical instances of symbols.

2.3 Cryptanalysis

With a full or partial transcription at hand, the cryptanalysis may be attempted. This may involve statistical analysis, cipher symbol clustering (Lehofer, 2022), hill-climbing, simulated annealing, heuristics, or educated guesses. Historical language models for the suspected time period and language can be of great help, especially if the ciphertext originates from the 18th century or earlier (Megyesi et al., 2023). Cryptanalysis components for historical ciphers including simple and homophonic substitution of various complexity are integrated (Lasry et al., 2021) and available within CrypTool (CrypTool Contributors, 2024).

CrypTool has been developed as a desktop version, and now also as an online version.

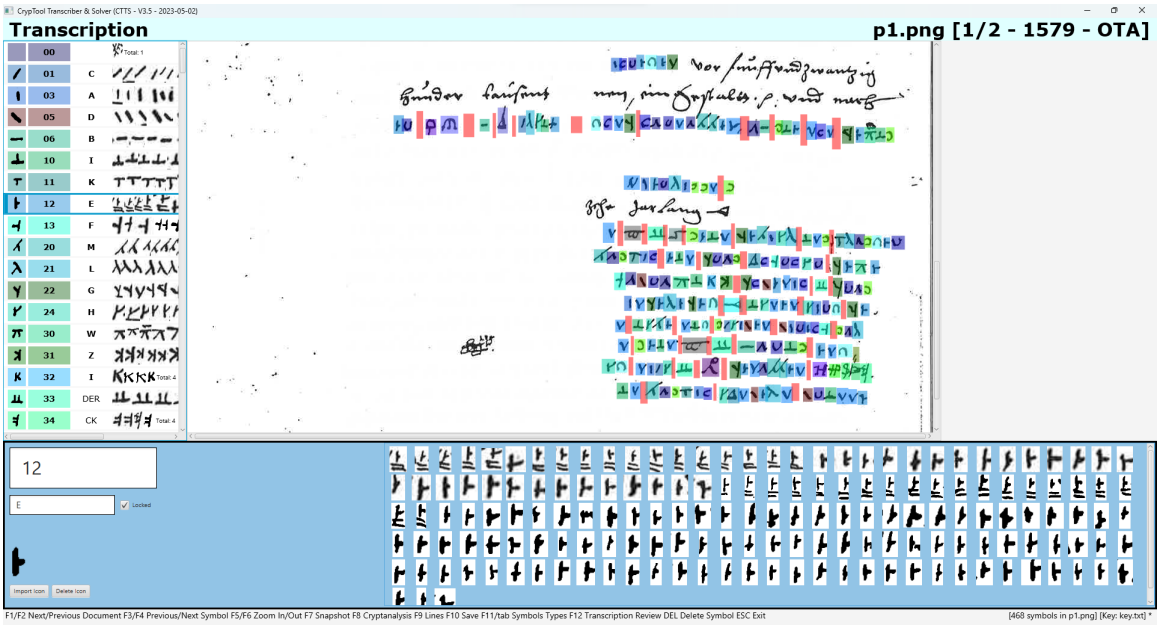


Figure 3: CrypTool Transcriber and Solver used to transcribe an image.

The desktop version, CrypTool 2 (Kopal, 2018), is aimed at more advanced users. It enables the creation of graphical programs, known as workspaces, which comprise various components for input, output, cryptography, and cryptanalysis working together to execute cryptographic or cryptanalytical tasks. A pre-created set of workspaces, referred to as templates, are included with CT2. These templates offer workflows for both cryptography and cryptanalysis, for instance, for cryptanalyzing simple substitution ciphers or homophonic substitution ciphers (Kopal, 2019). Additionally, templates designed for conducting statistical analyses (e.g., frequency analysis, Friedman test) on ciphertexts are part of the software. Currently, CT2 boasts over 250 such templates. It has been successfully utilized in cryptanalyzing different ciphertexts, such as written letters from Maximilian II from 1575 (Kopal and Waldispühl, 2022) or an encrypted letter from the Dutch East India Company (Dinnissen and Kopal, 2021). We are currently extending CT2 with the so-called DECRYPT editor, a specialized component of CT2 that facilitates easy access to the DECRYPT pipeline and provides a comprehensive set of cryptanalysis tools at a single point. Figure 5 illustrates the project overview on the left side and a selected project on the right. On the lower right side, the files associated with the project are displayed. Figure 6 depicts a frequency analysis of a ciphertext conducted using the DECRYPT editor. This tool allows for the flexible selection of the ciphertext's alphabet, enabling, for example, the analysis of digit distribution in digit ciphers. The CT2 DECRYPT editor is currently under development and will be published in the near future.

In addition to CT2, there is an online counterpart, CrypTool-Online (CTO). This web-based version does not require any installation, distinguishing it from CT2. CTO features a variety of cryptanalysis tools, notably including the new online homophonic substitution analyzer. This tool enables users to manually analyze homophonic substitution ciphers (automatic cryptanalysis is also planned) by assigning plaintext letters to ciphertext symbols. It highlights all instances of a selected symbol and automatically assigns the chosen plaintext letter to all occurrences of that symbol. A key advantage of the online version over CT2's analyzer is its support for the DECRYPT transcription guidelines (Megyesi, 2020). It accommodates not only single-letter homophones but also vowels and complex n-grams. Furthermore, the tool facilitates the annotation of nomenclature elements and inline plaintext elements, adhering to the DECRYPT guidelines. Figure 7 displays a ciphertext under analysis with the Homophonic Substitution Analyzer in CTO. Currently in beta, the online version is set to be launched in the coming months.

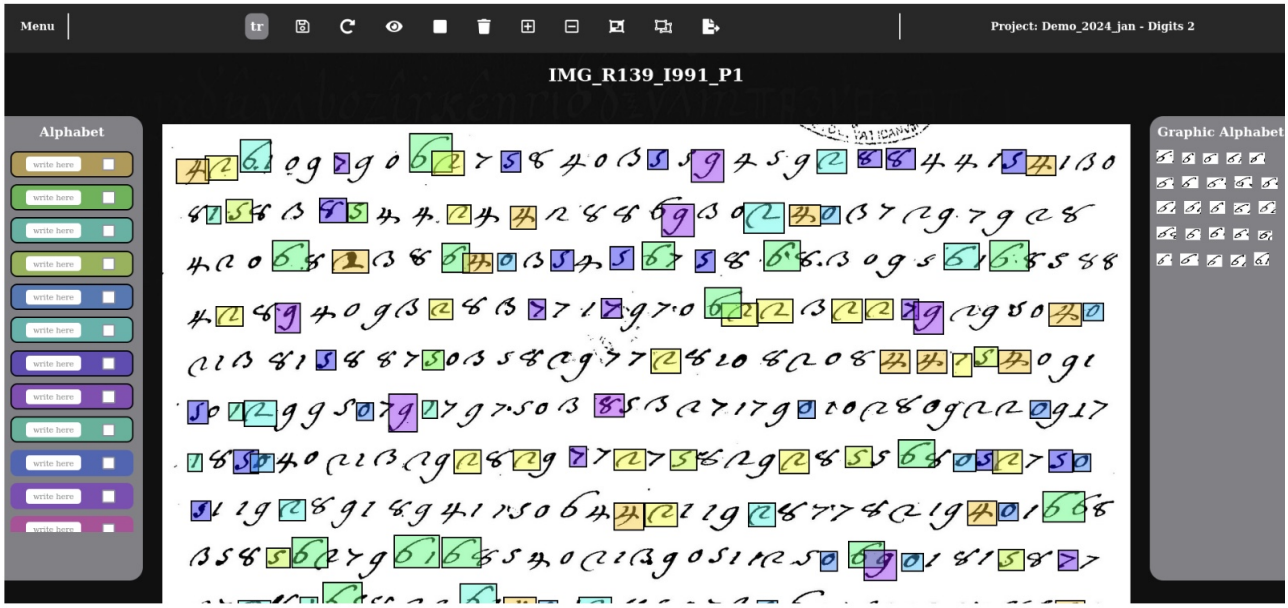


Figure 4: The TranscriptTool.

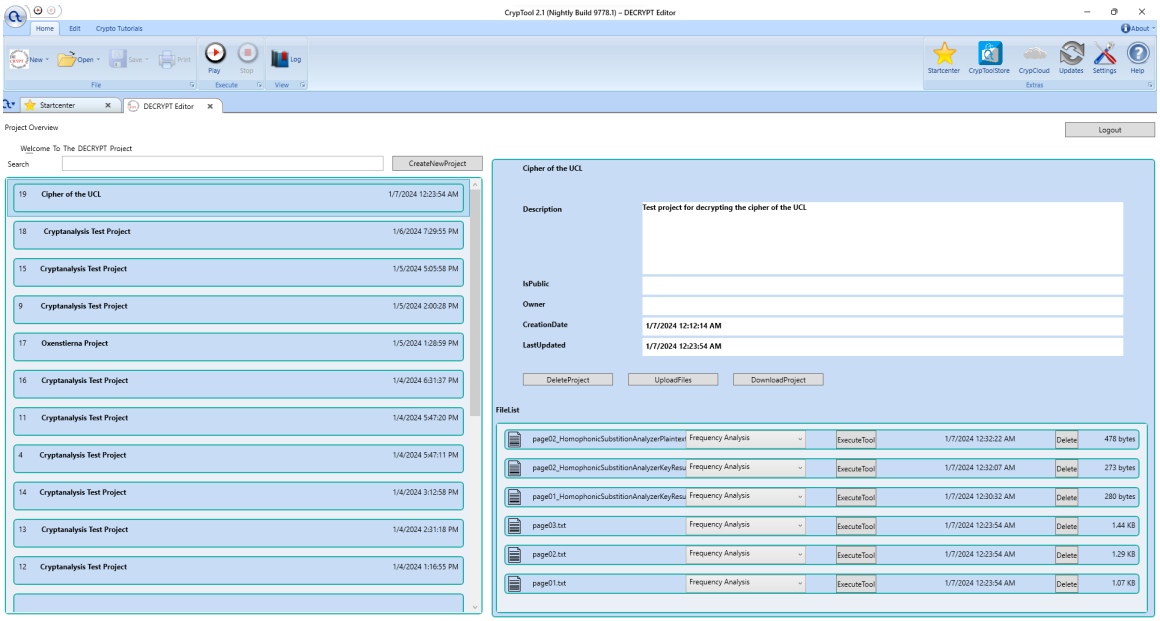


Figure 5: DECRYPT editor in CrypTool 2.

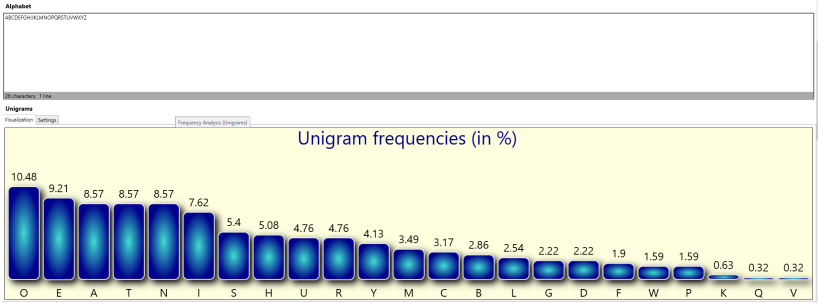


Figure 6: Frequency analysis conducted using the DECRYPT editor, which allows for the use of a user-defined alphabet (in this case, Latin), within CrypTool 2.

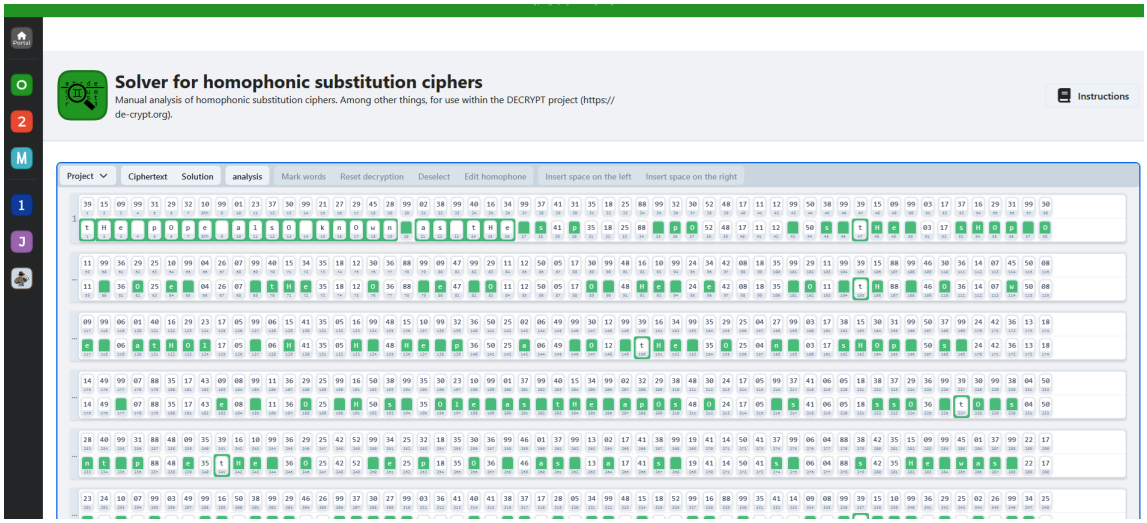


Figure 7: Cryptanalysis performed with CrypTool-Online's Homophonic Substitution Analyzer.

2.4 Historical and Linguistic Analysis

If the decryption is successful, the historian and the philologist may draw further conclusions. Usually, there is a correction and translation is necessary to make the content intelligible — sometimes even within the same language, i.e. from old Hungarian to the present day version. As the content, even in deciphered form, may lack context like sender, receiver or dates and locations, there is further work to be performed by the historian. The updated information, whether it is metadata or additional documents, can be uploaded to further enrich the Decode database. The upload can be made directly from the user-specific project or in connection to the record in the database.

3 Limitations and Lessons Learned

One main limitation of this pipeline, after evaluation by the Decrypt team, is the error rate of symbol recognition with image processing. Even with the best methods, the character error rate remains too high in order for the user to rely on the system without systematic verification error correction. This, in turn, makes document pre-processing and transcription similarly resource-intensive as manual processing, at least in case of shorter documents.

The Decrypt team also realized that while the manuscript transcription and cryptanalysis may be automatized to a large degree, the user's domain knowledge still needs to be present and the learning curve to that knowledge can be quite high.

3.1 The Connecting Fabric: API

Every action that can be conducted on the web interface of the DECODE database including the DECRYPT project can be managed via an application programming interface (API). This allows that elements of the pipeline store intermediate results, configurations and other information on the server. Moreover, the API also has been used to gather aggregated statistical data from about the several thousand records in the DECODE database. Figure 8 shows the auto-generated documentation and sandbox interface of the API.

4 Conclusion

The Decrypt project's pipeline represents a groundbreaking endeavor in the field of historical cryptology, offering a multi-faceted approach to the analysis of historical ciphers. Despite facing challenges related to the accuracy of automated transcription and the necessity for user involvement, the project illustrates the potential for technology to significantly aid in deciphering historical documents. The lessons learned underscore the importance of combining technological innovation with expert knowledge, paving the way for future advancements in the study of historical cryptology.

Acknowledgments

This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

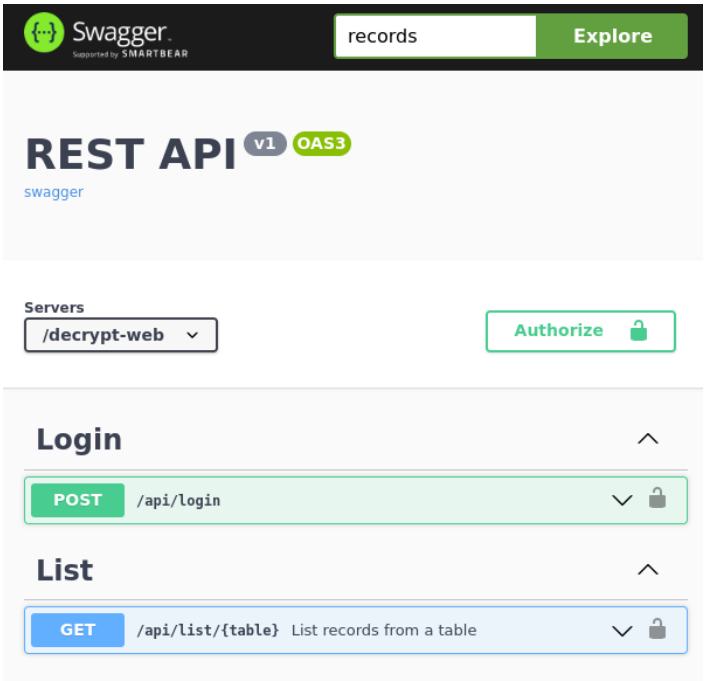


Figure 8: API.

References

Arnau Baró, Jialuo Chen, Alicia Fornés, and Beáta Megyesi. 2019. Towards a generic unsupervised method for transcription of encoded manuscripts. In *3rd International Conference on Digital Access to Textual Cultural Heritage (DATECH)*, pages 73–78.

Beáta Megyesi. 2022. Cracking Historical Ciphers with AI in the Loop. Invited talk at Colby College, USA May 6, 2022.

CrypTool Contributors. 2024. Cryptool 2 - cryptool portal.

Jörgen Dinnissen and Nils Kopal. 2021. Island Ramanacoil a Bridge too Far. A Dutch Ciphertext from 1674. In *Proceedings of the 4th International Conference on Historical Cryptology, HistoCrypt 2021*, pages 48–57.

Mihály Héder and Beáta Megyesi. 2022. The DECODE Database of Historical Ciphers and Keys: Version 2. In *Proceedings of the 5th International Conference on Historical Cryptology, HistoCrypt22*.

Nils Kopal and Michelle Waldispühl. 2022. Deciphering three diplomatic letters sent by Maximilian II in 1575. *Cryptologia*, 46(2):103–127.

Nils Kopal. 2018. Solving Classical Ciphers with CrypTool 2. In *Proceedings of the 1st International Conference on Historical Cryptology HistoCrypt 2018*, number 149, pages 29–38.

Nils Kopal. 2019. Cryptanalysis of Homophonic Substitution Ciphers using Simulated Annealing with Fixed Temperature. In *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt*, pages 107–16.

George Lasry, Beáta Megyesi, and Nils Kopal. 2021. Deciphering Papal Ciphers from the 16th to the 18th Century. *Cryptologia*, 45(6):479–540.

George Lasry, Norbert Biermann, and Satoshi Tomokiyo. 2023. Deciphering mary stuart's lost letters from 1578-1584. *Cryptologia*, 47(2):101–202.

Anna Lehofer. 2022. Applying hierarchical clustering to homophonic substitution ciphers using historical corpora. *Cryptologia*, 46(5):422–438.

Beáta Megyesi, Justyna Sikora, Filip Fornmark, Michelle Waldispühl, Nils Kopal, and Vasily Mikhalev. 2023. Historical Language Models in Cryptanalysis: Case Studies on English and German. In *International Conference on Historical Cryptology*, pages 120–129.

Beáta Megyesi, Alicia Fornés, Nils Kopal, Benedek Láng, Michelle Waldispühl, Vasily Mikhalev, and Bernhard Esslinger. 2024. Historical cryptology. *In Ed: Bernhard Esslinger Learning and Experiencing Cryptography with CrypTool and Sagemath*.

Beáta Megyesi. 2020. Transcription of Historical Ciphers and Keys. In *Proceedings of the 3rd International Conference on Historical Cryptology, HistoCrypt20*, Budapest, Hungary, June.

Societas Cooperativa Europaea. 2024. Transkribus.ai.

Mohamed Ali Souibgui, Alicia Fornés, Yousri Kessentini, and Beáta Megyesi. 2021. Few shots are all

you need: A progressive few shot learning approach for low resource handwritten text recognition. *arXiv preprint arXiv:2107.10064*.

Ferenc Szigeti and Mihály Héder. 2022. The TRANSCRIPT tool for historical ciphers by the DECRYPT project. In *Proceedings of the 5th International Conference on Historical Cryptology*, pages 208–211.

A new perspective on Dutch WWI codebreaking with its international ramifications

Bart Jacobs
iHub, Radboud University
Nijmegen – The Netherlands
bart.jacobs@ru.nl

Florentijn van Kampen
iHub, Radboud University
Nijmegen – The Netherlands
florentijn.vankampen@ru.nl

Abstract

During the First World War, the Netherlands maintained a stance of carefully guarded neutrality. International telecommunications in the form of telephone and telegraph were closely monitored and censored by so-called censorbureaus. In 2019 new files were declassified and released to the Dutch National Archive about these censorship bureaus at Amsterdam and Rotterdam, covering 1914 to 1918. They provide detailed insight in the day-to-day business, the codebreaking efforts and specific cryptanalytic results.

The material provides a completely new perspective on the genesis of modern Dutch codebreaking. This article gives a first survey of the development of these interception bureaus. It analyses their pioneering codebreaking activities and presents historic material on German diplomatic ciphers. Also, it provides new insight into the mysterious sale in 1919 of German codebooks from the Netherlands to the United States, as reported earlier in the literature.

1 Introduction

More than a century later, the First World War (WWI), also known as the Great War or la Grande Guerre, continues to fascinate from a cryptological perspective. It marked the systematisation and institutionalisation of cryptological activities in the belligerent countries — necessitated by the wide-scale adoption of wireless communication. It also involved a unique cryptanalytical achievement — the uncovering of the Zimmermann telegram — with geopolitical effects: it brought the United States to the battlefields in Europe, changing the balance of power. The recent overview book (Smoot, 2023) on American cryptology during WWI demonstrates this continued interest.

This article fits in the same line, yet from a Dutch perspective. It is based on an old dossier[1] that was made public recently. In this paper we refer to it as the 'GSIV dossier', where GSIV is the fourth section of the General staff. The dossier has been released in 2019 by the Dutch General Intelligence and Security Service (AIVD), to the Dutch National Archive. It covers reports by heads of the two military censorship bureaus, stationed at the central telegraph offices in Amsterdam and Rotterdam. Their task was to monitor telegraph and telephone communications. It also contains detailed cryptanalytic reports on German and American codes. Little was known about such activities in the Netherlands during WWI because the Dutch intelligence organisations destroyed their own archives in May 1940, as Nazi-Germany invaded the Netherlands. This remarkable dossier, that apparently survived and showed up recently, sheds new light on Dutch cryptological activities from those early years. It has the workfloor perspective of the military censors, how they started themselves to try and decrypt coded diplomatic messages that went through their hands, and how succesful they were in doing so. They systematically broke German encrypted communication and they succeeded in breaking some British, French and American codes as well. Moreover, the dossier contains a few orginal coded German telegrams, including their decryption, see for example Figures 4 and 5 below.

This article consists of four parts: Section 2 starts with a general description of the activities at the censorship bureaus of Amsterdam and Rotterdam, as described in the GSIV dossier. Subsequently, Section 3 will go into more detail on the

[1] Available via the Dutch National Archive: nationaalarchief.nl/onderzoeken/archief/2.13.70, Generale Staf, §8.A.1, GS IV.

cryptanalytic efforts together with historic examples. The main focus will be on German diplomatic communication. Further details, in particular about cryptanalysis of British and American codes will appear in follow-up publications. Section 4 presents some of the more anecdotal material in the GSIV file to provide some *couleur locale* to the codebreaking efforts. At the end, Section 5 will present new clues in the case of the mysterious procurement of German codebooks by the Americans in 1919 as described in book 'the American Black Chamber' (Yardley, 1931) and in Mendelsohn's study on German Diplomatic Ciphers (Mendelsohn, 1937).

As a general reminder to the reader, during WWI the Netherlands remained neutral. The Dutch army had been mobilised, but stayed out of the conflict. Maintaining this neutrality was a challenge. The two opposing sides in the war were keenly watching the Netherlands and could interprete any action as choosing sides and as a *casus belli* (Abbenhuis, 2006; Tuyll van Serooskerken, 2001). At the same time, the neutral territory attracted many spies, from all sides (Klinkert, 2013). In this situation the Dutch government acted cautiously and needed what is now called 'situational awareness'. Being able to decrypt secret diplomatic communications was definitely helpful.

2 WWI censorship at Amsterdam and Rotterdam: general findings

This section gives an overview of the military censorship activities at the central telegraph offices at Amsterdam and Rotterdam, as described in the GSIV dossier. After a general introduction, some specific findings are high-lighted in separate subsections.

The fourth section, GSIV, of the General Staff (GS) of the Dutch military organisation had a broad task, notably censorship, but also prevention and combatting smuggling. Intelligence gathering was done by GSIII. Immediately after the war broke out, on July 28 in 1914, two censorship teams of military officers from GSIV were formed and dispatched to the central telegraph offices of Amsterdam and Rotterdam. The newly formed teams started working on August 1, in close coordination with the local staff — which was under orders to cooperate and keep it all secret. The offices at Amsterdam and Rotterdam functioned as national hubs, through which 'suspicious' telegrams were routed from local offices.

The recently released GSIV dossier contains detailed reports[2] of (successive) heads of these military censorship bureaus, covering especially the first two years of the war. These reports were written for the General Staff and look like personal retrospects of the bureau chiefs. They are full of personal observations and remarks, and describe in a rather informal and casual manner what worked well and also what went wrong. They are a pleasure to read. The rapporteurs were Captain P. Schaafsma (at Amsteram) and Captain P.J.A. van Mourik, Lt. Colonel A.W.A. Michielsen and Captain G.W. Nyweide (at Rotterdam). The reports from Rotterdam are the most extensive and informative, covering about two hundred pages; they form the main basis for what follows. Initially, both censorship bureaus consisted of two (military) persons, but they grew during the war to 10 or 11 persons. They worked closely together with several more telegraph staff members.

2.1 Rules and regulations

The telegraph and telephone censorship operated under a special legal framework that was not available or announced to the public. This framework for the military was established by a secret Royal Decree (*Koninklijk Besluit*), dated July 31, 1914, which formulated a wide-ranging goal: to prevent any communication that forms a threat to national security. Telegrams could be withheld, changed, or partially deleted. Encrypted telegrams were not allowed: the contents should be formulated in an understandable language (in Dutch, or English, German, French) when submitted to a telegraph office. There was one diplomatic exception: consuls and chargé d'affaires of other countries were allowed to communicate in encrypted form. Such encrypted telegrams were copied, by the Dutch censors, for later analysis. Also, encrypted communication, in the form of cipher blocks, was red aloud by phone, for instance by the German consul stationed in Rotterdam, Martin Renner, talking to the German intelligence station (*Nachrichten Sammelstelle*) at Wesel (that covered the Netherlands). Such exchanges were also copied. Communications (via telegrams or phone) with relevance for national security were passed on to the General Staff in the Hague.

[2] Labeled with numbers 1937, 1938, 1939 within the 2.13.70 archive of footnote 1.

A part of the reports written by the censorship chiefs involved suggestions for improvement to the Royal Decree, based on experiences so far. In cases where the provisions of the Decree were unclear or incomplete, clarifying instructions were asked to the head of GSIV at the General Staff in The Hague, *e.g.* about whether or not to tap international phone calls only, or national calls as well (answer: yes). From today's perspective we notice that there is a legal framework in place, but no independent oversight. The surveillance was not universal but selective, driven by target information and by resource constraints. For instance, in mid 1916 the censorship bureau at Rotterdam had listed 55 individuals for phone taps, including the consuls of Germany, Britain, France and Belgium. Of course, in those days, phone calls were not so common, especially international calls. The report shows that the Rotterdam switch handled at the time almost 10.000 phone calls per day.

2.2 Origin of GSIV dossier

At the end of this general introduction we briefly discuss the surprising recent emergence of the WWI dossier on GSIV that forms the basis of this article. As mentioned, the Dutch intelligence dossiers were destroyed in May 1940, in order to prevent that they would fall in German hands. Why and how did this dossier on censorship and cryptanalysis by GSIV escape destruction? Frankly, we have no idea. What we can recover from the records in the Dutch National Archive is that after WWII the dossier existed, first at CCB (*Code Coördinatie Bureau*, 1944 – 1960) and at its successor NBV (*Nationaal Bureau Verbindingsbeveiliging*, 1960 – 2001); the latter organisation eventually merged into the AIVD, which transferred the dossier to the National Archive in 2019. This CCB and NBV had the role of national communication security organisations, see *e.g.* (Wiebes, 2001) for more information. As an aside, the CCB was first run by Colonel Jacobus Verkuijl, who worked on Japanese codes in the Dutch Indies in the 1930s and who was invited by the Americans to stay a year at Arlington Hall during WWII. There he worked (too) closely with J.S. Peterson, see (Wiebes, 2008), and learned that the Netherlands had to protect its communication better.

3 Cryptologic work in Amsterdam and Rotterdam

At the central telegraph offices in Amsterdam and Rotterdam, military censors were instructed to block all encrypted communication, with the exception of diplomatic ones. Copies of all ciphertexts had to be sent to the General Staff in The Hague. Soon, two months after the start of the censorship activities on August 1, 1914, the General Staff reported back that they should stop sending the ciphertexts because no-one was doing anything with them in The Hague. They were just piling up.

Interestingly, the military censors of GSIV at Amsterdam and Rotterdam then got interested and decided to give it a try themselves to break the encryptions. These officers were well-educated in general but not in cryptanalysis. Their reports clearly show an analytical mindset and are written in an almost academic style. The first reconstruction — of 2300 words, a substantial part — of a German code book (called 'code I' in the reports) happened in April 1915. What helped was that the code book was alphabetic in nature and that the Germans occasionally made mistakes in using it, and sometimes even duplicated messages (or phrases) in plaintext or in other codes. Also, the German consul in Rotterdam standardly reported about ships going in and out of the Rotterdam harbour. Thus, the contents of the encoded messages were often predictable[3].

The official top-down Dutch policy in 1914 was aimed at censorship (blocking 'dangerous' communication) not at uncovering secret, encrypted information. Once decryption succeeded, locally at Amsterdam and Rotterdam, and decrypted secret messages were sent to the General Staff, their value was recognised at the highest levels.

3.1 Cryptanalytic pioneering

The censorship officers at Amsterdam and Rotterdam were not prepared in any way for the cryptanalytical work that they chose to perform. They were autodidacts, who learned by doing, but also by studying. They did collect all the literature that they could find at the Department of War in The Hague. This included the following texts.

- A. Colon, *Étude sur la Cryptographie*, a Belgian text that appeared in *Revue de L'Armee*

[3] The censorship officers soon found out that the Germans used a mono-alphabetic substitution cipher for the names of ships, inside their code-book messages.

Belge and was apparently also known to the American WWI cryptographer Parker Hitt, see (Hatch, 2014).

- Eduard A. Fleissner von Wostzowitz, *Handbuch der Kryptographie*, Wien, 1881;[4]
- M. Muirhead, *Military Cryptography*, an article from 1912;[5]
- Rudolf Schmid Von Schwarzenhorn, *Universal geheimschrift*, and also *Neues Geheimschrift-verfahren*, two undated (and unfamiliar) manuscripts.

The Dutch censorship officers were cryptological autodidacts at a personal level. But one could say, the Netherlands, as a small neutral country without strategic partners, was also autodidactical as a nation. In contrast, Smoot (2023, p.2) writes:

> But the United States could not have developed its system so rapidly had it not been for the significant contribution of the United Kingdom (the Admirality's Room 40, the War Office's MI1(b), and the British Expeditionary Forces I(e) wireless and cryptologic staff), as well as France (the Deuxième Bureau's Bureau de Chiffre and subordinate army cryptologic units).

In Section 5 we shall see that the Netherlands also contributed to the cryptological position of the US.

3.2 Two teams of cryptologists

In the limited available sources, before the release of the GSIV dossier discussed here, one does find mention of Dutch cryptanalytical WWI successes, for instance, in (Klinkert, 2013) or in personal recollections, but without details. The achievements are always attributed to one single individual, namely to Henry Koot (1883 – 1959), an officer originally from the Royal Netherlands Indies Army. Koot is mentioned for instance in (Wiebes, 2008), as "considered to be one of the best Dutch cryptologists", and in (Kruh and Deavours, 2002). The *NSA Daily – History Today*, of August 24, 2011[6] writes about Koot :

> The Netherlands had its counterpart to Herbert Yardley … in Henri Koot, the "godfather" of Dutch military cryptology … one of the greats in cryptology, albeit little known outside of his homeland.

The censorship reports from Rotterdam and Amsterdam give a new, more nuanced picture. There were multiple people doing cryptanalysis, in a real team effort. They each had their own breakthroughs, with different codes. Successes are for example due to Rotterdam officers Bennewitz, Berenschot, Boomsma, Lettinga and Vis and to Amsterdam officer Van Tricht. It is interesting to note that also the acting station chiefs contribute to the succes as with Van Mourik and Nyweide. Koot was the most proficient in breaking codes, but definitely not the only one. Because of his skills, he was allowed to spend all his time on cryptanalysis and was freed from bureaucratic duties. Describing him as the sole Dutch WWI cryptologist is a misrepresentation.

3.3 Breaking German Diplomatic Codes

The GSIV dossier contains several sources with information about German diplomatic codes, their properties and the efforts of breaking them. First, there are the Rotterdam reports that describe successes but also *how* these were achieved, what mistakes were made and how information was gathered. In addition to the reports, there is also a separate file with descriptive and cryptanalytic articles about several German Diplomatic ciphers. These articles, or 'notes' as the Dutch called them, were used to summarise and archive the analysis of a certain code. These reports were exchanged between Amsterdam and Rotterdam to benefit from each others results and insights. Unfortunately, some of these notes are missing, for unclear reasons: there are some references in the Rotterdam reports to notes about German code systems with name, date and author that are not in the GSIV dossier.

The Rotterdam report describes in detail how the staff of the censorbureau had to bootstrap their codebreaking activities. Every aspect of the codebreaking metier had to be invented on the spot. When they broke[7] their first code in april 1915, it is simply referred to as 'code I'. After a while the Dutch codebreakers discovered more new codes with new systems and new variants, so they had to invent a scheme to order and catalogue the codes.

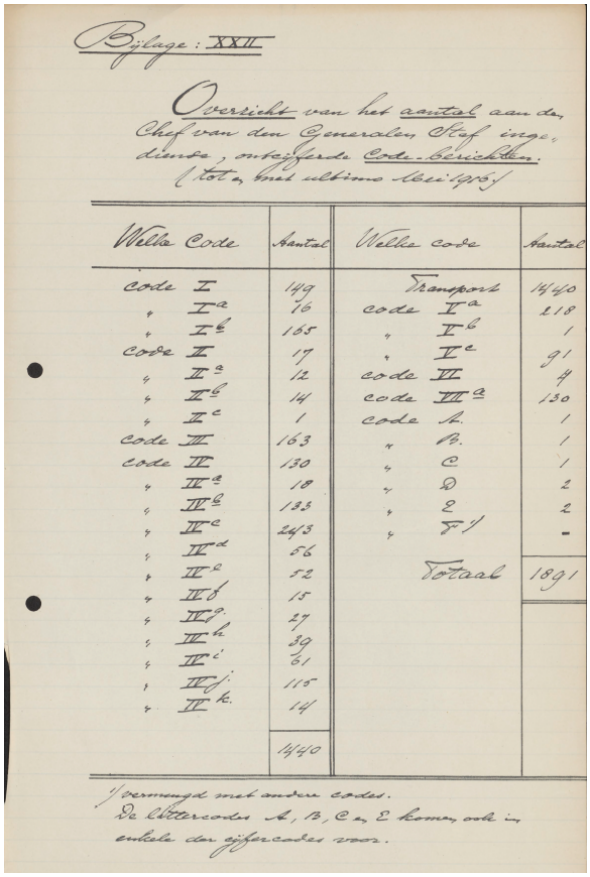
The Dutch codebreakers started to number different codes with a Roman number: code I, II, III, IV etc. After a while this system had te be

[4] See kryptografie.de/kryptografie/personen/eduard-fleissner.htm

[5] Republished as (Muirhead, 1912), see doi.org/10.1080/03071841209417859.

[6] Released in 2015, see pdf link

[7] The rapporteurs use a very peculiar but effective phrase for when a code is solved or broken; They would say that a code has "fallen"



Bijlage: XXII

Overzicht van het aantal aan de Chef van den Generalen Staf ingediende, ontcijferde Code-berichten. (tot en met ultimo Mei 1916)

Welke Code	Aantal	Welke code	Aantal
code I	149	Transport	1440
" I^a	76	code I^a	218
" I^b	165	" I^b	1
code II	17	" I^c	91
" II^a	12	code III	4
" II^b	14	code III^a	130
" II^c	1	code A.	1
code III	163	" B.	1
code IV	130	" C	1
" IV^a	18	" D	2
" IV^b	133	" E	2
" IV^c	243	" F	-
" IV^d	56		
" IV^e	52	Totaal	1891
" IV^f	15		
" IV^g	27		
" IV^h	39		
" IV^i	61		
" IV^j	115		
" IV^k	14		
	1440		

*) vermengd met andere codes.
De lettercodes A, B, C en E komen ook in enkele der gijfercodes voor.

Figure 1: From appendix XXII of the Rotterdam report by van Mourik an overview, in Dutch, of decryptions until May 1916. In the left column, in Roman numerals, the codesystem and variant. In the right column the number of decryptions.

expanded because the Germans were constantly modifying their codes with new variations, modifications and additions. This was probably done in an attempt to increase security. So code I expanded to variant Ia and Ib and code II was refined to IIa and IIb and so on. The Rotterdam report states that during the first two years of the war in total 76 German codes, including variations, were broken. That is a non-trivial achievement. The Rotterdam office kept detailed statistics on the number of decrypted messages and the kind of code used. An example can be seen in Figure 1 where one finds in the last column the number of decrypted messages in a particular code sent to the head of the General staff from the start of the bureau until May 1916 — totalling 1891.

One of the code families was of special value. Van Mourik writes in his report: (translation by the authors): "Code III (...) is a very important (code) because it is the consular code, that means the code that is used to discuss the important political matters". Code III, and some of its successors, are the family of German diplomatic codes.

Below we present some distinctive properties to describe and catalogue German diplomatic codes and are also used internationally. This will make it easier to describe results and put things in a larger context. Every German (diplomatic) code has a certain designating number. This is a number that (almost) always appears at the beginning of the message and it ifunctions as an indicator which codebook or system is used. This is of course necessary for the recipient to be able to decode the message. So one might talk about the 2500 code or the 29000 code. These designating numbers will be used in the rest of this article.

German Diplomatic codes were based at the time on large codebooks. These books contain thousands of words, names and places with a corresponding number. The combination of this number, the page in the book and sometimes some other ingredients, would lead to a translation from a word to a number and vice versa. The Germans would make variations of a codebook by reordering the pages, renumbering the words or a modification in how to construct the final number. These encoding variations would get a new designating number so that the communicating parties would know exactly what to use. More about such variations can be found in (Lasry et al., 2020).

In the reports from Rotterdam, but also in international sources, these variants are ordered in some form of hierarchy. One code would be considered the main code and other codes are seen as descendants. Depending on the reconstructed codebooks available, it is not always certain which codebook should be considered the main one and which one a variation. It will depend on the kind of messages that are intercepted and which codebook is solved first. These different family trees of German diplomatic codes will turn out to be of value in the last section of this article. Figure 2 from (Mendelsohn, 1937) shows a graphic representation of such a hierarchy and also shows designated numbers of German diplomatic codes.

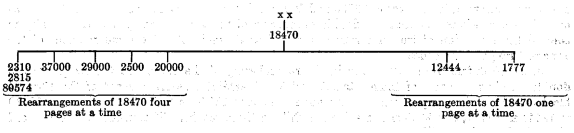


Figure 2: Cryptologic family tree of code 18470.

3.4 Examples of historic codebreaking

What makes the reports from Rotterdam so interesting is that they offer a peek at actual cryptanalysis during WWI, hidden in a back room at the Rotterdam Telegraph office between 1914 and 1918. Sometimes, results were achieved by statistics or logical reasoning. But often breakthroughs were made because of clever combination of operational possibilities.

Section 3 already mentioned code I. The break of code I started with a house search by the police in a case of suspected German espionage. During this search, the police found a note together with a letter to the German chief of naval Intelligence Prieger. The note is included in the report from Van Mourik and is reproduced in Figure 3.

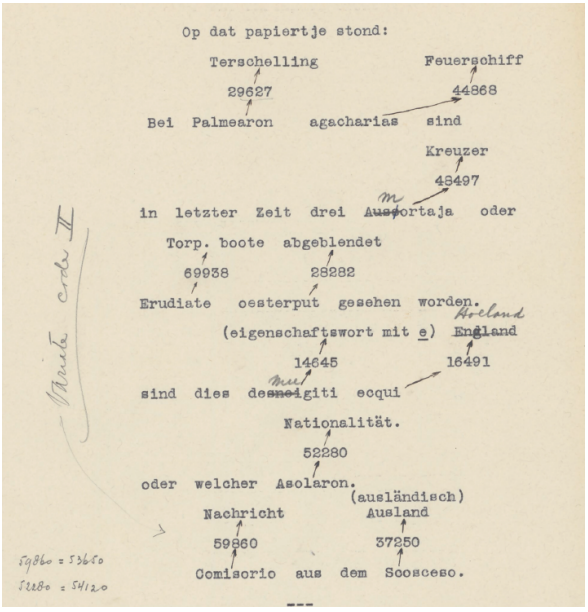


Figure 3: First clues to ‘code I’ in a note captured by the Dutch police from a suspected German spy.

These few codewords mark the start of the fall of code I. After a couple of days the Rotterdam bureau intercepts a phone call to the German intelligence station in Wesel, in which some words are replaced with code numbers. Van Mourik describes the process honestly: “We could still not draw any conclusions, since we had no idea how the German code was constructed”. They keep combining intercepted phone calls with the word list reconstructed so far. At the same time they had to figure out how the code actually works (the fact that there is a code book, with page numbers and alphabetically ordered words). Van Mourik explains why he describes the breaking of this particular code in such great detail : “The discovery of code I is so important, because with that experience (...), we were able to find all the other codes”.

The next example in Figure 4 shows an intercepted German coded telegram with added handwritten decryption. It starts with two numbers: 175 and 200. The handwritten notes above the numbers show that 175 is the number of the telegram and 200 is the *Erkennungszahl*, or designating number. As far as readable, it says:

> Dampfer aus England ist heute Vlissingen nicht parterop(?) ankommen. Feuer - schiff Nord - Hinder(?) doch nächtens nach Nord partirop(?) verlegen werden wegen substanti(?) Aus(?) English Meinen - feld bis 52 Grad Nord Breite und 2 Grad 22 Minuten Ost Länge. Genau Lage Feuer - schiff wird bei substantisch(?) Verlegen partizip bekannt. Gefahrlich is jetzt Fahrt zwischen 1 35(?) Ost und 3 18 Ost von 51 15 Nord bis 51 40 Nord Ferner zwichen 1 55 ost und 2 32 Ost von 51 40 Nord bis 52 Nord. Mueller.

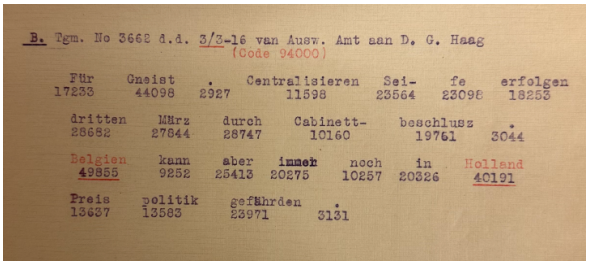


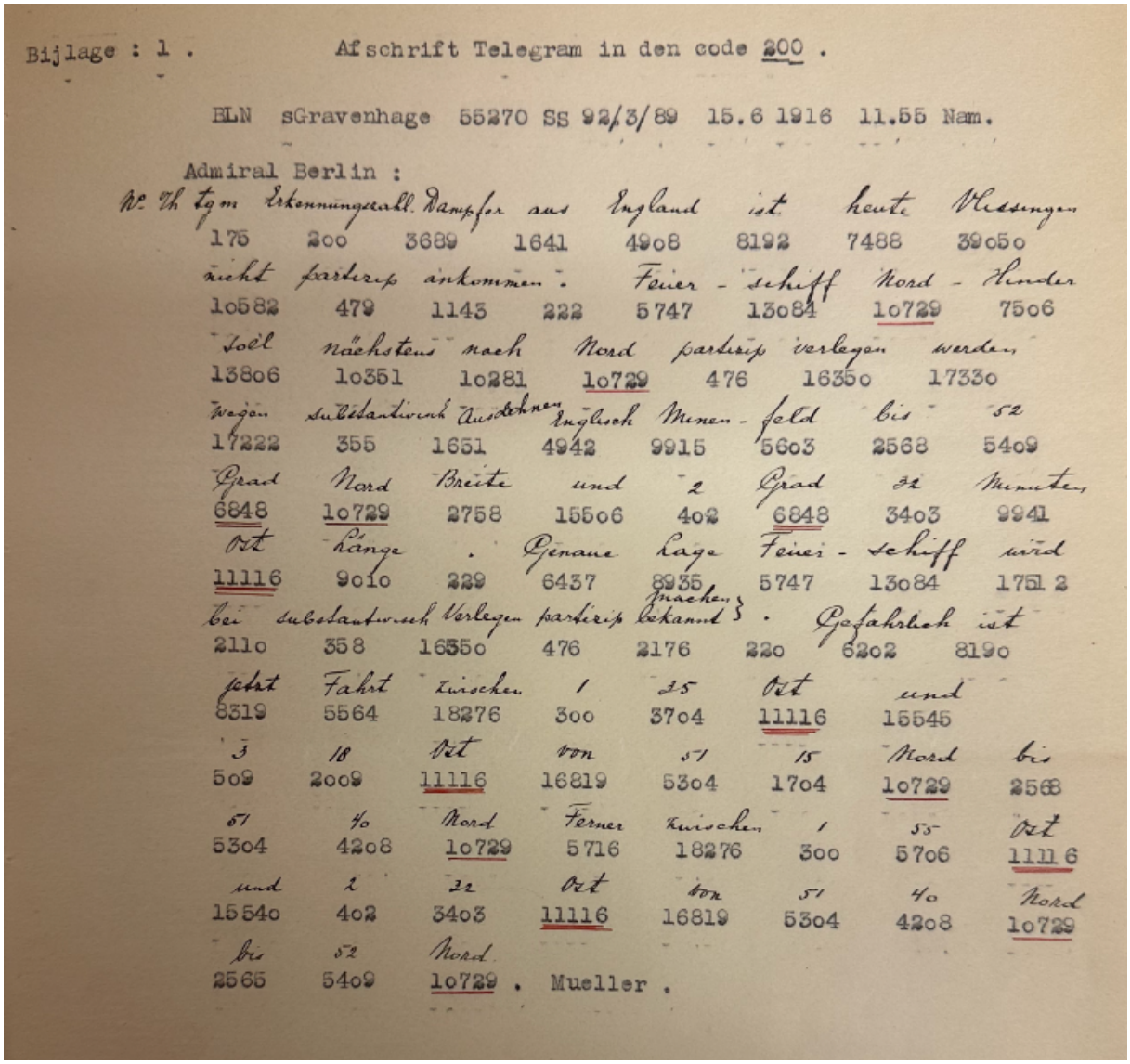
Figure 5: Decryption of a German telegram in code 94000 dated March 3rd 1916

Figure 5 shows another original German telegram, in code 94000. This code was used between the *Deutsche Gesandtschaft* in The Hague and *Auswärtig Amt Berlin* about matters of import and export. The censors in Rotterdam therefore called it the ‘trade code’. As in this telegram, the designating number is sometimes omitted when both communicating parties considered it to be obvious in which code they were communicating. The message is as follows.

> Für Gneist. Centralisieren Sei- fe erfolgen dritte März durch Cabinett- beschlusz. Belgien kann aber immer noch in Holland Preis politik gefährden.

4 Anecdotal observations

As previously mentioned, the telegraph station staff, under the instruction of the censors, monitored selected phone calls. Summaries of the calls


Bijlage : 1 . Afschrift Telegram in den code 200 .

BLN sGravenhage 55270 SS 92/3/89 15.6 1916 11.55 Nam.

Admiral Berlin :

W. Th tgm Erkennungszeall. Dampfer aus England ist heute Vlissingen
175 2oo 3689 1641 49o8 8192 7488 39o5o
nicht fahrplanmäßig ankommen . Feuer - schiff Nord - Hinder
1o582 479 1145 223 5747 13o84 1o729 75o6
soll nächstens nach Nord fahrplanmäßig verlegen werden,
13806 1o351 1o281 1o729 476 16350 17330
wegen substantiellen Ausdehnen englisch Minen - feld bis 52
17222 355 1651 4942 9915 56o3 2568 54o9
Grad Nord Breite und 2 Grad 32 Minuten,
6848 1o729 2758 15506 4o2 6848 3403 9941
Ost Länge . Genaue Lage Feuer - schiff wird
11116 9o1o 229 6437 8935 machen 5747 13o84 1751 2
bei substantiellen Verlegen fahrplanmäßig bekannt . Gefährlich ist
211o 358 16350 476 2176 22o 6202 819o
jetzt Fahrt zwischen 1 35 Ost und
8319 5564 18276 3oo 3704 11116 15545
3 18 Ost von 57 15 Nord bis
5o9 2oo9 11116 16819 5304 1704 1o729 2568
57 4o Nord Ferner zwischen 1 55 Ost
5304 4208 1o729 5716 18276 3oo 5706 11116
und 2 32 Ost von 57 4o Nord
15540 4o2 3403 11116 16819 5304 4208 1o729
bis 52 Nord
2565 5409 1o729 . Mueller .

Figure 4: Decryption in handwriting of a German telegram in code 200 dated June 15th 1916.

were recorded[8]. At the time, the telephone system worked via switchboards with cables. Connections were added so that the sensitive calls could be copied to an additional 'tap' phone in an adjacent room. The report dryly remarks that the censors soon found out that it was wise to remove the microphone from this second tap phone.

The censorship itself was meant to be secret and the military censorship officers worked in plain clothes, but pretty soon almost everyone in the telegraph offices of Amsterdam and Rotterdam knew about them. Also outside, many journalists and diplomats were soon aware of the censorship.

During the war there were few limitations for the Dutch press. Occasionally, telegrams from journalists working at the border were redacted by the censors, in order not to leak military details. Foreign journalists were monitored systematically, also because several spies worked under journalistic cover, see also (Klinkert, 2013).

Encrypted messages were also passed on in phone conversations, in which the wordcodes, like 90987, were red aloud, in sequence. This could easily go wrong. The intended German receiver could ask for a repetition when a sequence of numbers was unclear, but the Dutch copiers could not, to their frustration. Copy mistakes were a constant concern: such coded telephone calls could easily last an hour.

[8]These summaries are not included in the GSIV dossier. In general, the dossier mostly describes procedures together with a few highlights and personal reflections.

The Dutch noticed that the German military attaché Renner was very careful and, for instance, never by accident mentioned the cleartext instead of the ciphertext. They assumed that Renner had been trained in these matters. In contrast, the German consuls in Rotterdam, first Gneist and then Bosenick, made more mistakes — to the advantage of the censors. Moreover, they were contemptuous when they erred and would say things like: "never mind, those Dutch don't understand such matters anyway".

The reports about the censorship activities describe several times how enthousiastic the (self-taught) censors were about their cryptanalytical activities. For instance, van Mourik write (translated by the authors):

> ... working on these codes is extraordinarily captivating and interesting. Hours and hours in succession — usually during our spare time — have we dedicated our efforts to this; nothing was more satisfying, so we found, than having fully decrypted a message. It is noteworthy that we sometimes spent hours thinking about one word, and days about one short message.

Van Mourik thinks ahead about how to train codebreakers in the future. He proposes that the elite school for senior rank officers (*Hogere Krijgsschool*) should develop a course on cryptology. He vividly describes in his report the discussions he had with Koot about this (translation by the authors):

> At the Military Academy, mr Koot and myself learned a thing or two about cryptography, but we did not realise, that this was such an extensive and interesting study. We both have expressed multiple times, that it would be very worthwhile, if cryptography would be part of the standard curriculum of the Military Academy, for example in the first couple of years at least one hour every week. What beautiful puzzles would we provide the students of this course; Mr Koot often salivated at this idea.

Indeed, after WWI Koot teaches at the military school and educates a whole new generation of dozens of Dutch military cryptologists, including Verkuijl (Wiebes, 2001). Van Mourik shows himself to be quite the visionary when he thinks about how to institutionalise the cryptologic activities in Netherlands. Van Mourik writes (translation by the authors):

> The undersigned has - for a long time - considered the question of whether it is not desirable - we need not doubt its feasibility - to establish a 'decryption bureau' in our country during peacetime. (...) Especially for times of tension, this measure seems to me very desirable. Knowing what is going on in Europe during such times is - needless to say - extremely important. To only take this measure when there is some tension on the political horizon does not seem wise to me. At that time, the individuals who would then be charged with this task could not fully immerse themselves in it; they must solve the various codes and collect the necessary data during peacetime, so as to be able to use them at the right time.

Many countries, including the Netherlands, would forget this lesson between the two world wars. In the economic crisis of the 1930's, many government cipherbureaus were closed.

4.1 Ships, spies and smugglers

The Rotterdam report of van Mourik also elaborates on the contents of the German messages that were decrypted, especially from the military attaché Renner. They cover many newspaper articles, both from Dutch and international media, and also much shipping information especially about the Rotterdam harbour and about its continued trade with the UK. Also, many smuggle activities showed up, and were shared by the censors with Dutch authorities. There are also several spy stories, partly overlapping with (Klinkert, 2013). The latter source was written before the release of the GSIV dossier at hand, giving opportunities for further study.

5 Selling German codebooks to the USA

During WWI the United States were also breaking German codes (Smoot, 2023). Herbert Yardley founded MI8 as a so called *Black Chamber* to focus the US cryptanalytic efforts (Yardley, 1931). Charles J. Mendelsohn reveals in his report 'Studies in German Diplomatic Codes Employed During the World War' (Mendelsohn, 1937) an intriguing story about German codebooks stemming from the Netherlands. He describes that in April 1919, in the Netherlands, American officials had been offered German Codes books for sale. At Christmas 1919 the material was sent from the Netherlands to Washington for inspection and analysis to see if it was worth buying.

Mendelsohn describes the person selling these codes as "The Dutchman" He also describes the uncertainty surrounding the identity or nationality of this person. On the one hand there is (Yardley, 1931) stating that the codes were offered to the Americans in The Hague, that is the Netherlands,

but by a German spy. Mendelsohn, on the other hand, makes a case for the fact that the Dutchman was, in fact, actually truly a Dutch person. Both scenarios have their merits and drawbacks and at the end Mendelsohn is reluctant to draw a final conclusion.

The GSIV dossier about the Dutch code breaking efforts in Rotterdam and Amsterdam contains evidence, described below, that the material offered to the Americans actually came from GSIV. The evidence makes it plausible that the Dutchman from Mendelsohn's report was someone from or with access to our group of pioneering codebreakers at the telegraph offices.

5.1 The mysterious Dutchman

The first part of the puzzle is Mendelsohn's description of the Dutchman's material:

> This material contains (...) a skeleton of code known as 2500, with tables for changing this code into four encipherments called by the "Dutchman" (...) 37000, 29000, 20000 and 18400. The last turned out to be identical with 18470, although in the messages received by MI8 that designating number was never employed.

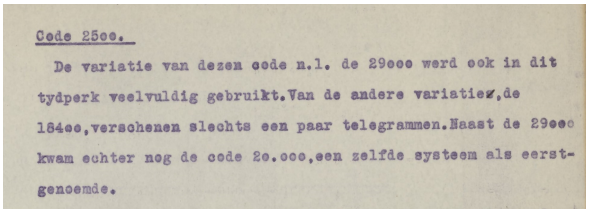
The list of codes is of course a clue, but more specifically the fact that one of the codes is referred to by the Dutchman with the designating number 18400. Apparently, MI8 never used that number, but used 18470 instead. This will be an important clue.

Mendelsohn writes in his report how the Dutchman thinks these codes relate to each other (which code is a variant of which code):

> Probably code 2500 is the original code book. From this code are derivated (sic!) the codes 18400, 29000, 37000 and 20000.

The Americans have a different codetree, see Figure 2, with 18470 as the main code with 2500, 29000, 37000 and 20000, and others, as variants. Now it is time to compare this with the material from Rotterdam to see how the three Rotterdam reports describe the various codes.

The reports from the Rotterdam bureau set out in quite some detail which codes are found, how they are broken, what name the bureau assigned to them and how the codes relate to each other.


Code 2500.

De variatie van dezen code n.l. de 29000 werd ook in dit tydperk veelvuldig gebruikt.Van de andere variatiew,de 18400,verschenen slechts een paar telegrammen.Naast de 29000 kwam echter nog de code 20.000,een zelfde systeem als eerstgenoemde.

Figure 6: Section about code 2500 in the Rotterdam report by Nyweide. Translation: "Code 2500: The variation of this code, namely the 29000, was also frequently used in this time period. From the other variation, the 18400, only a few telegrams appeared. In addition to the 29000, however, there was also the code 20,000, a similar system to the aforementioned."

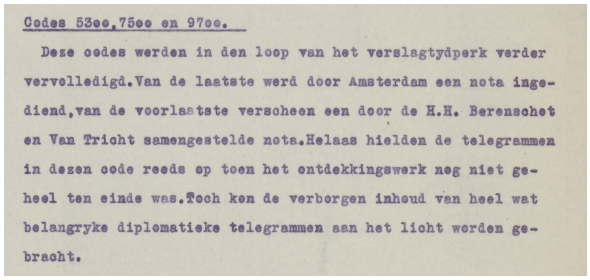
The third report from Rotterdam by Nyweide contains a separate section on code 2500. This section is shown in Figure 6. Here we clearly see that the group in Rotterdam consideres code 2500 to be the main code and the others (20000, 29000 and 18400) variants of that main code. We can also clearly see that they use the 18400 designator for what in the US and UK codebreaking literature mostly is referred to as the 18470 code.

These two elements, the code-tree with 2500 at the top and the designating number 18400 in stead of 18470, are a unique fingerprint for the Rotterdam and Amsterdam codebreakers. This makes it extremely likely that the mysterious Dutchman that Mendelsohn describes is, in fact, someone from or in the vicinity of our Dutch censor codebreakers.

5.2 How good were the Dutch Codebreakers?

As a final thought experiment, we try to position the Dutch codebreakers at an international stage. The most well known counterparts are Yardley's group MI8 in the United States, Room 40 in the United Kingdom, and the Deuxième Bureau in France. In no way do we present this as a systematic comparison of these different codebreaking group. But if we lift out one section from the last report from the Rotterdam bureau in the GSIV file, we obtain an interesting international perspective.

Nyweides report from Rotterdam mentions codes 5300, 7500 and 9300 quite casually. This short paragraph can be seen in Figure 7 .



Codes 53oo,75oo en 97oo.

Deze codes werden in den loop van het verslagtydperk verder vervolledigd.Van de laatste werd door Amsterdam een nota ingediend,van de voorlaatste verscheen een door de H.H. Berenschot en Van Tricht samengestelde nota.Helaas hielden de telegrammen in dezen code reeds op toen het ontdekkingswerk nog niet geheel ten einde was.Toch ken de verborgen inhoud van heel wat belangryke diplomatieke telegrammen aan het licht worden gebracht.

Figure 7: Description in Dutch from Nyweides report from Rotterdam about the cryptanalytic succes on German diplomatic codes 5300, 7500 and 9700. Translation: "These codes were further completed over the course of the reporting period. For the last one, a note was submitted by Amsterdam, and for the penultimate one, a note compiled by H.H. Berenschot and Van Tricht appeared. Unfortunately, the telegrams in this code already ceased when the discovery work was not yet completely finished. Nevertheless, the hidden content of quite a few important diplomatic telegrams could be brought to light."

One of the codes that immediately draws attention is 7500. It is remarkable to see that the Dutch were able to read it. This was the code that was used to encrypt the original Zimmerman telegram (Mendelsohn, 1938). Room 40 in the UK apparently was able to read this code and considered this to be a major achievement (Friedman and Mendelsohn, 1938). We dare to claim that, had the famous Zimmerman telegram been transmitted through telegraph offices in the Netherlands, the Dutch codebreakers would have decrypted it.

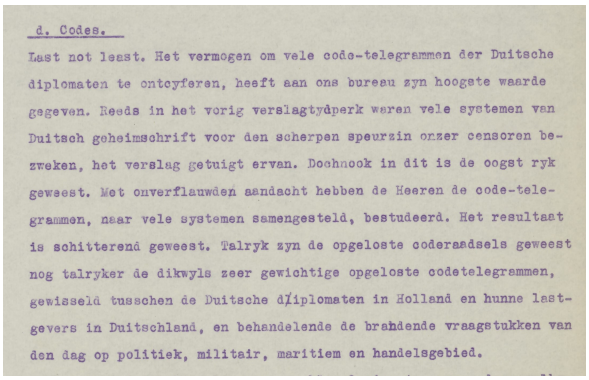
Mendelsohn remembers the Dutchman and what he had to tell about the other two codes: 9700 and 5300. MI8 also succesfully broke these codes, but considered them to be quite complex. Mendelsohn can almost not believe that the Dutch cryptologists would have been capable of breaking such complicated codes:

> In his description of codes 9700 and 5300, not belonging to the 18470 family of which he likewise furnished partial copies, the "Dutchman" has indicated certain additives, some of them running to many figures, which were used with these codes. To work out these long additives from the fractions of the codes at his disposal would have been a very rare cryptographic achievement.

For a group of self-taught enthusiasts, working in an improvised cipherbureau setup in a spare room at the local telegraph office in Amsterdam or Rotterdam, without any international collaboration, this is quite an achievement indeed.

6 Conclusions

A recently declassified GSIV dossier from the Dutch National Archives offers a novel perspective on the origins of twentieth century Dutch codebreaking, in particular during World War I. It reveals that codebreaking started as a bottom-up effort, initiated by two teams of intelligent, enthusiastic, and self-taught censor officers. This effort included the well-known cryptographer Henri Koot, but had many more contributors. The dossier demonstrates that the (isolated) cryptoanalytic achievements of the Dutch reached levels that are comparable to those of the British, French and Americans. In fact, the dossier also shows that German codebooks bought by the Americans in 1919 must have come from these Dutch teams.



d. Codes.

Last not least. Het vermogen om vele code-telegrammen der Duitsche diplomaten te ontcyferen, heeft aan ons bureau zyn hoogste waarde gegeven. Reeds in het vorig verslagtydperk waren vele systemen van Duitsch geheimschrift voor den scherpen speurzin onzer censoren bezweken, het verslag getuigt ervan. Dochnook in dit is de oogst ryk geweest. Met onverflauwden aandacht hebben de Heeren de code-telegrammen, naar vele systemen samengesteld, bestudeerd. Het resultaat is schitterend geweest. Talryk zyn de opgeloste coderaadsels geweest nog talryker de dikwyls zeer gewichtige opgeloste codetelegrammen, gewisseld tusschen de Duitsche diplomaten in Holland en hunne lastgevers in Duitschland, en behandelende de brandende vraagstukken van den dag op politiek, militair, maritiem en handelsgebied.

Figure 8: Michielsen reflects on the code breaking in Rotterdam.

6.1 "The result has been brilliant"

We end with one more (translated) quote illustrating the professional enthousiasm and pride of the Dutch codebreakers, see Figure 8.

> Last (but) not least. The ability to decipher numerous coded telegrams of German diplomats has given our office its highest value. Already in the previous reporting period, many systems of German secret writing had succumbed to the keen investigative sense of our censors, as the report testifies. But even in this period, the harvest has been ripe. With unwavering attention, the gentlemen have studied the coded telegrams, composed according to many systems. The result has been brilliant. Numerous have been the solved code puzzles, even more numerous the often very significant solved coded telegrams, exchanged between the German diplomats in Holland and their principals in Germany, addressing the burning issues of the day in political, military, maritime, and trade areas.

Acknowledgements

We thank the reviewers for their feedback, especially one of them who provided additional material which demonstrated that Code I in Subsection 3.3 is a variant of Code II and that Code II is actually the German naval code (the so called *Verkehrsbuch*). These connections will be further explored in future work.

References

Maartje Abbenhuis. 2006. *The Art of Staying Neutral. The Netherlands in the First World War, 1914–1918*. Amsterdam Univ. Press.

William F. Friedman and Charles J. Mendelsohn. 1938. *The Zimmerman Telegram of January 16, 1917 and its Cryptographic Background*. United States Government Printing Office.

David Hatch. 2014. The dawn of American communications intelligence: The Spanish-American war and after. *Cryptologic Quarterly*, 2024(1):17–29.

Wim Klinkert. 2013. A spy's paradise? German espionage in the Netherlands, 1914–1918. *Journ. Intelligence History*, 12(1):21–35.

Louis Kruh and Cipher Deavours. 2002. The commercial Enigma: Beginnings of machine cryptography. *Cryptologia*, 26(1):1–16.

George Lasry, Ingo Niebel, and Torbjörn Andersson. 2020. Deciphering German diplomatic and naval attaché messages from 1900-1915. *Cryptologia*, 45:1–43.

Charles J. Mendelsohn. 1937. *Studies in German Diplomatic Codes Employed During the World War*. United States Government Printing Office.

Charles J. Mendelsohn. 1938. *An Encipherment of the German Diplomatic Code 7500*. United States Government Printing Office.

Murray Muirhead. 1912. Military cryptography: A study of transposition cipher systems and substitution frequency tables. *Journ. of the Royal United Services Institution*, 56(418):1665–1678.

Betsy R. Smoot. 2023. *From the Ground Up: American Cryptology during World War I*. Series II, World War I, Volume 2. National Security Agency, Center for Cryptologic History.

Hans van Tuyll van Serooskerken. 2001. *The Netherlands and World War I. Espionage, Diplomacy and Survival*. History of Warfare, Volume 7. Brill Publishers, Leiden.

Cees Wiebes. 2001. Dutch sigint during the Cold War, 1945–94. *Intelligence and National Security*, 16(1):243–284.

Cees Wiebes. 2008. Operation 'Piet': The Joseph Sidney Petersen Jr. spy case, a Dutch 'mole' inside the National Security Agency. *Intelligence and National Security*, 23(4):488–535.

Herbert O. Yardley. 1931. *The American Black Chamber*. The Bobbs-Merrill Company.

Lost in Translation: Missing Background, Contextual Blindspots, and Editing Mishaps in Translated Intelligence Content

Stephen Jaskoski
Independent Researcher
Wellsville, Pennsylvania, USA
stevejaskoski@hotmail.com

Abstract

This article examines the content as rendered by a small sample of intelligence reports issued by the UK Government Code and Cypher School (GC&CS) at Bletchley Park that address the initial indications in Italian communications of Italian regime change after Mussolini and considers how higher meaning in the texts may not have been fully appreciated in the context when they were written. The episode offers lessons in grasping context and significance in translated material.

1 Introduction

Archival research of historic intelligence reports based on translation and interpretation of an adversary's communications occasionally reveals lessons in understanding the larger meaning than what is directly conveyed, as an individual intelligence report may be analogous to a single tile within a larger mosaic. As one reflects on the currently increasing sophistication of neural machine translation powered by Large Language Models (LLMs), it may be worth considering from historical examples the errors in translation and interpretation that may arise from incomplete contextual awareness or other human factors. Intelligence reporting from the UK's Government Code and Cypher School (GC&CS) at Bletchley Park in the summer of 1943 offer relevant examples of the pitfalls of overlooking deeper meaning conveyed in an adversary's communications.

Analysis and reporting of the contents of highly-encrypted Axis communications took place in Hut 3 at Bletchley Park, as soon as signals were successfully decrypted in Hut 6. Decrypted signals still in very raw form had to be emended to clarify the gaps and reception garbles in order to provide coherent German or Italian plaintext which would be immediately evaluated for urgency and provided to a Watchkeeper for translation. The Watchkeeper's translated text would be checked for accuracy by another Watchkeeper, then provided to an Advisor who would evaluate it for relevant intelligence content and draft an intelligence report. The production of the half-dozen or so Watchkeepers on a shift converged at the Hut 3 Duty Officer, who was the last Hut 3 authority to review each intelligence package for accuracy and distribution before release to a Signals Officer for ultimate transmission to field commands. Owing to the daily Hut 6 processing rhythm that started with the midnight crypto key change of the adversaries, followed by the key recovery process yielding decrypts for analysis in Hut 3, the evening shift in Hut 3 was the busiest.[1]

2 Examples

[1] Millward, in *Codebreakers*, pp. 20-23.

Despite the exemplary record of GC&CS against highly enciphered Axis communications during World War II, blindspots are inevitable in any human endeavor, including translating and assessing content of foreign communications. A GC&CS report issued in August 1943 referred to cargo aboard an ME-363 transport aircraft forced down in northern Corsica as including, *inter alia*, two "peoples cars"[2], clearly a reference to Volkswagen, the manufacturer of light, general purpose vehicles for the German military. This linguistic error of translating a proper name instead of rendering it in its original form is certainly excusable for its time, for in 1943, few people outside Germany or German-occupied Europe would have been acquainted with the firm that was to become the worldwide automotive giant decades later. Only an individual with an immersive German language experience after the founding of Volkswagen in 1937 could have been aware of the vehicle. The error is both reasonable for the context of the time, and fortunately insignificant.

On other occasions, insufficient contextual awareness may result in deeper meaning being missed in the translated text. On 30 July 1943, five days after Mussolini's fall from power, two reports were issued that provided insight into the shedding of fascist symbology by the newly established Badoglio government. The first report indicated that Italian aircraft markings were to be changed to large black circular spots, painting over the fasces symbol at the center of the black circle in the extant Italian Air Force marking.[3] Painting over the fasces carried political meaning greater than the narrow purpose of aircraft identification, for it removed the Fascist Party emblem from the ancient Roman insignia of the bundle of rods tied together with a protruding ax head, symbolizing the penal power of the Roman magistrate or other high official. The political act of removing the fasces in 1943 was certainly considered important enough to undertake, for repainting the wing and fuselage markings on an entire air force was no trivial task, particularly during wartime.

A second GC&CS report issued shortly thereafter stated that the Italian Admiralty had issued orders on 30 July that the battleship *Littorio* was to be renamed the *Italia*, and the destroyer *Camicia Nera* was to be redesignated as the *Artigliere*.[4] In contrast to the German translation example above in which a proper name was mistakenly translated *verbatim*, the significance of the Italian Navy vessel renaming is discernible only through awareness of the meanings of the proper names. *Littorio* is the Italian word for lictor, the ancient Roman official who was responsible for carrying the fasces for the Roman magistrate that symbolized the latter's authority. The vessel name similarly refers to symbols appropriated from ancient Rome by the fascists. Renaming the *Littorio* conveyed as much political content as the painting over of the fasces on Italian aircraft. Renaming the *Camicia Nera* was an even less subtle rejection of the Fascist Party, for *Camicia Nera* translates to Black Shirt, a reference to the fascist paramilitary militia. By contrast, the new name, *Artigliere*, translates more benignly as Gunner or Artilleryman.

Whether the political significance of these communications was fully grasped by Bletchley Park analysts at the time is uncertain. No commentary was appended to the report text to call attention to the political implications of the intelligence fact conveyed, but such treatment was not unusual for GC&CS reporting during the period. Comments were relatively rare in reports, and those that did appear were usually confined to interpreting ambiguous indicators of changes in military order of battle or force dispositions. Such comments were based on background material from Bletchley Park's own reference files or from input from intelligence liaison

[2] GC&CS report ML9605, 012128Z Aug 1943, HW 20/136, UK National Archives. ML9605 refers to a previous report, ML9464 of 311401Z Jul 1943 (HW 20/135, UK National Archives), identifying the downed aircraft.

[3] GC&CS report ML9380, 302114Z Jul 1943, HW 20/135, UK National Archives.

[4] GC&CS report ML9382, 302125Z Jul 1943, HW 20/135, UK National Archives.

officers from the War or Air Ministries. Certainly GC&CS was sensitive to the implications of Mussolini's sudden fall from power on 25 July, as it had published in sensitive channels for specifically named recipients on the 27th a report based on orders from German Navy communications on contingency measures to be taken in light of reports that the Italians were negotiating an armistice with the allies.[5] Nevertheless, the absence of a comment drawing attention to the political implications of an intelligence fact is not itself significant.

3 Editing Error

A clue suggesting that the significance of the Italian Admiralty orders of 30 July may have been missed lies in a handwritten, pen-and-ink change to the report. Occasional handwritten amendments, most likely by the Hut 3 Duty Officer, to the typewritten message drafts by the Advisor commonly added an address group not on the original draft or a missing word that the typist had overlooked, or corrected a typographical error before the report was taken to the communications center for encryption and transmission. In the 30 July message based on the Italian Admiralty order, pen-and-ink changes were made to correct the second spelling of the *Littorio* (proper names and locations were repeated likely to ensure clarity against garbled transmission) and in the same vein, to change the spelling of the vessel name, *Camicia Nera*, Black Shirt, erroneously to *Camilia Nera*, Black Camilia. This mistaken editing error was almost certainly committed by someone who knew no Italian, most likely the Hut 3 Duty Officer performing the last review of the report before release. The date-time-group (DTG) of the report, assigned by the Advisor typing the report as the time of origination, indicates this report was generated during the busy evening shift, and the DTGs of reports originated up to and following this report suggest the Duty Officer most likely had only a few minutes to review each package. Operational tempo was a likely factor in an all too human error, while the misjudgment nevertheless indicates the Duty Officer could not have grasped the significance of the name change nor considered why the Italian Navy would bother. Like the order to repaint Italian Air Force aircraft, changing a naval vessel name would not be a trivial matter of merely repainting the stern of the ship, but would entail changing numerous records at the Italian Admiralty and subordinate commands related to administration, communications routing, *et al*. The editing error on the 30 July Italian Admiralty report, most likely based on the mistaken assumption of a typographical error, is the clearest indication that the deeper political meaning of the Italian Navy order was missed at the initial production level.

4 Understandable Oversight

As with the instance with the Volkswagen translation, the initial oversight of fascist symbology is understandable. History is usually written when the ending is known and a certain perspective reached. In 1943, the history of Italian fascism had not yet been written. The references to the fasces and the ancient Roman lictor would likely have been recognized by a classical scholar, but the appropriation of these symbols and the naval reference to the Black Shirts, though discernible only to an Italian linguist, would more likely have been recognized by a specialist in contemporary Italian politics at that time. The generational dead zone between history and current events in any generation may have contributed to a contextual blindspot that inhibited understanding of the implications of these communications in the summer of 1943. Recipients most likely to have recognized the error would have been Ultra-cleared analysts responsible for maintaining continuity on the Italian Navy fleet inventory, most certainly at the British Admiralty, but also likely at the Combined Bureau Middle East in Cairo and elsewhere in the region. Analysts maintaining continuity on the

[5] GC&CS report ML9026, 271925Z Jul 1943, HW 20/134, UK National Archives.

Italian Naval order of battle would likely have recognized the editing error and made the necessary adjustment to the intelligence records on the Italian Navy order of battle. It would thus most likely have been a self-correcting error between intelligence producer and consumer concerned at the military and not political level.

5 Limited Impact

One must not overestimate the implications of such an oversight in a relatively small piece of intelligence in the immediate aftermath of Mussolini's fall from power. The possibility that the allied invasion of Sicily could precipitate Italian exit from the war was certainly envisaged by allied leaders. The British government had received feelers from Marshall Badoglio suggesting the latter was ready to desert Mussolini as early as December, 1942.[6] Badoglio's anti-fascist credentials formed a substantial part of a preliminary assessment by the British Joint Intelligence Committee (JIC) to the British chiefs of staff within days of Mussolini's fall. While the analytical report couched its findings as subject to revision as new intelligence became available, it asserted that Badoglio's accession to power was truly a change in regime, not a mere change in government.[7] To the larger question of the orientation of the new government, the shedding of fascist symbols in two military services as revealed in these Ultra reports could only be reinforcing evidence to what the JIC already believed to be true.

Regardless of its negligible effect on an already formed intelligence conclusion, the Ultra intelligence indicating the renaming of the two naval combatants was relevant to follow-on policy measures under consideration to respond to the change in Italian regime. At the same Chiefs of Staff meeting on 28 July, First Sea Lord Sir Dudley Pound circulated a memorandum proposing that propaganda be employed to discourage Italians from sabotaging their naval and merchant ships in the event of an Italian political collapse.[8] The removal of fascist connections in vessel names in the fleet as reported in Ultra would certainly have been a useful input to understanding the command climate in the Italian Navy pursuant to constructing such a propaganda campaign.

6 Conclusion

While not a critical piece of intelligence to assess the posture of the new Badoglio government toward turning the page on Italian fascism, Ultra intelligence on the Italian armed forces shedding their fascist symbology, if properly understood and conveyed to the intelligence customer, would have reinforced the existing assessment of the Badoglio regime and could have influenced policy response.

Besides the historical aspect, there are linguistic implications. In a pre-Artificial Intelligence environment, language immersion would be the remedy for contextual blindspots and oversights. As Large Language Models (LLMs) seek to emulate linguistic immersion, reaching the depth of analysis required for full understanding of information conveyed in a foreign language in these historical examples may pose an interesting challenge for LLMs.

Acknowledgments

The views expressed in this paper are those of the author and do not reflect the views of any governmental or academic institution with which the author has been associated.

References

GC&CS Reports. 1943. Record volumes HW 20/134, HW 20/135, HW 20/136. UK National Archives. Kew, Richmond. UK.

William Millward. 1993. Life in and out of Hut 3, in *Codebreakers: The Inside Story of Bletchley*

[6] Woodward, *British Foreign Policy in the Second World War*, p. 229.

[7] War Cabinet Chiefs of Staff Committee meeting minutes, 28 July 1943, CAB 79/63, UK National Archives.

[8] *Ibid*.

Park. F.H. Hinsley and Alan Stripp, eds. Oxford University Press. Oxford. UK.

War Cabinet Chiefs of Staff Committee. 1943. Record volume CAB 79/63. UK National Archives. Kew, Richmond. UK..

Sir Llewellyn Woodward. 1962. *British Foreign Policy in the Second World War*. Her Majesty's Stationery Office. London. UK

Fake or real? A mysterious metal book on the market

Levente Zoltán Király
Independent scholar
Vienna, Austria
lev.z.kiraly@gmail.com

Benedek Láng
Eötvös Loránd University
Budapest, Hungary
lang.benedek@gtk.elte.hu

Gábor Tokai
Museum of Fine Arts
Budapest, Hungary
gabor.tokai@szepmuveszeti.hu

Abstract

A newly emerged gilded folio metal book containing illustrations and unreadable character strings raises the question of authenticity. The article describes the object and examines on what grounds can be claimed with relative confidence that such a book is a forgery. The examination includes a structural analysis of the symbol sets and a comparison with similar, analogous metal books.

1 Introduction

In 2023 a mysterious metal book appeared on the book market. The small format (approximately 20cmx10cm) booklet is made of 16 metal plates embossed by various symbols. The folios are called by the owner gilded silver plates. The owners, who wished to remain anonymous, offered the object for sale to the Library of the Hungarian Academy of Sciences, claiming that it was a valuable rare object. The authors of this study got an informal request to analyze the book's authenticity. What follows below is not an official evaluation, but rather the results of our analysis based on historical analogies and the behavior of the writing system. Note that the authors were not in the position to examine the book as an object, they were only provided high-quality photos of all the folios.

2 The first impressions

Upon perusing the folios of the book under scrutiny, one is immediately struck by the presence of illustrations. Notably, discernible depictions include the Star of David, the seven-armed candlestick, the palm tree, the eye of God, the "prophet", and an oriental female figure (see Figure 1). The book's features are reminiscent of Phoenician or proto-Hebrew origins, while others evoke the imagery of Etruscan or early Greek signs. Additionally, interspersed within the book are characters reminiscent of the Etruscan and the Phoenician alphabets as well as other signs known as charaktêres, and encountered in Jewish magical manuscripts and Renaissance magic handbooks (Diringer 1948, Facchetti 2001, Le Pape 2006, Gordon 2014, Buda 2022).

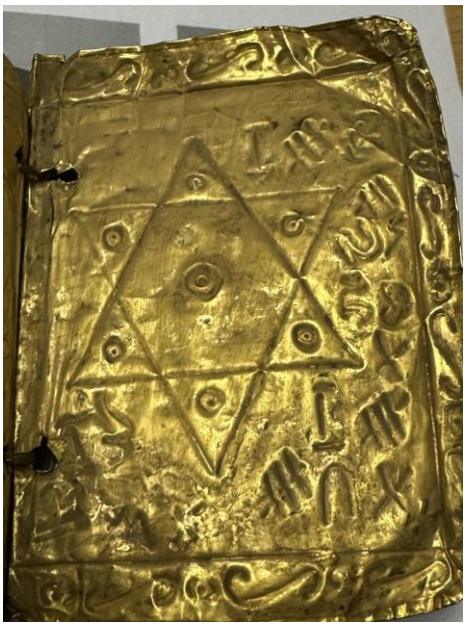


Figure 1. Folio no. 1 from the metal book with the Star of David

It becomes apparent that certain symbols, signs, and illustrations within the book convey an intentional effort by the author to impart the impression that an original Biblical or similar sacred text had been created. However, upon closer examination, the "text" contained within

the book reveals brevity and a notable absence of the structural elements characteristic of realistic texts, even those veiled in cryptic scripts (see Figure 2).

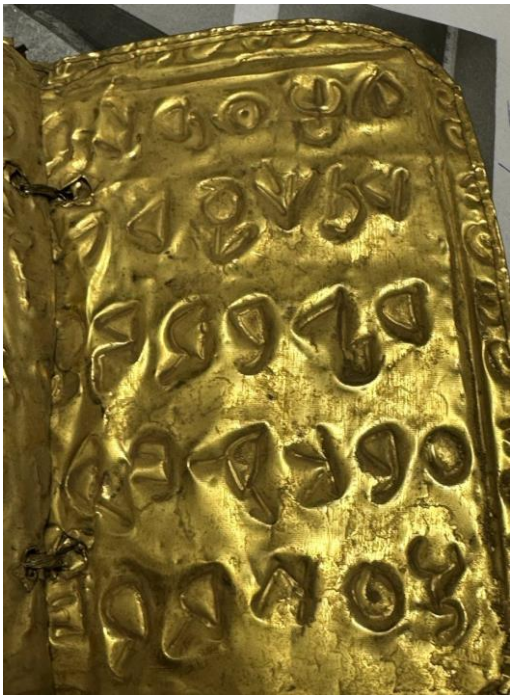


Figure 2. Folio no. 12. from the metal book with "text"

3 Analogies

In terms of visual resemblance, the book bears a striking similarity to the Pyrgi gold plates discovered in 1964 and gold-plated books recently unearthed in Turkey. Three golden metal plates discovered near Pyrgi contain bilingual Phoenician and Etruscan dedicatory text (Smith, 2016, Schmitz 1995.). While these languages were not unchallenging to the linguistic experts of pre-Roman Italy, they succeeded in decrypting and interpreting them, and the plates were never suspected to be fake.

Considering the book format of our source, even closer analogies emerged from Turkey in recent years. As one of the news titles put it: "In Turkey, counterfeit Jewish artifacts are commonplace – and often sloppy" (Klein 2021). Some of these are amulet-like books. A particularly similar object to our book is reported in the Turkish news as a fake.[1] As far as one can judge from the video, this second golden book is also relatively short, equally small in size, and contains some of the same images. Character strings are equally scarce, only making up for a few folios.

4 The writing system

The symbol sets of the whole book have been transcribed manually to analyze whether they show any structures that may refer to a natural language or any kind of intentional structure (including encryption) (see Figure 3).

We have tried to reconstruct the folio layout of the book on the basis of the photos. In the transcription, we considered it logical that the recto folio is the folio from which the inscription is embossed. It is more difficult to imagine that the intended writing is mirrored first and then embossed, but it is clear that the convex shapes are turned outside on the two cover folios of the book. This is probably because the 'author' realized that the convex writing is more readable and looks better.

As a result of the reconstruction, 16 folios (each with a recto and verso) were identified and numbered. The numbering of the figures in this article refers to this line-up.

Figure 3. Transcription of folio no. 12.

We could split the symbols of the script into two groups. In one group, the Phoenician alphabet seems to have been the model, and the folios that mostly have symbols of this group are mainly found at the beginning of the book. In the other

[1]https://www.youtube.com/watch?v=19dK4-9R8wI&t=13s

group, the Etruscan alphabet seems to have been the model. This is more typical towards the end of the book. For this reason, one could even think of two authors or two stages in the preparation. However, the two groups are not perfectly distinct. On one hand, there are characters that are common to both alphabets. On the other hand, there are some characters that appear in the other group in isolated instances. Among the repeated character strings there are also some that appear on a folio of the other type.

In the Etruscan alphabet (the order of which, by the way, has been preserved in ancient relics), there is hardly any missing character that has no equivalent in the metal book, but among those that are reminiscent of the Phoenician alphabet, there are many missing.

A peculiarity of the Etruscan letter group is that some of the signs are reminiscent of its archaic figures while others remind us of the later forms. These have never occurred simultaneously in history. Many are found only in the archaic alphabet (B, D, O, S); some are later developments (F). The vowels of the Etruscan alphabet (A, E, I, O, U) are extremely rare in the text. The author might have imitated the consonant-dominated Phoenician script.

It is plausible to imagine that the author has selected the symbols from a table of alphabets showing the different periods together. Browsing the internet for samples of "golden books" immediately brings up the Pyrgi gold tablets that also contain a Phoenician-Etruscan bilingual inscription. At the same time, however, there is no doubt that the Phoenician of Pyrgi is a later cursive (Punic) version, which did not necessarily provide a direct model for the present "Phoenician" characters.

In the metal book, there are altogether 462 characters, making up for a very short text in any writing system. Among these characters, there are 92 different ones (many of which have variations, which eventually could be different letters). In the “Phoenician corpus” there are 20 different Phoenician characters, and 27 further graphic signs.

In the “Etruscan corpus” (folios mainly in the second part of the book, where the Etruscan alphabet dominates), there are 27 different symbols reminiscent of the elements of the Etruscan-Latin alphabet, and 24 further signs.

There are also 11 Etruscan-Latin symbols in the “Phoenician corpus” on isolated positions, and the “Etruscan corpus” also contains 3 isolated Phoenician symbols. Note that there are 6 characters that are common to the two alphabets creating a low degree of ambiguity.

The statistical analysis of the text reveals that many of the symbols occur very rarely. Half of the 92 different symbols – precisely 42 – only occur once. 76% of the symbols (70) occur 1 to 4 times.

The recurrent symbols also exhibit an unusual pattern, with five of them recurring so frequently that they collectively add up to 180 characters, constituting a noteworthy portion (38%) of the text's total character count of 462. An additional set of 17 symbols appear at 147 places, amounting to 31% of the total. This implies that a subset resembling an alphabet, comprising 22 symbols, collectively furnishes two-thirds of the entire text.

Merely 92 signs would indicate syllabic spelling, but syllabic spellings do not have such sign distribution statistics. The above statistics make it improbable that the writing is a letter script in a natural language. The only probable option is that the text is based on an alphabet complemented by logograms (such as a monoalphabetic cipher system with nomenclators).

However, assuming for a while that this is the scenario (an alphabet script comprising 22 symbols enriched with infrequently appearing logograms), the scenario remains disorderly. Each folio exhibits a unique pattern. Folio 16 (a cover folio) features symbols occurring only once. Folios 3, 8, and 11 predominantly contain only a few uncommon symbols, making them applicable to an alphabet analysis. Folios 14 and 6 encompass numerous symbols occurring 1–4 times and almost lack "alphabet letters," yet even the infrequent symbols exhibit an unusual behavior: nearly all their instances in the book originate from these two folios. Numerous cases can be found where all occurrences of a rare symbol are confined to a specific folio. The distribution of symbols appears too arbitrary for a system combining an alphabetical framework with logograms.

Analysts and codebreakers of unknown character sequences first look for patterns and repetitions. Since both Phoenician and Etruscan were written from right to left, the transcription

followed the same direction. The repeated combinations of signs in the metal book's alphabet are usually no more than two characters long; each corpus has one repeated trigram. But perhaps there is a certain intentionality in the combination where two h's are repeated (there are 4 such places) because this cannot be accidental. In addition, on folio 4, two double signs are in the same order. It is also noteworthy that the bigram combinations mainly include the sign "samekh" in Phoenician and "theta" in Etruscan, which does not necessarily indicate the possibility of random repetition either.

The inclusion of "REX" on folio 14 adds complexity—unclear if coincidence or an intentional trace that is meant to prove that the text is meaningful. However, this term is to be read in the opposite direction of the assumed right-to-left writing direction in order to yield a meaningful interpretation. Additionally, considering the existence of mirror-symmetrical variants for multiple signs, one could posit that signs resembling early Greek script should be read in a system known as boustrophedon ('ox-turning,' alternating direction line by line), a phenomenon frequently cited in the history of Greek writing. The text on folio 14 does indeed appear interpretable as if written in a boustrophedon system.

This inference is supported by the presence of "directional arrows" between lines on four folios (2, 6, 9, 15), which, in any text, have an unusually striking effect—unless we consider them as auxiliary markers intending to illustrate the operation of boustrophedon in a modern study of writing history (see Figure 4).

While the assumption of boustrophedon writing direction could provide an explanation for several phenomena in the book, its application beyond the REX term has not been justified. (It is worth noting that, apart from theoretical considerations, clear indications regarding the direction of writing have generally proven elusive.)

5 Conclusions

Upon comprehensive examination of all the folios, a prevailing sentiment emerges—that of an assemblage of seemingly arbitrary and devoid-of-meaning signs, lacking any discernible underlying system. The impression garnered from the entirety of the book is one characterized by randomness, with the signs appearing haphazardly arranged, defying attempts to unveil a coherent pattern.

We do not know what exactly this object purports to be apart from the indication that it is precious. Therefore, we would not go so far as to label it a fake. Furthermore, a thorough analysis should include the workmanship of the book, the metal should be examined by a materials expert. However, the textual analysis raises serious doubts about its authenticity.

Figure 4, verso of folio no. 6 with the "directional arrows"

Finally, the authors' impression is that the creators of this metal book lacked historical perspective. Such deliberately crude, careless, "primitive" workmanship can only occur in the mind of someone who wants to give credibility to the antiquity of the object but has no nuanced knowledge of antique craftsmanship that may seem primitive to modern man but is carried out with careful and delicate workmanship, orderliness, thoughtfulness.

6 Acknowledgments

Benedek Láng's research has been supported by the Swedish Research Council, grant 2018-06074: DECRYPT—Decryption of historical manuscripts.

References

Zsofia Buda. 2022. “Speaking to angels: Charaktêres in Jewish magical manuscripts” blog entry in Rylands blog, https://rylandscollections.com/

David Diringer. 1948. *The Alphabet. A Key to the History of Mankind.* Scientific and Technical Publications, London, 213.

Giulio M. Facchetti. 2001. *L’enigma svelato della lingua etrusca*. Newton & Compton editori, Roma, esp. 18–24.

Richard Gordon. 2014. “Charaktêres between Antiquity and Renaissance: Transmission and Re-invention,” in *Les savoirs magiques et leur transmission de l’Antiquité à la Renaissance*, ed. Véronique Dasen and Jean-Michel Spieser, Florence, 253–300.

Gilles Le Pape. 2006. *Les ecritures magiques: aux sources du „Registre des 2400 noms”d’anges et d’archanges de Martines de Pasqually.* Arche, Milano.

David Ian Klein. 2021. “In Turkey, counterfeit Jewish artifacts are commonplace – and often sloppy” *Forward* portal https://forward.com/news/467503/in-turkey-counterfeit-jewish-artifacts-are-commonplace-and-often-sloppy/ accessed: 2023.12.14.

Philip C. Schmitz. 1995. The Phoenician text from the Etruscan sanctuary at Pyrgi. *Journal of the American Oriental Society*, 559-575.

Christopher John Smith. 2016. “The Pyrgi Tablets and the View From Rome" in *Le Lamine di Pyrgi* eds V. Bellelli and P. Xella, Verona, 203–221.

Decipherment of an Encrypted Letter from 1724 Found in UCL Special Collections' Brougham Archive

Nils Kopal
University of Siegen
Siegen, Germany
nils.kopal@uni-siegen.de

Katy Makin
UCL Special Collections
London, United Kingdom
k.makin@ucl.ac.uk

Abstract

This paper shows the decipherment of a 1724 encrypted letter, discovered recently in the Brougham Archive at University College London (UCL) Special Collections. The letter's content hints at political intrigue and possibly relates to the Jacobite movement during George I's reign in Great Britain. However, as all individuals mentioned in the letter are referred to by code names, except for Madame de Prie, their true identities remain unknown to the authors. Therefore, any connection to the Jacobites remains speculative. The paper covers the cipher's security, historical context, and unresolved inquiries surrounding the letter.

1 Introduction

Members of the DECRYPT research project (Megyesi et al., 2020) assist other researchers and historians when faced with encrypted texts discovered in archives and libraries. Often, these individuals find themselves unable to decipher the found scripts on their own (Megyesi et al., 2024). This is where the expertise of the DECRYPT project comes into play. An example of successful collaboration between DECRYPT experts and a historian is the decryption of the Ramanacoil cipher, a ciphertext from the Dutch East India Company from 1674 (Dinnissen and Kopal, 2021). Other examples for fruitful collaborations are the decipherment of the Codex Copiale (Knight et al., 2011), an encrypted manuscript from the 18th century, the decipherment of letters of Holy Roman Emperor Maximilian II written in 1575 (Kopal and Waldispühl, 2022), as well as the recent breakthrough in deciphering newly found Mary Stuart ciphers (Lasry et al., 2023).

Members of the DECRYPT project are actively engaged in the search for such encrypted manuscripts within archives and libraries. Occasionally, encrypted manuscripts also find their way to DECRYPT researchers through other ways. In 2023, the first author of this paper came across a call for assistance from UCL Library on X (formerly Twitter). The call sought help in deciphering a letter discovered in one of their archives, the Brougham Archive. The first author undertook the task of deciphering the pages of the letter presented on X but realized that these pages were not the complete document. Subsequently, the author contacted the archivist, who is also the co-author of this article, and obtained scans of all available pages, along with valuable background information. This paper discusses the decipherment and content of this letter.

2 UCL Special Collections and Brougham Archive

The letter was found in a box of family documents in the archive of Henry Brougham, 1st Baron Brougham and Vaux (1778-1868), one of the founders of UCL. The collection is a true 'family archive', containing both Henry Brougham's working papers and extensive correspondence, and also letters, deeds and estate papers belonging to his siblings and ancestors. At around 160 linear meters, it is the largest and one of the most important archives at UCL Special Collections. It is currently in the process of being cataloged, which is how the letter came to light. As the encrypted letter predates Henry Brougham by over 50 years, it seems likely that it belonged to an ancestor. The other items in the box where it was found provide no clues to its origins, being a miscellaneous assortment of 19th century notes, deeds and bills. However, in the 18th century the Brougham family had connections to the Dukes of Norfolk, prominent Catholics and Jacobite supporters who were involved in the uprising of 1715.

3 The letter

The encrypted letter is written on paper that has aged to a yellow hue over time, with noticeable ink bleed-through visible on the pages. Figure 1 shows the first page of the letter. The letter comprises a total of 16 pages of ciphertext, written in numbers separated by dots and dashes. In some places between these numbers, there are a few simple words written in cleartext. Examples are ”and”, ”the”, and ”whatever”. Also, there are names like ”Mons Garnet and Gee” and ”Mons Grandy and Gay”. The letter’s last five pages are damaged, so we were not able to transcribe and decipher everything with 100% accuracy, but were able to deduce the missing text parts with high confidence. Further analysis indicates that the letter is incomplete, suggesting the existence of additional pages that, regrettably, were not found in the UCL Brougham archive. At the top of the first page, the letter is dated February 24, 1724. Additionally, the first letter includes a unique form of signature we are unable to identify its meaning. The pages are sequentially numbered, ranging from 2 on the second page to 14 on the fourteenth page. These numbers are located either on the top left or right side of each page. Pages 15 and 16 likely had page numbers, but due to the damage to the papers, these no longer exist.

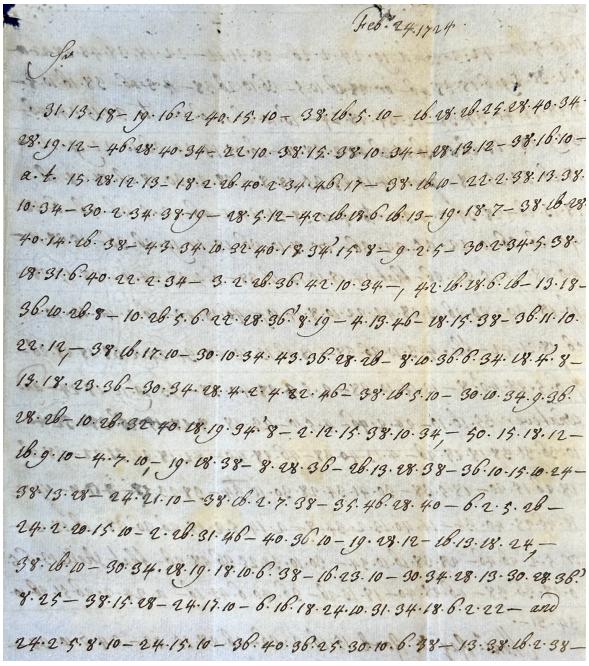


Figure 1: First page of the encrypted letter. UCL Special Collections, Brougham Archive [uncatalogued] (Cipher ID-6317, 1724)

4 Transcription

For the transcription of the letter, a manual process was conducted. All the numbers were meticulously entered along with the dots and dashes between them into a text editor. Cleartext elements were also accurately transcribed. At points where the ciphertext was damaged, question marks were inserted. Throughout the transcription process, care was taken to maintain the original line and page layout.

5 Cryptanalysis

For the cryptanalysis of the letter, we employed our open-source tool, CrypTool 2 (Kopal, 2018) and its Homophonic Substitution Analyzer component, which uses heuristics (hill climbing; simulated annealing) for automated cryptanalysis (Kopal, 2019). Given the large number of different numbers present in the ciphertext (totaling 52), we hypothesized that the cipher used is a homophonic substitution cipher. We also worked under the assumption that the plaintext was in English, as the letter originated from University College London and the readable cleartext parts of the letter are written in English. Consequently, we set our analysis components to English for the cryptanalysis.

Initially, the automatic analysis with the Homophonic Substitution Analyzer yielded no results (Figure 3 shows the cryptanalysis in CrypTool 2). After several restarts, we began to see English words emerge, but they were separated by incorrect letters. Upon closer examination of these separations and the corresponding ciphertext numbers, we noticed that characters separating words were always represented by odd numbers in the ciphertext. This led us to hypothesize that all odd numbers were ’nulls’, meaning they were characters without meaning, intended to confuse a cryptanalyst during cryptanalysis. We then marked all odd numbers as nulls, instructing the Homophonic Substitution Analyzer to ignore them during automated analysis. Additionally, we configured the initial key of the cryptanalysis algorithm to assign only one letter for each ciphertext symbol in the letter distribution. These adjustments allowed us to decipher the full text, revealing that it was encrypted using only a simple monoalphabetic substitution. Figure 2 presents a graph depicting the

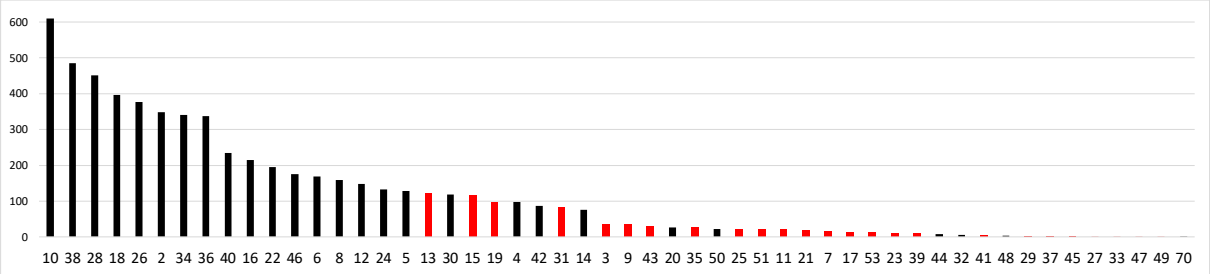


Figure 2: Ciphertext number frequencies: black bars for symbols, red bars for nulls.

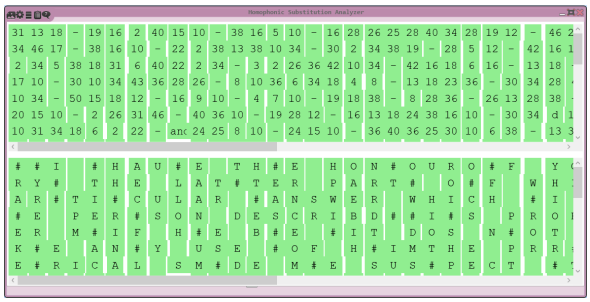


Figure 3: CrypTool 2's Homophonic Substitution Analyzer analyzing the ciphertext

distribution of numbers in the ciphertext. The graph utilizes black bars to represent actual ciphertext symbols, while red bars are used to denote nulls. In the deciphered text, the letter pairs 'i' and 'j', as well as 'u' and 'v', are each encrypted with the same ciphertext symbol, a characteristic typical of ciphers of that period.

In the following, we present the cipher's key:

```
nulls = odd numbers; A = 2; B = 4; C = 6;
D = 8; E = 10; F = 12; G = 14; H = 16; I = 18;
K = 20; L = 22; M = 24; N = 26; O = 28;
P = 30; Q = 32; R = 34; S = 36; T = 38;
U = 40; W = 42; X = 44; Y = 46; Z = 48; & = 50
```

As can be observed in the shown key, the distribution of ciphertext symbols (numbers) onto plaintext characters (letters) follows a highly regular pattern: 'A' starts with 2, 'B' follows with 4, 'C' with 6... and 'Z' ends with 48. In other words, a simple consecutive numbering scheme (using even numbers) was employed for the creation of the cipher. It's evident that this approach is quite careless and even in the time of its creation was not a recommended method for generating a cipher. The consistent pattern simplifies the cryptanalysis process once the cryptanalyst recognizes it. Significantly, the word 'and' is the only word assigned its distinct ciphertext symbol (50).

We also closely examined the cleartext words appearing within the ciphertext numbers. The simple words among these fit syntactically into the sentences where they are placed, acting as regular parts of these sentences. But the code names that appear also integrate smoothly into the flow of the letter's text. Based on their positioning, we suspect that these names might refer to either people or places, even if only personal names are used as code words. Interestingly, most of the code words in the text appear as pairs and as alliterations, meaning each name of a code word starts with the same initial letter, such as in the example "Waller and Wall."

We also concluded that these code words serve as the so-called 'nomenclature elements' of the letter. They still represent the most significant mystery of the letter yet to be solved. Without additional context or the original key used, we cannot ascertain the meaning of these names. Historians should look at the deciphered letter's cleartext to possibly learn more about these code names and give more meaning to them.

6 Edition of the letter

This section presents an edited version of the first two pages of the decrypted letter. All decrypted plaintext letters are capitalized for clarity. The code words found in the text are written in their original form. All nulls were removed from the text. Additionally, punctuation (full stops for endings of sentences) has been added to enhance readability. Typos were left in the text:

Page 1:

```
Feb 24 1724
I HAUE THE HONOUR OF YOUR LETTER OF THE a.t.
OF IANUARY THE LATTER PART OF WHICH I THOUGHT
REQUIRED A PARTICULAR ANSWER WHICH I SEND
ENCLOS BY ITSELF. THE PERSON DESCRIBD IS
PROBABLY THE PERSON ENQUIRD AFTER AND IF HE
BE IT DOS NOT SEEM TO ME THAT YOU CAN MAKE
ANY USE OF HIM. THE PROIECT HE PROPOSD TO ME
CHIMERICAL and MADE ME SUSPECT THAT
```

Page 2:
HE WAS a MAN OF LITTLE CAPACITY OR ONE SENT
BY THE COURT FROM HENCE TO TRY TO INSINUATE
HIMSELF INTO YOUR GOOD OPINION IN ORDER TO
BETRAY ANY CIUNSELLS OR DESIGNS OF YOURS THAT
MIGHT COME TO HIS KNOWLEDGE YOU HAUE. NOBA
THE BEST INFORMATION JCOUD GET ABOUT HIM and
whatever HE MAY BE I DONT QUESTION BUT YOU
WILL THINK IT PROPER TO BE upon YOUR GUARD
AGAINST HIA OR ANY OTHER PERSON THAT SHALL
COST YOU IN SUCH A MANNER.
Mons Ray & Rook HAS NOW MET

To facilitate reading the complete letter, the letter as well as its plaintext and all the photos of the ciphertext pages have been uploaded to the DECODE database (Héder and Megyesi, 2022), a repository for historical ciphers and keys (see (**?**)).

7 Content of the letter

Here are some interesting points mentioned in the letter. The use of code names makes it unclear who the involved individuals are and who is being referred to.

- An unknown individual proposed an unrealistic project, raising suspicions of potential betrayal.
- Mention is made of the political situation, with Mons. Ray & Rook (unidentified individuals) being inactive for six weeks.
- The Mons. Garnet and Gee, and the Mons. Grandy and Gay (unidentified – perhaps organs of government?) have committed 4,000 men to the cause.
- The Monsieur Waller & Wall group (unidentified individuals) had become cautious and hesitant in discussions.
- Success relied on public sentiment and external support, not just prominent allies.
- Caution was urged when dealing with loyal yet doubting friends.
- Nevertheless, the current political situation looked hopeful and it would soon be a good time to take action.
- Mr Echard and Mr Patington (unidentified individuals) could be persuaded if they were offered terms not prejudicial to the English interests.
- It is mentioned that Madame De Prie is inclined to support the cause mentioned but may require encouragement through financial incentives, which should be offered discreetly and skillfully.

Both the sender(s) and the recipient(s) of the letter remain unclear to us. The letter lacks salutations, with the mentioned individuals referred to only by code names (except for the mentioned Madame de Prie). Additionally, the letter ends abruptly due to missing pages, leaving us unaware of the identity of the author(s).

8 Historical Context

The letter, composed in 1724 during the reign of George I of Great Britain, coincided with a period marked by the looming threat of the Jacobite movement. The Jacobites, loyal to the exiled Stuart dynasty, were actively striving to reinstate a Stuart monarch on the British throne. This era witnessed significant Jacobite uprisings, most notably the rebellions of 1715 and 1745. The contents of the letter hint at potential connections to this movement, as the sender seeks French assistance, particularly from Madame de Prie, in their endeavors. Madame de Prie (1698-1727), also known as Jeanne Agnès Berthelot de Pléneuf, was a prominent figure in the French court during the 18th century, being the mistress of Louis Henri, Duke of Bourbon (1692-1740; "Monsieur le Duc"), who was prime minister of France. She held significant influence on him and played a crucial role in the political affairs of her time.

9 Conclusion

In summary, we can conclude that the letter was relatively easy to decipher, as the cipher used was weak even for its time. However, the letter leaves many unanswered questions, as we do not know the sender or the recipient, and the referenced individuals cannot be identified due to the use of code names. We suspect that the letter was sent within the Jacobite movement. Nevertheless, it is certain that the letter was neither written nor received by Henry Brougham (in whose archive the letter was found) since he was born in 1778, long after the letter was sent. It is possible that one of his ancestors had a connection to the Jacobites, leading to the letter's presence in his archive. Now, historians should analyze the letter, conduct background research, and solve the remaining mysteries.

Acknowledgments

This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Cipher ID-6317. 1724. Reproduced image from Ciphertext found in the UCL Special Collections' Brougham Archive. DECODE ID 6317, link: https://de-crypt.org/r/6317.

Jörgen Dinnissen and Nils Kopal. 2021. Island Ramanacoil a Bridge too Far. A Dutch Ciphertext from 1674. In *International Conference on Historical Cryptology*, pages 48–57.

Mihály Héder and Beáta Megyesi. 2022. The DECODE Database of Historical Ciphers and Keys: Version 2. In *International Conference on Historical Cryptology*, pages 111–114.

Kevin Knight, Beáta Megyesi, and Christiane Schaefer. 2011. The Copiale Cipher. In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web*, pages 2–9.

Nils Kopal and Michelle Waldispühl. 2022. Deciphering Three Diplomatic Letters Sent by Maximilian II in 1575. *Cryptologia*, 46(2):103–127.

Nils Kopal. 2018. Solving Classical Ciphers with CrypTool 2. In *Proceedings of the 1st International Conference on Historical Cryptology HistoCrypt 2018*, number 149, pages 29–38. Linköping University Electronic Press.

Nils Kopal. 2019. Cryptanalysis of Homophonic Substitution Ciphers Using Simulated Annealing with Fixed Temperature. In *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt*, pages 107–16.

George Lasry, Norbert Biermann, and Satoshi Tomokiyo. 2023. Deciphering Mary Stuart's Lost Letters From 1578-1584. *Cryptologia*, 47(2):101–202.

Beáta Megyesi, Bernhard Esslinger, Alicia Fornés, Nils Kopal, Benedek Láng, George Lasry, Karl de Leeuw, Eva Pettersson, Arno Wacker, and Michelle Waldispühl. 2020. Decryption of Historical Manuscripts: the DECRYPT Project. *Cryptologia*, 0(0):1–15.

Beáta Megyesi, Alicia Fornés, Nils Kopal, Benedek Láng, Michelle Waldispühl, Vasily Mikhalev, and Bernhard Esslinger. 2024. Historical Cryptology. In *Chapter 3 of Bernhard Esslinger, Learning and Experiencing Cryptography with CrypTool and SageMath*, pages 97–138. Artech House, Norwood.

Sources of Alchemical Cryptography

Sarah Lang
Centre for Information Modelling
University of Graz
Elisabethstraße 59/III
8042 Graz, Austria
sarah.lang@uni-graz.at

Sergei Zotov
Centre for the Study
of the Renaissance
University of Warwick
Coventry CV4 7 AL, UK
sergei.zotov@warwick.ac.uk

Megan Piorko
Falvey Library
Villanova University
800 Lancaster Ave.
Villanova, PA 19085, USA
megan.piorko@villanova.edu

Abstract

This paper presents an initial overview of cryptographic sources relating to alchemy, an area that remains largely unexplored. Alchemists and chymists frequently encrypted short passages relating to recipes and experiments, obscured content using exotic foreign languages or custom shorthand, and created unique symbol codes. A survey of manuscripts reveals the diversity of sources in over 100 instances of ciphering in alchemical contexts, where ciphers were only one of several methods traditionally used to maintain secrecy. It serves as a starting point for further research, demonstrating the wealth of archival material related to alchemical cryptography – a goldmine yet untapped.

1 Introduction

The alchemical tradition is rife with practices of secrecy, but its cryptographic and steganographic habits are not yet well understood. This article builds upon the publication by Lang (2023), which contextualized ciphers within alchemical secrecy techniques. Working within the framework posed by Lang, this survey focuses on the scribal practices of encryption within alchemical secrecy. The current state of the field lacks comprehensive bibliographical information on extant alchemical ciphers discovered in early modern manuscripts, posing a roadblock preventing scholars from a comprehensive understanding of early modern encryption practices and scribal knowledge-production among natural philosophers. This article aims to address this by presenting and contextualizing alchemical ciphers from over 110 manuscript sources located in libraries across 10 countries.[1] This survey is an early attempt to systematically identify sources of alchemical cryptography to lay the groundwork for future scholarship.

2 Literature Review

The realm of alchemical ciphers represents a largely unexplored area in the study of alchemy. These ciphers, though known to some extent within the field, have not been subjected to the comprehensive analysis they merit. Alchemy's communication traditions are deeply embedded in sophisticated secret-keeping methods, with ciphers being just one aspect of these practices. David Kahn's seminal work, *The Codebreakers*, briefly acknowledges the use of enigmatic symbols in astrology and alchemy, noting that, similar to ciphers, these symbols may appear nonsensical but are, in fact, laden with concealed meaning (Kahn, 1996, 91). However, this reference is cursory at best.[2]

Benedek Láng observes, as others have also noted, that the encipherment methods in alchemy and chymistry seem to operate differently compared to other scientific secrecy techniques (Láng, 2018, 163, 165–166). He points out that only a limited number of ciphers from alchemical texts predating 1600 are documented (Láng, 2018, 165), such as examples from the *Libro del Tesoro* in Madrid, Martin Roesel von Rosenthal's recipes (˜1586), the mid-16th century diary of Cluj artisan Johannes Cementes of Kolozsvár, in addition to the ciphers used in the 17th century by chymist Robert Boyle (1627–1691) in his laboratory notebooks (Principe, 1992; Hunter, 2016). Focused research is largely limited to 17th-century

[1] While we acknowledge there exists a plethora of printed examples of encryption, this paper focuses explicitly on manuscript examples as a particular method of knowledge dissemination with the shared goal of encrypting and decrypting the secrets of nature.

[2] Agnieszka Rec highlights the scarcity of research on alchemical ciphers, despite their widespread presence in alchemical writings (Rec, 2014).

chymists like Robert Boyle, who publicly championed transparent communication in chymistry.

Alchemists and chymists traditionally used a whole range of methods to hide their knowledge from the uninitiated, often employing multiple layers of technologies of textual concealment. Alchemical secrecy devices can include, but are not limited to, the following types: *Decknamen* and specialist terminology (Newman, 1996), word/name substitution (Principe, 1992), dispersion of knowledge (*dispersio*) across multiple sources (Principe, 1992, 65), *parathesis* and *syncope* (Newman, 1996), monoalphabetic ciphers (Principe, 1992, 67), polyalphabetic ciphers (Bean et al., 2022), trade symbols and codes (Gaede, 2017), alphanumeric knowledge charts (Forshaw, 2005; Clucas, 2017), astrological horoscopes (Piorko et al., 2023), cabbalistic mysticism (Forshaw, 2013), Lullian diagrams (Forshaw, 2013), emblems (Bilak, 2020), (mythoalchemical) allegory (Forshaw, 2020), omitted or enciphered publication information (Purš and Hausenblasová, 2005; Piorko, 2019), pseudonomia (Newman, 1991) or even the use of multiple languages (Principe, 1992).

An intriguing case study is the cipher cast into Emperor Rudolf II's 'Alchemical Hand Bell' (Bean et al., 2023). However, the lack of contextual information makes deciphering this particular cipher challenging, leaving its purpose and solvability uncertain.

The decryption results of an early example of a Bellaso/Porta/Vigenère cipher outside of contemporary cipher manuals indicates specific alchemical secrecy techniques embedded in alchemical scribal culture and textual recipe traditions (Piorko et al., 2023). The discovery and subsequent decryption of the polyalphabetic Bellaso cipher in Sloane MS 1902 published by Bean et al. (2021) is a notable and early instance of this style of encipherment that illustrates the deep knowledge networks inherent in copying, decrypting, and circulating alchemical secrets in the form of a specific encrypted message: *Hermeticae Philosophiae Medulla*. Such scribal networks, which span international manuscript collections, can only be detected through scholarly collaboration. Thus, it is imperative that scholars continue interdisciplinary collaboration to fully grasp the scope of alchemical scribal ciphering.

3 Results: Circulating scribal secrets through encryption

This overview of sources indicates that in alchemical contexts, extensive pages of ciphertext are rare. More commonly, we find short ciphers, single lines, small text blocks, or coded words within recipes. Alchemists also employed other modes of secrecy or obfuscation, including the use of foreign languages like Arabic, Hebrew (quite prevalent), and Latin, which served as obfuscation for the non-erudite (Principe, 2018, 143), though many scholarly alchemists likely had Latin proficiency. They frequently used symbols, not limited to typical alchemical ones, and were inventive in creating their own symbol codes. Shorthand, particularly the Ashmole shorthand (Josten, 1967) found in numerous manuscripts, was widely utilized, likely due to Elias Ashmole's (1617–1692) prolific writing.

Ciphers often relate to alchemical experiments and recipes. In collections featuring multiple authors, certain figures like Paracelsus, Basil Valentine, Roger Bacon, Arnald of Villanova, and other medieval alchemical authorities frequently appear in conjunction with ciphers. This may be due to the large scope of these collections and the finite pool of respected authorities. Paracelsus, for instance, is often associated with ciphered texts because he propagated that there was power in *characteres* (Gannon, 2019, 84).

Cryptographical tradition authors such as abbot Johannes Trithemius (1462–1516), who blended cryptography with occult writings (Gamer, 2022, 1–36), and Giambattista della Porta (1535–1615), known for both alchemy and ciphering (della Porta, 1563; Kodera, 2021), are also commonly mentioned. However, other known cryptographers not engaged in alchemy are less frequently mentioned, with exceptions like John Willis (ca. 1575–1625), whose manual influenced Ashmole's shorthand (Josten, 1967). Ashmole, following the tradition of figures like Dee, showed interest in mathematics and ciphering from a scholarly perspective, suggesting an evolution towards greater cipher literacy (Ellison, 2016) in alchemy.

Regarding existing literature on alchemical ciphers, this survey contributes significantly, revealing several ciphers not previously noted. Benedek Láng, in his broader study of cryptography, mentions only a few known alchemical ciphers (Láng, 2018, 163–166). However, as our survey shows,

alchemical ciphers are more common than previously thought. Rec's assertion of their infrequency in alchemy, due to alchemist's differing intentions (Rec, 2014), is contested here. While they are not as ubiquitous as alchemical *Decknamen* and serve a different function, ciphers are nonetheless a frequent and under-researched aspect of alchemical texts, often overlooked by researchers due to the difficulty in deciphering them without a key or cipher table, and the substantial effort required for decryption.

Alchemical ciphers manifest in diverse forms and complexities, ranging from basic to highly intricate, with some accompanied by their cipher tables and keys, while others are designed to conceal content effectively. Instances where cipher tables seem intentionally corrupted to hinder decryption are notable (Piorko et al., 2023), requiring additional knowledge to rectify such manipulations. The practices of alchemical ciphering likely drew from standard ciphering methods in correspondence and diplomacy, particularly evident in ciphers found in letters, suggesting a cultural overlap with diplomatic or political practices. Alchemy's potential for politics and the economy (Nummedal, 2007) might have influenced the use of ciphers to protect valuable information.

While Benedek Láng points to scientific priority in some instances (Láng, 2018, 163–166), other motivations, like the pursuit of esoteric knowledge, are also significant (Forshaw, 2013). The Dees' work, for example on angel names and Enochian tables, stems from a scholarly pursuit of occult knowledge, incorporating elements from traditions like the Kabbalah (Harkness, 1999; Gannon, 2020). This pursuit aligns with discussions around universal languages (Strasser, 1989; Strasser, 2011), as seen in attempts to create languages like Enochian and the use of Kabbalistic number magic or multilingual ciphering, i.e. using little-known foreign languages as a means of 'encipherment'. In this vein, alchemical ciphering intersects with distinct alchemical techniques of hiding knowledge, such as *Decknamen* or knowledge dispersion. Often, multiple secrecy layers and ciphering are employed together, sometimes combined with visual elements adding to the knowledge transmitted via language.

Cipher tables should also be viewed in relation to other contemporary knowledge tables (Forshaw, 2005; Clucas, 2017), not merely as technical tools for concealing information. Lang (2023) underscores the multifaceted reasons behind alchemical ciphering, highlighting the need for further research on the topic. An immediate objective for future work is to decrypt and interpret the ciphers listed in this survey, followed by developing a comprehensive taxonomy of alchemical ciphers, as suggested by Lang (2023). Such a taxonomy would elucidate the varied motivations and methods in alchemical ciphering. Future work should also address the use of sympathetic inks in the context of alchemy, which has been noted (Macrakis and Lye, 2014; Macrakis, 2014; Wentrup, 2023), but not yet been the focus of an in-depth study.[3]

4 Survey: Sources of alchemical cryptography

What follows is an extensive, albeit surely not comprehensive, list of sources containing alchemical ciphers and related secrecy devices.[4] In order to create this working list of extant materials containing alchemical ciphers, the following collecting institutions were surveyed by the authors between 2016 and 2024:

1. Allard Pierson Library, Amsterdam, NL
2. Embassy of the Free Mind, Amsterdam, NL
3. University Library, Bamberg, Germany
4. State Archive, Bamberg, Germany
5. University Library, Barcelona, Spain
6. State Library, Berlin, Germany
7. State Library, Bremen, Germany
8. University Library, Bremen, Germany
9. University Library, Brescia, Italy

[3] Macrakis and Lye (2014) mention the use of sympathetic ink following Dorothea Juliana Wallich's (1657–1725) discovery of the element bismuth-cobalt in 1705 (Kraft, 2019): "The cobalt mineral [discovered by Wallich] also displayed remarkable visual qualities: Its color changed from rosy red to grassy green to sky blue when heat was applied. When the cobalt was prepared and turned into a solution with which to write, it was clear, but it produced a fabulous blue-green color when heated. The writing disappeared when cooled (Macrakis and Lye, 2014, 199; cf. 71–72)." Robert Boyle, for instance, also used invisible ink (Macrakis, 2014, 55–58).

[4] While substantial, this contribution cannot address in detail all ciphers known from published literature. This task remains a desideratum. For instance, Timmermann (2015) describes a number of ciphers in Cambridge *alchemica* that are not discussed here. This study is based on physical library visits, focusing on ciphers found in alchemical manuscripts or those pertinent to alchemical contexts. As a result, it uncovers many ciphers that are not widely known and might not be mentioned in bibliographical descriptions, although most are. However, the scope of this work is limited to manuscripts; it does not include printed materials nor does it involve discovering new sources for alchemical cryptography through database searches. These aspects – covering print sources and database exploration – are next steps for future work in this area.

10. Cambridge University Library, UK
11. University Library, Coburg, Germany
12. Royal Library, Copenhagen, Denmark
13. State Archive, Darmstadt, Germany
14. University Library, Darmstadt, Germany
15. State Library, Darmstadt, Germany
16. University Library, Edinburgh, UK
17. Royal College of Physicians Library, Edinburgh, UK
18. National Library (Matenadaran), Erevan, Armenia
19. State Archive, Erfurt, Germany
20. State Archive, Frankfurt am Main, Germany
21. Glasgow University Library, UK
22. Gotha Research Library, Germany
23. Gotha State Archive, Germany
24. State Library, Hamburg, Germany
25. University Library, Hamburg, Germany
26. National Library, Jerusalem, Israel
27. Topkapı Palace Library, Istanbul, Turkey
28. State Library, Kassel, Germany
29. Murhard Library, Kassel, Germany
30. Leiden University Library, NL
31. British Library, London, UK
32. Wellcome Library, London, UK
33. John Rylands Research Library, Manchester, UK
34. University Library, Marburg, Germany
35. State Archive, Meiningen, Germany
36. Russian State Library, Moscow, Russia
37. Russian State Archives of Ancient Documents, Moscow, Russia
38. Roudnice Lobkowicz Library, Nelahozeves, Czech Republic
39. Beinecke Rare Book & Manuscript Library in Yale (New Haven), US
40. National Library, Oslo, Norway
41. Bodleian Library, Oxford, UK
42. National Library, Prague, Czech Republic
43. Strahov Library, Prague, Czech Republic
44. The Accademia dei Lincei in Rome, Italy
45. National Library, Sarajevo, Bosnia and Herzegovina
46. University Library, Sarajevo, Bosnia and Herzegovina
47. State Archive, Rudolstadt, Germany
48. St Andrews University Library, UK
49. Russian State Library, St Petersburg, Russia
50. National Library, Stockholm, Sweden
51. Leopold-Sophien-Library, Überlingen, Germany
52. Marciana Library, Venice, Italy
53. Austrian National Library, Vienna, Austria
54. Herzogin Anna Amalia Library, Weimar, Germany
55. State Archive, Weimar, Germany
56. Herzog August Library, Wolfenbüttel, Germany
57. State Archive, Wolfenbüttel, Germany
58. National Library, Zagreb, Croatia
59. University Library of Zurich, University Library of Zurich, Switzerland
60. Othmer Library, Philadelphia, United States
61. Houghton Library, Cambridge, United States
62. Huntington Library, Pasadena, United States
63. Lehigh University Library, Bethlehem, United States
64. Library Company of Philadelphia, United States
65. Lilly Library, Bloomington, United States
66. Newberry Library, Chicago, United States
67. Van Pelt Library, Philadelphia, United States
68. Memorial Library, Madison, United States
69. New York Academy of Medicine, United States
70. National Library of Medicine, Bethesda, United States

Alchemical ciphers were not found in every collection surveyed.[5] In the following section, the sources found to contain ciphers are listed and described according to country and institution.[6]

4.1 Denmark

4.1.1 Copenhagen National Library

MS. 1717: Circular diagram containing a few letters. 16th Century. Paper. Alchemical miscellany (6 vols) including Thomas Aquinus, *Clavicula Raymundi Lullii cum declaratorio*, Arnald of Villanova, and *Turba Philosophorum*.

MS. 238: Ciphered line close to the spine of the book; three pages with puzzling illustrations and diagrams, amongst which one definitely seems to be a cipher alphabet related to the seven planets/metals. 15th century German *Buch der heiligen Dreyvaltigkeit* [Dreifaltigkeit] with illustrations, dated 1415–1417.

4.2 Germany

4.2.1 Darmstadt Bibliothek

Ms 2625: Page titled 'Ophir Salomonis' containing text in Hebrew that may have been used as a means of encryption.[7] 1746 by Karl Zimmermann.

Ms 3259: Three pages of pigpen ciphertext in a collection of alchemica.

Ms 3266-2 (*Opera philosophica*) Three letter-based cipher tables, two named 'clavis minor' and 'clavis major' (looking like a Trithemius table); ciphertext using symbols.

[5]However, the following survey includes some libraries that were not thoroughly surveyed but contain single manuscripts with alchemical cryptographic content known to the authors from research literature. These were included to make the list as comprehensive as possible. However, libraries not mentioned above were not extensively surveyed for additional alchemical ciphers they may hold. Furthermore, it is, of course, possible that each of the libraries surveyed contains additional sources of alchemical cryptography we may simply have missed.

[6]A brief investigation revealed that the great majority of these ciphers does not yet seem to be included in the DECODE database (Megyesi et al., 2019; Megyesi et al., 2020; Héder and Megyesi, 2022). We noted the DECODE records in the footnotes where applicable. As hardly any of our examples are from explicitly cryptographical works or contexts, they do not seem to be included in the Shulman bibliography either (Shulman, 1976). Of course, many historical ciphers remain unbroken, suggesting there may be many more ciphers whose alchemical context might only be revealed upon decryption. For instance, letters exchanged between Emperor Ferdinand III and his brother, Archduke Leopold Wilhelm reveal an unexpected level of exchange on alchemical practices amidst the atrocities of the Thirty Years' War. Notably, extensive sections of these letters detail the progress of alchemical experiments (Soukup, 2023, 52–54). Thus, DECODE record 1579 (`https://de-crypt.org/decrypt-web/RecordsView/1579`) could theoretically relate to alchemical matters as well.

[7]Pseudo Solomonic grimoires such as *Clavicula Salomonis* are discussed in Gannon (2020, 100–101). The Pseudo-Salomonic tradition attempted to implement angel magic using methods and symbols loosely related to the Kabbalah and was popular, for instance, in Rudolphine Prague. Symbols and sigils related to Hebrew, but interpreted as a celestial or angelic script are described in Agrippa of Nettesheim's *De occulta philosophia*, sometimes under the name of a 'Malachim script'. Marsilio Ficino also discusses astromagical images in *De vita* – both are sources that would have been known in alchemical circles.

4.2.2 Darmstadt Staatsarchiv

D 4[8] **Nr. 585.1:** Cipher alphabet, text passages in Greek and Hebrew. Archival Material of Prince Christian about Rosicrucians, containing *Arcana Collectanea*, *Sperma astrale*, etc., ca. 1730. Mainly late 18th century correspondence of Prince Christian, including alchemical drawings.

D 4 Nr. 588.2: Huge page filled with tiny pigpen-ciphered text, at the bottom signatures (amongst which 'Philippus Melanchton'), accompanying cipher table.

D 4 Nr. 75.9: Hebrew-looking ciphertext on one page, something that looks like a cipher-poem on another; dedications, writings, and compositions sent by various individuals to Landgrave Philipp of Hessen-Butzbach from 1611–1630, 1637 & 1641. Amongst which cabbalistic writing and a book chapter overview by a Jew named Abraham from Worms (magical text).

D 4 Nr. 76.6: Multiple pages of partially ciphered text, related alphabet/key and nomenclature; but also a page that seems to be a pigpen cipher for/using Hebrew characters. Something that looks like a Trithemius table. A number-based nomenclature including related ciphertext; more pigpen ciphertexts and plaintext pages with lines in code. All in all, a plethora of cipher keys, either for the individuals' personal use or potentially for learning purposes. In the 1641 *varia* of Landgrave Philip of Hessen-Butzbach, including a writing calendar for 1641, mottos, alchemical and kabbalistic writings (with a drawing of the hemisphere), cipher tables for names, tables for developing a secret script, notes on war events, a draft for a sundial, and an illustration of the town and fortress S. Mauro; belongs to D 4 Nr. 106/3.

4.2.3 Gotha Forschungsbibliothek

Chart. A 1014: Two lines that either use Hebrew as a steganographic device or use Hebrew characters as a code. In *Der alchemistische Nachlaß Friedrichs I. von Sachsen-Gotha-Altenburg*, 1727. 169 sheets in folio (Moeller, 1826, 382,1).

Chart. A 1017: Possible cipher table. Medieval alchemical manuscript in Latin and French (Moeller, 1826, 382,3), mostly texts by Raimund Lull, 167 sheets; described in detail in Wunderle (2002, 65ff.).

Chart. B 1156: Six pages with some coded lines and a list resembling a cipher alphabet. Circa 1455 from Johann Baptist von Seeback's collection, mostly short recipes for alchemical processes, partially ciphered (Moeller, 1826, 81); described in Wunderle (2002, 370ff.).

Chart. B 1188: Numbers 0–9 associated with pigpen-style symbols. Late 17th-century manuscript with several processes and authors named, 376 sheets, multiple hands (Moeller, 1826, 384,r).

Chart. B 1393: A few lines of code, possibly pigpen, but faded, in a copy of Kenelm Digby, *Auserlesene philosophische Geheimnisse* (Hamburg 1684), 334 pages including index, one hand, bound in parchment (Moeller, 1826, 383,6).

Chart. B 246: Partially coded and symbol-laden two-page text about an alchemical furnace. 17th and 18th-century alchemical collection, 124 numbered sheets (Moeller, 1826, 379,2).

Chart. B 255: Two cipher alphabets in the margins of a plaintext letter. 17th/18th-century Italian manuscript with a code key, 119 sheets, referencing Lull, Geber and Paracelsus (Moeller, 1826, 380,6).

Chart. B, 256–257: Partially ciphered lines which seem to be part of the explanation of a code system. 18th-century manuscript stating to be copied from an unreadable template (Moeller, 1826, 380,7–8).

Chart. B 368: The inside of the book cover and flyleaf are filled with partially faded notes, amongst which seems to be a short cipher. *Kunstbuch*, 1571, in German. *The hidden art and work of preparing the tincture*, including: Rupecissa, apparatus sketches, Mediolan, drawings, *Practica Ruperti* (of Constantinople). (Moeller, 1826, 383,3)

Chart. B 370: Nomenclature of alchemical characters (although mostly not the common symbols). Process book, early 16th century, approximately 180 pages with registers, apparently also older fragments (Moeller, 1826, 383,5)

- **104:** Diverse papers (1 box, roughly sorted) containing text in Friedrich's hand.
  - **2:** 8 pages of text in which many lines are coded. Booklet 10 x 16 cm, 111 pages. Inc. 1681 *Timor Domini*. Contains a compilation of alchemical authorities on the dry path, dated June 22, 1682.
  - **18:** Table beside text on *Aurum potabile*. Small booklet, 9 x 15 cm. Various alchemical processes, with 3 slips of paper attached.
  - **22:** *Clavicula salomonis* containing a few ciphered sections. Kabbalistic texts and horoscopes. 3 loose, but related sheets.
  - **23:** Number table, likely for divination. Small geomantic booklet, 17 x 20 cm, 11 written pages.
  - **24:** Alphabet for monoalphabetic substitution. 4 unbound writings, including the cabbalistic text: *Semi Phoras*, 4 pages, 17 x 20.5 cm. *Semi Phoras*, [6] pages, 21 x 34 cm. Inc. *Nomina imposuit Adam*, 2 written pages (numbered 26–27), 17 x 21 cm. Slip of paper, folded to 17 x 20.5 cm. Inc. 'ob die LAMBDA sache gegen Michaeliß werde ihren anfang genomen'.
  - **26:** Nomenclature linking number combinations to words, seems to be for included correspondence. In a small book, 9.5x16 cm, bound in leather, incl. key to outgoing correspondence.
  - **27:** Nomenclature in small book, 10x16 cm, bound in leather, incl. key to outgoing correspondence.

- **73:** Upside down page with potential cipher alphabet. In *H. Friederichs I Chymische Correspondenz* Vol. II, Sammelband, 20 x 34 cm, 399 sheets, containing 8 unbound sheets. Correspondence concerning curiosities and alchemical texts, primarily 1688–1690, including drafts for fireworks, note fragments by Friedrich I on alchemical processes, expenditure, astrological notes, code keys, and laboratory plans.

- **73'''**, 1 and 9: multiple cipher alphabets, probably as instruction or an exercise.

- **79:** *Nihil occultum quod non reveletur* ('There is nothing occult, which will not be revealed') followed by two lines of ciphertext. In *De origine et compositione Lapidis Philosophorum* (1598–1603), a manuscript bound in parchment musical notation, 16 x 20.5 cm,

[8] **D4:** `https://arcinsys.hessen.de/arcinsys/detailAction.action?detailid=b3916`

contains various alchemical processes and theories, including work by Wilhelm Ouerlacker (probably Georg Schwalenberg in Fritzlar). Notable entries include the first chapter on the origin of metals and their generation (dated 1603), *Theorica de auro vel lapide Philosophico* and *Practica Rogerii Baconis de Sole*.

- **80:** Two pages with a few coded lines and many alchemical symbols. In *Prozeßbuch* (1580), process book bound in cardboard, 17 x 21.5 cm, containing excerpts from the works of Johann Marcelli Hessen van Regensburg and mentions of various alchemical practitioners including C. F., Peter Senffeneder, Balthasar Rennschaff, and Sebastian Eibenauer.

- **95:** Multiple cipher alphabets. In Herzog Ernst Ludwigs zu S. Meiningen *Chymica* Vol. II (1694–1709), a folder of bound correspondence, max. 23 x 35 cm, containing 140 pages of various alchemical processes, calculations on expected profits, and curious chemical writings. It includes contracts with alchemists and inventories of laboratory equipment.

4.2.4 Hamburg, Staats- und Universitätsbibliothek

Cod. alchim. 651: A German poem entitled *Thorough Summary of All Celestial Influence*, making intriguing claims about alchemists at the court of Emperor Rudolf II (1552–1612). The supposed author, Martinus de Delle, a fictional court poet and adept, is likely a result of textual corruption. Both the prose introduction and the verse contains ciphered passages (Prinke and Zuber, 2020, esp. 418).[9]

4.2.5 Heidelberg Universitätsbibliothek

Codex Palatinus Germanicus 597: The manuscript, *Alchymey Teuczsch*, was composed over several years starting in 1426 by a group of alchemists in eastern Bavaria, possibly under the patronage of the Bishop of Passau (Rec, 2014, 4). *Alchymey Teuczsch* includes texts on medicine—some involving magical practices—and astrology. However, only the alchemical sections are encrypted (single words or entire recipes).[10]

[9]Some of the ciphered sections are included as images in Prinke and Zuber (2020). The similarity of many symbols used in the code to standard alchemical symbols and each other suggests that copyists may not have rendered them distinctly, complicating efforts to decode them. For example, well-known symbols like the Luna symbol (☽) were sometimes mixed up with letters 'd' by later scribes. Since Prinke and Zuber will not continue publishing on this, they passed down their research materials to the authors of this survey. From the HistoCrypt community, George Lasry and Richard Bean have consulted on these ciphers in the past but, up to this point, to no avail. The plaintext language is likely early modern German, which may have complicated the decryption process. We have not yet attempted employing historical language models (Megyesi et al., 2020; Sikora, 2022).

[10]The group was led by Nicholas Jankowitz, assisted by at least two collaborators, Michael von Prapach and Michael Wülfing, and a laboratory assistant named Friedrich. Jankowitz and his team employed three different cipher alphabets to encode parts of their work, concealing individual words and sometimes entire recipes. Eis (1982) has suggested that these ciphers were developed to ensure that the perfected recipes remained confidential within their laboratory, thereby preventing them from falling into the hands of rival practitioners.

4.2.6 Kassel Bibliothek

2° Ms. chem. 4: One page in which symbols are correlated to numbers, however, the mechanism seems not to be completely trivial. 2° Ms. chem. 4: *Donum dei*, German, 1520–1527.[11]

4° ms. philol. 10: Cipher alphabets explained in this copy of *Steganographia nova* by Friedrich von Öttingen-Wittgenstein, wrongly attributed to Johannes Trithemius (1462–1516).[12]

2° Ms. chem. 19: About 50 pages of code, possibly for practice. Some words in recipes are in code, according to the catalogue in order to 'make them more mysterious' (Wiedemann and Broszinski, 2011). It is suggested German or Latin words were written using Greek or Hebrew letters as replacements.[13]

4.2.7 Rudolstadt Staatsarchiv

Kanzlei Arnstadt 404: Multiple pages where plaintext is interspersed with words in code, some pages are partially ripped out or words blackened. Letters addressed in German to 'Monsieur', with salutations in French. On some pages, alchemical symbols are drawn over the codes in pencil, likely an attempt to solve the code. Letters to Prince Anton Günther, 1689–1700s.[14]

Kanzlei Arnstadt 407: Some coded words within plaintext in letters from alchemists to Count (Prince) Anton Günther II, dated 1684–1714.

4.2.8 Weimar HAAB

Fol 96: One page with a pigpen cipher, another contains a table in a different system. Nuremberg curiosities (*Nürnberger curiosa*), 1663. Extensive collection of remarkable things, curiosities, theological, historical, philosophical, legal, alchemical notes, etc., including a series of songs and poems, collected with special reference to Nuremberg, presumably by Hanns Münchner of Nuremberg.

Oct 106: Multiple pages in which little drawings seem to be correlated to letters as a sort of alphabet or nomenclature. The drawings are repeated, albeit with different coloured backgrounds.

Q 454.4: A square table containing circles, symbols and letters, titled *Tabella Rabellina Solomonis*. Magical manuscript, *Three Books of Magical Wisdom*, the date is ciphered in Hebrew.[15]

Q 456: On a page titled 'Archeus', an acrostic poem spells out 'ALCHIMIA' through the first letter of each line (in red ink) in an alchemical composite manuscript (*Sammelhandschrift*).

[11]**2° Ms. chem. 4:** `https://www.handschriftencensus.de/23834`

[12]**4° ms. philol. 10:** `https://orka.bibliothek.uni-kassel.de/viewer/fullscreen/1486550062531/29/LOG_0027/`

[13]**2° Ms. chem. 19:** `https://orka.bibliothek.uni-kassel.de/viewer/fullscreen/1486550062531/27/`

[14]**Kanzlei Arnstadt 404:** `http://www.archive-in-thueringen.de/de/findbuch/view/bestand/25960/systematik/132824`

[15]On the Pseudo-Solomonic, Pseudo-Hebrew tradition of Kabbalah-inspired magic, see Gannon (2020).

Q 458.3,.4,.5,.6: Related text to Weimar Staatsarchiv, A XIV 20, alternating between Arabic and Latin, mentioning 'AUGUSTUS ERNESTUS' and 'Sachsen'; may be the translation of a dedication. One line in Hebrew or code related to Hebrew letters. Dated 1740, contains *Clavicula Salomonis*.

4.2.9 Weimar Staatsarchiv

A XIV 20: Latin-Arabic text, might just be a translation. Since the Latin text below each line is about the prince, it may just be a dedication of a book given as a present from abroad. Letters by Heinrich Gottlieb Reime, Paul Valtin Reiss, Carl Philipp Raidel and related correspondences, 1738–1746. Großherzogliches Hausarchiv (Ernst August).[16]

A XIV 24: A few symbols on a title page, likely Cabbalistic. Correspondence 1738–1747.

A XIV 4: Ciphertext in geometric style (1 page), drawings on the opposite page make it seem like a pigpen variant. *Collectanea alchemica*, 1493–1747, 313 pages.

A XIV 5b: Strange code made up of five characters, repeated multiple times on one page. Alchemical signs to prevent theft, ca. 1700–1750, 11 pages.

4.2.10 Wolfenbüttel HAB

80-4-aug-8f: One line of ciphered text. Alchemical composite manuscript, 15th century, paper, 105 pages, 14.5 × 10.5 cm.[17]

4.2.11 Wolfenbüttel Staatsarchiv

2 Alt 2211: Simple substitution alphabet dubbed 'philosophisches Alphabet'. Inquiry of Hermann Sprenger's clandestine visit to Wolfenbüttel and related alchemical notes, 1573–1574.[18] He had been invited by Philipp Sömmering, an invidivual known from debates revolving around the notion of 'alchemical fraud' (Nummedal, 2007).

4.3 Israel

4.3.1 National Library Jerusalem

Ms. Ed. 7: Page with German text related to Basilius Valentinus contains a pigpen-style enciphered word. Since the word before it is 'Hungarian', the enciphered term probably denotes a place and/or a substance to be found there. Neatly written manuscript copy, on 330 pages, with several pen-and-ink drawings in the text. Folio, contemporary marbled wrappers. Copied by Franciscus Cling in Berlin, 1747. Alchemical anthology from 1554, *Testament oder Morgen Roethe und Heiligen blutigen Steinem in demn alla glanzenden Crystallingen Sulfurigen durchscheibe und Salis Petra Philosophorum von Niter*, 329 pages.

Ms. Ed. 7.2: Chemical notes from Germany, dated 17th century, include three pages of partially enciphered text in simple symbols.[19]

[16] **A XIV 20:** `http://www.archive-in-thueringen.de/de/findbuch/view/searchall/Gro\%C3\%9Fherzogliches+Hausarchiv+/bestand/27222/systematik/130157`

[17] **80-4-aug-8f:** `https://diglib.hab.de/?db=mss&list=ms&id=80-4-aug-8f`

[18] **2 Alt 2211:** `https://www.arcinsys.niedersachsen.de/arcinsys/detailAction.action?detailid=v3961504`

[19] **Ms. Ed. 7.2:** `https://www.nli.org.il/en/manuscripts/NNL_ALEPH990038366820205171/NLI`

4.4 Italy

4.4.1 Brescia Library

I.IV.31: Cipher alphabet in multiple lines (starting with the plaintext letters in the first row), some symbols for specific terms. In *Secreta secretorum philosophorum*, a manual of alchemy from the late 16th or early 17th century, featuring various recipes for metal transmutation and herb distillation.

ms.I._V.13: Contains a symbol code, potentially a form of shorthand, interspersed with alchemical symbols (1 page) from the second half of the 17th century. In what is described as one of the oldest and most complete examples of the *ABRAMELIN*, a famous esoteric grimoire attributed to a legendary Egyptian mage Abraha-Melin, or to Abraham of Worms or Würzburg, a German Talmudic Jew from the 12th or 13th century; likely composed during the 14th century in the Balkan region, only a few known copies.

4.4.2 Marciana Library

Italian II. 152 [5046.]: 1 page containing hard to read text (not ciphered), combined with symbols (potential cipher). Line drawings accompanied by 4-letter combinations each, labelled 'litera egittiaca' and 'Numeri Egittij Hyeroglifici', however, they are definitely not Hieroglyphic numerals (2 pages). Paper, quarto, 16th century. *Opusculum alchimiae auri et argenti*.

4.5 Netherlands

4.5.1 Allard Pierson

PH344: 5 pages of partially enciphered text, all with the same alphabet. One page contains a table called 'Die erste Tabelle des Königs Xophor' ('The first table of King Xophor') Basilius Valentinus' *Das große Geheimniß der Egyptischen Konige*, dated to the 1700s.

4.5.2 Leiden UL

Cod. Voss. Chym. Q. 51: This *Liber de magna alchymia* lists several alphabets, existing and fictional, ranging from Greek and Egyptian to Chaldean and even angel symbols. Foreign scripts are heavily adapted through scribal modifications or transmission errors (Gannon, 2019).[20]

VCQ 17: Cipher alphabet for monoalphabetic substitution with symbols often found in the context of alchemical ciphers. A 16th-century manuscript from 1588–95, 337 folios on paper, dimensions 212x164mm, featuring various works on alchemy including poems, operations, and dialogues in both Latin and German, including *Rosarium philosophorum sive Donum Dei*, *Splendor solis*, Pseudo Roger Bacon, Pseudo Thomas Aquinas, Bernard Trevisan, Johannes Aurelius Augurellus, Pseudo Alexander von Suchten, Pseudo Raimon Lull, Paracelsus and Pseudo-Paracelsus.

4.6 Norway

4.6.1 National Library Oslo

Ms.8° 32321 1702: A manuscript (*Libellus singularis*) that includes a substitution cipher, in which letters seem to be substituted for numbers, on the back side of the title page, 1702.[21]

[20] The Viennese ÖNB Cod. 11133 also contains comparable alchemical alphabets.

[21] **Ms.8° 32321 1702** `https://beta.nb.no/dhlab/privatarkiv_navn/`

4.7 Russia

4.7.1 State Library Moscow

183.1082: Pigpen cipher table within masonic manuscript, written in French from the mid-18th century.[22]

183.1426: French 18th-century manuscript with a cipher table and a partially enciphered passage about Roger Bacon (where ingredients would follow: 'il faut prendre…'), 281 folios. *Recueil de divers traits de la Philosophie Hermetique, composes par divers maitres dans cette science*. Contains Arnald of Villanova, Paracelsus (on *Electrum*, see Gannon (2023)) and P.-J. Fabre.

183.943: Page mentioning 'Theophrastus' with symbol-encoded text in a manuscript titled *Tinctura universalis* in German Gothic script from the 18th century, 78 fol.

183.998: German text (*Tabula Smaragdina*, 2 pages total) with a portrait of Hermes Trismegistus holding a ciphered table. In *Reichversammlung und Stimmen der berühmtesten Philosophen von der Sündflut an bis auf unsere Zeit über die höchste Geheimnis der Natur, sonderlich ihres großen Steins.* 1663.

4.8 Sweden

4.8.1 National Library Stockholm

F.a. 14: 5 pages of plaintext containing runes, some aligned in a way that makes them look like potential substitution alphabets. Different alphabets arranged in table formats (5 pages), the page heading mentions Johannes Trithemius. In a manuscript related to the idea of Goths as keepers of ancient wisdom with runes having 'double' meanings and sacred significance, similar to the Hebrew Cabala and Egyptian hieroglyphs.[23]

F.a. 2: Titled *Azotica Astronomia*, the text contains a few enciphered words. Beside it there is a cipher table, potentially for polyalphabetic substitution but also containing many symbols familiar from other ciphers in alchemical contexts. Reflects Bureus' interest in alchemy from early 1604, noting key ideas from medieval classics and works of Paracelsus, Gerhard Dorn, and Andreas Libavius.[24]

Huseby 78: An Old English manuscript from 1550 with various recipes includes an alphabet for simple substitution cipher and a few enciphered words.

Råd. 9: A text titled 'mysteries of the alphabet', appears to contain cipher instructions within Swedish text. In Johannes Bureus, *Adulruna Rediviva seu Sapientia Sveorum Veterum*.[25]

X113: List and table of syllables (2 pages), could be just a writing exercise. From a late 15th-century German medical book.[26]

[22] **183.1082:** https://viewer.rsl.ru/ru/rsl01004721310?page=1&rotate=0&theme=white

[23] **F.a. 14:** https://lucris.lub.lu.se/ws/portalfiles/portal/2285913/3809108.pdf

[24] **F.a. 2:** https://lucris.lub.lu.se/ws/portalfiles/portal/2285913/3809108.pdf

[25] **Råd. 9:** https://lucris.lub.lu.se/ws/portalfiles/portal/2285913/3809108.pdf

[26] **X113:** https://kortkataloger.kb.se/hsnominal/20647/

4.9 UK

4.9.1 Edinburgh Royal College of Physicians

DEP ERG 5 1-5-1-23: Potential cipher table below text in Italian. Alchemical manuscripts of George Erskine, titled *Arbatel – The magik of the auncient Philosophers*, dated 13 Feb 1602, with aphorisms and possibly in Erskine's own writing.[27]

4.9.2 Edinburgh University Library

MS. Dc.1.30: *Medulla* cipher discussed in Piorko et al. (2023) containing the Bellaso/Vigenère/Della Porta cipher table, enciphered text, key and plaintext; plus one number wheel to calculate the days of the alchemical work.

4.9.3 Glasgow University Library, Ferguson Collection

Ferguson Ms. 114: Brief note in cipher or shorthand on one page. 18th-century alchemical manuscript with various tracts including *The Praxis of Meriam* and *Sir George Ripleys Epistle to King Edward unfoulded*.[28]

Ferguson Ms. 130: Ciphertext based on alchemical symbols (two pages with each one paragraph, one line on another), cipher table with one symbol per letter. 18th-century French manuscript, *Oeuvre du philosophe Solidanius*, featuring colored hermetic figures and sections in code, 55 folios, 205x156mm.

Ferguson Ms. 19: Nomenclature alphabet encoding alchemical processes. 17th-century Spanish manuscript with Italian section headings, including *Cedula ritrovata* and *Elucidarius Christophori Parisiensis*, 52 folios, 215x142mm.

Ferguson Ms. 191: One coded block in the margin of one page. 17th-century English manuscript with Latin verses, including Thomas Norton's *The Ordinall of Alchimy*, 207x147mm, 52 folios.

Ferguson Ms. 262: French text contains a cipher alphabet under the heading 'Alphabet cryptographique'. The two pages seem to refer to another text (mentioning Nicolas Flamel) in which this cipher seems to have been used. 17th-century French manuscript, a copy of (Pseudo-)Nicolas Flamel's writings, featuring horoscopes and planetary arrangements. ii folios + 36 pages + 4 folios on vellum. 159x108mm.

Ferguson Ms. 323: One page contains what seems to be a nomenclature. 16th-century English manuscript with alchemical receipts, diagrams, *Raymond Lullye's alphabeta*. 12 folios. 294x198mm.

Ferguson Ms. 4: Familiar circular drawing typical of *Buch der Heiligen Dreifaltigkeit*.

Ferguson Ms. 77: Caballistic table, relating letters to numbers. The heading contains the symbol for Mercury. 18th-century French manuscript with various alchemical treatises, including a depiction of Copernican cosmology and writings by Bernard Trevisan, Isaac Hollandus, Raymund Lull, Nicolas Flamel and *Traité de l'Or Potable*. 132 folios. 210x168mm.

[27] **DEP ERG 5 1-5-1-23:** http://archives.rcpe.ac.uk/CalmView/Record.aspx?src=CalmView.Catalog\&id=DEP\%2fERG\%2f1\%2f5\%2f23

[28] **Ferguson Ms. 114:** https://www.gla.ac.uk/collections/\#/details?irn=265726\&catType=C\&referrer=/results\&q=GB+114+MS+Ferguson

Ferguson Ms. 94: Cipher table (1–2 symbols per letter), about 20 pages with partially or entirely encrypted text amongst French plaintext. 17th-century French alchemical and astrological notebook of A. Mereau, including sections in code, pen portraits, and drawings of alchemical apparatus. 190 folios (of which 54 are blank). 201x145mm. The manuscript features sections in code, with the key affixed to the inside front and rear covers. An astrology-focused section presents several horoscope charts. The cover bears coded words, followed by the inscription: "Elvoh a servez a la gloire de Dieu et au salut de [ ] ame A.M."

MS Hunter 110 (T.5.12): Number and letter tables on one page. Alchemy compendium, late 14th century, in English hand, illuminated initials, pen and ink drawings, diagrams, and numerous marginalia. Includes works such as *Synonyma Alchemiae* and *Mappae Clavicvla* (unknown authors) and Albertus Magnus.[29]

4.9.4 London British Library

Sloane MS 1902: Astrological medical manuscript produced by John and Arthur Dee (Lang and Piorko, 2021; Bean et al., 2022; Piorko et al., 2023), containing the ciphertext *Hermeticae Philosophiae Medulla*, adjacent *tabula recta*, and cipher key (ff. 12–14).

Harley Ms. 2407: 110011101001010101 code on the left of one page. 15th and 16th century manuscript including various alchemical texts, poems, and treatises in Latin and English including Arnald of Villanova, and numerous alchemical drawings and preparations.

Sloane MS 3189: *The Book of Enoch* contains the fortune telling tables of John Dee, as scribed by Edward Kelley, later owned by Elias Ashmole. The angelic conversations (Shumaker, 1983; Harkness, 1999; Reeds, 2006) were communicated via encrypted angelic languages unique to each angel. Dee and Kelley used alpha-numeric *tabula recta* containing Enochian language symbols, which are also recorded in the codex.

Sloane 3604: 4 pages containing cipher wheels and multiple cipher tables. 16th Century manuscript in the hand of Robert Frelove, containing various alchemical treatises such as by Raymund Lull, 293 folios.

Sloane Ms 1118: One page with different cipher alphabets, including rectangles. End of the 15th Century manuscript with 33 items; paper, small quarto, 154 folios.

Cotton MS Vespasian A II (ff. 2–10, 27–40 1): Numeric/alphabetic cipher tables corresponding to the months of the year; geometric computational charts, and letter ciphers are all present in this Arabic astrological manuscript owned and annotated by John Dee.

Sloane MS 3687: Contains the alchemical processes of George Marrowe and others mentioned in Sloane MS 1902 related to the *Medulla* cipher (Piorko et al., 2023). After Villanova is referenced, the rest of the MS contains laboratory notes that include numeric ciphers referencing encrypted days of the week.

[29] `https://www.gla.ac.uk/collections/\#/details?irn=296480&catType=C\&referrer=/results\&q=MS+Hunter+110+`

4.9.5 London Wellcome

Ms. 164: Circular drawing in which words are linked in mysterious ways in this copy of *Buch der Heiligen Dreifaltigkeit*.

Ms. 259: Text in Latin and Ancient Greek with religious contents, a two-line code of alchemical ciphers, marginalia with what could be Ashmolean shorthand; mapping of letters to numbers. 1649 manuscript by Petrus Almerigus Encherchz, written in a cipher of semi-Greek characters, with content focused on the secret chimica and the importance of discretion in alchemical pursuits. 1649 manuscript containing semi-Greek cipher recommending secrecy in the Work.[30] Further names follow in an unresolved cipher (pointing to members of an order).

Ms. 309: Several pages encoded by a monoalphabetic cipher consisting of 51 symbols (most letters are replaced by more than one symbol). The ciphertext spans the following pages: 7–10, pages 11–12 are missing, 13–20, 21, 26, 27, 28, 29. The rest of the text is in German. The cipher key is included. Some of the ciphertext pages contain drawings of alchemical vials and furnaces. In Johann Gerlach's *Occulta scripta chymica* written in German at Gottersdorff in 1572 and Naunburg in 1575–1576. The 18th-century manuscript includes secret alchemical scripts, prayers, lists of ingredients, and alchemical and medicinal receipts. It is illustrated with water-color symbolic alchemical drawings, some heightened with gold, and is written partly in cipher. The relevant section is pp. 7–178 *Secreta alchemica* in cipher and German. There follows an ingredients list (called *Elixir vitae*) and another partly coded section pp. 179–209. A manuscript key for the 51-symbol code by Julius Kohn is included.[31]

Ms. 424: French text (seemingly an alchemical recipe) broken up by musical notes, which seem to be used as symbols for encoding words (2 pages). *LIBVRE accomply de secrets chimiques pour distillations,* [...] , a mid-17th century manuscript, 20 x 15 cm, containing astrological tables. It includes a section in Italian with some words in alphabetical cipher and features astrological diagrams and tables.[32]

Ms. 447: Tables (structured as grids or lines), a circle mapping numbers to concepts (4 pages). A 16th-century manuscript (circa 1575), including small pen-drawn diagrams, figures, and drawings of alchemical apparatus,

[30] `https://wellcomecollection.org/works/m4ka7d4g` Fol I (In a cipher made up of semi-Greek characters): Jesvs Et Maria Et / Ioseph. / (red) Uerba propria philosophorum auree cur/cis super secreta chimica Dei gratia / a fideli et dilectissimo amico mihi huani [?] / ter elargita pro cuius anima semper Deum / rogare teneor quoniam mihi testauit. Deum ideo Precor ut mihi concedat gratiam eadem / uerba cognoscere pro eius laude et gloria / pro salute anime mee, pro auxilio paupe / rum.et pro commoditate mee familie pau / pere [sic]. Amen. Translation: Jesus, Mary and Joseph. The peculiar words of the Wise of the Golden Cross about the secret alchemy of God's grace were donated to me in a humanistic manner by a faithful and highly esteemed friend, for whose soul may God always care, as he has testified to me, which is why I adhere. Therefore, I pray to God to bestow the grace to understand these words for His praise and glory and for the salvation of my soul, to aid the poor, and for the benefit of my poor family. Amen.

[31] `https://wellcomecollection.org/works/qb3carav`

[32] `https://wellcomecollection.org/works/ja2s52se`

with the text within red rules and written in 'caractres de civilie'.[33]

Ms. 3563: *Miscellanea Alchemica XXI*, 1746. This composite manuscript is written in a personalized alchemical shorthand containing alchemical and astrological symbols to be used for practical experimentation. The text of *Smaragdina Hermetis Tabula* is partially coded (ends p. 58).[34]

4.9.6 Manchester Rylands

German MS 1: A 15th-century manuscript titled *Alchemica* illustrated with colored drawings and texts of alchemical subjects, written in Bavarian dialect. Contains a cipher alphabet on a page with faded text.[35]

German MS 3: An 18th-century collection of alchemical writings titled *Sammlung Alchymistischer Schriften* originating from Germany. Features partially enciphered text with a key (*clavis*) and encoded parts highlighted in red.[36]

Latin MS 65: A 15th-century Italian alchemical miscellany with a cipher table (possibly polyalphabetic), a page of nomenclature, and a *tabula Mercurii*.[37]

4.9.7 Oxford Bodleian

Ms. Ashmole 1408: Ciphertext and two lines of substitution alphabets. An early 17th-century manuscript from Richard Napier's collection, containing excerpts from classical authorities like Geber, Roger Bacon, Paracelsus, notes on Lull and more; with a ciphertext and substitution alphabets. Includes various alchemical experiments, poems, and treatises in English and Latin. Also containing an alleged chymical experiment by Johannes Trithemius in his own hand.

Ms. Ashmole 1420: Lines and marginalia in code, possibly Ashmole shorthand (Josten, 1967). Lines and marginalia in code, possibly Ashmole shorthand. The 17th-century manuscript includes works by Avicenna, Maria Prophetissa, Edward Kelly, George Ripley's *Medulla*, *Rosarium philosophorum* and more.

Ms. Ashmole 1459: Enciphered paragraph and marginalia, possibly Ashmole shorthand (Josten, 1967). The 16th and 17th-century manuscript includes Ripley, Lull, and English alchemy; 324 folios.

Ms. Ashmole 1440: Contains two pages titled 'riddles' with a paragraph in code, likely Ashmole shorthand (Josten, 1967). The manuscript includes notes and tracts on alchemy in Latin and English, including Quercetanus and Dee's *Monas Hierogylphica*.

MS. Ashmole 1441: Brief notes in symbols, probably Ashmolean shorthand. The manuscript *Liber collectaneorum de arte alchemica* consists of a collection of alchemical papers and fragments by Ashmole, the Napiers, and others.

Ms. Ashmole 1445: Enciphered notes in the margins of multiple pages; words on one page and complete sentences are meticulously cut out of the paper. The manuscript includes Ripley's *Compound of Alchemy*, *Coelum Philosophorum*, Arnald of Villanova, Raymond Lull and various alchemical works and poems.

Ms. Ashmole 1459: Enciphered paragraph and marginalia, may be Ashmole shorthand (Josten, 1967).

Ms. Ashmole 1479: Contains a cipher table and wheel, and one page of hard-to-read handwriting that may be partially enciphered. The 16th-century manuscript scribed by Rychard Walton includes works by George Ripley (*Twelve Gates*, *Marrow*), Raymond Lull, and others on alchemical topics.

Ms. Ashmole 1487: Features enciphered words and marginalia in a passage on Paracelsus. The 16th-century manuscript includes various alchemical texts, such as works by Paracelsus, Hermes Trismegistus, Mary the Prophetess, and others, encompassing theories, practices, and poems in English.

Ms. Ashmole 1490: Italian text headed 'Zifra', paragraph looking like Ashmole shorthand (Josten, 1967), cipher table (listing mostly different synonymous alchemical symbols) and seemingly a Trithemius table. The 16th and 17th-century manuscript includes a diverse collection of medical, chemical, and astrological pieces, alchemical texts, and dialogues. The manuscript is a rich compilation of various recipes, tracts, and treatises in Latin, Italian, and English, including Lull, Paracelsus, Kelley; 358 folios.

Ms. Ashmole 1492: Enciphered page in Dee manuscript. 16th and 17th century manuscript, 109 pages + 18 + 14 folios + 211 pages including notes by John Dee and Richard Napier, with text in English, Latin and Dutch. Alchemical works such as Lull and Ripley, chymical recipes partly in German.

MS. Marshall 15: Table aligning letters with numbers, under the heading 'To knowe whether a person doe tell the truth or not', thus probably a divination context rather than ciphering. This late 16th-century manuscript from England comprises medical, chemical, and astrological pieces written in various hands.[38]

4.10 US

4.10.1 University of Indiana, Lilly Library

LMC 2450, bound 11: Latin translation of eighth-century Jewish astronomer, Messahala, containing astrological tables, drawings of di, and multiple alpha-numeric tables and circular magical letter charts.

4.10.2 New Haven, Beinecke Rare Books Library

Mellon MS 108: 18th-century alchemical miscellany featuring four lines of symbol code. Compiled by Jan Pieter Rathlaan, the manuscript includes poems, teachings, and illustrations on alchemical subjects in German, Dutch, and Latin, including authors like von Hellwig. 140 folios, 312x197mm.

[33] https://wellcomecollection.org/works/c827npxm

[34] **Ms. 3563:** https://wellcomecollection.org/works/upqex4fa/items

[35] https://www.digitalcollections.manchester.ac.uk/view/MS-GERMAN-00001/1

[36] https://www.digitalcollections.manchester.ac.uk/view/MS-GERMAN-00003/1

[37] https://www.digitalcollections.manchester.ac.uk/view/MS-LATIN-00065/1

[38] **MS. Marshall 15:** https://archives.bodleian.ox.ac.uk/repositories/2/archival_objects/175318

Mellon MS 309: Alchemical text from 17th-18th century Southern Germany, includes four lines of substitution ciphertext made up of symbols. The manuscript features a wide range of texts, recipes, and alchemical illustrations, including Andreae, Arnald of Villanova, Morienus, Rases, Lull, Paracelsus, and Geber.[39]

Mellon MS 34: 16th-century manuscript (ca. 1550), includes a Hebrew *Tabula Smaragdina* side-by-side with probably the same text in different symbols (possibly an unfamiliar language). Compiled by Johannes Baptista, it encompasses a collection of alchemical works by authors such as Arnald of Villanova, Geber and Rupecissa in Latin, Italian, and Spanish. 155 folios, 210x140mm.

Mellon MS 74: 18th-century manuscript with Old French text in which metals seem to be substituted by three-letter codes. The manuscript, titled *Livre de la très Sainte Trinité*, is a French version of *Buch von der Heiligen Dreifaltigkeit*, incorporating 19th-century illustrations. 212 folios, 396x261mm.

Mellon MS 86: 18th-century manuscript with additions from c. 1740, featuring two sets of two lines of symbol code within an Old French text. Authored by Salomon Trissmosin, it includes *La toison d'or* (Golden Fleece) and *La splendeur du soleil* (*Splendor solis*). 219 folios, 203x153mm.

Mellon MS 27: Alchemical compilation by Martin Roesel von Rosenthal (˜1586), containing texts and recipes in Latin and German, partially ciphered (Rec, 2014).[40] This manuscript on paper comprises three parts, containing numerous practical procedures primarily alchemical, but also medical, along with standard medieval alchemical texts by Khalid ibn Yazid, Theodoric, and Albertus Magnus. Occasionally, passages in cipher, added by Martin Roesel of Rosenthal around 1586 – well after the main content was written – are present. The cipher appears to be a simple number-substitution type.

References

Richard Bean, Megan Piorko, and Sarah Lang. 2021. Deciphering the philosophers' stone: how we cracked a 400-year-old alchemical cipher. *The Conversation Media Group*.

Richard Bean, Sarah Lang, and Megan Piorko. 2022. Solving an alchemical cipher in a shared notebook of John and Arthur Dee. In *Proceedings of the 5th International Conference on Historical Cryptology HistoCrypt 2022*, number 188, pages 12–21. Linköping University Electronic Press.

Richard Bean, Corinna Gannon, and Sarah Lang. 2023. The cipher of Emperor Rudolf II's "alchemical hand bell". In *Proceedings of the 6th International Conference on Historical Cryptology (HistoCrypt 2023)*. Linköping University Electronic Press.

[39]**Mellon MS 309:** `https://pre1600ms.beinecke.library.yale.edu/docs/pre1600.ms309.htm`

[40]`https://de-crypt.org/decrypt-web/RecordsView/2874`

Donna Bilak. 2020. Chasing Atalanta. Maier, steganography, and the secrets of nature. In *Furnace and Fugue. A Digital Edition of Michael Maier's Atalanta fugiens (1618) with Scholarly Commentary*.

Stephen Clucas. 2017. The royal typographer and the alchemist: John Dee, Willem Silvius, and the diagrammatic alchemy of the *Monas Hieroglyphica*. *Ambix*, 64/2:140–156.

Giambattista della Porta. 1563. *De Furtivis Literarum Notis vulgo. De ziferis Libri IIII.* Scotus, Naples.

Gerhard Eis. 1982. Alchymey teuczsch. In *Medizinische Fachprosa des späten Mittelalters und der frühen Neuzeit*, pages 307–315. Rodopi, Amsterdam.

Katherine Ellison. 2016. *A Cultural History of Early Modern English Cryptography Manuals*. Routledge.

Peter J. Forshaw. 2005. The early alchemical reception of John Dee's *Monas Hieroglyphica*. *Ambix*, 52/3:247–269.

Peter J. Forshaw. 2013. Cabala Chymica or Chemia Cabalistica – early modern alchemists and Cabala. *Ambix*, 60/4:361–389.

Peter J. Forshaw. 2020. Michael Maier and mythoalchemy. *Furnace and Fugue. A Digital Edition of Michael Maier's* Atalanta fugiens *(1618) with Scholarly Commentary*.

Jonathan Gaede. 2017. Zur Verwendung astrologischer und alchemistischer Symbole in frühneuhochdeutschen Fachtexten. In Wolf Peter Klein, Matthias Schulz, Sven Staffelt, and Peter Stahl, editors, *Würzburger elektronische sprachwissenschaftliche Arbeiten (WespA) 19*, Würzburg.

Maximilian Gamer, editor. 2022. *Die Polygraphia des Johannes Trithemius nach der handschriftlichen Fassung (Band 1)*, volume 56/1 of *Mittellateinische Studien und Texte*. Brill, Leiden.

Corinna Gannon. 2019. The alchemical hand bell of Rudolf II: A touchstone of art and alchemy. In Štěpán Vácha and Sylva Dobalová, editors, *Studia Rudolphina 19. Bulletin of the Research Center for Visual Arts and Culture in the Age of Rudolf II*, pages 81–97, Prag. Artefactum.

Corinna Gannon. 2020. The amulet of Rudolf II – Kabbalistic talisman and pansophic collectible. In Štěpán Vácha, editor, *Studia Rudolphina 20. Bulletin of the Research Center for Visual Arts and Culture in the Age of Rudolf II*, pages 83–101, Prag. Artefactum.

Corinna Gannon. 2023. Electrum in the Kunstkammer of Rudolf II. objects made from seven metals. In Sarah Lang, editor, *Alchemical Laboratories: Texts,*

Practices, Material Relics, pages 114–131. Universität Graz, Graz.

Deborah E. Harkness. 1999. *John Dee's Conversations with Angels: Cabala, Alchemy, and the End of Nature*. CUP, Cambridge.

M. Héder and B. Megyesi. 2022. The decode database of historical ciphers and keys: Version 2. In C. Dahlke and B. Megyesi, editors, *Proceedings of the 5th International Conference on Historical Cryptology HistoCrypt 2022*, pages 111–114, Linköping, Sweden. LiU E-Press.

Michael Hunter. 2016. Robert Boyle and Secrecy. In Elaine Leong and Alisha Rankin, editors, *Secrets and Knowledge in Medicine and Science, 1500–1800*, pages 87–104, NY. Routledge.

Kurt Josten, editor, 1967. *Biographical Introduction, III. ASHMOLE'S CIPHER*, volume 1. Clarendon Press, Oxford.

David Kahn. 1996. *The Codebreakers: The comprehensive history of secret communication from ancient times to the internet*. Simon and Schuster.

Sergius Kodera. 2021. Giambattista della Porta. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy (Fall 2021 Edition)*.

Alexander Kraft. 2019. Dorothea Juliana wallich (1657–1725) and her contributions to the chymical knowledge about the element Cobalt. In *Women in Their Element*, pages 57–69.

Sarah Lang and Megan Piorko. 2021. An alchemical cipher in a shared notebook of John and Arthur Dee (Sloane MS 1902) [work in progress]. In *Proceedings of the 4th International Conference on Historical Cryptology HistoCrypt 2021*, number 183, pages 90–93. Linköping University Electronic Press.

Sarah Lang. 2023. Situating ciphers among alchemical techniques of secrecy. In Carola Dahlke and Matthias Göggerle, editors, *Proceedings of the 6th International Conference on Historical Cryptology HistoCrypt 2023*, Linköping Electronic Conference Proceedings 195, pages 93–104.

Benedek Láng. 2018. *Real Life Cryptology. Ciphers and Secrets in Early Modern Hungary*. Amsterdam University Press, Amsterdam.

Kristie Macrakis and Jason Lye. 2014. The hidden past of invisible ink. *American Scientist*, 102(3):198–205.

Kristie Macrakis. 2014. *Prisoners, Lovers, & Spies: The Story of Invisible Ink from Herodotus to Al-Qaeda*. Yale University Press.

B. Megyesi, N. Blomqvist, and E. Pettersson. 2019. The DECODE database: Collection of historical ciphers and keys. In *Proceedings of the 2nd International Conference on Historical Cryptology HistoCrypt 2019*, volume 37 of *NEALT Proceedings Series*, page [specific pages]. Linköping Electronic Press.

Beáta Megyesi, Bernhard Esslinger, Alicia Fornés, Nils Kopal, Benedek Láng, George Lasry, Karl de Leeuw, Eva Pettersson, Arno Wacker, and Michelle Waldispühl. 2020. Decryption of historical manuscripts: the DECRYPT project. *Cryptologia*, 44(6):545–559.

Johann Heinrich Moeller. 1826. *Catalogus librorum tam manuscriptorum quam impressorum*, volume 2 Bände. Glaeser, Gotha.

William Newman. 1991. *The* Summa Perfectionis *of Pseudo-Geber. A Critical Edition, Translation, and Study*. Collection de Travaux de l'Académie Internationale d'Histoire des Sciences 35. Brill, Leiden.

William R. Newman. 1996. "*Decknamen* or pseudochemical language"? Eirenaeus Philalethes and Carl Jung. In *Revue d'histoire des sciences*, volume 49, pages 159–188.

Tara E. Nummedal. 2007. *Alchemy and Authority in the Holy Roman Empire*. University of Chicago Press, Chicago.

Megan Piorko, Sarah Lang, and Richard Bean. 2023. Deciphering the *Hermeticae Philosophae Medulla*: Textual cultures of alchemical secrecy. *Ambix*, 70:2.

Megan Piorko. 2019. Seventeenth-century chymical collections: A study of unique copies of fasciculus chemicus. *The Papers of the Bibliographical Society of America*, 113:409–445, December.

Lawrence M. Principe. 1992. Robert Boyle's alchemical secrecy: Codes, ciphers and concealments. *Ambix*, 39/2:63–75.

Lawrence Principe. 2018. *The Aspiring Adept: Robert Boyle and His Alchemical Quest*. Princeton University Press.

Rafał T. Prinke and Mike A. Zuber. 2020. 'Learn to restrain your mouth': Alchemical rumours and their historiographical afterlives. *Early Science and Medicine*, 25:413–52.

Ivo Purš and Jaroslava Hausenblasová. 2005. Kontakty Michaela Maiera s Rudolfem ii. v praze roku 1609. In Lumbomir Konecny and Beket Bukovinska, editors, *Studia Rudolphina 5.Bulletin of the Research Center for Visual Arts and Culture in the Age of Rudolf II*, pages 51–65, Prag. Artefactum.

Agnieszka Rec. 2014. Ciphers and secrecy among the alchemists: A preliminary report. In *Societas Magica Newsletter*, volume 31 (Fall), pages 1–6.

Jim Reeds. 2006. John dee and the magic tables in the book of soyga. In Stephen Clucas, editor, *John Dee: Interdisciplinary Studies in English Renaissance Thought*, International Archives of the History of Ideas/Archives internationales d'histoire des idées 193, pages 177–204, Dordrecht. Springer.

David Shulman. 1976. *An Annotated Bibliography of Cryptography*. Garland Publishing, Inc., New York.

Wayne Shumaker. 1983. *Renaissance Curiosa: John Dee's Conversations with Angels, Girolamo Cardano's Horoscope of Christ, Johannes Trithemius and Cryptography*, volume 8 of *Medieval and Renaissance Texts and Studies*. University of Michigan Press, Ann Arbor.

Justyna Sikora. 2022. The influence of language models on decryption of German historical ciphers. MA thesis, Uppsala University.

Rudolf Werner Soukup. 2023. Alchymistische Kunststücke am kaiserlichen Hof: Alchemie unter den Habsburgerkaisern Rudolf II., Ferdinand III. und Leopold I. In Sarah Lang, editor, *Alchemical Laboratories: Texts, Practices, Material Relics*, pages 43–78. Graz.

Gerhard F. Strasser. 1989. Lingua realis, lingua universalis und lingua cryptologica: Analogiebildungen bei den Universalsprachen des 16. und 17. Jahrhunderts. *Berichte zur Wissenschaftsgeschichte*, 12(4):203–217.

Gerhard F. Strasser. 2011. Von der Lingua Adamica zur Lingua universalis: Theorien über Ursprachen und Universalsprachen in der Frühen Neuzeit. In Herbert Jaumann, editor, *Diskurse der Gelehrtenkultur in der Frühen Neuzeit. Ein Handbuch*, pages 517–592, Berlin. De Gruyter.

Anke Timmermann. 2015. Alchemy in Cambridge. An annotated catalogue of alchemical texts and illustrations in Cambridge repositories. *Nuncius*, 30(2):345–511.

Curt Wentrup. 2023. The alchemist, metal-divider and transmuter Carl F. Wenzel and his 1776 award from the Royal Danish Academy of Sciences through Professor CG Kratzenstein. *ChemPlusChem*, 88(5).

Konrad Wiedemann and Hartmut Broszinski. 2011. *Alchemie am Kasseler Hof. Zwischen Spekulation und Experiment. Ausstellung kostbarer Alchemie-Handschriften der Universitätsbibliothek Kassel*. Romanistischer Verlag, Kassel.

Elisabeth Wunderle. 2002. *Katalog der mittelalterlichen lateinischen Papierhandschriften*. Handschriften der Forschungsbibliothek Gotha, Band 1. Wiesbaden.

Deciphering Historical Syllabic Ciphers

George Lasry
The DECRYPT and CrypTool Projects
george.lasry@gmail.com

Abstract

Historical ciphers with syllabic elements are significantly more challenging for cryptanalysis than regular homophonic ciphers. We present here a novel computerized technique which recovers significant parts of the keys, allowing for the remaining parts to be manually completed. We solved several previously undeciphered French, Spanish, and Italian syllabic ciphers, and we also evaluated the performance of this method against a series of additional historical syllabic ciphers.

1 Introduction

The ciphers used in Europe from the 15th century and until the 18th century were primarily homophonic, with a nomenclature of varying size. In recent years, several computerized techniques have been developed and successfully applied to the deciphering of historical documents encrypted with homophonic ciphers. Those techniques, however, are ineffective against syllabic ciphers. In this article, we describe various types of syllabic ciphers in Section 2, and the challenges in deciphering them in Section 3. In Section 4, we present the new algorithm which can recover significant parts of the key. With this initial partial solution, a cryptanalyst familiar with the language can easily recover most of the remaining key elements. In Section 5, we provide several case studies of successful decipherment. In Section 6, we evaluate the performance of the algorithm against several historical syllabic ciphers. We conclude our results in Section 7.

2 Syllabic ciphers

Homophonic ciphers consist of a list of symbols representing letters of the alphabet – more than one per letter, as well as a nomenclature with symbols representing common words, persons, places, punctuation signs, signs for doubling consonants, for repeating, or for deleting the previous symbol, or nulls. Syllabic ciphers are an extension of homophonic ciphers, adding dedicated symbols to represent various types of syllables, such as: [1]

- **Consonant-vowel (CV)** syllables, such as MA/ME/MI/MO/MU or TA/TE/TI/TO/TU.
- **Vowel-consonant (VC)** syllables, such as EB/EC/ED/EF etc.
- **Consonant-consonant-vowel (CCV)** syllables, such as PRA/PRE/PRI/PRO/PRU.
- **Consonant-vowel-consonant (CVC)** syllables**,** such as PAR/PER/PIR/POR/PUR.

We refer to the letters of the alphabet, the syllables, and the words, persons, places which are part of the nomenclature as the **plaintext vocabulary**. With homophonic ciphers, there was usually only one way to decompose a given word before encryption, into plaintext vocabulary elements. We illustrate this with the English the word ESTABLISHED, which first needs to be decomposed into E-S-T-A-B-L-I-S-H-E-D, then enciphered.

With a cipher with CV syllables, there are additional options, such as E-S-**TA**-B-**LI**-S-**HE**-D, E-S-**TA**-B-**LI**-S-H-E-D, or E-S-**TA**-B-L-I-S-H-E-D.

With a cipher with also VC syllables, we have additional options, such as **ES**-**TA**-B-**LI**-S-H-**ED.**

With more complex syllables (e.g., CCV), we have even more options, such as E-**STA**-**BLI**-S-**HE**-D.

[1] In some cases, VCC syllables were encoded, e.g., EST, as well as elements composed only of consonants, such as TR or STR.

In the ciphers we examined, we saw two main patterns of how words are decomposed before encryption:

- **Random decomposition**: Any of the options for decomposing a word could be used, and often, the same word can be decomposed differently within the same ciphertext.
- **Systematic decomposition**: Decomposition is systematic and predictable, as described below.

We illustrate the case of systematic decomposition with the word ESTABLISHED, assuming the cipher has CV, VC, and CCV syllables.

- There are two options to start, E and ES. We select the longest one, ES.
- Next, we decompose TABLISHED. There are two options to continue, T or TA. Again, we select the longest, TA.
- Next, we decompose BLISHED. There are two options, B and BLI. We select the longest one, BLI.
- Similarly, we decompose the rest, obtaining **ES-TA-BLI-SHE-D**.

In some historical ciphers, we also see syllables spanning two adjacent words, for illustration purposes, the expression AN IDEA can be decomposed as A-NI-DE-A, the syllable NI including the last letter of the first word and the first letter of the second word.

The set of symbols for syllabic ciphers is usually significantly larger than for homophonic ciphers. For example, to represent all the CV syllables, with 17 consonants (B, C, D, F, G, H, L, M, N, O, P, QU, R, S, T, V, Z) and the 5 vowels, 85 additional symbols are needed. A similar number is required to represent VC syllables. For CCV and CVC syllables, dozens of additional symbols were needed.

For that purpose, instead of adding totally new symbols into the cipher key tables, diacritics were employed to alter the meaning of other symbols. For example, if the numerical symbol ***36*** represents the letter T, ***36:*** (***36*** with a colon on the right) could represent the syllable TA. Similarly, ***36.*** could represent the syllable TE, etc.[2] Such diacritics were added either on the top, bottom, left or right side of the symbol. Furthermore, two diacritics could be added, usually to represent CCV or CVC syllables. For example, ***36:'*** would represent TRA (the added ***'*** means that the letter R should be inserted between T and A). If diacritics are used consistently, e.g., ***46:*** is CA, ***46'*** is CE, we denote such sets of syllabic symbols as **regular syllabic symbols**.

There were cases, which we denote as **irregular syllabic symbols**, in which diacritics were not employed in a systematic manner. For example, ***36:*** means TA, but ***46'*** means CA (rather than ***46:***). Furthermore, in fully irregular syllabic ciphers, ***36:*** could be TA, but ***52'***, with a diacritic added to another unrelated numerical code (***52***), would represent TE. There were also cases in-between, with partial regularities.

Ciphers with regular syllabic symbols are significantly easier to solve. For example, if we know that ***36:*** represents TA, and that ***46.*** represents CI, then ***36.*** is likely to represent TI. In the algorithm we present in this article, we did not take advantage of such regularities in some ciphers, as we wanted to implement a solution applicable to the more general case.

This description is not comprehensive as we did not conduct a systematic survey of historical syllabic ciphers, and this paper is instead focusing on cryptanalysis techniques. The sample syllabic ciphers we analyzed are from Italy (15th and 16th centuries), Spain (16th and 17th centuries), and France (17th and 19th centuries).

3 Cryptanalytic challenges

The cryptanalysis of syllabic historical ciphers is significantly more challenging than of regular homophonic ciphers, such as a much larger key space. The set of cipher symbols most often consists of a few hundred distinct symbols. The size of key space is exponentially related to this number. Also, it may not be practical to compute n-gram statistics for very large vocabularies.

The types of syllables (CV, VC, CCV, CVC) the cipher employs and the decomposition scheme may vary and are generally unknown upfront.

[2] In some cases, diacritics were added to a new symbol, rather than to the one representing the base letter. For example, ***36*** could be T, but TA would be ***72:*** rather than ***36:***.

Existing computerized codebreaking algorithms for regular (non-syllabic) homophonic ciphers could only provide the meaning of most of the letter homophones, and a manual process was most often required to interpret the remaining symbols. Given the additional challenges with syllabic ciphers, our goal was to develop an algorithm that would be able to recover enough parts of the key and the plaintext, so that the remaining parts of the key may be reconstructed, and the ciphertext fully decrypted, with manual interactive work by a cryptanalyst.

4 The codebreaking algorithm

The algorithm is an extension, with substantial adaptations, of a simulated-annealing algorithm developed to solve homophonic ciphers (Kopal, 2019).

Due to the size of both the key space and of the plaintext vocabulary, the algorithm typically requires **extensive computing power**, e.g., a computer with dozens of cores, or multiple computers, running parallel instances of simulated annealing, to obtain useful results in minutes rather than in hours. In our tests, we ran the algorithm on a 64-core Windows 10 Pro PC with 256Gbytes of RAM memory. The algorithm may also require extensive trial-and-error to fine tune its parameters, which include:

- The expected pre-encryption word **decomposition scheme**. The algorithm we developed supports two schemes, **systematic** (deterministic) and **random**.
- **The set of syllables** expected: **CV, VC, CCV, CVC**, or any combination thereof.
- The **maximum number of homophones** per type of vocabulary element: Per vowel, per consonant, and per each type of syllables.
- Specifying letters that are **interchangeable**, e.g., U and V in French, Spanish, and Italian, or I, J, and Y in French.
- Letters that should be **replaced** with other letters (e.g., K with C, W with V), or **ignored** (e.g., X or Y in Italian).
- Whether **repeated letters** (e.g., LL, SS, TT) are represented by dedicated symbols.
- **A set of reference texts** of reference texts in the expected language, to compute n-gram statistics. While the algorithm is language-agnostic, some assumptions must be made on the plaintext vocabulary and the decomposition scheme before creating a database of n-gram statistics, which will be based on sequences of elements in the expected vocabulary, rather than just letter n-grams. These assumptions are formulated using the previously listed parameters. In addition to counting the occurrences of n-grams of single letters like E-S-T, S-T-A, or T-A-B, we also must count n-grams such as ES-TA-B or E-STA-B. As a result, n-gram statistics must be computed ad-hoc based on specific vocabulary parameters, rather than relying on pre-computed statistics.
- The **n-gram size**: 4-grams were empirically found to be the most effective, while 3-grams or 5-grams might be useful in some cases.

Other parameters are optional and can help the algorithm to converge better and faster:

- A small **set of common words** expected to be found in the nomenclature.
- Some **limitations** on the set of symbols allocated to letter homophones, such as allowing only symbols without diacritics, or numerical codes within a certain range, to be assigned as letter homophones.
- The **maximum number of distinct ciphertext symbols** to be considered. The algorithm discards the less frequent ones. This effectively reduces the size of the search key space.
- **Tentative key assignments** of symbols to vocabulary elements, as they are being identified with the semi-automated work described later in this section. If correct, those allow simulated annealing to produce a better and more complete solution, quickly and more reliably.

Simulated annealing starts by randomly allocating ciphertext symbols to plaintext vocabulary elements, prioritizing the most frequent ones. This means that the less frequent vocabulary elements will be ignored if there are not enough distinct ciphertext symbols, or if the number of processed ciphertext symbols has been limited.[3]

During simulated annealing, the only allowed key change is swapping the assignment of any two cipher symbols (for example, if ***32:*** was assigned to TA, and ***43'*** was assigned to CE, after the swap, ***32:*** is assigned to CE, and ***43'*** is assigned to TA). As a result, and in contrast with other solvers of homophonic ciphers, the number of symbols allocated to each vocabulary element is constant. This was found to provide for a more stable and more effective algorithm.

A score measuring the quality of the deciphered is computed as follows:

- Before starting simulated annealing:
    - Parse all the reference plaintexts, decomposing words into the specified vocabulary elements, according to the specified decomposition scheme.
    - Compute F_g, the relative frequencies for every combination g of four successive elements of the vocabulary, a.k.a. *4-gram*, such as E-S-T-A, or CO-N-TRO-L, which appear in reference plaintexts.[4]

- During the search with simulated annealing, evaluate a candidate key as follows:
    - Decipher the ciphertext using the candidate key.
    - Compute N_g, the number of occurrences in the decrypted text of each 4-gram g.
    - Compute N_c, the number of occurrences in the decrypted text of each vocabulary element c.
    - The score S for the tentative decipherment is computed as follows:

$$S = \sum_g N_g \log F_g / \sum_c N_c^2$$

When the algorithm starts producing tentative decryptions, it also highlights (in capital letters) plausible segments of vocabulary elements, if those can be found in the specified reference texts. This is especially useful if the cryptanalyst is not familiar with the plaintext language. It is expected that as more elements of the key are correctly recovered, there will be more of those highlighted plausible segments. Furthermore, the algorithm counts how many times each symbol occurs in such a plausible segment, and those occurring dozens of times are listed as likely to have been correctly assigned.

Working with the tool is done iteratively. At first, the program may produce only a few highlighted segments, and a few key assignments suggestions. The cryptanalyst manually reviews those segments and suggestions, and if they look plausible, they can be entered as parameters for the next run.

An example of an initial run is given in Figure 1. It is possible to discern several highlighted plausible French words or expressions, such as EN PRENDRE, GRANDEMENT LE, and IMPRIMER LE. The correct assignment of several ciphertext symbols (e.g., those representing N, E, T, R) composing those expressions are also be validated by the statistics listed below the decrypted text, and those symbols may be safely assigned accordingly for subsequent runs, by setting the tentative key assignments' parameter.

Running the algorithm again with the revised parameters will reveal additional assignments. Non-highlighted segments that are plausible may also reveal additional plausible assignments. When enough elements have been recovered, it may neither be necessary nor useful to run the automated algorithm again and the remaining work can be completed manually.

[3] By setting the parameters which specifies the maximum number of distinct ciphertext symbols to be processed.

[4] Or 3-grams, or 5-grams.

```
b-o< 14 U^ L S H :- S -8 r& S i- r& b-o_ 18 :- O. S d& 20 d. b U_ S u&
sv   h  ti c N T E  N P  RE N D  RE SO   vv e  me n ce z  ca r so n de
svhticNTENPRENDRESOvvemencezcarsonde

D< a b-o& a -i oio \o r. S u& n& S H U& + L 6 : ii r& + m_ i= b n. g
m  i se   i a  y   G  RA N DE ME N T LE a c f d i  re a po v  r ma l
miseiayGRANDEMENTLEacfdireapovrmal

ii ap -8 r^ n& b U& + n^ O^ + H r& + u& !  x^ b-o& S b-o_ S H :- ap m& +
I  M  P  RI ME R LE a mi ni a t re a de rr i  se   N SO  N T E  M  PE a
IMPRIMERLEaminiatreaderriseNSONTEMPEa

135 times: S    -> N
 89 times: t&   -> E
 79 times: H    -> T
 75 times: b    -> R
 60 times: i=   -> V
 36 times: :-   -> E
 35 times: +    -> A
 35 times: r&   -> RE
 33 times: a    -> I
 28 times: u&   -> DE
 23 times: =    -> O
 22 times: U&   -> LE
```

Figure 1– Sample printout of the algorithm for syllabic ciphers

5 Decipherments

With this technique and the semi-automatic process described in Section 4, we deciphered several documents encrypted with historical syllabic ciphers, for which the key and the plaintext were not known in advance.

5.1 Archivio di Stato di Milano - Visconteo Sforzesco Segnatura

A letter in Italian from Ottone de Carretto from 1457. We analyzed about 3,900 ciphertext symbols, with 113 distinct ones. The cipher features symbols for VC syllables assigned in an irregular manner, and words are decomposed randomly. The reconstructed key is shown in Appendix 1. It later turned out that a copy of the cipher key is held in the Archivio di Stato di Milano.[5]

5.2 Simancas EST LEG 1381 – 143

A letter in French, from 24 August 1551. After deciphering the letter, it turned out to be using the same cipher used between Charles V and his ambassador Jean de Saint-Mauris (Pierrot et al., 2023). We analyzed a total of 4,300 ciphertext symbols. There are 148 unique cipher symbols, some with diacritics to represent CV syllables and a few CCV syllables, assigned in a regular manner. Word decomposition is random.

[5] ASMi Carteggio Visconteo Sforzesco Segnatura1598 f.89.

5.3 Simancas EST LEG 1381 - 180

A letter in Spanish from 15 September 1551.[6] It contains about 2,300 ciphertext symbols, with 123 distinct ones. It features CV syllables, as well as a few CCV syllables, marked with diacritics, and assigned in a regular manner. Word decomposition is systematic.

5.4 BnF Clairembault 421 f. 160

A letter from Henri Brasset, a French resident in the Hague, to Cardinal Mazarin, at the time the chief minister of infant King Louis XIV, written on 30 March 1649 and held in the Bibliothèque Nationale de France.[7] The ciphertext has about 1,800 ciphertext symbols, and 156 unique symbol types. Some archive images are damaged, several parts missing or illegible. Non-numerical symbols represent letter homophones. The other elements (CV syllables, nomenclature) use two-digit numerical codes with optional diacritics, but the syllable symbols are mostly assigned in an irregular manner. Word decomposition was random. Overall, cryptanalysis was quite

[6] The decipherment includes some indications on the possible sender and recipient: "Copiado loque Su Magestad scrive a [Principe] Doria a v de setienbre presinte", "Al seno Ferando", "Al enbaxador Figueroa." More details on this cipher and other Spanish syllabic ciphers in (Tomokiyo, 2023, direct link).

[7] This cipher was first presented as an unsolved cipher by Satoshi Tomokiyo in (Tomokiyo 2023, direct link).

challenging. After decrypting the ciphertext, we were able to find the original plaintext in French archives,[8] and to complete most of the key assignments, as shown in Appendix 2.

5.5 Dresden - Militärhistorisches Museum

An unpublished letter, recently discovered in the archives of Russian Field Marshall Michail Andreas Barclay de Tolly. The letter was written on 4 August 1813 by General Rapp to Napoleon's headquarters, during the siege of Danzig. The cipher consists of about 900 ciphertext symbols, with 109 unique symbols, composed of one or more digits, mostly in the range up to 200, without any diacritics. After cryptanalysis, we established that the cipher features CV syllables, assigned irregularly, and that words are decomposed randomly. It later turned out that the cipher key was a known version of Napoleon's Small Cipher.[9]

6 Performance evaluation

We also analyzed the performance of the algorithm against additional syllabic ciphers for which the key was already known.[10]

- **Baldassare Castiglione to Niccolò Schomberg**, 25 March and 3 April 1527.
- **Simancas EST LEG 1386 - 1**. From 15 March 1577, from Pedro Gonzalez de Mendoza to King Philip II.
- **ARA Brussels SEG 2559.** A series of letters in Spanish from 1674-1678 also sent to Balthazar de Fuenmayor.
- **ARA Brussels SEG 2559.** A letter in French, 31 July 1676, sent to Balthazar de Fuenmayor, Spanish ambassador in Denmark.
- **KHA Amsterdam, Willem II/XIII-I**. A letter, from 9 January 1684 from le Comte d'Avaux, the French ambassador in Holland, to King Louis XIV.

In Figure 2, we summarize the performance of the new algorithm, tested against those ciphers and the other five listed in Section 5. The accuracy numbers refer to the **decrypted text accuracy** – the percentage of the ciphertext symbols in the documents correctly decrypted, and the **reconstructed key accuracy** – the percentage of the symbol types (distinct ciphertext symbols) correctly assigned. The accuracy of the decrypted text is always higher than the accuracy of the reconstructed key, as lower frequency symbol types are often ignored by the algorithm (and therefore, not assigned, thus reducing the key accuracy), and sparsely used symbols are more likely to be incorrectly interpreted. More importantly, the accuracy of the decrypted text is the main factor affecting the ability to make further progress manually.

Achieving those promising performance numbers with the automated algorithm required extensive trial-and-error and tweaking of the parameters, especially for those ciphers for which the types of syllables used were not known. The algorithm turned out to be highly sensitive to the parameters specifying the maximum number of homophones per vowel or per consonant, and the word decomposition scheme – a wrong selection would often prevent the algorithm from converging.

We also tested the performance of the algorithm when limiting the number of ciphertext symbols to 1,000, so that the performance may be compared across the various ciphers, as shown in the two rightmost columns. The performance is strongly affected by the number of distinct symbol types, especially for ciphers with more than 150 symbol types, which require longer ciphertexts to obtain satisfactory results. In general, an initial accuracy of 40% or above is most often enough to decipher a ciphertext with the semi-automated process described in this Section 4.

[8] BnF Français 17901 f.230.

[9] More details in (Tomokiyo, 2023, direct link).

[10] We recovered the keys for the first two items based on plaintext inscribed on the margins. The key for the Comte d"Avaux cipher was recovered based on a similar key (Lasry, 2019). The key for the fourth item was recovered by Carlos Köpte (Tomokiyo, 2023), and the key for the fifth one by Norbert Biermann (Simonetta, 2023).

Reference and origin	Language and year	Length	Symbol types	Syllables	Regular syllables	Word decomposition	Accuracy - Decrypted text and key		Accuracy with only 1000 symbols	
ASM – Ottone de Carretto	Italian 1457	3,900s	113	VC	No	Systematic	**75%**	58%	**27%**	13%
Castiglione to Schomberg	Italian 1527	900	134	CV	Partially	Systematic	**35%**	18%		
Simancas EST LEG 1381 180	Spanish 1551	2,300	123	CV CCV	Yes	Systematic	**78%**	51%	**49%**	30%
Simancas EST LEG 1381 143	French 1551	4,300	148	CV CCV	Yes	Random	**81%**	44%	**<5%**	<5%
Simancas EST LEG 1386 1	French 1577	1,200	156	CV VC CCV	Yes	Systematic	**78%**	47%	**<5%**	<5%
BnF Clairembault 421 f. 160	French 1649	1,800	156	CV	Yes	Random	**80%**	48%	**<5%**	<5%
ARA SEG 2559 - French	French 1674	2,300	101	CV	No	Random	**96%**	68%	**86%**	51%
ARA SEG 2559 - Spanish	Spanish 1674-1678	3,200	178	CV	Yes	Systematic	**73%**	33%	**35%**	16%
KHA Prins Willem II/XIII-I	French 1684	4,200	219	CV	Partially	Random	**78%**	39%	**<5%**	<5%
Dresden	French 1813	900	109	CV	No	Random	**76%**	50%		

Figure 2 – Performance evaluation of the new codebreaking algorithm for syllabic ciphers

7 Conclusion

Before this work, the cryptanalysis of a syllabic cipher was highly challenging and most often possible only if the structure of the syllabic symbols was regular, or if a matching plaintext could be found. The new algorithm presented here can provide an initial breakthrough, which, with manual analysis, can most often lead to a successful decipherment of historical syllabic ciphers, as exemplified here.

In addition, the algorithm was designed for the general and more challenging case of irregular syllable symbols. It may be easily improved to take advantage of possible regularities in the assignment of the ciphertext symbols to syllables.

Acknowledgments

The author would like to thank Jessika Novak for providing a copy of the Ottone de Carretto cipher, Klaus Schmeh, Wolfgang Schmidt, and Erik Zimmermann for publishing the Dresden cipher, Satoshi Tomokiyo for bringing the Brasset-Mazarin cipher to light and for reviewing an early version of the paper, Paolo Bonavoglia for providing a copy of original the Ottone de Carretto cipher key and helping with the decipherment, and Norbert Biermann for reviewing the paper and helping to fine-tune decipherments.

Funding

This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Nils Kopal, 2019. *Cryptanalysis of homophonic substitution ciphers using simulated annealing with fixed temperature*. HistoCrypt 2019.

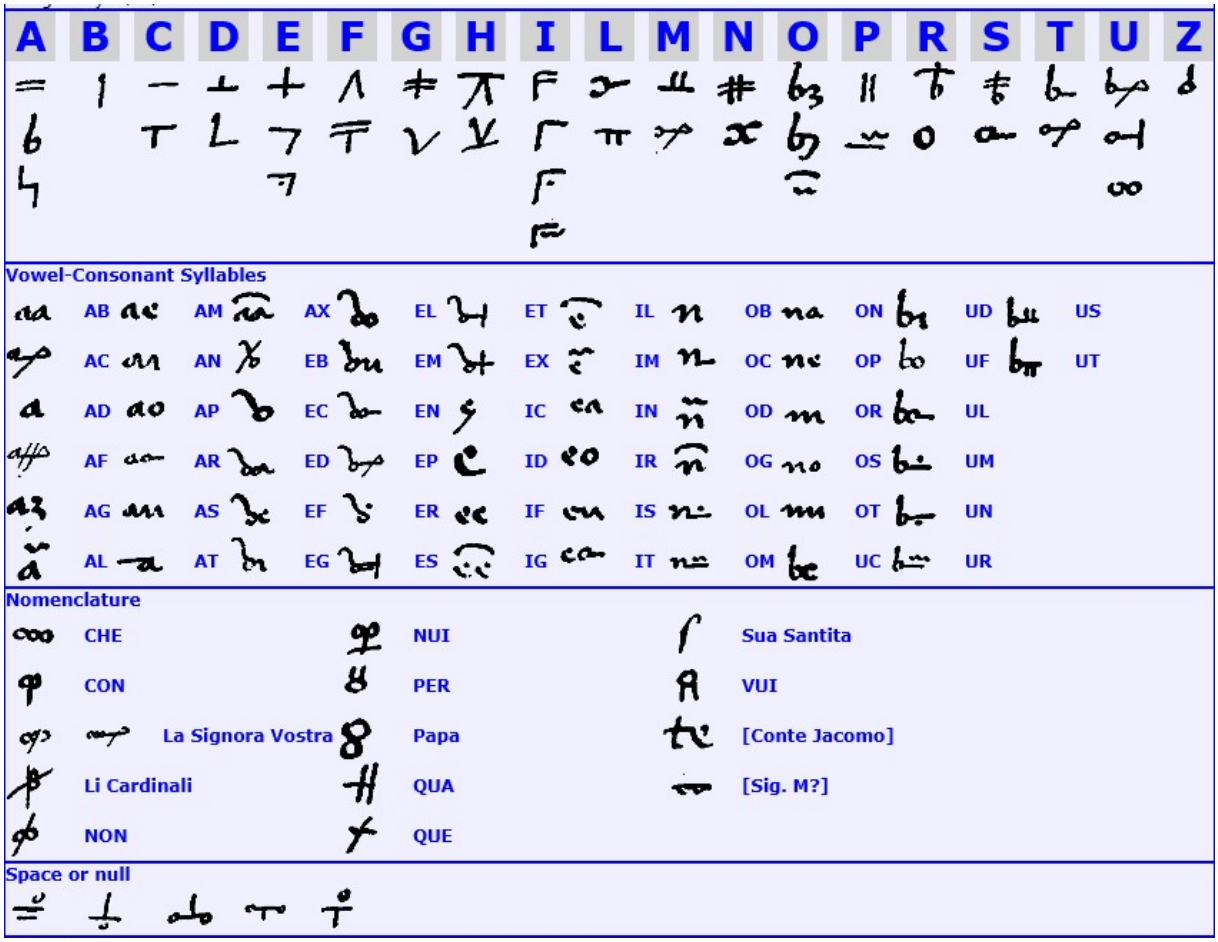
George Lasry, 2021. *Deciphering a Letter to Louis XIV from his Ambassador to the Dutch Republic, le Comte d'Avaux, 1684*. HistoCrypt 2021.

Cécile Pierrot, Camille Desenclos, Pierrick Gaudry, and Paul Zimmermann, 2023. *Deciphering Charles Quint (A diplomatic letter from 1547)*. HistoCrypt 2023.

Marcello Simonetta, 2023. *Svelati i segreti delle lettere di Castiglione alla vigilia del Sacco di Roma* [Accessed: November 2023]. Storia in Rete. https://storiainrete.com/svelati-i-segreti-delle-lettere-di-castiglione-alla-vigilia-del-sacco-di-roma/

Satoshi Tomokiyo, 2023. *Cryptiana, Articles on Historical Cryptography* [Accessed: November 2023]. http://cryptiana.web.fc2.com

Appendix 1 – The reconstructed key for the Ottone de Corretta cipher



Appendix 2 – The reconstructed key for the Brasset-Mazarin cipher

A Typology for Cipher Key Instructions in Early Modern Times

Beáta Megyesi[1], Benedek Láng[2], Nils Kopal[3], Vasily Mikhalev[3], Crina Tudor[1], Michelle Waldispühl[4]
[1] Stockholm University, Sweden
[2]Eötvös Loránd University, Hungary
[3]University of Siegen, Germany
[4]University of Oslo, Norway

Abstract

We present an empirical study on instructions found in historical cipher keys dating back to early modern times in Europe. The study reveals that instructions in historical cipher keys are prevalent, covering a wide range of themes related to the practical application of ciphers. These include general information about the structure or usage of the cipher key, as well as specific instructions on their application. Being a hitherto neglected genre, these texts provide insight into the practice of cryptographic operations.

1 Introduction

Historical cipher keys have been studied empirically and during the past years rather extensively, as evidenced by works such as (Rockinger, 1892; Meister, 1902; Meister, 1906; Kahn, 1996; Láng, 2018; Lasry et al., 2020; Megyesi et al., 2024; Megyesi et al., 2022), along with the examination of ciphertexts. Despite this attention, our understanding of how cipher keys were practically applied and used remains limited. This gap in knowledge arises because ciphertexts and cipher keys are seldom stored together in archives and libraries, posing challenges to researching key usage practices. To address this issue, one approach is to investigate the instructions written by the creators of cipher keys.

In this paper, we aim to explore whether and how instructions were utilized in historical cipher keys, and what information they contained about their structure, application, and usage. The following questions guide our inquiry:

- How prevalent are instructions in historical cipher keys from early modern Europe?
- What types of information do instructions in historical cipher keys encompass?
- What insights can we gain about the practical usage of cipher keys from these instructions?

In the subsequent section, we will review prior work on the subject. Following that, we will define the term ”instructions” and outline our investigation and methodology. We will then present the characteristics of the instructions found in historical cipher keys, accompanied by illustrative examples, and discuss our discoveries. Finally, we will conclude the paper and identify further research directions.

2 Previous Work on Instructions in Keys

Instructions such as notes, explanations, and cipher rules are widely known among cryptohistorians studying the early modern era, yet explicit secondary literature on them is scarce. Alois Meister called attention to the importance of cipher instructions given to individual envoys already in his earlier book on the beginnings of diplomatic cryptography (Meister, 1902), and cited a variety of them in his second book as sources for the actual application of ciphers (Meister, 1906). He added that these brief written texts attached to the ciphertexts were probably accompanied by oral explanations from the cipher secretary (p. 59).

In the 120 years following Meister’s works, instructions were not systematically studied as a genre but as individual examples that facilitate understanding a specific, unconventional key (Láng, 2022). Lately, Camille Desenclos devoted more systematic attention to how the 16th-century scribes in France acquired cryptographic knowledge. In her unpublished study, she is particularly interested in the actual practice using the instructions as precious sources regarding the practical reality of cryptography (Desenclos, 2023).

3 Instructions

Before we explain how we investigate instructions in cipher keys, we need to comment on the actual term ”instruction” as it allows for various interpretations.

By an instruction we usually mean a set of directions or orders given to guide someone in performing a task or carrying out a particular action. It typically provides step-by-step information or guidance on how to achieve a specific goal or complete a task. Instructions are designed to be clear, concise, and informative, helping the recipient understand what needs to be done and how to do it optimally. Instructions can be found in various contexts, such as manuals, guides, procedures, or verbal communication, and they play a crucial role in conveying information and facilitating understanding for effective execution of tasks.

In our study, we adopt a broad interpretation of the term in the context of historical cipher keys. We focus on identifying descriptions within keys that elucidate the operational use of a cipher key, encompassing both the key's content and intended application in clear text, regardless of the language. Moreover, instructions are contemporaneous with the keys they accompany and, just like the keys themselves, are oftentimes anonymous. Our inquiry remained open to the length of the instruction, ranging from a single word to several pages of documents.

Notably, instructions and explanations occurring in the cipher treatises of named cryptographers are certainly relevant but not the subject of our present inquiry.

4 Method

Our study originates from an examination of a sample of cipher keys to understand what was encoded in historical cipher keys and how this was done (Megyesi et al., 2024). We use the same set of cipher keys, consisting of 1610 keys, all sourced from the DECODE database (Héder and Megyesi, 2022). We identified and extracted cipher keys containing instructions, resulting in a total set of 235 cipher keys. This means that 15% of the keys contained some kind of instructions, which is not a result of representative research, but we believe it gives a realistic image of how many cipher keys survived together with some instructional notes. The keys originate from a wide range of geographic areas and time periods, as illustrated in Figures 1 and 2.

Each key in the sample was manually analyzed and cross-validated in cases where there was ambiguity. The parameters that we marked were length (short or long), the language of the cleartext (e.g., Latin, French, English, German, Spanish), and the content described as free text. The content description encompassed details about the usage of individual code elements and their application, such as explanations of how to apply nullities. It also incorporated notes about the interpretation of the cipher key table or the key's application in various contexts. On the basis if these parameters and and descriptions, we categorized the types of instructions and their functions. Below, we provide a summary of our findings.

5 Types of Instructions

5.1 Structural Information

The great majority of cipher keys feature concise and short explanations of various functions represented within the key. These explanations often include section titles such as ”Ad scribendum” (to write) and ”Ad legendum” (to read), as seen in two co-stored keys, depicted in parts in Figure 3. The titles indicate that the key information presented in the ”Ad scribendum” part shall be used for encryption while the section ”Ad legendum” shall be used for decryption. The plaintext elements and the code elements are sorted in reversed order in the two sections facilitating the respective cryptographic operation. As illustrated in Figure 3, the code elements follow the order of the plaintext elements (alphabet letters in Figure 3) in the ”Ad scribendum” section, and vice versa, the plaintext elements are arranged according to the code elements' order in the ”Ad legendum” section (numerical order in Figure 3). Other titles indicating the same function occurring in our material are ”Zum Chiffriren” vs. ”Zum Dechiffriren”, ”pour Chiffrer” vs. ”pour Dechifrer”.

Structural information given as titles may also include explanation of specific types of code elements, such as nullities, cancellation signs, days of the month, or grammatical categories like ”Signa Casuum” (signs of cases) and ”Numeri nihil significantes” (numbers of no significance), as shown in Figure 4.

Additionally, it is common to find descriptions of specific sections of the cipher key table, providing clarity on the content of the key. An illustrative

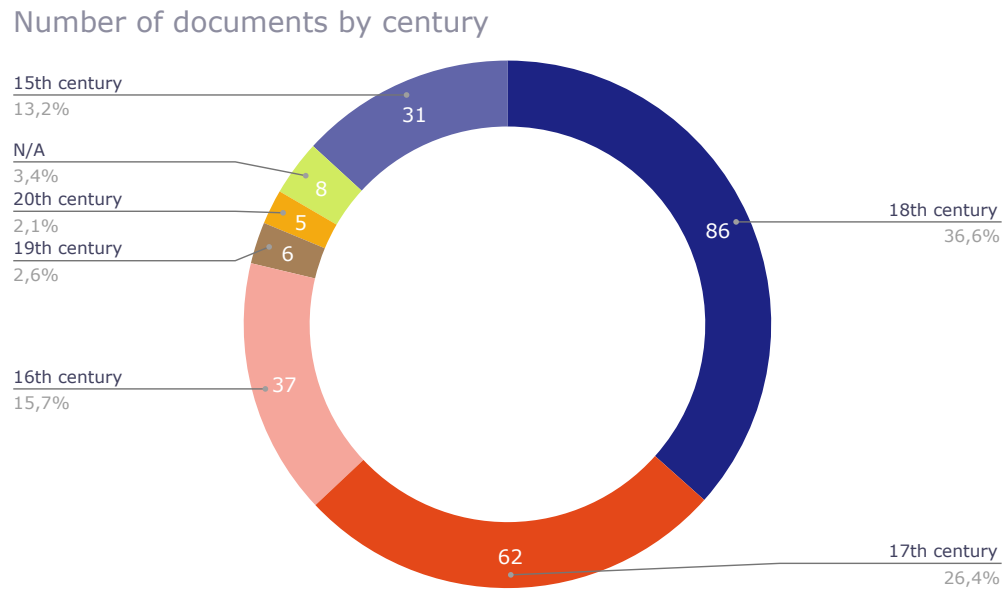


Figure 1: Time period of the key sample containing instructions

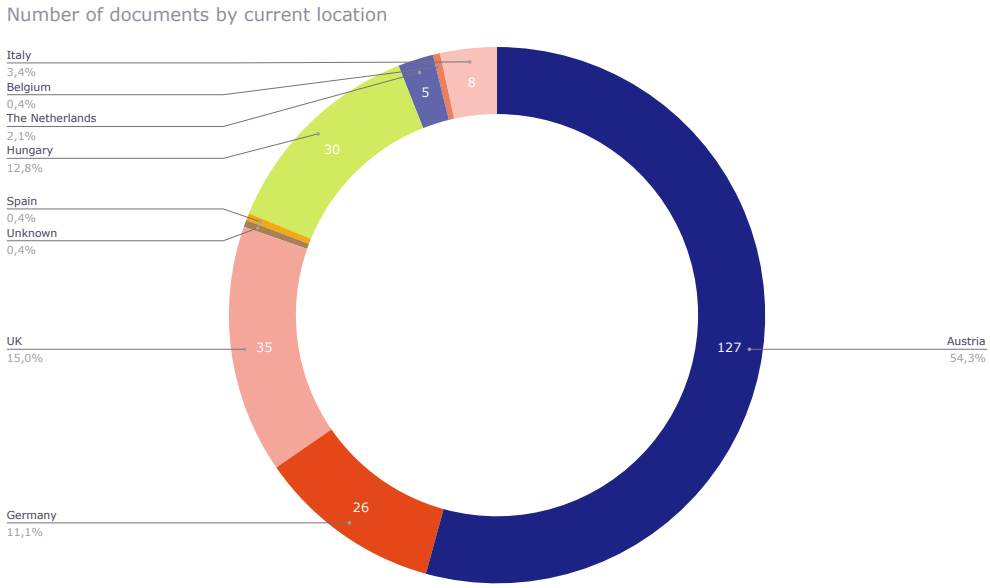


Figure 2: Geographic areas of the key sample containing instructions

example is presented in Figure 5.

5.2 Information about Code Elements

Typically, cipher keys also contain information about various groups of code elements with certain functions such as code elements for nullities, cancellation signs, double plaintext letters, punctuation marks or paragraph markers, to mention a few. An example of such a cipher key is illustrated in Figure 6 containing the cleartext ”Nulles 9, 99, 999; deux points 909 - ad lineam ou Commencement de Chapître 959 - Point 995” with translation of the Latin and French sequences to English as: Nulls 9, 99, 999; two points 909 - to the line or Beginning of the Chapter 959 - Point 995.

In several keys, we find a section with code elements that do not have any value and, in practice, function as nulls. Typically, such sections are in-

Figure 3: Section title in a cipher key (Key ID-676, 1664–1668)

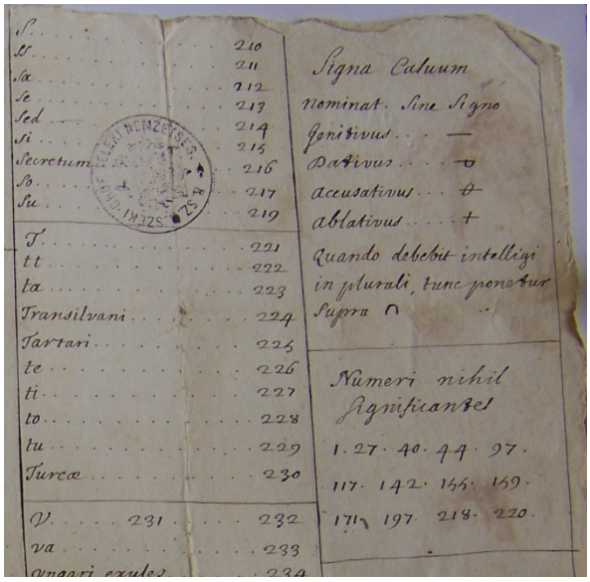


Figure 4: Explanation of code groups in a cipher key (Key ID-670, 1660–1690)

dicated with a short explanation about these missing code elements, such as ”Nonvaleurs depuis 1201 jusqu’à 1250” (No value from 1201 until 1250) (Key ID-613, 1703–1711). Instructions also contain definitions of the nullities or cancellation signs by one or several intervallum such as in ”Nullae a 1° usque ad 9vum, item a 450 usque ad infnitum”, meaning Nulls from 1st to 9th, likewise from 450 to infinity in Figure 7, or explain how to construct or recognize them.

Not surprisingly, it is also common to find descriptions of specific parts of the cipher key table, not only providing clarity on the content of the key but also containing explicit instructions about their application and usage. An example is depicted in Figure 8 where on the top of the key document, the title describes the first table as a key for a homophonic cipher for single letters. Below, the use of more homophones for the vowels is indicated, ”Vocales quarum notæ alternatim cum superioribus adhibendæ ”(The vowels, whose codes should be used alternately with the preceding ones) and on the second page, nulles are described ”Errantes sed nihil significantes interdum interserent” (Nulls that signify nothing and that may be inserted sometimes).

5.3 Explanation of the Key

As explanatory instructions we classify texts consisting of a few sentences that are included in the cipher key table and longer texts that were written on separate slips of paper. In several instances, we find these short instructions titled as ”Nota” or ”NB” which abbreviates ”nota bene”. Hence, they were introduced to the intended user of the key by metalanguage.

Diacritics such as dots, lines or other markers placed on top or around a symbol are oftentimes used in order to change the original significance of the symbol, which can also be explained and exemplified in some of these short instructions. One instance of this is the use of dots above the code element to signify double letters. Another instance is shown in Figure 9, where placing a line on top of a code element that initially stands for the digraph ”ba” reverses the order of the letters to ”ab” instead. In the example given in Figure 10, the instruction indicates that a line above a code element signifies a cancellation of this sign while two or more lines mean the same as no line. The instruction is given in the second half of the right most column of a one-page cipher key shown in Figure 10, transcribed and translated as:

> `Linea supra vel infra numerum, seu hic litteram, seu syllabam, seu nomen denotet, ducta significat errantem. Si uero duæ, vel plures supra vel infra numerum ductæ sint lineæ, numerus significat idem, quod significat, si nulla desuper ducta sit linea.`
> (A line above or below a number, whether it marks a letter, or a syllable, or a name, signifies an errant one. If indeed two or more lines are drawn above or below the number, the number signifies the same as it signifies if no line were drawn above.)

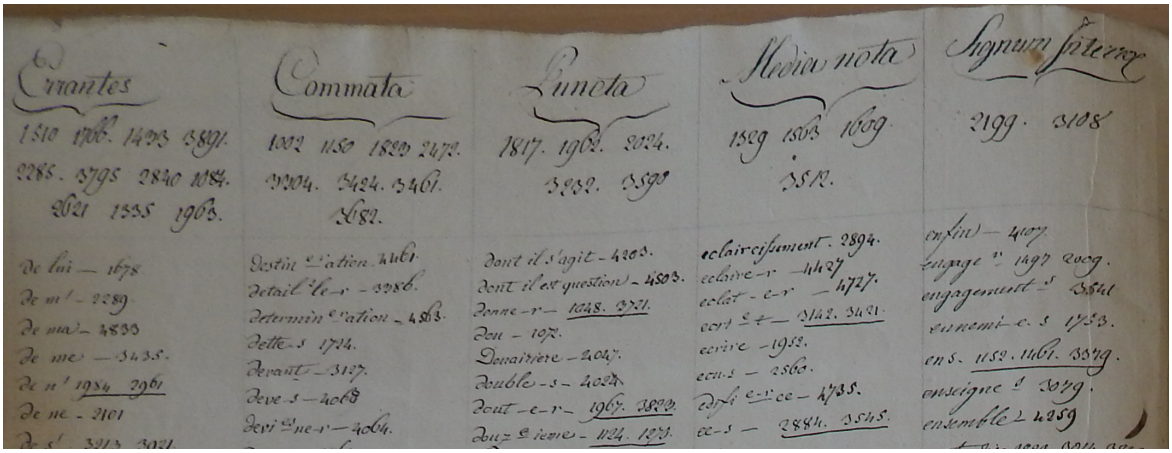


Figure 5: Explanation of different sections in a cipher key (Key ID-1728, 1805)

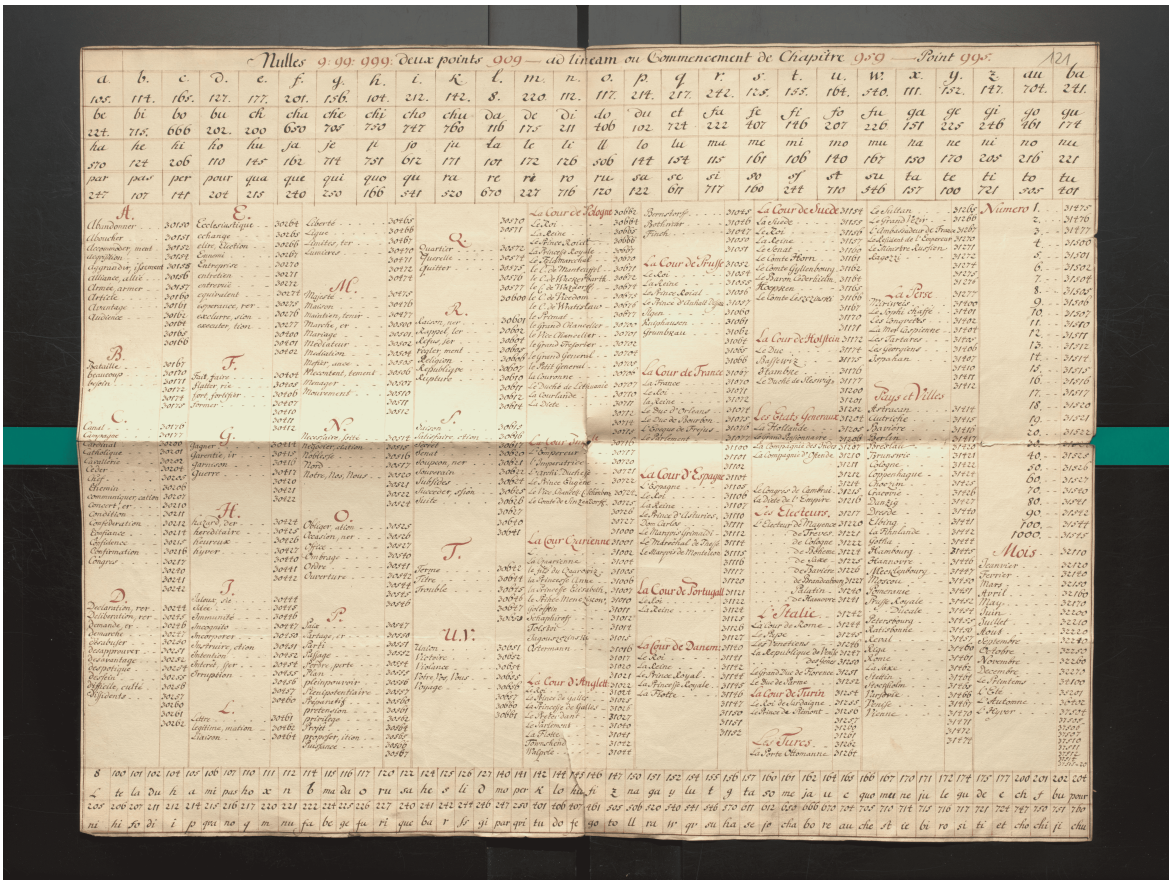


Figure 6: Explanation of code elements types in a cipher key (Key ID-2330, year unknown)

In a number of short instructions for digit ciphers, the use of sign separators is explained. An example of such an instruction is shown in Figure 11. The note is given in rather small writing between curly brackets at the top of the cipher key table. Interestingly, the instruction to this key from 18th-century Saxony, is given in German while the key here is in French. The German text says:

`NB: die Chiffres mit zwey Zahlen sind ohne Punckt, die Chiffres aber mit einer Zahl müßen einen Punckt . nach sich haben; die Zahl 8. gilt nichts,.`

(The code elements with two digits are without dot, however, the code elements with one digit need to have a dot . after themselves; the digit 8 has no value).

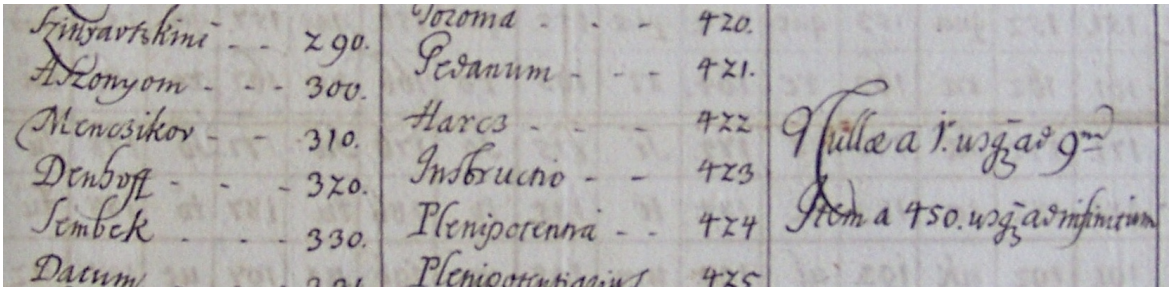


Figure 7: Explanation of an interval in a cipher key (Key ID-613, 1703–1711)

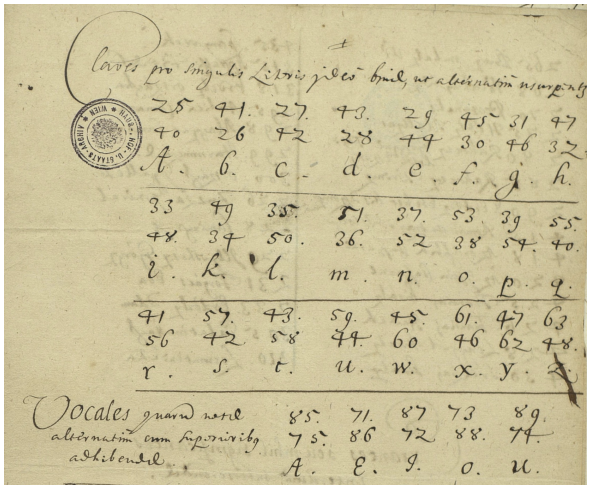


Figure 8: Explanation of groups of code elements in a cipher key (Key ID-700, 1664–1668)

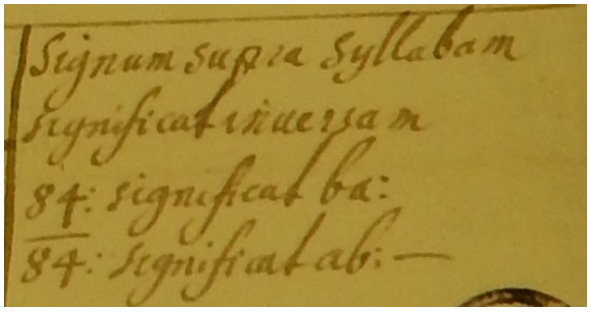


Figure 9: An example of diacritics explained in the instructions (Key ID-2061, 1666)

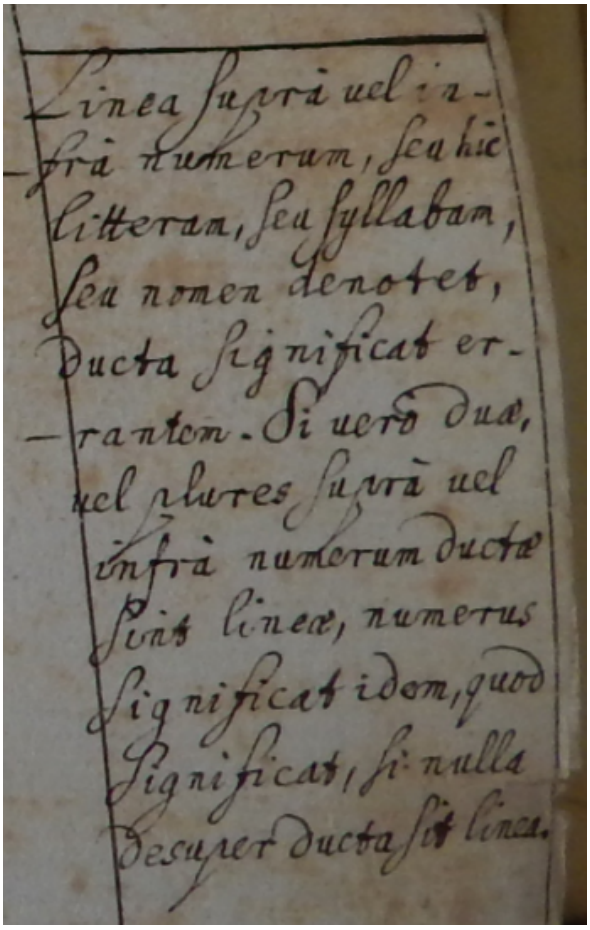


Figure 10: Explanation of diacritics in a cipher key (Key ID-1295, 1500–1699)

Longer instructions are one to several pages long, structured texts, often on separate slips of paper attached to the cipher key. They are easy to recognize, as they often are composed of numbered paragraphs, which we exemplify in Figure 12. We differentiate two classes in this category of longer instructions.

The first kind explains an unconventional cipher key that could not be otherwise intuitively applied. Some of these ciphers are particularly complicated, inventions of their authors who do not suppose that the system can be appropriately used without the explanations. In such cases, sample encryptions are also included to help the learner (Láng, 2022) and (Key ID-531, 1731).

In other cases, the cipher might be a simpler but still not very widespread type, such as a matrix cipher, that requires explanations. In the case illustrated in Figure13, the instructions continue for the rest of the page as well as the two following

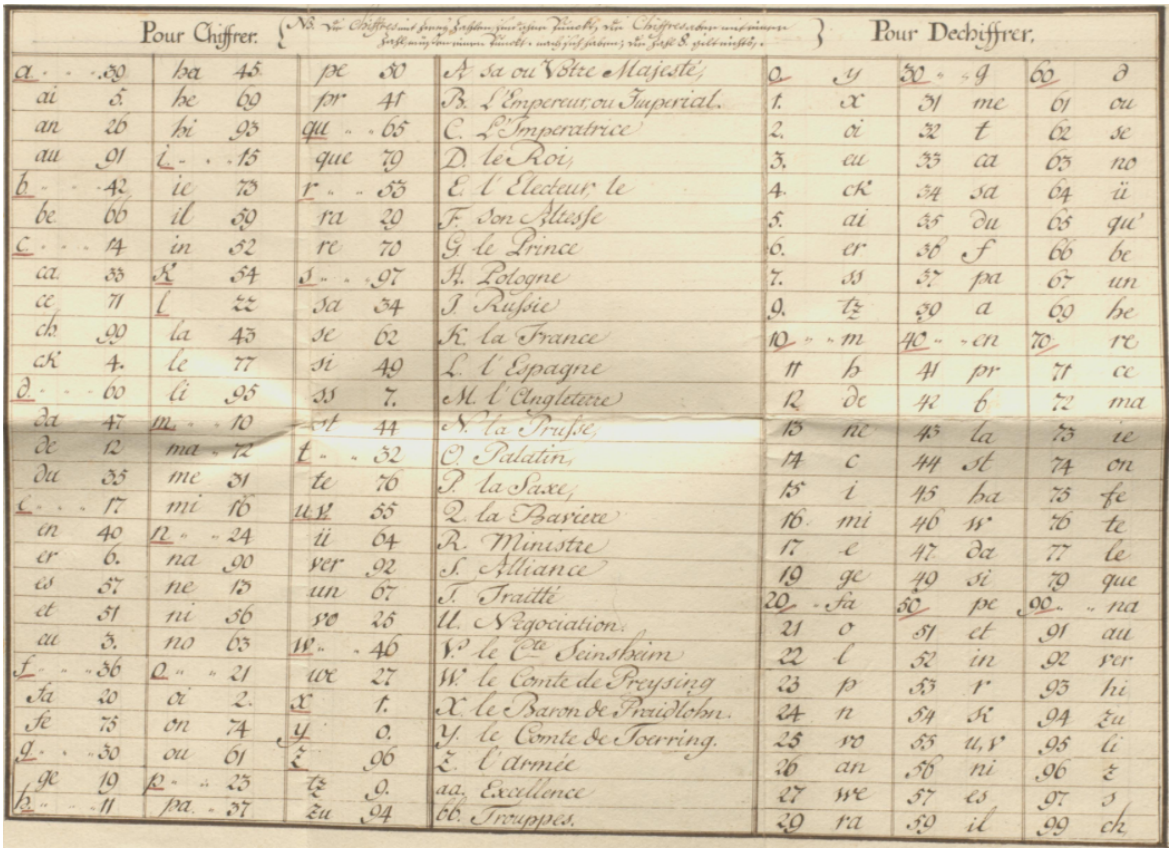


Figure 11: Explanation of sign separators in a cipher key (Key ID-935, 1761)

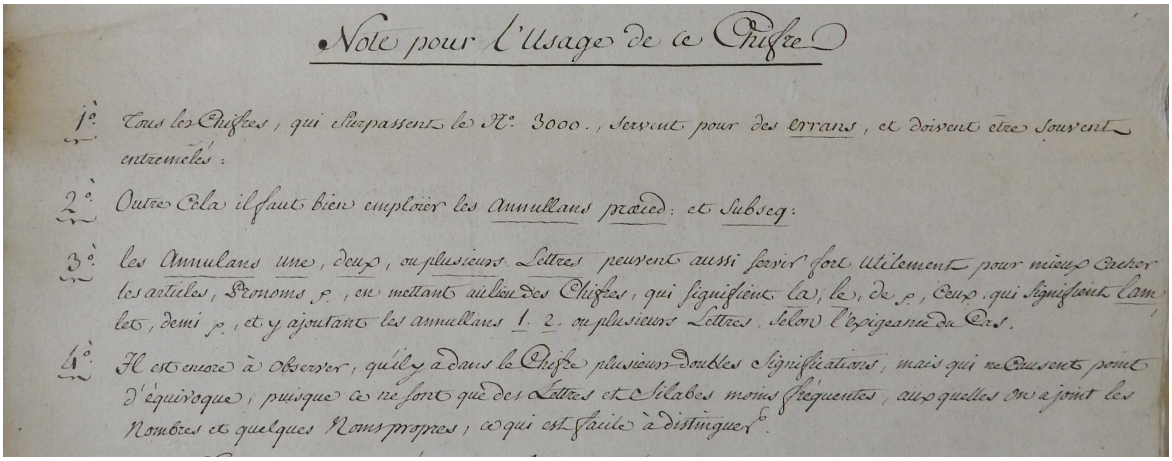


Figure 12: An example with numbered paragraphs of instructions (Key ID-1722, 1790)

ones, providing insights on the usage of the key.

The second sub-category of long instructions includes a list of notes that usually accompany homophonic cipher keys and give advice on how to put them into practice. This second type does not provide us with new information about how homophonic ciphers function, rather, they show what typical mistakes the inattentive or lazy scribes made. They typically stress the importance of alternating the homophones (not always using the first code element for each letter), to make full and varied use of the code elements signifying syllables, to avoid using letters or syllables for words which have a nomenclator equivalent, or they stress the importance of incorporating nulls throughout the ciphertext, as exemplified in the bottom paragraph in Figure 14.

They might furthermore instruct the user that

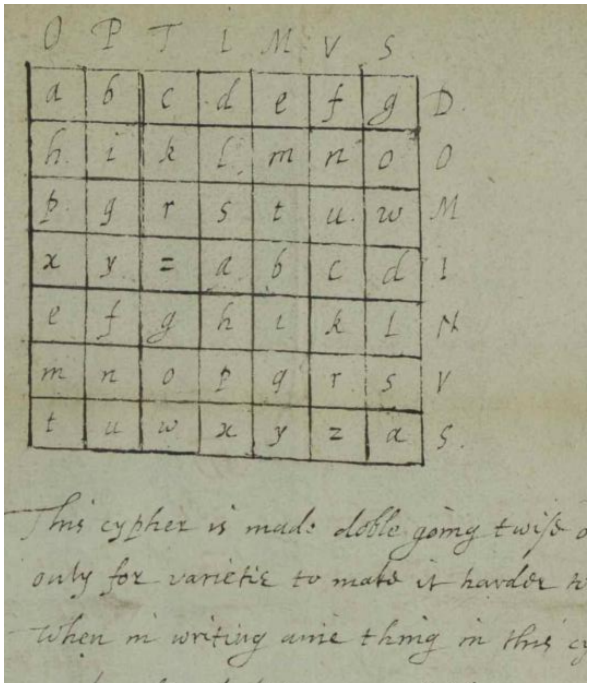


Figure 13: An example of instructions for ciphers beyond the most widespread substitution-based types (Key ID-415, 1628)

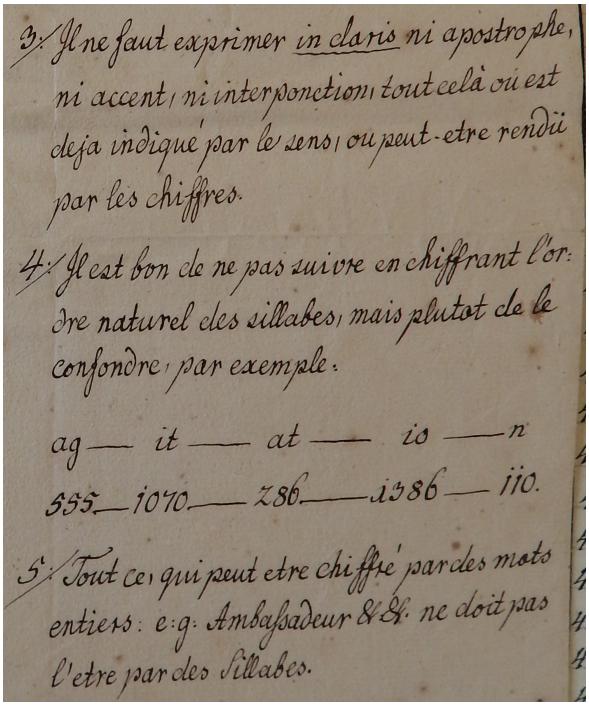


Figure 14: An itemized list of instructions (Key ID-1704, 1764)

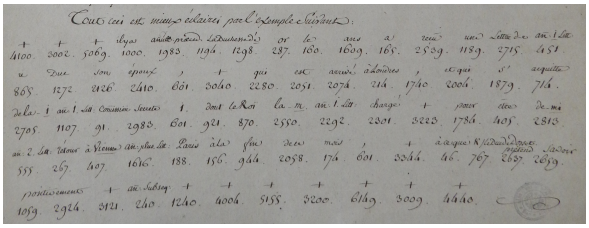


Figure 15: An example on how to use the key (Key ID-1722, 1790)

"words should not be distinguished from each other, so that in this way, the writing becomes more difficult to read.", which means that the user should not mark word boundaries by space as we usually do in writing. Some instructions also refer to the use of punctuation. The top paragraph in Figure 14, for example, instructs the user to avoid using any apostrophes, accents or punctuation - especially if the meaning can be understood from the context.

5.3.1 Exemplifications

Key instructions sometimes also contain examples with plaintext and its corresponding ciphertext to clarify the application of the key, see Figure 15.

One of the keys with instructions we analyzed is a type of polyalphabetic homophonic cipher stored at the Haus-, Hof- und Staatsarchiv in Vienna, which utilizes a system of multiple encryption tables used to encrypt messages. (Key ID-1601, Year unknown) The key's cipher is unique in that it involves a special code element, termed "numeris indicans Tabellam", which signals a change in the encryption or decryption table being used. Only after reading and understanding the instruction of the key and the provided exemplifications, were we able to understand how the cipher works.

These examples offer insight into how scribes were taught to apply ciphers.

6 Discussion

In this study, we analyzed, to the best of our knowledge, the hitherto largest sample of cipher keys containing instructions. In order to do so, we manually labeled and described the instructions, without relying on pre-defined categories, or any existing typology. This paper presents our initial attempt to characterize the nature of instructions within cipher keys.

The scope of our study is of course constrained by the limitations of our data sample. Including more diverse sources from various locations and time periods would undoubtedly enhance the richness of our study.

Analysing the key sample, we found big overlaps in the use of language in the cipher key in-

Figure 16: A note attached to a cipher key (Key ID-2044, 1672)

structions. Oftentimes, in groups of keys from the same place and time period, similar expressions and sentences were applied to describe certain operations. A case in point is the sample of cipher keys from eighteenth-century Saxony. This fact might be interpreted as a sign that instructions were copied between different cipher keys. Similarly, the consistent choice of Latin for certain types of instructions also in Early Modern times (e.g., for nullities), see (Waldispühl and Megyesi, 2024), might be explained by the transfer of knowledge between creators of cipher keys and a preference for formulaic language in the genre of instructions.

Noteworthy is the fact that keys often are accompanied by notes written on separate pieces of paper. These notes often document some kind of metadata associated with the key, such as the corresponding entities or places. Examples of a note detailing the key's users are depicted in Figures 16 and 17. However, we do not treat these types of notes as instructions as they do not convey any information about the operational use of the keys.

7 Conclusion

Historians of cryptography have a wealth of knowledge about how specific ciphers were supposed to be used, and through the instructions they can gain insight into how they were actually used. Repeated advice and guidelines about what should and should not be done inform us of the fears of the cipher's constructors, who worried about the ignorance or negligence of users. In the history of cryptography, knowledge transfer was primarily

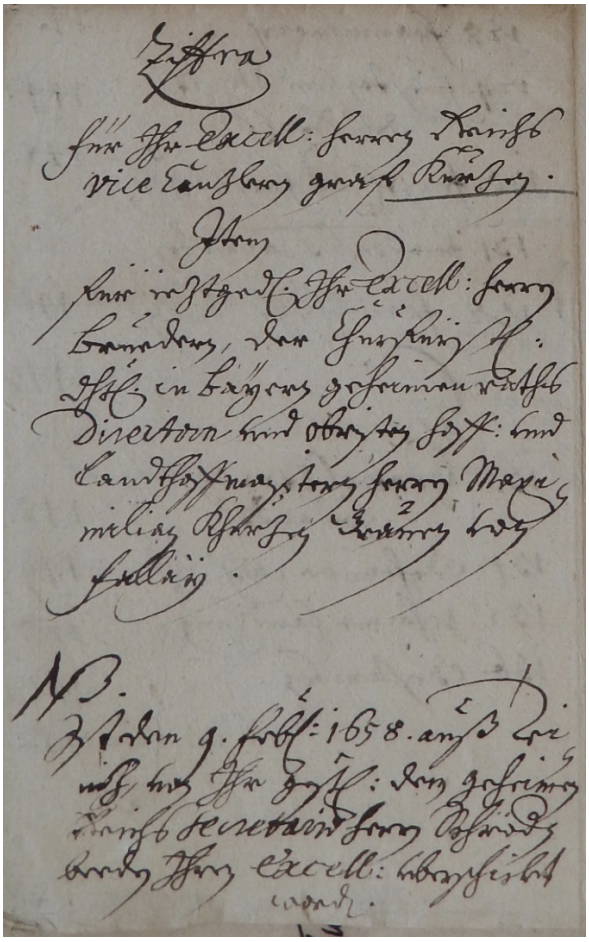


Figure 17: A note attached to a cipher key with information about the users (Key ID-1266, 1658)

oral, which we no longer have access to. Thanks to the instructions and the accompanying sample encryptions appearing in shorter or longer form, we can gain a limited but still relevant insight into how this knowledge was taught and acquired, the usual frustrations and worries of the constructors, and the mistakes of the users. To put it shortly: we get a glimpse of everyday practical cryptography.

In this work, we investigated 235 historical cipher keys that included some kind of instructions. We can conclude that ca 15% of our sample of cipher keys, all originating from early modern times in Europe, actually contained some kind of instruction with explanation about the operational application of the keys. We categorized the instructions and presented a typology based on their content. We distinguish between instructions containing structural information about the key table, information about various groups of code elements and their functions, and longer or shorter sequences of explanations about the key to help

the encrypter or decrypter to apply the key as intended by the creator of the key.

What more can we do to gain more insights into the practical usage of the cipher keys from the instructions? An obvious research direction is to widen the time period including the late 19th and early 20th centuries and/or to collect sources from a wider geographic area (including Spain, or the United States to mention a few).

A second line of attack - planned to be carried out by the present authors - is to analyse the content of the 1-2 page-long instructions more into detail. This would include text editions in the most typical languages (Latin, Italian, German, French and English), with examples in the DECODE database.

Text editions will prepare the ground for a third research direction, namely to analyse in detail, how the mostly anonymous instructions take over text elements and repeat messages from the late medieval and early modern cryptographers' advice. These would include the shorter lists of instructions by the Italian cipher designers such as Alberti, the Argentis, the Amadis, and Simonetta, as well as the thick handbooks by Della Porta, Vigenère and others.

Acknowledgments

We are wholeheartedly grateful to Paolo Bonavoglia, Camille Desenclos, George Lasry, and Klaus Schmeh for valuable input about cipher key instructions and the anonymous reviewers for helpful comments on our submission draft.

This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Camille Desenclos. 2023. La cryptographie, langue universelle de la diplomatie? In *Talk at "Apparati, tecniche, oggetti dell'agire diplomatico (secc. XIV-XIX)" workshop*. German Historical Institute in Rome.

Mihály Héder and Beáta Megyesi. 2022. The DECODE Database of Historical Ciphers and Keys: Version 2. In *Proceedings of the 5th International Conference on Historical Cryptology, HistoCrypt22*.

David Kahn. 1996. *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*. Scribner, New York, NY.

Key ID-1266. 1658. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 13. Fasc. 20. f 168., DECODE ID 1266, link: https://de-crypt.org/decrypt-web/RecordsView/1266.

Key ID-1295. 1500–1699. Reproduced image from Österreichisches Staatsarchiv, Haus-Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 13. Fasc. 20. f 217., DECODE ID 1295, link: https://de-crypt.org/decrypt-web/RecordsView/1295.

Key ID-1601. Year unknown. Key form the Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 15. Fasc. 21. f 125-137., DECODE ID 1601, link: https://de-crypt.org/decrypt-web/RecordsView/1601.

Key ID-1704. 1764. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 16. Fasc. 23. iv f 51-56., DECODE ID 1704, link: https://de-crypt.org/decrypt-web/RecordsView/1704.

Key ID-1722. 1790. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 17. Fasc. 24. f 105-112., DECODE ID 1722, link: https://de-crypt.org/decrypt-web/RecordsView/1722.

Key ID-1728. 1805. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Staatskanzlei Interiora, Chiffrenschlüssel, Kt. 17. Fasc. 24. f 140-147., DECODE ID 1728, link: https://de-crypt.org/decrypt-web/RecordsView/1728.

Key ID-2044. 1672. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Österreichische Geheime Staatsregistratur (Rep. N) 61 Fasc. 46. Pars. 6. f 31., DECODE ID 2044, link: https://de-crypt.org/decrypt-web/RecordsView/2044.

Key ID-2061. 1666. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Österreichische Geheime Staatsregistratur (Rep. N) 61 Fasc. 46. Pars. 6. f 45., DECODE ID 2061, link: https://de-crypt.org/decrypt-web/RecordsView/2061.

Key ID-2330. year unknown. Reproduced image from the Saxon State Archive ,Great Cipher of Saxony, HStAD, 10024, Loc. 8236/11, f 121, DECODE ID 2330, link: https://de-crypt.org/decrypt-web/RecordsView/2330.

Key ID-415. 1628. Reproduced image from the National Archives in Kew, State Papers, inv.nr. 106 box 5. , DECODE ID 415, link: https://de-crypt.org/decrypt-web/RecordsView/415.

Key ID-531. 1731. Reproduced image from The National Archives in Kew (UK),State Papers, inv.nr. 106 box 7., DECODE ID 531, link: https://decrypt.org/decrypt-web/RecordsView/531.

Key ID-613. 1703–1711. Reproduced image from the National Archives of Hungary, G15 Caps. C. Fasc. 43. 59., DECODE ID 613, link: https://decrypt.org/decrypt-web/RecordsView/613.

Key ID-670. 1660–1690. Reproduced image from National Archives of Hungary, P1238 Mihály Teleki Collection. Miscellaneous documents. Cipher keys. 10., DECODE ID 670, link: https://decrypt.org/decrypt-web/RecordsView/670.

Key ID-676. 1664–1668. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Ung. Akt. Spec. Verschwörerakten VII. Varia, Fasc. 327. Konv. D. Chiffres 1664-1668. fol. 1-2., DECODE ID 676, link: https://decrypt.org/decrypt-web/RecordsView/676.

Key ID-700. 1664–1668. Reproduced image from Österreichisches Staatsarchiv, Haus-, Hof- und Staatsarchiv, Ung. Akt. Spec. Verschwörerakten VII. Varia, Fasc. 327. D. Chiffres 1664-1668. fol. 56-57. , DECODE ID 700, link: https://decrypt.org/decrypt-web/RecordsView/700.

Key ID-935. 1761. Reproduced image from the Saxon Main State Archive Dresden, Calenberg-Saxony 1761, HStAD, 10024, Loc. 08236/11, Bl. 3, DECODE ID 935, link: https://de-crypt.org/decrypt-web/RecordsView/935.

Benedek Láng. 2018. *Real Life Cryptology: Ciphers and Secrets in early modern Hungary*. Atlantis Press, Amsterdam University Press.

Benedek Láng. 2022. Colonnele Frank's Indecipherable Chiffre. In *Proceedings of the 5th International Conference on Historical Cryptology, HistoCrypt22*.

George Lasry, Beáta Megyesi, and Nils Kopal. 2020. Deciphering Papal Ciphers from the 16th to the 18th Century. *Cryptologia*, pages 479–540.

Beáta Megyesi, Crina Tudor, Benedek Láng, Anna Lehofer, Nils Kopal, and Michelle Waldispühl. 2022. What Was Encoded in Historical Cipher Keys in the Early Modern Era? In *Proceedings of the 5th International Conference on Historical Cryptology, HistoCrypt22*.

Beáta Megyesi, Crina Tudor, Benedek Láng, Anna Lehofer, Nils Kopal, Karl de Leeuw, and Michelle Waldispühl. 2024. Keys with Nomenclatures in the Early Modern Europe. *Cryptologia*, 48(2):97–139.

Aloys Meister. 1902. *Die Anfänge der modernen diplomatischen Geheimschrift.* Paderborn: Ferdinand Schöningh.

Aloys Meister. 1906. *Die Geheimschrift im Dienste der Päpstlichen Kurie von Ihren Anfängen bis zum Ende des XVI. Jahrhunderts*, volume 11. F. Schöningh.

Ludwig von Rockinger. 1892. Über eine bayerische Sammlung von Schlüsseln zu Geheimschriften des sechzehnten Jahrhunderts. *Archivalische Zeitschrift*, pages 18–92.

Michelle Waldispühl and Beáta Megyesi. 2024. Language choice in Eighteenth-century diplomatic ciphers from Europe. In Vladislav Rjáoutski and Gleb Kazakov, editors, *Languages of Diplomacy in the Eighteenth-Century World*, page in press, Amsterdam. Amsterdam University Press.

Cryptanalysis of Hagelin M-209 Cipher Machine with Artificial Neural Networks: A Known-Plaintext Attack

Vasily Mikhalev[1*], Nils Kopal[1], Bernhard Esslinger[1], Harald Lampesberger[2], Eckehard Hermann[2]
[1]University of Siegen, Germany
[2]University of Applied Sciences Upper Austria, Hagenberg, Austria
[*]vasily.mikhalev@gmail.com

Abstract

This paper introduces a machine learning (ML) approach for cryptanalysis of the cipher machine Hagelin M-209[1]. For recovering the part of the secret key, represented by the wheel pins, we use Artificial Neural Networks (ANN) which take as input the pseudo-random displacement values generated by the internal mechanism of the machine. The displacement values can be easily obtained when ciphertext and plaintext are known. In particular, we are using several distinct ANNs, each recovering exactly one pin. Thus, to recover all the 131 pins, we utilize 131 models each solving a binary classification problem. By experimenting with various ANN architectures and ciphertext lengths, ranging from 52 to 200 characters, we identified an ANN architecture that outperforms others in accuracy. This model, inspired by the architecture by Gohr used for attacking modern ciphers, achieved the following accuracies in recovering the pins of the first wheel of the machine: approximately 71% for 52-characters sequences, 88% for 104-characters, 96% for 200-characters. The first wheel has the largest size and hence represents the most complicated case. For the other wheels, these accuracies are slightly higher. To the best of our knowledge, this is the first time when ANNs are used in a key-recovery attack against such machines.

[1]The code used for this paper and instructions of how to download the data are available at `https://github.com/CrypToolProject/M209KnownPlaintextAttackML`

1 Introduction

The Hagelin M-209 is a mechanical cipher machine designed by Boris Hagelin and used by the United States military extensively during World War II. This device encapsulates the era's challenges and advancements in secure communications. The M-209, based on a stream cipher mechanism, encrypts messages using several wheels and bars whose interaction produces pseudo random numbers that are combined with plaintext for encryption.

Despite its historical significance, the M-209 has been the subject of various cryptanalytic efforts, showcasing vulnerabilities typical for mechanical encryption devices. ML presents new opportunities for cryptanalysis of such historical ciphers. This paper explores the application of ANNs in a known-plaintext attack (KPA) scenario, where the plaintext and corresponding ciphertext are known, enabling the study of the machine's internal key generation mechanism.

Our research aims to demonstrate the efficacy of using ANNs to cryptanalyze the Hagelin M-209. By treating the sequences of pseudo random numbers generated by the machine as data inputs for ML models, we train each of them to recover only one specific bit of the key represented by the wheel pin. This approach establishes a novel technique in cryptanalysis that could be extended beyond the M-209 to other cipher machines.

The remainder of this paper is organized as follows: Section 2 reviews related work in the field of attacks against M-209. Section 3 gives a short description of the machine. Section 4 details our methodology, including data preparation and neural network architectures used. Section 5 presents the results of our experiments. Finally, Section 6 concludes the paper with a summary of our findings and potential future research directions.

2 Related Work

Modern cryptanalytic efforts on the Hagelin M-209 used a variety of methodologies to break its cipher mechanism.

Morris's approach (Morris, 1978) involved a manual KPA. By analyzing and refining the displacement sequence, Morris's method allows to recovery of the internal key from 100-character-long messages and corresponding ciphertexts. Barker (Barker, 1977) proposed a ciphertext-only attack (COA) based on analysis of letter frequency distribution patterns. It requires from 2000 to 4000 letters to recover the key. In the work (Beker and Piper, 1982) a COA was presented using techniques to classify pins and solve ambiguities. Their method is claimed to be effective if 2500 characters are available. Sullivan's approach (Sullivan, 2002) for ciphertext-only recovery of M-209 key used a divide-and-conquer approach, incrementally recovering pin and lug settings, requiring 2500 letters to obtain the key.

Lasry, Kopal, and Wacker made notable advancements in applying heuristic algorithms to the cryptanalysis of M-209. Their initial contribution was an automated approach for known-plaintext cryptanalysis of short messages, having only 50 characters, in (Lasry et al., 2016a). They later proposed a COA in (Lasry et al., 2016b), which was improved in (Lasry et al., 2018), enabling the recovery of the machine's internal settings from approximately 500 characters of ciphertext.

Most recently, the exploration of ML techniques in the classification of World War II era ciphers, including the M-209 was done by Dalton and Stamp (Dalton and Stamp, 2023).

3 The Hagelin M-209 Machine

M-209 operates as a stream cipher[2]. The main component is a pseudo-random keystream generator (KSG) which is used to produce sequences of displacement values, which modify the plaintext to generate the ciphertext, as illustrated by Eq 1:

$$c_i = (25 - p_i + d_i) \; mod \; 26 \qquad (1)$$

In this equation, p_i and c_i represent the positions of the plaintext and ciphertext characters in the Latin alphabet[3], respectively, while d_i is the displacement value generated for encryption of the i-th character of the plaintext.

The KSG comprises a 27-bar cage and six wheels with varying numbers of letters (26, 25, 23, 21, 19, 17). Each letter on a wheel has a pin that can be set into active or inactive state. Every bar in the cage has two adjustable lugs, which can be set against any of the six wheels or remain in a neutral (zero) position. When both lugs on a bar are placed in an active position, it results in a *lug overlap* involving the bar and two wheels. The number of total overlaps is an important property for analysis of the key complexity as shown in the Section 5,

With each encryption step, the cage makes the full rotation around its 27 bars. If a bar lug interacts with an active wheel pin, the bar shifts to the left. The displacement value d_i is determined by the number of shifted bars and can range from 0 to 27. Before the next encryption cycle, all wheels advance by one position.

The machine's secret key consists of the initial wheel positions, the pin settings (active/inactive), and the lug positions. These configurations are not arbitrary and followed the specific operating instructions, which were changing over time. This study focuses on the keys that correspond to 1944 operating instructions (War Department, 1944).

4 Methodology

This study focuses on the KPA scenario which allows to reverse Equation 1 to derive sequences of displacement values, which are utilized as inputs for our ANN. The ANN's goal is to predict M-209 pin values from a sequence of displacement values[4]. This problem can also be regarded as a binary classification problem. Instead of predicting all 131 pins at once with a single model, we consider each pin as an individual target. Therefore, 131 binary classifier models that share the same ANN architecture are used to predict the states of all pins.

If different keys used in a cipher for encrypting the same plaintext lead to the same ciphertext, these keys are called equivalent. An important M-209 property used in our methodology is that for any initial position of the wheels, there are pin settings that result in the equivalent keys (Lasry

[2]We only discuss the most important parts of the machine, directing readers to Lasry's PhD thesis (Lasry, 2018) for detailed explanations and to Chapter 2.5 in (Esslinger, 2024) for additional Hagelin Machine models.

[3]A is encoded by 0, B by 1 etc..

[4]The ANN doesn't get information about the lugs settings.

et al., 2018). This property is crucial as it implies that we can assume any initial position of the wheels, such as the default position "AAAAAA", and find proper pin configurations which constitute an equivalent key to the one used during the actual encryption process. When such an equivalent key is found, we can decrypt the remainder of the ciphertext without the need for plaintext.

For training the models, we generated millions of random keys following the operational instructions provided in the 1944 technical manual. Utilizing these keys, we created sequences of displacement values which are 52, 104, and 200 numbers long. The choice of length 52 ensures that the first wheel completes two full rotations, such that every pin contributed at least twice in the displacement generation process. Similarly, length 104 corresponds to four full rotations. In experiments, this approach led to a slight but noticeable improvement in accuracy compared to sequences of 50 and 100 characters. Extending the sequence length to 208 did not bring a notable difference in the results but challenges in the training process; therefore, a maximum length of 200 was considered in the experiments.

We conducted experiments with several ANN architectures using a range of hyperparameters, including Feedforward Neural Networks, LSTM, and Transformer models. The best results were obtained using a residual-networks-like architecture, similar to the one employed by Gohr in his work on cryptanalysis of modern ciphers through ML (Gohr, 2019). This specific architecture incorporates a dual-layer convolutional residual tower (DLCRT), having a dense-prediction layer as shown in the Figure 1.

Residual networks (ResNets) use shortcut connections, allowing data to skip layers and directly connect to deeper layers. This addresses the so-called vanishing gradient problem and enables training of deeper networks efficiently.

The implemented DLCRT model processes input displacement values by first normalizing them through division by 25 during preprocessing. These values are then fed into the initial convolution layer, which is succeeded by five residual blocks, each comprising two convolutional layers. Subsequently, a dense layer, in which every input is connected to every output, transmits the values to the output layer which is responsible for generating the final prediction for a given pin.

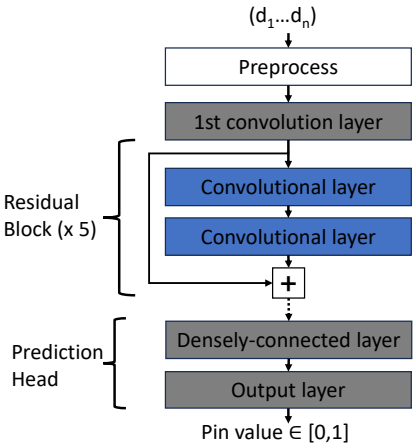


Figure 1: Used DLCRT architecture

5 Results

We now present the results[5] of testing our trained models with the randomly generated data, distinct from the data utilized during the training phase. For brevity, our discussion centers on the scenario of reconstructing the pins of the first wheel, which comprises 26 pins. This wheel is the largest among its counterparts, thus posing the most challenging scenario. The results for the other wheels are marginally better.

For testing, we generated 1000 random keys for each of the key classes based on their complexity which depends on 2 parameters. The first one is the number of lugs not involved in the lug overlap (non-shared lugs) positioned against the wheel, the pins of which are being recovered. The second one is the number of overlaps with the other wheels. To get an intuition on the significance of the number of non-shared lugs, consider a scenario where a wheel is set against 13 lugs. In this case, if the total displacement value is between[6] 15 and 25, it guarantees that the current pin of this wheel is active; otherwise, it's inactive. Such a property is easy to detect for an ANN, as opposed to situations where only one or two lugs are set against a wheel, which introduces a greater level of uncertainty.

For a single test, accuracy is defined as the ratio of correctly identified pins out of the total 26.

The accuracy distributions for sequence lengths n=52, n=104 and n=200 are depicted in figures: Figure 2, Figure 3 and Figure 4 respectively. In

[5]Initial measurements were done in (Landrichinger and Mikhalev, 2023)

[6]The values 26 and 27 are reduced modulo 26 in accordance with the Equation 1.

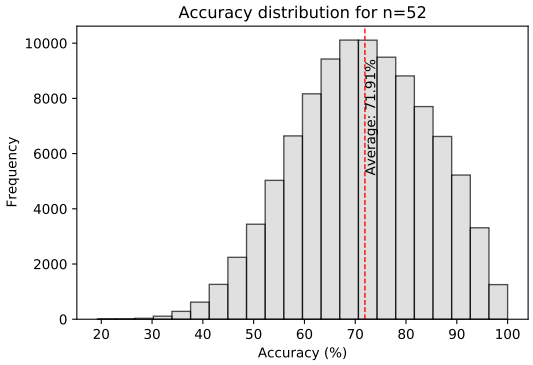


Figure 2: Accuracy distribution for displacement sequence length 52 (n= 52)

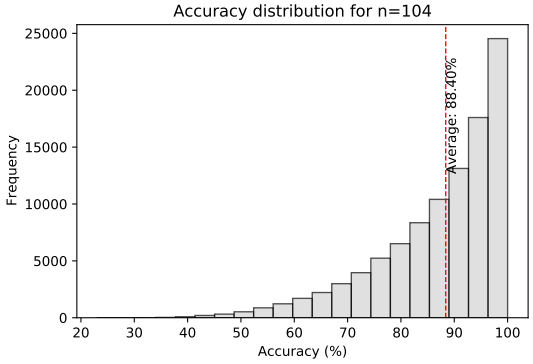


Figure 3: Accuracy distribution for displacement sequence length 104 (n= 104)

terms of average accuracy, sequences of 52 characters yield 71%, those with 104 characters reach 88%, and for sequences encompassing 200 characters, the accuracy stands at 96%. This behavior was expected because longer displacement sequences carry more information and the average accuracy therefore increases.

Additionally, our study includes an examination of the ANN's performance relative to the varying complexities of the keys.

The results related to the most difficult scenario, the 52-letter sequences, are shown in Table 1. These results suggest that in simple cases when a key has many non-shared lugs, the ANN can accurately recover the pins using only 52 characters of ciphertext and corresponding plaintext.

6 Conclusion and Future Work

This study showcased the potential of ANNs in cryptanalysis of the Hagelin M-209 cipher machine using a KPA. Our experiments with various ANN architectures, particularly a custom ResNet-like model, highlight the effectiveness of ML in the cryptanalysis of such machines.

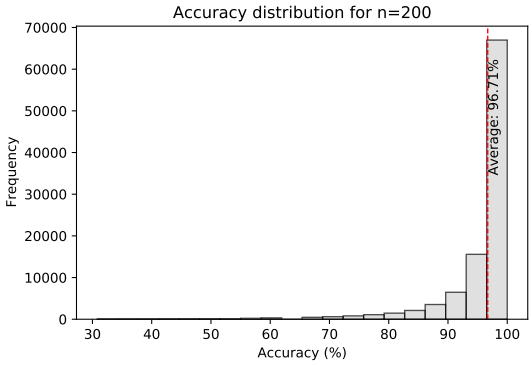


Figure 4: Accuracy distribution for displacement sequence length 200 (n= 200)

N\O	0	1	2	3	4	5	6	7	8	9	10	11	12
0	-	51	51	53	55	57	58	59	61	63	65	65	68
1	54	54	55	57	59	61	61	63	65	66	68	69	71
2	55	57	59	60	62	64	64	66	68	69	70	72	-
3	59	60	62	64	66	67	68	70	71	72	74	-	-
4	63	64	66	68	70	72	71	72	74	75	-	-	-
5	66	69	70	72	73	74	74	76	78	-	-	-	-
6	69	71	73	75	76	78	79	81	-	-	-	-	-
7	73	75	76	79	82	82	83	-	-	-	-	-	-
8	76	78	81	83	85	86	-	-	-	-	-	-	-
9	81	83	85	87	89	-	-	-	-	-	-	-	-
10	85	87	89	90	-	-	-	-	-	-	-	-	-
11	88	90	92	-	-	-	-	-	-	-	-	-	-
12	91	93	-	-	-	-	-	-	-	-	-	-	-
13	93	-	-	-	-	-	-	-	-	-	-	-	-

Table 1: Mean accuracy in % for the different number of non-shared lugs (represented by rows) and overlaps (represented by columns). These results are for sequences of 52 characters long.

The next steps of our research will include:

- Enhancing the accuracy of pin recovery, especially in shorter sequences, and designing the techniques to improve the errors.
- Recovery of the bar lugs.
- Investigating ciphertext-only cryptanalysis approaches.
- Applying ML to other cipher machines.

Acknowledgments

We are grateful to George Lasry who proposed the idea of this attack and for his valuable comments. This work was supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Wayne G. Barker. 1977. *Cryptanalysis of the Hagelin Cryptograph*, volume 17. Aegean Park Press, Laguna Hills, CA.

Henry Beker and Fred Piper. 1982. *Cipher Systems: The Protection of Communications*. Northwood Books, London.

Brooke Dalton and Mark Stamp. 2023. Classifying World War II Era Ciphers with Machine Learning. *arXiv preprint arXiv:2307.00501*.

Bernhard Esslinger. 2024. *Learning and Experiencing Cryptography with CrypTool and SageMath*. Artech House, Norwood. https://us.artechhouse.com/Learning-and-Experiencing-Cryptography-with-CrypTool-and-SageMath-P2378.aspx.

Aron Gohr. 2019. Improving attacks on round-reduced speck32/64 using deep learning. In *Advances in Cryptology–CRYPTO 2019: 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2019, Proceedings, Part II 39*, pages 150–179. Springer.

Robert Landrichinger and Vasily Mikhalev. 2023. Cryptanalysis of Cipher Machines with the Help of Artificial Neural Networks. Talk at Security Forum, Hagenberg im Mühlkreis, Austria. Retrieved from https://www.securityforum.at/cryptanalysis-of-cipher-machines-with-the-help-of-artificial-neural-networks-tag-2. Accessed: 25 January, 2024.

George Lasry, Nils Kopal, and Arno Wacker. 2016a. Automated Known-Plaintext Cryptanalysis of Short Hagelin M-209 Messages. *Cryptologia*, 40(1):49–69.

George Lasry, Nils Kopal, and Arno Wacker. 2016b. Ciphertext-only cryptanalysis of Hagelin M-209 pins and lugs. *Cryptologia*, 40(2):141–176.

George Lasry, Nils Kopal, and Arno Wacker. 2018. Ciphertext-only cryptanalysis of short Hagelin M-209 ciphertexts. *Cryptologia*, 42(6):485–513.

George Lasry. 2018. A Methodology for the Cryptanalysis of Classical Ciphers with Search Metaheuristics.

Robert Morris. 1978. The Hagelin Cipher Machine (M-209) Reconstruction of the Internal Settings. *Cryptologia*, 2(3):267–289.

Geoff Sullivan. 2002. Cryptanalysis of Hagelin Machine Pin Wheels. *Cryptologia*, 26(4):257–273.

War Department. 1944. TM-11-380, Technical Manual, Converter M-209, M-209A, M-209B (Cipher). https://deweger.net/apparaten/downloads/M209%20manual.pdf. Accessed: 25 January, 2024.

Bringing Cryptology into the Secondary Education Classroom

Catherine Murphy
University of Portsmouth
Winston Churchill Avenue
Portsmouth
PO1 2UP
catherine.murphy@port.ac.uk

Aaron Wootton
University of Portland
5000 N. Willamette Blvd
Portland, OR 97203
USA
wootton@up.edu

Abstract

Cryptology is becoming increasingly commonplace in undergraduate mathematics curricula as a way to motivate abstract mathematics. However, it is still typically absent in secondary education (students aged 11 to 18). In the following, we discuss why we think it would be advantageous to bring cryptology, both historical and modern, to the secondary education classroom. Additionally, we discuss some of the barriers we perceive to doing so and suggest, in our opinion, how they might be overcome. We illustrate implementation with a specific example of a topic in cryptology that could be included in a secondary education classroom.

1 Introduction

How does one motivate a group of recalcitrant teenagers to learn abstract mathematics on a wet Friday afternoon in January? Students need a 'hook', something that will engage and challenge their thinking. A concept that is relevant both to their lives and to their future careers.

In our experience, mathematics is viewed by some students as dry and boring, a subject with little real-life relevance and one which must be endured rather than enjoyed. However, it is fundamental to most (if not all) adults in the modern world. Without a grasp of mathematical systems, we would find ourselves unable to navigate the vista of technological advancement we find ourselves facing.

With its rich history and modern utility, we believe cryptology is an ideal vehicle for engaging students in complex mathematical problems. It is already used, with increasing regularity, in the undergraduate curriculum. There have been attempts to mimic this in secondary education, (Caballero-Gil & Bruno-Castaneda, 2006), and evidence suggests its use improves student perception of mathematics and assessed performance (Özdemir, Güler & Aydın, 2011). Additionally, both authors have experience using cryptology to engage mathematics students, including the topic described below, and are convinced of its efficacy.

Unfortunately, the idea of using cryptology in secondary education has gained limited traction. So, what are the underlying issues and how can they be overcome?

2 Cryptology and mathematics

The overarching idea and practice of cryptology can be understood by suppressing the technical details in a black box process. Specifically, cryptology is about securely sharing information, for example, military strategy or financial data. A process, illustrated by the black box in Figure 1, is used to translate this information into an unintelligible piece of text, or ciphertext, which is then transmitted. Intended recipients have the tools necessary to recover the original information, but anyone else who intercepts the communication does not (at least in principle).

Figure 1: Black Box Process

It is in this black box that cryptology and mathematics are intertwined. Every cryptosystem, from the Caesar wheel to the RSA algorithm, can be described and analysed using mathematical principles. For example, the classical Caesar cipher, which translates the Roman alphabet forward by three letters, can be described using modular arith-

metic. Specifically, we assign both plaintext and ciphertext letters “A” to “Z” to the numbers 0 to 25 respectively. To encipher a plaintext letter, we add 3 to the number corresponding to that letter with the convention if the result is 26 or larger, we subtract 26 creating a wrap around, see Figure 2. An alternate correspondence often used in classrooms is to assign “A” to “Z” to the numbers 1 to 26, their position in the alphabet, (Lewand, 2004), with the mechanics of encryption being the same.

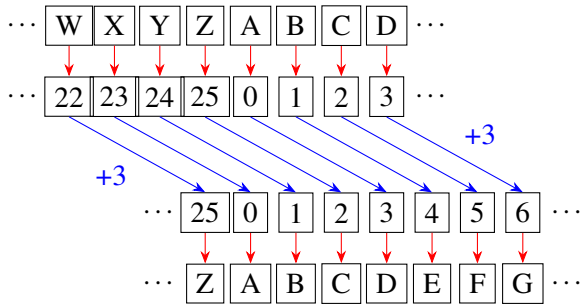


Figure 2: Caesar Cipher and Modular Arithmetic

Though significantly more complicated, the black box underlying many other cryptosystems can also be translated into some mathematical process on the integers to some base. This makes cryptology an ideal tool to motivate mathematics in the undergraduate curriculum. Modular arithmetic is a standard topic covered early in a degree program, and there are an increasing number of cryptology teaching resources for university faculty, see for example (Boersma, Christensen & Millichap, 2023). Unfortunately, this is also a significant barrier to cryptology being included in secondary education: modular arithmetic is not typically taught in secondary education.

3 Cryptology as a tool to motivate mathematics

Some of the motivations for using cryptology as a tool to enhance the teaching of mathematics, in our opinion, are given below.

First, it is about sparking interest in the subject. People of all generations continue to be fascinated with the idea of secret sharing and are well aware of its historical importance. For example, though highly questionable in their historical accuracy, movies such as U571 and The Imitation Game, both of which portrayed the Allies breaking the Enigma machine, were box-office hits. More generally, espionage is exciting, with James Bond and Mission Impossible films still holding box-office thrall. Furthermore, ‘escape rooms,’ which utilise encryption strategies, have become a pastime of the youth. In a nutshell, cryptology is sexy, and the considerable historical and political importance of cryptology gives it a frisson of excitement. So while some aspects of mathematics can appear dry, or indeed to have little purpose other than the mathematics itself, embedding the mathematics in a cryptological puzzle helps students to, not only see purpose, but also engage them in the mathematics.

A second reason is that the study of cryptology can be used to develop and link problem-solving skills and technological prowess, both of which are increasingly applicable to modern life. Moreover, understanding the digital world is ever more vital as a report by Dell Technologies (Dell Technologies, 2018) highlighted, suggesting that we are on the cusp of a revolution in terms of how we work with technology. The roles our students will play in the working world will almost certainly involve the use of computers; how our data is protected and shared may be central to their careers. Additionally, as responsible citizens, we should know the questions to ask about how our own (and our family’s) personal data (financial, medical etc.) is stored, protected, and shared.

A third reason is that there is the advantage of the multi-disciplinary nature of learning mathematical concepts through cryptology. Discussions can be had on the history, politics, language, and structure behind any cryptological technique. Patterns and algorithms thus cease to be the unique domain of the mathematics classroom: instead students are encouraged to view, through the lens of mathematics, other curriculum areas.

Finally cryptology can be taught using methods that are supported by the pedagogical literature. For example, students learn effectively when posing their own questions and finding solutions (Calder, 2013). Challenges offered in the form of cryptological puzzles offer the opportunity for students to do just this; question, challenge, reflect, and discuss. Additionally, there is the element of gamification. Solving puzzles can feel like a game, and students are increasingly familiar with this method of learning through the advent of new technology. While cryptology does not necessarily automatically offer all the underlying dynamics of successful games-based learn-

ing (Stott, 2013), there are nonetheless aspects of this (progression and storytelling) that are naturally present. By approaching a concept in mathematics as a challenge, a puzzle to solve, students are likely to become engaged in their own learning, thus neatly sidestepping the problem of the recalcitrant teenagers attempting to study abstract mathematics on a wet Friday afternoon in January.

4 Barriers to cryptology in secondary education and how can they be overcome

There are many convincing arguments for incorporating cryptology into secondary mathematics education when viewed through the lens of mathematics educators. However, in our opinion, there are also significant barriers to doing so. The following we consider to be the most noteworthy.

The confidence (and qualifications) of some mathematics teaching staff may be an underlying issue. A report by the United Kingdom (UK) Government in 2023 on the school workforce in England showed that 12.8 percent of the total hours of mathematics taught in the secondary classroom was by teachers without suitable degree-level qualifications. Additionally, mathematics was one of the three subjects with the highest number of vacant teaching positions in the UK (Department for Education, 2023). This situation is not unique to the UK; a brief look at the mathematics teaching jobs available globally tells the sad story of the lack of mathematics teachers around the world. While the lack of suitable degree-level qualifications does not preclude teachers from being highly competent mathematicians, it is more likely that some may lack confidence and content knowledge. As a result, it may be less likely that a teacher will stray from a traditional mathematics course as set out by publishers, schools, and colleges.

Additionally, much of the underlying mathematics behind cryptology is not covered in the secondary education curriculum. This means students do not have the preparatory material to understand many cryptosystems. In addition, teachers without suitable degree-level qualifications may not have had exposure to this type of mathematics.

Another barrier is the pressure teachers feel to 'teach to an exam'. Measures such as PISA and other international comparisons, as well as school and area-level comparisons and league tables within countries, look simply at results and do not consider the journey to knowledge. Teachers are often so anxious about their requirement to perform that there is little space left for autonomy (Ball, 2003). The idea of teaching a mathematical concept through cryptology may be perceived as inefficient and untried and, therefore, too risky to contemplate.

How can these barriers be overcome? In our opinion, they may, in part, be overcome through the creation of carefully curated resources that educators can utilise in their classrooms. These resources should be available to secondary education students and educators. Additionally, we believe it is essential that they focus on cryptosystems that make sense in classical number systems, such as the integers or real numbers. By restricting to classical number systems, less confident mathematics teachers would be working within comfortable bounds of knowledge and so, we hope, would feel more comfortable using these resources. These resources should also provide clear links between curricula and exam specifications which would allow teachers to see how this material will help prepare their students for exams and beyond.

A drawback to restricting to cryptosystems defined over classical number systems is that for many cryptosystems, the underlying security will be lost. However, the goal here is not to train teenagers to be experts in computer security, but rather to spark their interest in mathematics through real-life applications. Students will still understand the concept of cryptosystems: all that has been modified is the black box process. Specifically, for each cryptosystem studied, students will learn to encrypt and decrypt within the system enhancing both their understanding of the underlying mathematical concept, and the cryptosystem itself.

5 A sample topic for the classroom

Shamir's secret sharing scheme (SSS) is a perfect example of a cryptosystem that preserves the underlying cryptographic structure over a classical number system and can be adapted to motivate and assess key learning objectives on a standard topic in secondary education: linear equations. In this instance, the learning outcomes would include recognising that a straight line is uniquely determined by two points, finding its gradient (or

slope), and finding the y-intercept.

For motivation, SSS can be adapted to a contemporary espionage problem like the following:

"MI5 wish to secure the combination to a safe so that no spy can access it alone, but any two can access it together. How can this be achieved?"

SSS stores a combination as the y-intercept c of a linear equation $y = mx + c$. 'Keys' are distributed to spies in the form of points on the line, one unique point for each spy. Since a linear equation is completely determined by two distinct points, see Figure 3, any two spies can use their keys to find the linear equation and reveal the y- intercept. However, an individual key, or point, alone does not reveal the linear equation.

For example, suppose two spies are given the points $(8, 262)$ and $(17, 361)$. Plotting these two points on a graph and drawing the unique line passing through them, see Figure 3, we see that the y-intercept is approximately 170.

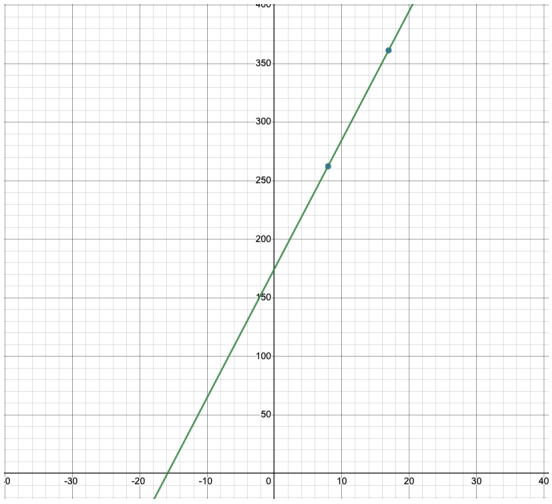


Figure 3: Line passing through two points

To find the exact combination, we need to find the linear equation. The equation of a straight line can always be given in the form $y = mx + c$, where m is the gradient and c is the y intercept. The gradient, or slope, is found by calculating the change in y with respect to x:

$$m = \frac{\text{change in } y}{\text{change in } x} = \frac{361 - 262}{17 - 8} = \frac{99}{9} = 11,$$

and so the linear equation is of the form $y = 11x + c$. To find the y-intercept c, we can use the fact that either of the points $(8, 262)$ and $(17, 361)$ lie on the line and so satisfy the equation. For example, using $(8, 262)$ we get,

$$\begin{aligned} y &= 11x + c \\ 262 &= 11 \cdot 8 + c \\ 262 - 11 \cdot 8 &= c \\ 174 &= c \end{aligned}$$

giving a combination of 174.

This example is a highly specific illustration of how SSS can be used to highlight features of a linear function. However, it also offers a springboard for further discussion. For example, what other pairs of parameters could be used rather than two points? How can SSS be extended to higher degree polynomials? SSS is a perfect vehicle for exploring these questions as it can be implemented with higher degree polynomials to create higher threshold systems. For example, a three threshold system, where three points are needed to recover the secret, can be created using a quadratic function. Such exploration can be aided with the use of the open-source software, JavaCrypTool, (CrypTool, 2023) where SSS can be implemented and visualised on polynomials of any degree.

6 Conclusion

In our opinion, bringing cryptology to the secondary education classroom could significantly enhance student enthusiasm for mathematics, which may lead to increased student engagement. It is not, of course, without barriers. However, we believe these are eminently overcomeable with sufficient development of thoughtful and well researched resources appropriately adapted to ensure the mathematical content is of a suitable level for secondary school students and teachers.

We believe that if such resources were made available to educators, it would draw together the rich history of cryptology with the modern application of problem-solving and mathematical skills. Thus, learners would not only develop the ability to tackle real-world problems of the 21st century, but do so with historical perspective.

References

Stephen J. Ball. 2003. The teacher's soul and the terrors of performativity, *Journal of Education Policy*, 18:2, 215–228. https://doi.org/fcvnnv

Stuart Boersma, Chris Christensen and Christian Millichap. 2024. Introduction to the Special Issue: Resources for Undergraduate Cryptology, *PRIMUS*, 34:1, 1-5. https://doi.org/k9b9

Nigel Calder. 2013. Mathematics in student-centred inquiry learning: Student engagement, *Teachers and Curriculum*, 13. https://doi.org/k64w

Pino Caballero-Gil and Carlos Bruno-Castañeda. 2006. A cryptological way of teaching mathematics, *Teaching Mathematics and its Applications*, 26. https://doi:10.1093/teamat/hrl008

CrypTool Portal. 2023. Cryptography for Everybody. https://rb.gy/1unnk1

Dell Technologies. 2018. *Realizing 2030 a Divided Vision of the Future.* https://rb.gy/rl9awk

Department for Education. 2023. *School Workforce in England.* https://rb.gy/gqgyl5

Robert E. Lewand. 2004. *Cryptological Mathematics.* Classroom resource materials, Mathematical Association of America. https://rb.gy/oahy28

Ahmet Özdemir, Enes Güler and Nuh Aydın. 2011. Effects of Cryptographic Activities on Understanding Modular Arithmetic, *Turkish Journal of Computer and Mathematics Education,* 2(3), 247–256. https://rb.gy/73ra38

Andrew Stott and Carmen Neustaedter. 2013. Analysis of gamification in education, [Unpublished manuscript] School of Interactive Arts and Technology, Simon Fraser University. https://rb.gy/r7kbx1

Musician Cryptologists: The Band of the USS *California* at Pearl Harbor and Beyond

Kyle Prescott, D.M.A.
Florida Atlantic University
Boca Raton, Florida, USA
Kpresco2@fau.edu

Abstract

Working from the basement of US Pacific Fleet Headquarters in Pearl Harbor from 1941, a small team of navy cryptanalysts and linguists known as the Combat Intelligence Unit (CIU) provided the US Pacific Fleet Command with timely details of the capabilities and intentions of the Japanese Imperial Navy (IJN) in the Pacific. A substantial portion of the CIU was comprised of 20 enlisted musicians of US Navy Unit Band 16, the band of the USS *California* (BB-44*),* who survived the sinking of their ship in the attack of December 7, 1941. The musicians were recruited to Combat Intelligence, retrained to perform tasks related to the deciphering of Japanese Naval Code JN-25b, and contributed to that unit's celebrated intelligence successes of 1942. As the war waged on, several musicians from Band 16 transferred to the Naval Communications Complex at Nebraska Avenue in Washington DC, and three eventually joined the National Security Agency (NSA) and served through the height of the Cold War.

1 Introduction

Band Unit 16 of the United States Navy, originally assigned to the battleship *USS California*, survived the attack on Pearl Harbor to become an important component in the effort that led to the breaking of the enciphered operational code of the Imperial Japanese Navy which provided the US Navy the opportunity to defeat the opposing fleet at the pivotal Battle of Midway. While USS *California's* Band 16 has been widely mentioned in previous literature, these references have been limited in scope, constrained to their work in the Combat Intelligence Unit in early 1942. This research describes the career arc of the bandsmen including their music training and shipboard life, experiences during the Battle of Pearl Harbor, their recruitment to Combat Intelligence and the attack on JN-25b, and notably work in intelligence later in the second world war, into the Cold War and the shifting priorities of the United States intelligence effort.

2 US Navy Unit Band 16

The 21-member band, led by Bandmaster Lovine "Red" Luckenbach, trained together at the newly formed US Navy School of Music (USNSOM) in Washington DC. This institution was formed in 1935 at the Navy Yard in Washington DC, and by 1939 was turning out highly regarded units for ships and shore installations. Bands at the school received rigorous training through a two-year course of musicianship, including music theory, solfège, and ensemble preparation in a wide variety of musical styles. Bands were formed up as single entities in their second year at the school to train together, deploy together, and perform together for two-year tours. The largest ships in the fleet were assigned 21-piece bands, as were most admirals afloat, while ensembles of 17 and nine pieces were assigned to smaller warships and troop transports, respectively (Jones, 2002). Admission to the school was by audition, generally taken at the age of 18 or 19. Prospective sailor-musicians travelled to the Washington Navy Yard to perform this live audition, and if admitted would immediate begin a brief 2-week course of basic training in Virginia, followed by admission to the school's curriculum (Bolen, 1976; Harding, 1980; Luckenbach, 1978).

After graduating from the USNSOM in May of 1941, Band Unit 16 was sent to San Diego, where they, along with two other newly graduated bands destined for the Pearl Harbor Naval Air Station and *USS West Virginia*,

boarded the Seaplane tender *USS Curtiss* for the territory of Hawaii, Pearl Harbor, and their new homes *(Muster Roll of the San Diego Receiving Station*, 1941). Once received aboard the *California*, the band began a surprisingly varied life. Navy bands of the time were cross trained at the Navy Yard school in a variety of musical styles, and each musician was expected to at double on at least one other instrument, such as trombonists playing violin, clarinetists playing oboe and saxophone, and trumpet players doubling French horn. This versatility gave the band the flexibility to meet their daily duties at sea and in port. The band had to be prepared to perform ceremonial music at one moment, then play on the march, perform patriotic selections, give formal concerts, and perform the popular dance music of the day for entertainment of the sailors (Jones, 2002). Bandmaster Luckenbach, a Navy Musician since 1929, had completed several tours aboard the USS *Saratoga*, as well as the Bandmaster School, a separate course at the Washington Navy Yard for band leaders and conductors. His 1939 rating of Bandmaster and the equivalent rank of Chief Petty Officer gave Red Luckenbach a bunk with the other Chiefs aboard ship (Luckenbach, 1978). As Bandmaster he arranged and composed music for concert performance, including a dramatic original work "Julius Caesar" and a march he would later title "Pearl Harbor Survivors' March" (Harding, 1988). He assembled star jazz soloists such as trumpet player Michael Palchefsy, whose Bunny Berigan-like sweet tone was well-admired, and jazz arranger Russell 'Sig' Shelley, a tenor saxophone player with the aural acumen to hear a Glenn Miller phonograph, then write out the parts for each player in a single sitting. With this talent, the band developed a reputation throughout the Pacific Fleet (Bolen, 1976).

2.1 Unit Band 16 Daily Life

Each morning Chief Luckenbach would meet with Executive Officer Earl E. Stone, future Director of the Armed Forces Security Agency (AFSA) (Manecki & Redacted, 2003) to determine the performance schedule for the band, such morning and evening colors from the fantail, ceremonial music for an Officer's Dining-in, swing dance music before and after the evening movie on the quarterdeck, a string quartet for a visiting Admiral, and a noon concert while the men ate lunch (Luckenbach, 1978). This adaptability was demonstrated in the equipment manifest of navy band units like number 16, including full sets of 34 string and wind instruments, and 1100 sets of music for concert band, orchestra, and jazz band (Jones, 2002). The entertainment of the sailors and officers, with the resulting improvement of morale, meant that the band spent substantial time performing as a swing/ dance band, including a fleet-wide competition among the bands of the Pacific, the 1941 Battle of Music (Harding, 1980; Luckenbach, 1978).

The Battle of Music was a series of competitions spread over the fall and winter of 1941, with an elimination round held every two weeks in the Bloch Arena, a new recreation center on the Pearl harbor base. Each competition included four bands performing a series of four selections for the audience. The event was organized by Chief Luckenbach, as the senior Bandmaster in the Pacific Fleet (Luckenbach, 1978).

Image 1: Band of the USS *California* at the 1941 'Battle of Music'. Bloch Arena, Pearl Harbor Navy Yard, Territory of Hawaii (NSA Library, Chiles/Weber Collection)

Battle of Music rounds were scheduled for November 22 and December 6, with the finals set for December 20. The November contest was won by Navy Band Unit 22 of the *USS Arizona*, led by Bandmaster Frederick Kinney. The *Arizona* band attended the contest in the Bloch Arena on the night of December 6 to hear the bandsmen of the *USS Tennessee*. Some attendees later recalled that the *Arizona* musicians played themselves that night, between rounds, though not as official contestants. The contest ended just after 1:00 am in the morning, when the crews left the arena

and headed for the liberty boats to take them back to their ships. It was the early hours of December 7, 1941.

3 The Battle of Pearl Harbor

The duty band of the *California* –one half of the ensemble –reported to the main deck aft at 7:30 am to perform, as they did each morning, the Star-Spangled Banner. By 7:50 am the group stood ready to play and watched the stern of the fleet flagship *USS Pennsylvania* for the signal to begin (Bolen, 1976).

Aboard the *California* Musician Second Class (Mus2c) Warren G. Harding, a trombonist from rural Indiana who was not on duty that morning, waited for his friend Mus2c Frank Wanat to catch the 7:30 am liberty boat for shore, where they were to meet friends of Harding's family for a traditional Hawaiian feast, or luau. Wanat, tired from the late night at Bloch Arena, was not ready in time, telling Harding they could catch the 8:30 am boat instead. Harding waited on the main deck in the beautiful tropical morning, admiring a graceful PBY flying boat taking off from the smooth waters. At 7:55, the band members heard aircraft of a different type approach. Harding first assumed that a movie was being filmed until the alarm rang out sending the ship the General Quarters (Harding, 1980). The bandsmen on deck left their instruments and headed below deck for various duty stations, including ammunition loading, first aid, and repair (Bolen, 1976).

Aboard the *Oklahoma*, moored outboard of *USS Maryland* along battleship row off Ford Island, the attack was swift and violent. Eight torpedoes opened the hull to too much water too quickly, and the great ship rolled over and capsized in 50 feet of water within ten minutes. Twelve members of the *Oklahoma* band did not survive (*Muster Rolls of USS Oklahoma, 1941)*.

Fifteen minutes after the attack began, a Japanese Aichi D3A dive bomber descended on the *USS Arizona*, releasing a converted 15-inch naval shell over the number two turret. The shell penetrated the armor deck of the *Arizona*, detonating in a magazine, instantly dooming the ship and 1,177 sailors aboard. Manning battle stations mere yards from the explosion, all 21 members of Navy Unit Band 22 perished in that moment. They would be remembered by their musician colleagues weeks later when, after the cancellation of the final round of the 1941 Battle of Music, the bands of the Pacific Fleet voted unanimously to award first prize to Unit Band 22, renaming it The Arizona Trophy. The cup resides now in the Pearl Harbor National Memorial.

Aboard the *California*, tuba and bass player Edgar Manley, a towering man, joined trombonist Harding four levels below the main deck, in a small room named 'Port Repair Four.' Here they sealed themselves in by dogging the watertight hatches fore and aft and waited. A torpedo hit near the bow twisted the ship's frame and brought the first signs of trouble as water began to seep in through the now wrenched door. Once the water reached their ankles, the men moved one room amidships, closing the hatches behind them, and waited. A second torpedo hit near the first, followed by a bomb at the midships hatch in the main deck. Harding plugged a headset into the communication system and asked for instructions. What he was told he chose not to repeat to the other men in the room. The man on the intercom informed Harding that the order had been given to abandon ship, but to preserve watertight integrity he was to remain at his station. "We will come back for you when we can," he was told. Manley had at this point passed out from exhaustion and the pain of two sprained ankles and was lying in a growing pool of oil and water now entering their new refuge. Working together, the men pulled Manley upright, and finally moved up through the ship to the main deck. Their sinking battleship was moored nearly 100 feet from the shore, the water thick with oil from the sunken vessels in the harbor. Swimming through this they made their way with other survivors to a hanger on Ford Island, where they received fresh clothing, a cup of tomato juice, and various weapons to defend against the anticipated island invasion (Bolen, 1976; Harding, 1988).

Bandmaster Luckenbach quickly went to work organizing the men of the *California* to save their ship. First he opened locked ammunition boxes near the antiaircraft guns, then led firefighting crews to rescue those trapped by the midships bomb explosion, and finally helped haul 50 lb. 5-inch shells up a series of ladders from the magazine to the main

deck. Recalling the morning 30 years later, Red Luckenbach remembered taking a launch over to assist the *USS Arizona*, only to find the fires too hot to approach, and no sign of survivors. On reboarding the *California,* he removed his neckerchief to cover the face of a man he knew only as 'Keys', the mortally wounded shopkeeper of the ship's store (Luckenbach, 1978).

Within a day, the *California* settled to the harbor bottom, her keel resting in the mud. With 102 total casualties, the ship fared better than many that day. However, once they were assembled again on shore, Red Luckenbach took a head count of his band, and found only nineteen of his twenty men. Musician Second Class Russell K. Shelly Jr, the 24-year-old saxophonist and celebrated arranger from Sellersville, PA, did not make it off the *California*. His duty station had been in the bow, near both torpedo impacts.(Harding, 1980; Luckenbach, 1978)

4 Combat Intelligence

Across the harbor, located in the basement of the Navy District 14 Administration Building, Commander Joseph Rochefort listened to the sounds of the attack, and felt a personal guilt. He was in command of the seven-month-old Combat Intelligence Unit (CIU) at Pearl Harbor, one of three communications intelligence stations operated by the United States Navy. Since arriving in Hawaii Joe Rochefort had been tasked with deciphering the presumably information-rich Flag Officers Code of the Imperial Japanese Navy (Carlson, 2011). This code had resisted penetration, though his unit had been effective using Traffic Analysis and Direction Finding through the massive radio antenna located at He'eia (Winton, 1993) to gather valuable information on Japanese naval movements. As he watched the battle unfold on December 7, Rochefort felt that he had failed in this duty. In hindsight, even if JN-25b had been his assigned decryption target, the *Kido Butai* force under Admiral Nagumo had maintained remarkable radio discipline for weeks and left very little to intercept. (Parker, 1994) But Commander Rochefort, like many in the navy and the nation following the attack, wanted to get into the fight.

4.1 Joseph Rochefort

Rochefort entered cryptology in late 1925. As navigation officer aboard an aging oil tanker, Rochefort would pass the time playing auction and contract bridge with his captain and gained a reputation as a master of crossword puzzles. Both skills were regarded as factors identifying aptitude in codes and ciphers, and he was soon assigned to the small group of Navy cryptologists in Washington DC. His mentors at the Washington Navy Yard included Lawrence Safford, a major figure in the field, and Agnes Meyer Driscoll, whose sharp mind and innovative thinking placed her at the center of the Navy's efforts to break several ciphers during the interwar period. Rochefort next attended a multi-year Japanese language School at the US Embassy in Tokyo, where he interacted with his counterparts and soon-to-be enemies in the Japanese Navy (Carlson, 2011).

Shortly after the attack on Pearl Harbor, Rochefort and his men were re-tasked from the Flag Officer's Code to breaking the primary Japanese Navy Operations Code designated JN-25b by the US Navy. By December 10, 1941, Navy Communications Intelligence in Washington DC and the new Commander-in-Chief of the Pacific Fleet had made this new effort a priority, and Rochefort and his team worked long hours to decipher, decode, and translate JN-25b (Wright, 1982).

4.2 The Dungeon

The small group in the damp, poorly ventilated basement, colloquially referred to as "The Dungeon," needed additional staff to meet this new challenge, particularly as active war meant a massive increase in the frequency of intercepted messages in JN-25b (Showers, 1998). Rochefort turned to his talented administrative aid, Senior Chief Petty Officer (CPO) Durwood "Tex" Rorie. Tex Rorie was the quintessential Chief for Combat Intelligence, getting Rochefort's men what they needed through his series of connections on Oahu. Rorie could often get the impossible and was now asked to find men to staff Combat Intelligence. He personally met each transport ship arriving into Pearl Harbor and was given free rein to have any enlisted men he wanted assigned to Rochefort's command. Tex Rorie had a sense of what he was looking for: dedicated men capable of long hours of detailed

work, bright sailors capable of finding patterns within millions of seemingly identical data points, and men who could hold close vital secrets of the United States Navy (Carlson, 2011; Rorie, 1984).

4.3 Unit Band 16 to the CIU

Red Luckenbach and the musicians of band 16 had lost nearly everything in the attack. Beyond the devastating loss of their shipmates, their belongings, including many of their instruments, were entombed within the hull of the *USS California*, and with each capital ship of the fleet damaged or sunk there was no post or vessel in need of a band. Sent to the Fleet Receiving Station Pearl Harbor, they waited and wondered what the future might hold (Harding, 1980).

The musicians of Band 16 were in the Receiving Station when intrepid Chief Tex Rorie arrived on December 10, 1941searching for additional manpower for Combat Intelligence. As Rorie recalls in his oral history from 1984, "I just walked out there and asked, 'Any of you keep your damn mouth shut?'" A unanimous show of hands was followed with "We can get even with the Japs… and we can give you a job that will give you some satisfaction, [since] your instruments have blown up." This air of mystery was enough, and Band 16 became a part of Rochefort's operation in the CIU (Rorie, 1984).

An unforeseen stumbling block came from the FBI, who determined that several members of the band could not be cleared to work in this Top-Secret field. Rorie was told that men with the last names of Palchefsky, DeStwolinska, and Garbuschewski sounded too close to those of the enemy and would not be granted clearance. Chief Rorie brought the news to Red Luckenbach, who did not accept it. Rorie, a senior chief, was not used to being told no from a subordinate, but the men of Band 16, having been through years of training and the horrifying experience of December 7 together, would not be broken up. Rorie took the dilemma to his Commanding Officer, and Joe Rochefort, in a moment of decisive efficiency, told his CPO, "to hell with the FBI. Bring those guys in here." (Carlson, 2011; Palchefsky, 2023) Musician Harding, in several published recollections starting with the *USS California* Cruise Book and through his own 1988 memoire *Band of Secrecy*, notes that the Navy School of Music commander Lt. Charles Benter required all students to be second generation Americans and to pass an FBI background check, noting that FBI agents travelled to his hometown to interview friends and family as part of the Secret clearance Benter required. Harding believes this FBI clearance in the bandsmen's personnel files allowed them to remain together in their new highly secret assignment (Harding, 1980; Harding, 1988). Documentation of such a USNSOM requirement remains elusive.

Luckenbach recalls the recruitment differently. Upon checking in to the receiving station around December 10 and not finding his band, he learned from the Fleet Pooling Officer that they had been transferred away. The Navy School of Music had specific rules governing Unit Bands, which by design could not be broken up for two years. A note to this effect from the Navy Bureau of Personnel was in the personnel file, or 'jacket', of each musician. With this in-hand, Luckenbach approached Senior Chief Rorie to demand that the entire band be accepted, as well as their Bandmaster. He accepted the complete roster of Unit Band 16 (Luckenbach, 1978). The retraining of this group was entrusted entirely to Chief Rorie, though in practice the bandsmen looked to Luckenbach as their effective leader. (Bolen, 1976; Harding, 1980; Rorie, 1984)

The band settled in quickly and were given vital if repetitive assignments. The JN-25b code consisted of a book of 50,000 5-digit groups, each with a meaning. A group might designate a ship, a location, or a letter (Budiansky, 2000). The crew of station NEGAT in Washington DC, including Lawrence Safford and Agnes Driscoll, had largely duplicated this book and sent a copy to Rochefort, who received it on December 15 (Carlson, 2011). But each JN-25b code group had a level of encryption: a five-digit number that was 'false-added' to each code group. False addition was simply adding the numbers together without carrying over any integers. An original code of 95832 might signify "Arrived port," but encryption meant a randomly selected number, for example 32401, would be false added, thus the radio messenger would send out 27233. To accurately read a code group, one had to have the original book,

the encryption key designator given as column and line, and the encryption key book. Combat Intelligence had to find the answers using only one of these three (Holtwick, 1971). Built-in repetition in radio messages, such as "to your excellency" or "to the honorable" helped the team find patterns in swarms of 5-digit codes, as did the time delay of sending updated code books out to each post of the rapidly expanding Japanese Empire. A final challenge was that the messages, once decrypted and decoded, were in Japanese. With terrible news coming in daily from American bases at Wake Island, Guam, and the Philippines, the solutions had to be found quickly.

4.4 The Machine Room

In a cinderblock room created in one corner of the Pearl Harbor basement, Lt. Jack Holtwick ran a series of IBM tabulating and punch card machines, used to find repetition in the millions of code groups recovered by the Combat Intelligence Unit. The musicians were assigned to a specific machine that would generate hundreds of punch cards per day with the number groups from the radio intercepts, and to sort and count the cards as the machines processed them (Showers, 1998). The sorted cards could then be used to find commonly used code groups which in sufficient quantity could be cross-referenced to break the encryption. The sectioned off area of the dungeon with these noisy and pungent machines was referred to as 'Holtwick's Boiler Factory,' and became the new home for many of Band 16 (Holmes, 1979).

4.5 Unit Band 16 Function in CIU

Chief Rorie was impressed by the bandsmen, who had been turned over to him for training by Commander Rochefort who told him, "100%, that's your band" (Rorie, 1984). Mus2c Frank Wanat worked directly with Rochefort's second-in-command Lt. Thomas Dyer, a gifted cryptanalyst, and became a 'gifted cryppie' himself (Ferguson & Foundation, 2001). Others were assigned to Lt. Jasper Holmes, a previously retired submariner who worked the plotting tables marking ship locations and liaised with his former colleagues at the Submarine Base. Others worked in Lt. Holtwick's Machine Room, where IBM Tabulators sorted millions of punched cards with five-digit groups. Several of the bandsmen, including trombonist Peter Panyon of Minnesota, worked to create these punch cards (Rorie, 1984). Panyon, speaking to a group from the National Cryptologic History Foundation in 2001, said "we had no idea what we were doing, or what the result would be. We just typed in the punch cards and sent them on to the next guy" (Ferguson & Foundation, 2001).

Chief Luckenbach and Mike Palchefsky worked closely with the IBM sorting and tabulating machines themselves. In Luckenbach's case, the expertise he gained in Holtwick's machine room led him to an officer's commission and a post-war career with IBM, first in Seattle and then in Washington DC, where he was a civilian consultant to the Naval Security Group (Luckenbach, 1978) working in the 3810 Nebraska Avenue facility.

There is conflicting evidence regarding the degree to which the bandsmen of the *California* were involved in actual cryptanalysis at this point. Jasper Holmes in *Double Edged Secrets*, an early declassified account of WWII cryptology in the Pacific, records that the musicians were not only involved in cryptanalysis but that as a result, a link between musical and cryptanalytic ability was suggested, an idea often repeated by Chief Rorie (Holmes, 1979; Rorie, 1984). John Prados goes further, suggesting that other Navy bandsmen were sought out to fill similar roles due to the *California* Band's success (Prados, 1995). Supporting this observation, Mus2c John Klaboe Engen of Unit Band 17, from the capsized *USS Oklahoma*, was later assigned to the Wahiawa radio intercept station, and finally given a Radioman rating and sent to a forward US Navy Airbase on the forward island of Roi. Both Palchefsky and Panyon noted working with and admiring Engen (Ferguson & Foundation, 2001). Recently, Mus2c Adone Calderone of the *USS West Virginia* related in his final years related a compelling story of Unit Band 17 serving in Communications Intelligence in Hawaii after the Pearl Harbor attack, while the band also maintained an active performing career at the Royal Hawaiian Hotel in Waikiki (Calderone, 2016). It is an interesting coincidence that the band of the *West Virginia* reported to their Executive Officer (XO) Captain (later Admiral) Roscoe Hillenkoetter, who would eventually become the first Director of the Central Intelligence Agency, in the same

manner Band 16 reported to the XO of the *California* Commander (later Admiral) Earl E. Stone, eventual head of the Armed Forces Security Agency which evolved intothe National Security Agency (NSA). Musician Calderone references working in cryptography rather than cryptanalysis in Hawaii (Center, 2016) and contemporary muster rolls attach the surviving bandsmen of the *West Virginia* to the Navy Yard, Pearl Harbor for the duration of the conflict (*Muster Roll of Pear Harbor Navy Yard, Territory of Hawaii*, 1941).

In a conflicting observation, Forrest Biard, an Ensign in the basement unit until he joined the *USS Enterprise* for sea duty in early 1942, addressed the cryptanalytic acumen of the bandsmen directly, calling the evaluation of them as expert cryptanalysts "horse feathers," while simultaneously praising the work of the band and describing them as "wonderful help," and "we turned them into IBM Machine punchcard operators, and they were excellent at it, and they loved it" (Biard, 1992).

5 Operational Code JN-25b

Within weeks of joining Combat Intelligence, the code groups that had been backing up were processed: punched, sorted, and analyzed. Of the 50,000 base code groups in the JN-25b code, the cryptanalysts in Hawaii and Washington were able to break 10% to 15% of these by February 1942. By mid-April, enough of the code had been broken to allow for regular reading of the Imperial Japanese Navy (IJN) operational code, at least by Rochefort's intuitive team. (Parker, 1993) The experienced cryptanalysts and linguists Rochefort, Thomas Dyer, Wesley 'Ham' Wright, Joseph Finnegan, and Alva 'Red' Laswell took the connections and patterns within the millions of five-digit groups and made leaps of intuition, coupled with long hours of diligent effort. The men of the unit worked on a 6-hour on, 6-hour off schedule, getting up from their desks or machines only once or twice per shift to get water. With the memories of the Pearl Harbor attack so vivid, and with the wreckage of battleship row still visible from the front lawn of the building where they worked, they were well-motivated (Carlson, 2011; Layton, 1983).

In early 1942, Commander Rochefort nearly lost his band to their former Admiral. Rear Admiral William S. Pye had commanded the Battleships of the Pacific Fleet at Pearl Harbor in late 1941, from his flagship USS *California*. Here he was well-aware and impressed by the musicianship of Unit Band 16, who provided both ceremonial and entertainment music for him at that time. With the rapid removal of Pacific Fleet commander Admiral Husband Kimmel following the attack of December 7 Admiral Pye was briefly tasked as Commander in Chief of the Pacific Fleet (CINCPAC) until the arrival of his replacement Admiral Chester Nimitz on December 31, 1941. As the Admiral known for ordering the end of the Wake Island relief effort, Pye was ordered away from the war zone and given command of the remaining older battleships in San Francisco Bay (Spector, 1989). From the mainland Admiral Pye had sent word that he needed his old band, Unit Band 16, to provide entertainment at his headquarters. A request from an Admiral, particularly the past Commander in Chief, for 20 enlisted men would normally be approved and executed in short order. But the bandsmen in question were already integrated into Combat Intelligence. Commander Rochefort, through his liaison to the fleet Edwin Layton, informed Admiral Nimitz that such a transfer was not possible, as the musicians were better serving the war effort working on the JN-25b code than playing for the "Market Street Commandos," a colloquial term for the Task Force on the west coast of the US. Nimitz finally agreed, and the band was permanently transferred Rochefort (Carlson, 2011; Palchefsky, 2023).

It was in April of 1942 that the unit made its most significant breakthrough. Traffic analysist made it clear that the IJN was planning an operation of unprecedented size, including the four combat effective aircraft carriers of the First Air Mobile Fleet, responsible for the Pearl Harbor attack in December (Prange et al., 1990). High-ranking members of the Communications section of the Office of the Chief of Naval Operations (designated OP-20-G) in Washington D.C. were convinced this was a move on the west coast of America, or perhaps the Panama Canal. But the analysts in Combat Intelligence had determined through reading the JN-25b code that the target was Midway Island, at the far end of the Hawaiian chain (Parker, 1993). The recovered messages provided

Admiral Chester Nimitz with precise data about the size, location, and timing of the coming attack. He would later tell Rochefort that he was within "Five miles, five degrees, and five minutes" of knowing exactly where the Japanese were. (Prange et al., 1990) With this information, Admiral Nimitz boldly moved the last remaining American aircraft carriers into place northwest of Midway and waited (Whitlock, 1995). Surely the greatest credit for the decisive victory that followed must be given to the brave sailors and airmen who fought it, and the Admirals who acted with trust and 'calculated risk'. But it would not have been possible without Combat Intelligence who gave Admiral Nimitz just the information he needed, just when he needed it. The naval cryptologists of the Dungeon deserved credit for their accomplishments, but due to the secretive nature of the work, they could not tell others of the success (Harding, 1988).

The bandsmen had found a true calling. Even after Rochefort's controversial departure from CIU in late 1942, they continued to work in the field. Tuba player and violinist Edgar Manely, bass clarinetist Horst Garbuschewski, and trombonist Peter Panyon transferred to the newly constructed barracks near the radio antenna at Wahiawa in 1943 and trained in signals reception (*Muster Rolls of Pear Harbor Navy Yard, Territory of Hawaii*, 1943). The Wahiawa facility, completed soon after the Pearl Harbor attack, centralized Radio Interception efforts on Hawaii with an unobstructed antenna network among the pineapple fields in the wide valley in the center of Oahu (Layton et al., 1985). Here musical skills would be more directly employed, as they could perceive the subtle differences in the 'fists' of the Japanese radio operators, identifying their personal style at the telegraph in the same way they could identify the nuanced phrasing of a particular jazz soloist. (Ferguson & Foundation, 2001; Panyon, 2017)

6 Late War Developments

In October of 1942 Combat Intelligence was redesignated Fleet Radio Unit Pacific, or FRUPAC, and moved to a larger building on the Makalapa crater near both Fleet Admiral Nimitz' headquarters and the other operations of the Joint Intelligence Center Pacific Ocean Areas (JINCPOA), such as photographic reconnaissance (Showers, 1998).

Image 2: Fleet Radio Unit Pacific (FRUPAC), machine tabulators and reproducers. (Photo RADM J. N. Wenger)

By late 1943, the band began to separate for the first time since they met at the Navy School of Music in 1940. Two moved to the mainland to become naval aviators (Harding, 1988; *Muster Roll of Pear Harbor Navy Yard, Territory of Hawaii*, 1941). Joseph Bolen took the Navy offer to complete his college degree in Texas (Bolen, 1976). Tex Rorie was transferred to the radio unit on Bainbridge Island in Washington State before being transferred back to Hawaii and finally Washington D.C., where he continued working with the Communication Intelligence community including Thomas Dyer and John Harper at the Navy Communications Supplemental Activity Washington facility, CSAW, at 3801 Nebraska Avenue (Rorie, 1984).

7 Unit Band 16 into the Cold War

Peter Panyon, while serving in the Communications Supplemental Annex Washington in 1945, was asked if he studied any languages in school, to which Panyon responded no, but that his family conversed exclusively in Slovenian. This prompted his immediate enrollment in Russian language school in Colorado (Ferguson & Foundation, 2001). Panyon served in the Navy and NSA into the late 1970's, with postings including the Antarctic during *Operation Deep Freeze* (Panyon, 2017). Michael Palchefsky, whose family spoke Polish at home, was similarly

admitted into Russian language school, with subsequent postings near Hamburg, Germany and Hokkaido, Japan (Palchefsky, 2023). Palchefsky credited his skill at transposing music in applying the process of transposition to code recovery, and to a lesser extent his experience using phrase substitution in improvised solos with number substitution. Musical transposition is a skill particularly evident in jazz musicians and orchestral trumpet players, of which Palchefsky was both. (Bolen, 1976; Ferguson & Foundation, 2001; Palchefsky, 2023)

7.1 Project Venona Connection

Of the nine bandsmen who moved to the Nebraska Avenue Complex later in the war (*Muster Roll of Washington Navy Yard*, 1943), Frank Wanat leaves a particularly intriguing trail. Musician First Class Wanat sat between Peter Panyon and Warren Harding in the trombone section of Unit Band 16 (Panyon, 2017). Along with others from the band he transferred to Washington DC in 1943, and like Pete Panyon and Mike Palchefsky would remain in Communication Intelligence for decades (Administration, 1941-1945; *Muster Roll of Washington Navy Yard*, 1943; Palchefsky, 2023).

At the same time the bandsmen were at the Nebraska Avenue facility from late 1943, the US Army Signals Intelligence Service (SIS) was housed in Arlington Hall, a similar site near Washington D.C. In a back hallway of Arlington Hall in 1943, US Army Lieutenant and accomplished Assyriologist Richard Treadwell Hallock assembled a small team to tackle a highly classified project, the reading of Soviet encoded and enciphered messages sent from trade missions in the US back to Moscow (Benson, 2001). Known by several codenames in the 1940s, this effort would be collectively referred to as *Project Venona*. *Venona*, a tightly held secret throughout the Cold War and declassified in 1995, revealed the code-names of Soviet agents and influencers within the United States Government, including those holding political, military, or technological secrets that would compromise national security if shared. In his official *History of Venona*, NSA Historian Robert 'Lou' Benson lists the names of Hallock's team of cryptanalysts working on this 'Russian Problem'. The list includes Frank Lewis, Genevieve Grotjan Feinstein, and Frank Wanat (Benson, 2001). Frank Wanat of Band 16 was a sailor of the Navy during the war, not a soldier of the Army, but the connection is intriguing.

The 'Russian Trade Problem' Richard Hallock's team was working on in secret was a cryptanalytic challenge requiring techniques similar to those used on JN-25b, with four- and five- digit code groups and related additives. A Soviet codebook held by the Finnish Government and acquired by the United States showed similarities between the two systems (Benson, 2001). Wanat's experience with punch card and tabulating machines, and his years of experience with the similar JN-25b systems, make him an interesting candidate for this important Cold War effort. While the political climate between the US Army and Navy intelligence services at the time make a cross-service assignment before 1944 implausible, soon Arlington Hall and Nebraska Avenue would indeed begin collaborating, and sharing both ideas and staffing. In his Early History of the NSA, George F. Howe mentions that US Army Signals Intelligence was willing to begin merging with the smaller Naval Communications Intelligence element OP-20-G as early as 1945, particularly in matters requiring a united front toward American allies (Howe, released 2007). By 1943 Admiral Joseph Redman of the Navy's OP-20-G and Colonel Clarke of the SIS began investigating methods of collaboration, with the extended goal of holding on to the national communications intelligence (COMINT) capabilities following the war (Burns, 1990).

Frank Wanat continued in Communications Intelligence as a member of the US Navy, the AFSA, and from 1952 the National Security Agency until his death in 1967. Wanat's wife Better Rogers Wanat served in the NSA for two decades after his death. This interesting relationship between Unit Band 16, the intelligence successes in the early War in the Pacific, and the *Venona Project* revelations, is the next stage of investigation for the author.

7.2 Future study

Other musicians of this mid-20th Century found success crossing into cryptology. Agnes Meyer Driscoll was the daughter of conductor

Gustav Meyer, and an accomplished pianist who taught at a Music Conservatory in Texas before becoming a pioneer cryptologist of the US Navy. Lambros Callimahos was a lauded flute soloist who gained a Flute Professorship in Salzburg's *Mozarteum* in 1935, then joined the Signals Intelligence Service and eventually edited the NSA training text *Military Cryptanalytics*. Such musical associations are worthy of further study both for historical value and perhaps to aid in identifying potential cryptologic perspicacity in the future.

References

Muster Rolls of the US Navy, WWII. National Archives and Record Administrration, NARA 1, Washington, DC, accessed via Fold3.com October – Decemebr, 2023.

Robert. L Benson (2001). *The Venona Story*. Ft. George Meade, MD: Center for Cryptologic History.

Forrest Biard (1992). *University of North Texas Oral History Collection* [Interview]. Denton, TX; The Board of Regents of the University of North Texas.

Joseph Bolen (1976). *North Texas State University Oral History Collection Number* [Interview]. The Board of Regents of North Texas State.

Stephen Budiansky (2000). *Battle of wits : the complete story of codebreaking in World War II*. Free Press.

Thomas. L. Burns (1990). *Origins of the National Security Agency, 1940 - 1952*. Ft Meade, MD: Center for Cryptologic History

Elliott Carlson (2011). *Joe Rochefort's War : the odyssey of the codebreaker who outwitted Yamamoto at Midway*. Naval Institute Press.

Adone Calderone Interviewed by Greg Carumbus, "Veterans Chronicles", *Pear Harbor Survivor Adone Calderone, USS WestVirginia*. (November 3, 2016). https://youtu.be/FnPSxZV83Wc?si=EP8bJNBry0baiZ6y (acccesed 11/15/2023),

William Ferguson & N. C. H. M. Foundation (2001). *And the Band Played On, vice to punchcards not instruments* Ft Meade, MD, National Cryptologic History Museum Foundation.

Warren. G Harding Interviewed by Ronald Marcello (1980). *North Texas State University (*now the *Univeristy of North Texas) Oral History Collection*, UNT Library, Denton, TX .

Warren. G Harding (1988). *Band of Secrecy: A Sea Story from the Battle of Pearl Harbor*. Tuscon, AZ, The Academy of Real Estate.

Wilfred J. Holmes (1979). *Double-edged secrets : U.S. naval intelligence operations in the Pacific during World War II*. Naval Institute Press.

Jack. S. Holtwick Jr. (1971) *Naval Security Group history to World War II. Part 1C* University of Southern California Digital Library (USC.DL). https://doi.org/10.25549/loureiro-c12-15149

George F. Howe (released 2007). *The Early History of NSA*. Ft Meade, MD.

Patrick. M. Jones (2002). *A History of the Armed Forces School of Music* (Publication Number UMI 3051676) [Ph.D., Pennsylvania State University].

Edwin. T. Layton interviewed by Robert D. Farley (1983). *Declassified Oral History Collection,* OH-1983-02. National Security Agency, Center for Cryptologic History. Ft. Meade, MD.

Edwin. T. Layton, Rodger Pineau, R & John Costello (1985). *"And I was there": Pearl Harbor and Midway--breaking the secrets* (1st Quill ed.). W. Morrow.

Lovine B. Luckenbach interviewed by Ronald Marcello (1978). *North Texas State University (*now the *Univeristy of North Texas) Oral History Collection*, UNT Library, Denton, TX.

Sharon Manecki & Redacted. (2003*). Rear Admiral Earl Everett Stone: A Convert to Cryptologic Centralisation.* In Cryptologic Almanac 50th Anniversary Series: Center for Cryptologic History. Ft. Meade, MD

Muster Rolls of Pear Harbor Navy Yard, Territory of Hawaii. (1941). National Archives and Record Administrration,, Washington, DC, accessed via Fold3.com July – Decemebr, 2023.

Muster Roll of Washington Navy Yard (1943). National Archives and Record Administrration, Washington, DC, accessed via Fold3.com October – Decemebr, 2023.

Kirk Palchefsky (2023). Email communication between Kyle Prescott and Kirk Palchefsky, son of Michael Palchefsky.

Peter J.Panyon Jr. (2017). Email communication between Kyle Prescott and Peter Panyon, Jr., son of Peter Panyon.

Frederick D. Parker (1993). *A priceless advantage: U.S. Navy communications intelligence and the battles of Coral Sea, Midway, and the Aleutians*. National Security Agency, Center for Cryptologic History.

Frederick D. Parker (1994). *Pearl Harbor revisited: United States Navy communications intelligence, 1924-1941*. National Security Agency, Center for Cryptologic History.

John Prados (1995). *Combined fleet decoded : the secret history of American intelligence and the Japanese Navy in World War II* (1st ed.). Random House.

Gordon. W. Prange with Donald M., Goldstein & Katherine V. Dillon (1983). *Miracle at Midway*. St. Paul, MN, HighBridge.

Durwood G Rorie interviewed by Robert Marcello (1984). *North Texas State University (*now *Univeristy of North Texas) Oral History Collection*, UNT Library, Denton, TX

Donald. M. "Mac" Showers interviewed by Bill Alexander (1998). *The National Museum of the Pacific War,* Fredericksburg, TX.

Ronald. H. Spector (1989). *Eagle against the sun : the American war with Japan* (Collector's ed.). Easton Press.

Duane. L Whitlock (1995). The Silent War against the Japanese Navy. *Naval War College Review*, *48* (Autumn 1995), 43 - 52.

John Winton (1993). *Ultra in the Pacific : how breaking Japanese codes & cyphers affected naval operations against Japan 1941-45*. Naval Institute Press.

Wesley A."Ham" Wright interviewed by Robert Farley and Henry F. Shorreck (1982). *Declassified Oral History Collection,* OH-1982-11. National Security Agency, Center for Cryptologic History. Ft. Meade, MD.

Appendix A: Roster of US Navy Unit Band 16

Last, First	Nickname	Jazz Instrument	Concert Instrument	1941 rating	1943 rating	1946 Rating/ Rank	1949 rating/ Rank	AFSA/ NSA
Bolen, **Joseph**	Duke	Banjo	French Horn	Mus2c	Y2c			
Carroll, **Robert**	Bob	Trumpet	Trumpet	Mus1c		CTRC	CTC	
Carroll, **Raymond**	Ray	Drums	Drums	Mus2c	Y2c	CTRC	CTC	
Conley, **Lawrence**	Larry	Alto sax	Clarinet	Mus2c	Y2c			
De Stwolinska, **Adelbert**	De	Double bass	Double Bass	Mus2c	Y1c	CWO2	CWO2	
Garbuschewski, **Horst**	Garbo	Tenor sax	Bass Clarinet	Mus2c		CTA-CM	CTC	
Harding, **Warren**		Trombone	Trombone	Mus2c		CTRC	CTC	
Kennedy, **David**	Grandma	Guitar	Horn/ arr.	Mus2c	Y2c			
Leonard, **Percy**	Perc	Trumpet	Trumpet	Mus2c	Y1c			
Luckenbach, **Lovine**	Red	Leader	Conductor	Bmstr AA	Bmstr PA	LCDR		
Manley, **Edgar**	Jug Butt	Tuba	Violin/	Mus2c	Y2c	CTC	CTC	
Marnette, **Richard**	Dick	Baritone sax	Clarinet	Mus2c	Y2c	CTRC	CTC	
Palchefsky, **Michael**	Mike	Trumpet	Trumpet	Mus2c	Y2c	CWO2	CWO2	Yes
Panyon, Peter, Jr.	Pete	Trombone	Trombone	Mus2c	Y1c	LT	LT	Yes
Parker, **Robert**	Bob	Alto sax	Clarinet (Eb)	Mus1c	Y1c	CWO2	CTC	
Rutledge, **John**		Piano	Bass drum	Mus2c	Y1c			
Shelly Jr, **Russell**	Sig	Tenor saxophone	Clarinet/ arranger	Mus2c	KIA 12/7/41			
Sumpman Jr., **Russell**	Baby	Trombone	Euphonium	Mus2c	Y3c			
Tallbacka, **Verner, Jr**	Finn	unknown	Flute/ piccolo	Mus2c	Y2c			
Theis, **Robert**	Bob	Trumpet	Trumpet	Mus2c	Y2c			
Wanat, **Frank**		Trombone	Trombone	Mus1c	Y1c	CTC	CTC	Yes

Key

Bmstr (AA) (PA) – Bandmaster (Acting and Permanent)
Mus2c/ Mus1c – Musician 2nd class and 1st class; *Y2c/ Y1c* – Yeoman 2nd class and 1st class
CTA – Cryptologic Technician (administrative); *CTA-CM* – Cryptologic Technician (administrative), Master Chief Petty Officer
CTC – Cryptologic Technician (tech.); *CTRC* – Communication Technician (collection), Chief Petty Officer
CWO2 – Chief Warrant Officer 2; LT – Lieutenant; LCDR – Lieutenant Commander

The Keys of Diplomacy.

The encrypted correspondence of Saxon-Polish Ministers Wackerbarth and Flemming 1700-1720.

Anne-Simone Rous
State Palaces, Castles and Gardens of Saxony, non-profit
Ostrauer Str. 4
D-01277 Dresden
asrous@gmail.com

Abstract

In the Great Northern War (1700-21), the two most important ministers of August II of Poland, Count of Flemming and Count of Wackerbarth, regularly exchanged reports and communications. Several passages of their correspondence are encrypted. Based on examples from the years 1700, 1706, 1715, 1717, and 1720s, this paper presents tentative results of the first phase of a project aimed at analyzing the entire correspondence from 1698 to 1728. Key questions concern differences in the structure of the ciphers and the efforts involved in decryption. New research perspectives are outlined, such as how ciphers are reused and how cryptography and steganography were combined. The examination of encrypted passages provides insights beyond the facade of 'August the Strong'.

1 Background and Agenda

During the Great Northern War, two ministers of August II of Poland (also known as 'August the Strong'), Count Jakob Heinrich von Flemming and Count August Christoph von Wackerbarth, regularly exchanged reports and communications. Their correspondence is of great value and is nearly complete in the Saxon State Archive.[1] Several passages are encrypted. The use of ciphers and their decryption or the identification of the nomenclator will be the focus of the first part of the paper. Examples from the years 1700, 1706, 1715, 1717, and 1720 will be detailed to demonstrate why these passages were encrypted and others were not. This highlights the scientific significance and the necessity of deciphering historical documents. The second part explores the extent to which individual ciphers differ in their structure, opening new research perspectives.

2 Introduction of the actors

Count August Christoph von Wackerbarth (1662-1734) and Count Jakob Heinrich von Flemming (1667-1728), key ministers under August the Strong, were masters of diplomacy, adept at collecting, validating and encrypting information for their ruler's benefit.

[1] Cf. Saxon State Archive - Saxon Main State Archive (SächsHStAD), 10026, Loc. 00711/07, 00711/08, 00712/01 bis 00712/07, 00713/01 bis 00713/08, 02869/09, 03425/01 bis 03425/03; 11237, Loc. 10890/01. The counterpart records in Polish archives will be additionally consulted later in cooperation with Warsaw University.

Fig. 1: Flemming[2]

Fig. 2: Wackerbarth[3]

Wackerbarth served in various military and diplomatic roles, including as envoy in Vienna and governor of the residence city Dresden (Löffler 2002). Flemming, always close to the king, was the gatekeeper in the information network of the court and involved in the cabinet affairs (Czok 1990). Their collaboration formed a communication triangle between Dresden, Warsaw, and Vienna, leading to extensive correspondence. A research project conducted by the author is underway to transcribe and decipher approximately 6,500 folio pages of their letters from 1699-1728, offering insights into both professional arguments and personal sentiments.

3 Research questions and methods

The dense transmission allows for an almost day-to-day reconstruction of the actors' knowledge. Due to their high offices, research interest focuses on the encrypted parts of the correspondence: What was encrypted and why? How strong were the codes? But the source corpus opens up the horizon to further questions. Because the effort to create new ciphers was very high, older ciphers were often reused.This aspect has been scarcely addressed by research so far. To recognize a pattern in reuse, a statistical analysis is necessary. Furthermore, invisible ink was also used, which additionally complicates the processing of the source. Partially readable sections may possibly be supplemented by AI. Behind this lies the question: Did secret inks offer better confidentiality than ciphers? Starting from these ciphers the research not only offers a new perspective on the court's operations and the "Saxon Sun King," but also on the practices of communication during the Great Northern War. The workflow is as follows: after transcription and translation, the letters are analyzed for their encryption method and then decrypted. This is done either using the vast number of transmitted cipher tables or decryption programs.

4 Research Status

The prevailing opinion in Saxony that everything has already been researched and said about this famous ruler is based on numerous publications. However, previous works are consistently pro-Prussian in color or remain very superficial because archival research into noble history was difficult in the GDR. Current research on the court of August II mainly focuses on art historical topics, as the saxon bibliography indicates.[4] Polish sources, on the other hand, are intensively evaluated (recently Kosińska 2019, 2023). Research in Saxony is gradually gaining momentum as well. A scientifically sound biography of the queen, Christiane Eberhardine, was recently published (Herz 2020). The cipher of August II has been deciphered, and the secret diplomacy of the court has been analyzed (Rous 2022). A Saxon-Polish-Lithuanian network (PLUS18) was initiated. The Saxon State Archive recently restored and digitized the letters of August II as well as all Saxon cipher tables. These are being gradually incorporated into the DECODE database (www.de-crypt.org/decrypt-web/RecordsList)database.[5] With the 300th anniversary of August II's death in 2033, increased research on this king and his court is expected. The processing of the Flemming-Wackerbarth correspondence is an important pillar, with the encrypted passages being of particular importance for research due to their highly relevance and brisance.

[2] Portrait of Jakob Heinrich Reichsgraf von Flemming (1667–1728), Anonymus, 1720er, Belarusian National Arts Museum ЗЖ-4, CC 1.0, URL: https://commons.wikimedia.org/wiki/File:Jok%C5%ABbas_Henrikas_Flemingas.jpg?uselang=de [30.01.2024].

[3] Portrait of August Christoph of Wackerbarth. Frontispiz. Gesprek tussen August Christoph von Wackerbarth en Magnus Stenbock, Anonymous, 1737, Rijksarchiv Amsterdam, CC 1.0, URL: https://www.rijksmuseum.nl/en/collection/RP-P-2016-1197 [30.01.2024].

[4] Cf. Sächsische Bibliographie Online (Saxon Bibliography Online). URL: https://swb.bsz-bw.de [15.04.2024]

[5] Cf. DECRYPT. URL: https://www.de-crypt.org/decrypt-web/RecordsList]15.04.2024]

5 Encryption in Correspondence

So far more than 130 letters with twenty cipher documents have been analyzed. In these documents, simple substitution ciphers, nomenclators, syllable ciphers, matrices and symbol ciphers were used. The ciphers were regularly exchanged and renewed, because after a few years, a new suitable digit panel always has to be found. In addition to small, very short nomenclatures, large cipher books were created. Decryption allows for answering the question of what contemporaries considered sensitive. Not only were military details in war encrypted, but also diplomatic information or opinions on intrigues. Preliminary results show that Wackerbarth and Flemming wanted to leave the court in the face of defeats in 1706. This fact is hitherto unknown but of the utmost importance because Wackerbarth, as Minister of Construction later, organized the king's major projects. Without this brilliant mind, August II might not have been able to realize the Dresden Zwinger, which is a famous architectural complex near the castle with great pavilions for courtly games and well known as a main cultural attraction in Germany. Thus, the question of what was encrypted can be answered clearly. In the future the question of why needs to be separately considered for the three areas: military, diplomacy, and court internals. The strengths or weaknesses of the various encryption methods will emerge from further analysis of the cipher tables.

6 Deciphering

Some encrypted letters from the year 1700[6] could be made completely readable thanks to the support of computer expert George Lasry. The challenge was that the few snippets of text did not contain the most common letters of the German alphabet. O, C, H, G, and B often appear, while K is only included once.

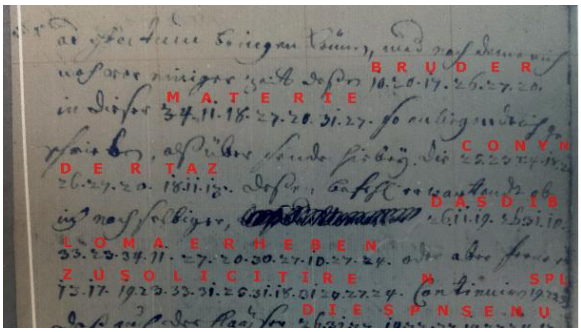


Figure 3: Letter from Wackerbarth to Flemming, 27th June 1700, SächsHStAD, 10026, Loc. 00711/06, f. 11-12, decrypted and image edited by George Lasry

Despite the small text basis, it was possible to decrypt the simple substitution by sending it through the decryption process several times. Later in the archive, I discovered the exact cipher for it – it was labeled "Cipher avec le Cour de Dennemarc" (Cipher with the Danish Court).[7] This cipher is a numerical cipher with a nomenclator in German language. The numbers are two- and three-digit, and the nomenclator extends up to 300. Non-values were not included. The nomenclator contains terms, names, and numbers. Simple substitution is mostly used in Saxony, but often changed. That was supposed to ensure security.

The letters from 1706, now written in French, were interlinearly resolved.[8] Finding the nomenclator among the tens of thousands of undated and unnamed nomenclators was difficult, but successfully. The number range extends from 57 to 999. It also has no instructions and no fixed nonvaleurs. So an "unknown" cipher, that is traditioned without any hint of its owner, date or localization, could be identified and assigned to an owner.

In 1715 another simple substitution cipher was used. The letters contain an interlinear resolution.[9] The matching nomenclator had an earlier year on the envelope paper.[10] This proves, that we cannot always rely on labelling.

[6] Cf. SächsHStAD, 10026, Loc. 00711/06, f. 11-12.

[7] Cf. SächsHStAD, 10026, Loc. 03233/02.
[8] Cf. SächsHStAD, 10026, Loc. 00711/07.
[9] Cf. SächsHStAD, 10026, Loc. 0712/05, eg. f. 5-42.
[10] Cf. SächsHStAD, 10026, Loc. 03233/02.

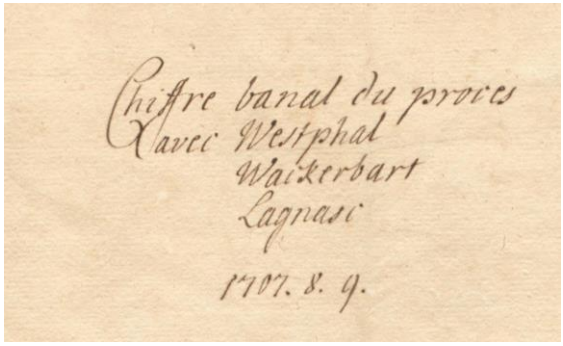


Figure 4: Cover sheet of a Nomenclator, matching to letters of 1715, SächsHStAD, 10026, Loc. 03233/02.

Why was this cipher specifically used again in 1715 and not another cipher? The cipher's selection for the reuse was certainly not random. To find out the pattern of reuse, statistical analysis will be applied in a later stage of the project.

Later the letters were apparently deciphered on a separate sheet, as the numbers were only crossed out and no longer deciphered interlinearly.[11]

A letter from the 24th July 1717 contains a substitution cipher with single- to three-digit numbers, with most numbers being two digits. Rarely do high numbers like 355 or 1000 appear and refer to a nomenclator. Deciphering or finding the nomenclator used in this letter has not yet succeeded.

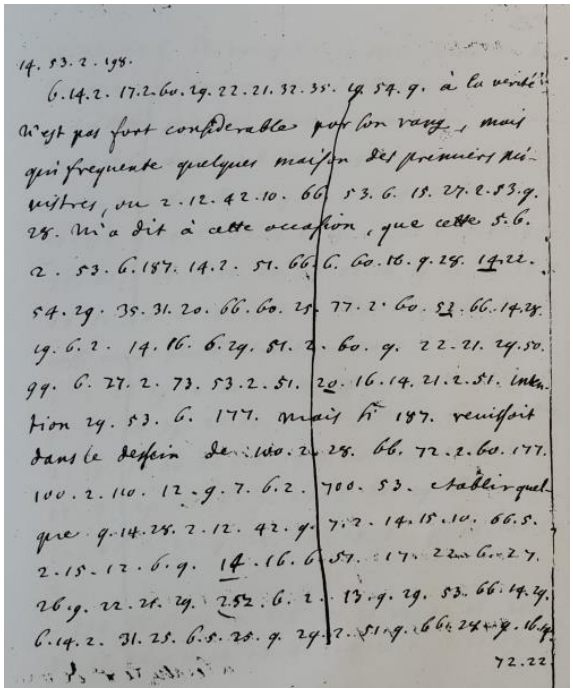


Figure 5: Letter from Wackerbarth to Flemming, 24th July 1717 (detail), SächsHStAD, 10026, Loc. 00713/01

In 1720, the Saxon-Polish court utilized a substitution cipher based on syllables in combination with special ink.[12] Thus, they had moved away from simple substitution and sought innovative methods for enhanced secrecy. Parallel to deciphering the syllable cipher, a method is to be developed to render invisible ink readable through chemical-physical processes. Since the hidden messages are still not readable today, the ink evidently provided better security than the encryption.

7 Summary and Research perspectives

The project highlights the increasing efforts towards improved communication security. It demonstrates the relative carelessness at the beginning of the Northern War and the learning curve within 20 years to change from simple substitution with nomenclator to a syllable cipher combined with special ink. The clerks abandoned doing interlinear deciphering. The hidden content proves to be extremely delicate for Saxon history.

However, the project also reveals research deficits. Three research perspectives emerge from the project. Firstly, a pattern of cipher reuse can be sought through statistical methods. Secondly, the use of invisible ink comes into focus and demands interdisciplinary methods, especially since it is also possible that sympathetic ink and cryptology were combined. Thirdly, the search for nomenclators among the mass of unassignable tables highlights the importance of describing the typical structure of a specific cipher. The analysis of the strengths and weaknesses of a cipher can be derived from the structure of the cipher. Each cipher has a specific structural image, comparable to a fingerprint. This 'fingerprint' consists of the representation, the rule system, the scope, and the specific non-values. In particular, the number of digits, the lowest and highest assigned number, and the exact specification of the non-values provide important clues to the pattern of a cipher.

[11] Cf. SächsHStAD, 10026, Loc. 00713/01, eg. f. 29-33.

[12] Cf. SächsHStAD, 10026, Loc. 03424/02, eg. f. 41-42, f. 100.

References

The DECRYPT project. Decryption of secret historical manuscripts URL: https://www.de-crypt.org/decrypt-web/RecordsList [15.04.2024].

Karl Czok (1987). August der Starke und Kursachsen. Leipzig.

Karl Czok (1990). Am Hofe Augusts des Starken. Stuttgart.

Paul Haake (1902). König August der Starke. Eine Charakterstudie. Berlin.

Paul Haake (1926). August der Starke. Berlin.

Silke Herz (2020). Königin Christiane Eberhardine. Pracht im Dienst der Staatsraison: Kunst, Raum und Zeremoniell am Hof der Frau Augusts des Starken. Berlin.

Jürgen Heyde, Tagungsbericht: Historikertag 2023: Transnationale Verflechtungen in der polnisch-litauisch-sächsischen Union. H-Soz-Kult, 18.11.2023, <www.hsozkult.de/conferencereport/id/fdkn-139981>. [30.01.2024]

Urszula Kosińska (2019). Jacob Heinrich Flemming; Mémoires (1696-1702). Originaltitle: Pamietniki (1696-1702). Warsaw.

Urszula Kosińska (2023). W kręgu mitów, czyli o tym, co nie zadecydowało o wyborze Augusta II na tron polski w 1697 roku. Warsaw.

Józef Ignacy Kraszewski (1873-1885): August der Starke, Gräfin Cosel, Flemmings List. Graf Brühl. Dresden.

Fritz Löffler (2002). Das Alte Dresden. Leipzig[17].

PLUS18 – Poland, Lithuania and Saxony in the 18th century (2019). Project of the Institute of Saxon History and folklife studies. URL: www.isgv.de/projekte/saechsische-geschichte/polen-litauen-sachsen [29.01.2024].

Rouven Pons. 2006. „Die Dame ist romanesque und coquet…“. Catharina Gräfin von Wackerbarth (1670-1719) als kursächsische Gesandte in Wien. *Mitteilungen des Instituts für Österreichische Geschichtsforschung* 114 : 65-95.

Portrait of August Christoph of Wackerbarth. Frontispiz. Gesprek tussen August Christoph von Wackerbarth en Magnus Stenbock, Anonymous, 1737, Rijksarchiv Amsterdam, CC 1.0, URL: https://www.rijksmuseum.nl/en/collection/RP-P-2016-1197 [30.01.2024].

Portrait of Jakob Heinrich Reichsgraf von Flemming (1667–1728), Anonymus, 1720er, Belarusian National Arts Museum ЗЖ-4, CC 1.0, URL: https://commons.wikimedia.org/wiki/File:Jok%C5%ABbas_Henrikas_Flemingas.jpg?uselang=de [30.01.2024].

Anne-Simone Rous (2022). Geheimdiplomatie in der Frühen Neuzeit. Stuttgart.

Sachsens Glanz und Preußens Gloria (1983/84). Six-part television series. Regie Hans-Joachim Kasprzik. GDR. Production: DEFA. Vertrieb: DFF.

Sächsische Bibliographie Online. URL: https://swb.bsz-bw.de [15.04.2024]

Subtle Signs of Scribal Intent in the Voynich Manuscript

Andrew Steckley, PhD
QuantumLynx Research
andrew@quantumlynxresearch.com

Noah Steckley
QuantumLynx Research
noah@quantumlynxresearch.com

Abstract

This study explores the cryptic Voynich Manuscript, by looking for subtle signs of scribal intent hidden in overlooked features of the "Voynichese" script. The findings indicate that distributions of tokens within paragraphs vary significantly based on positions defined not only by elements intrinsic to the script such as paragraph and line boundaries but also by extrinsic elements, namely the hand-drawn illustrations of plants.

1 Introduction

The Voynich Manuscript, with its inscrutable script and peculiar illustrations, has been extensively studied by amateur and professional researchers for almost a century. Despite this, there is no consensus as to its origin, authorship, or the meaning of its unusual script. There is even disagreement as to whether it has meaning at all.

Many reasonable arguments have been put forth supporting divergent scenarios: either that it conveys meaningful content or that it is meaningless, crafted only for visual appearance. Most researchers who have studied the script have implicitly assumed that it contains meaning, focusing their efforts on finding grammatical structures or statistical signatures that would identify the most likely known language of origin from which Voynichese may have been derived or encoded. Those efforts that have aimed objectively at the more primary question —whether it does or does not have meaning in the first place— have predominantly looked for evidence of the script sharing various statistical properties with known languages, the implication being that it is therefore not just a meaningless imitation.

A few studies have explored the possibility of the script being meaningless. Rugg demonstrated a technique, inspired by a Cardan table and grille cipher, that could explain the script's general appearance and feasibly produce a manuscript of similar size manually, within a reasonable timeframe (Rugg, 2004). Zandbergen further analyzed the technique to show how it could produce some of the apparent statistical structure observed in the Voynichese script (Zandbergen, 2021). Neither researcher sought, nor claimed, to explain all of the observed structure, but they did show that a script produced using a language simulation device, although meaningless, could still exhibit some linguistic structure as an artifact of the simulation process. Gaskell and Bowern showed that samples of "gibberish" script also shared certain statistical properties with the Voynichese script (Gaskell and Bowern, 2022). The majority opinion, however, remains that the Voynichese script has meaning and will eventually be deciphered.

This broad-brush summary describes not only the historic published research, but also the many blogs and forum discussions within the community of Voynich enthusiasts that has been growing continually since the Internet made images and details of the Voynich Manuscript more widely available. That community includes a large number of amateur researchers who have bought wholesale into the assumption of meaning, believing they have found concrete connections to known languages or that they have even deciphered specific words or sections of the manuscript. None of these claims, however, have withstood even modest levels of critical review.

It should be acknowledged that an unexpected solution could still emerge from the online community; it would be shortsighted to overlook the role of guesswork and intuition in scientific discoveries. Nevertheless, advances in understanding the Voynichese script appear to have plateaued over the past decade, despite increased scholarly involvement and public interest, and it seems clear

that fresh perspectives in attacking the problem are needed.

We suggest that a paramount objective ought to be the determination of whether the script contains meaningful content that can be deciphered or whether it presents merely the appearance of such content. A determination, irrespective of its outcome, would significantly sharpen and enhance the effectiveness of ongoing research efforts.

Toward these ends, our research seeks to uncover evidence of the scribe's intentions; the supposition is that subtle signs may be found in overlooked patterns in the script. In the present study, we focus on the statistical distribution of "word" tokens and their placement relative to structural features like line beginnings and endings, and adjacency to intricate plant drawings that disrupt the script's flow.

In the following section, we describe further how such statistical evidence might indicate the scribe's purpose. Following that, in Sections 3, 4, and 5, we describe how the transliterated Voynichese data were prepared for use in this study. Sections 6 and 7 then describe the two major analyses performed and discuss their results. Section 8 summarizes the conclusions drawn directly from the analytical results. Finally, in Section 9, we discuss the implications of these conclusions, particularly with respect to whether the scribe intended the Voynichese script to convey meaningful or meaningless content.

2 An Overlooked Feature of the Manuscript

One prominent feature of the manuscript is that the lines of Voynichese script appear to have been intentionally written out to visually conform to the outlines of previously drawn illustrations. In most cases, when encountering the intrusion of a drawing, the scribe has skipped over it to resume writing on the far side of the drawing, or within space between drawing elements when it is of sufficient size. Various researchers have noted this contouring of the script's margins, but it has received minimal investigative effort.

Figure 1 shows an example of this shape conformity, which is consistently seen across all folio pages featuring large drawings in the manuscript. This phenomenon prompts an intriguing question: Did the scribe deliberately select tokens of specific lengths to create this visual effect potentially altering —or simply without regard to — meaningful content in the process?

Gaskell and Bowern looked briefly at whether student volunteers, instructed to compose meaningless text wrapped around plant drawings in this way, were inclined to self-select the lengths of their made-up words to better achieve the desired result (Gaskell and Bowern, 2022). Their assessment, although worth reporting, was not surprising; it suggested that this feature of the manuscript may be common to gibberish documents. Beyond this, no other analytical efforts have focused on this visual aspect of the script.

Whether the tokens represent words in the manner of common written languages, or by some other structured means of converting thoughts into written script, it would seem to be a difficult task to achieve this level of visual conformity while preserving prescribed semantic content. We use the term "prescribed" here to refer to the idea that the meaning of the text is established before pen meets parchment to form each word. This content could originate from the thoughts of the scribe himself as he formulates words during the writing process, or it might involve the direct transcription of someone else's words.

In the first case, he must frequently reconcile the word choices he would naturally make with what may be accommodated by the space remaining before a drawing. In doing so, he must then continually adjust for any drifts in the prescribed meaning introduced by his contrived word choices. This is all possible, but at a cognitive cost. Achieving both the visual and semantic objectives is even more challenging if the scribe is an amanuensis, tasked with faithfully transcribing words from another document or taking dictation from another person.

On the other hand, if the scribe's aims are to produce a merely visual result, while attempting only to simulate script appearing to have meaning and structure, then he is not constrained by prescribed content; he is free to choose words of suitable length at each opportunity.

Some researchers have noticed and analyzed the positional aspect of word tokens with respect to the beginning and ending of lines (Bunn, 2022; Bowern and Lindemann, 2021). The present study, however, looks in detail at token positions not only at line boundaries, but adjacent to the hand-drawn illustrations. One would not expect

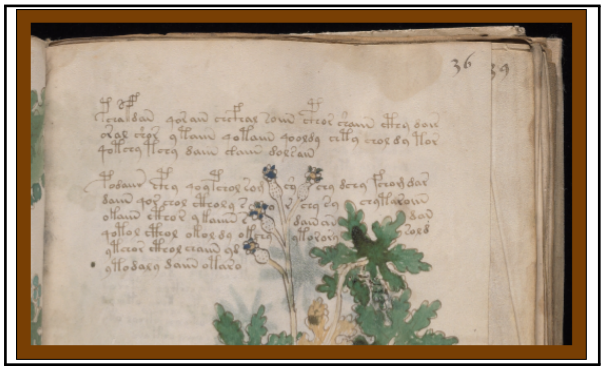


Figure 1: Voynichese Script on Folio f36r. The script exhibits a noticeable conformity to the outline shapes of the drawn illustrations.

the drawings to be coupled to the syntactic content of the script, as might be conceivable in the case of tokens at the beginning and end of lines.[1]

However, this study found distinct and statistically significant differences in the populations of tokens that appear immediately before and after the intrusion of the drawings, as well as at the beginning and ending of lines.

3 Transliterations

Several transliterations of the Voynichese script have been compiled by various researchers as far back as the 1940s using various alphabets designed to represent the Voynichese glyphs. Zandbergen provides an excellent history and description of these transliterations (Zandbergen, 2023). He also designed a standard format called Intermediate Voynich Transliteration File Format (IVTFF) and has made several of the most comprehensive and reliable transliterations available in this format on his website. These standardized transliterations are invaluable for performing detailed analyses.

For this study, we have used the "Zandbergen-Landini" transliteration,[2] which is described and available on Zandbergen's website. This transliteration contains comprehensive specifications of the location of the Voynichese elements, as well as indications of uncertain glyph identifications and uncertain token delimiting spaces. It also indicates when two tokens within the same apparent line of text are separated by a portion of a drawn illustrations.

[1] More conceivable would be an impact on the spacing between tokens and anomalies in the hand-written width of the glyphs. A separate study analyzing these features is in progress.

[2] https:voynich.nu/data/ZL3a-n.txt

4 Study Corpus

It is apparent that the manuscript contains several different sections differing in layout style and illustration type. Several researchers (Newbold, 1928; D'Imperio, 1978; Zandbergen, 2022) have grouped the manuscript's folio pages according to these features and it is generally assumed that each group deals with a different topic.

In addition, it is believed that several individual scribes were involved in the creation of the manuscript. Using digital paleographic techniques, five separate scribes and the particular folios written by each have been proposed (Davis, 2020).

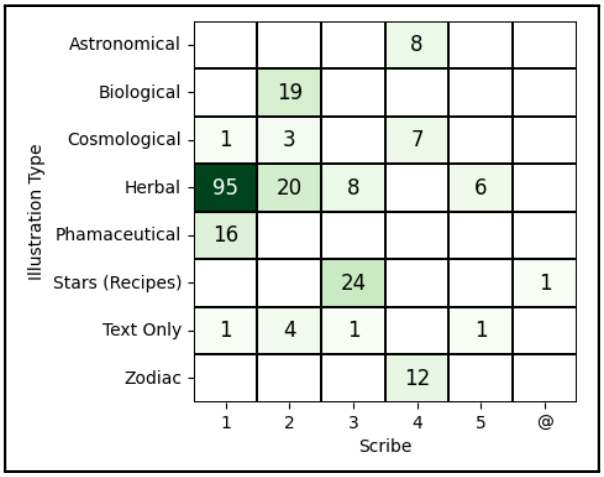


Figure 2: Folio Page Count by Scribe and Illustration Type

To reduce variations in vocabulary that may result from the different topics being covered, and from variations in the mannerisms and styles of different scribes, we wanted to limit the analysis to a single topic and a single scribe. Our reasoning for this is as follows: any analysis of a corpus consisting of heterogeneous folios increases the likelihood of significant attributes being obscured by the noise of multiple contributory factors. On the other hand, any conclusion once drawn from a more homogeneous corpus (e.g. that of a single scribe dealing with a single topic) may subsequently be given independent and explicit consideration as to whether it likely applies to the rest of the manuscript (e.g. other scribes dealing with other topics).

The set of 95 folios of "Herbal" sections attributed to Davis' "Scribe 1" was selected. As seen in Figure 2, this is the largest homogeneous collection of folio pages. There are other topics

containing more tokens per page, but this selection also contains the majority of drawings with the conforming script feature described earlier.

We also wanted to restrict our analysis to the 'paragraphs' wherein a concept of position could be well defined, and to omit tokens that might only contribute statistical noise.

These considerations resulted in the following criteria, which were applied to the full Landini-Zandbergen transliteration to produce the target corpus for this study.

Include:

- Folios with 'Herbal' illustrations
- Folios penned by 'Scribe 1'
- Lines of the script that were in 'paragraph' form (as opposed to labels or floating phrases)

Exclude:

- Final token in each paragraph
- Any token that is ambiguous in the sense that the transliteration indicates more than one possibility for any of its glyphs
- Any tokens where there is uncertainty as to whether the space before or after it is meant to delimit it from an adjacent token

Applying these criteria substantially reduced the data from the original transliteration — by roughly 80% — but this sacrifice of quantity for quality ensured that the study corpus[3] was comprised of tokens that were all well defined and unambiguous.

5 Token Cohorts for Analysis

Several sets of tokens were compiled from the study corpus, in order to provide separate cohorts for analytical comparisons. These are summarized in Table 1 by the reference label used for each cohort throughout this paper.

The reference cohort (MIDDLE) and the subject cohorts (TOP, FIRST, LAST, BEFORE, and AFTER) provided the main targets for the analyses. Mutual exclusivity was enforced on these cohorts, meaning that any token that would otherwise be in more than one of these cohorts was excluded from all of them. Two supplementary positional cohorts (SECOND and FOURTH), along with several randomized cohorts (RAND 1 through RAND 6), were used for additional validations. The random cohorts correspond in size to the smaller subject cohorts, and were formed by drawing random selections of tokens from the MIDDLE cohort.

Note that any transliteration uncertainty of spacing to the left of a token on a line results in a range of uncertainty of its ordinal position. The net effect of this on these data is that 8.5% of the tokens designated as being in the second position, and 20.8% of those in the fourth position, may in fact belong in a higher position. It is for this reason that the SECOND and FOURTH cohorts were excluded from the set of subject cohorts, thereby ensuring that none of the main subject cohorts were affected by such spacing uncertainties.

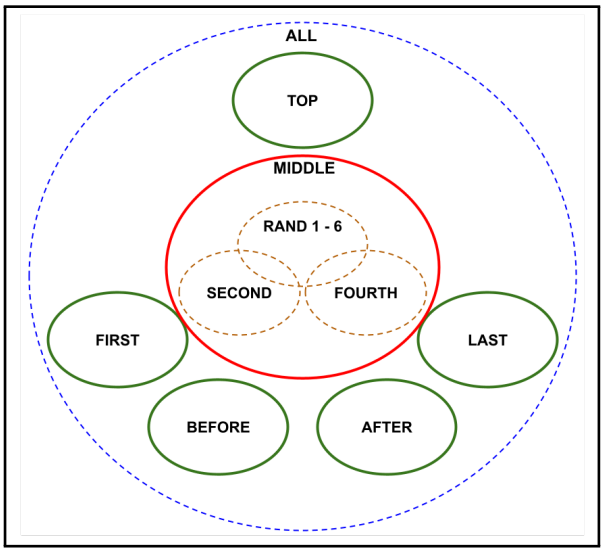


Figure 3: Venn Diagram of Cohorts. *Note that the reference and subject cohorts (shown with solid borders) are all mutually exclusive.*

One may gain a more precise understanding of what tokens are represented by each of the positional cohorts by studying the Venn diagram in Figure 3 and the schematic diagrams in Figure 4.

6 Analysis of Token Lengths

Table 1 also includes the mean and standard deviation of the token lengths observed in each of the cohorts, where the length is defined by the glyph count within the token.[4] We can see some distinct differences in these mean glyph counts. There are thirteen cohorts that each comprise a subset of the

[3] From the 39,020 tokens arranged in 5,389 lines on 227 folio pages, the study corpus extracted 7,660 tokens arranged in 1,223 lines on 95 folio pages.

[4] Multiple glyphs denoted as ligatures in the transliteration were counted individually.

Cohort Type	Cohort Reference Label	Description	Counts				Basic Statistics	
					Tokens		Token Length	
			Folios	Lines	Obs.	Unique	Mean	St. Dev.
Total Corpus	**ALL**	Entire set of tokens in the Study Corpus	95	1,223	7,660	2,355	4.72	1.79
Reference Cohort	**MIDDLE**	Tokens in the middle area of a paragraph, excluding Subject Cohorts	95	1,002	3,807	1,115	4.58	1.69
Subject Cohorts	**TOP**	Token in the top line of a paragraph	95	178	847	506	5.04	1.81
	FIRST	Tokens in the first position on a line	95	998	998	532	5.13	1.75
	LAST	Tokens in the last position on a line	95	777	777	427	4.47	1.91
	BEFORE	Tokens immediately preceding a drawing intrusion	71	330	349	215	4.24	1.96
	AFTER	Tokens immediately following a drawing intrusion	63	264	278	160	4.66	1.63
Validation Cohorts	**SECOND**	Tokens in the second position on a line	95	970	970	433	4.66	1.65
	FOURTH	Tokens in the fourth position on a line	95	691	691	339	4.68	1.72
	RAND 1	Random Tokens selected from MIDDLE	~	~	970	444	4.62	1.66
	RAND 2		~	~	970	435	4.58	1.64
	RAND 3		~	~	970	441	4.62	1.75
	RAND 4		~	~	349	197	4.61	1.76
	RAND 5		~	~	349	221	4.60	1.78
	RAND 6		~	~	349	209	4.64	1.68

Table 1: Summary of Cohorts

MIDDLE cohort and of these, ten have mean token lengths that are clustered within 2.5% of the mean token length for the MIDDLE cohort. The TOP and FIRST cohorts, however show outlier means that are 10% and 12% greater than that of the MIDDLE cohort respectively. Interestingly, the BEFORE cohort is also an outlier, showing a mean that is 7.5% lower.

To be clear, none of these mean token length differences between cohorts are very great. Furthermore, the variance of token length in each cohort is relatively broad. Therefore, little can be concluded from these measurements alone, but they are sufficient to indicate that there may be something more going on, and that a deeper look comparing the distributions of these token lengths is warranted.[5]

The data were therefore analyzed further by applying a statistical test of independence to the token length distributions of each pair of cohorts. A test of independence attempts to quantify the probability, or 'p-value', of a 'null hypothesis'—that is, the hypothesis that the differences between two observed distributions can be attributed to chance alone. The alternative hypothesis is that the two distributions are different, not by chance, but due to some underlying causal mechanism.

[5]Plots overlaying the probability mass distributions of token lengths for each pairs of cohorts can be found in the Supplemental Online Material.

Although it has been observed that the Voynichese token lengths follow binomial distributions very closely (Stolfi, 2000), we have avoided assuming that this, or any other particular distribution, holds for every cohort. Consistent with this, we have applied a χ^2 (Chi-squared) test of independence, which makes no assumptions regarding the form of the probability distributions of the sampled populations being compared.

The test does assume that the observations within each category of the distribution are independent. In our case, this would be the assumption that each value of token length within a distribution is independent. We would expect this to be the case since, given a set of unique tokens employed by a scribe (presumably representing words in a language vocabulary), there is no reason to believe that the usage of tokens of any particular length would be related to the usage of tokens of any other length.

The χ^2 test also depends on there being several observations within each value category; having an expected count of more than 5 in at least 80% of the value categories is considered sufficient (Bewick et al., 2003). This condition was met by all of our cohort distributions, but we have still combined the highest couple of value categories (glyph counts of 9 and 10) when needed to achieve at a minimum of 5 counts for each value of token length.

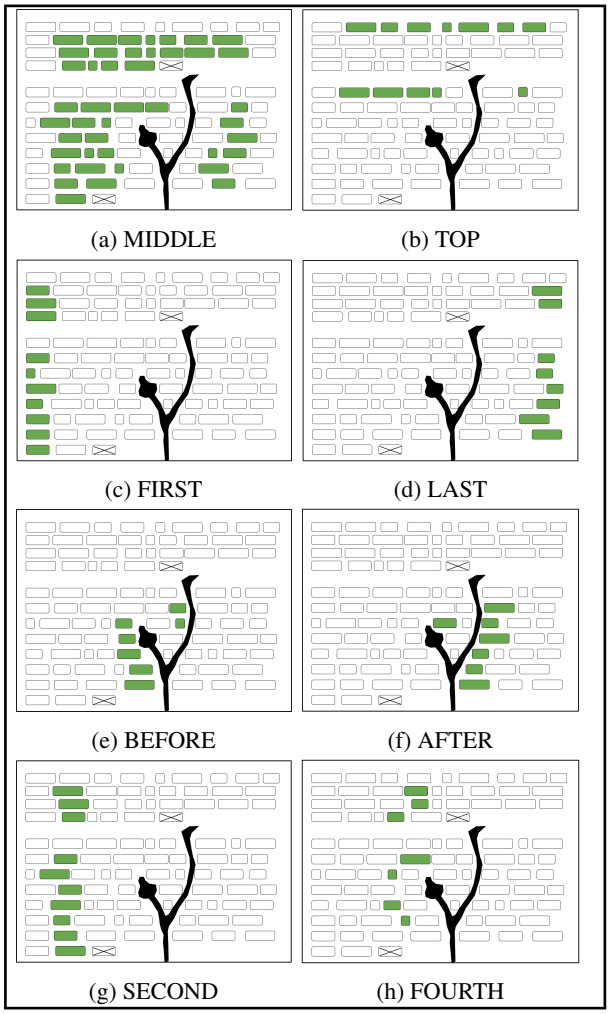


Figure 4: Schematic of Cohort Tokens. *Each diagram shows a hypothetical folio page with two paragraphs of tokens, the second paragraph conforming around a drawing. The shaded tokens indicate cohort members.*

Figure 5 presents the symmetrical matrix of p-values showing which pairs of cohorts pass the test. This, of course, depends on the p-value threshold chosen. A value of 0.05 is very common when applying statistical tests but we used a more stringent value of 0.01. In the figure, the p-value is shown to five significant digits for each test pair. Darker cells indicate pairs of cohorts that pass the null hypothesis test. This means that the two cohorts are similar enough that their differences may be explained by chance sampling. In other words, we cannot infer that they represent different underlying populations. Conversely, the lighter cells indicate a pair of cohorts governed by different causal phenomena, resulting in statistically different distributions of token length.

This matrix shows some interesting results. It indicates that the cohorts of MIDDLE, AFTER, SECOND, and FOURTH are all similar enough to each other that differences observed between them cannot be considered significant. It also indicates that the FIRST cohort is unique, showing a statistically significant difference from *all* the other cohorts. We also see that the TOP cohort is different from most of the others although it passes the test against the AFTER cohort (meaning we simply cannot rule out that the differences are due to chance sampling). None of this is too surprising—differences in the tokens in the top line of paragraphs and in the first position of lines have been noted by many other Voynich researchers, although the observation has generally been regarding the greater presence of certain glyphs in these locations, not their token lengths.

What is notable from these tests, however, is the difference of the token lengths of the BEFORE and LAST cohorts from all other cohorts along with their similarity to each other. We had anticipated, for reasons described earlier, that the token choices immediately prior to a drawing and to a lesser extent at the end of a line or script, may contain trace evidence of the scribe's intentions. Nevertheless, we were surprised that these results, using formal tests of statistical significance, were so pronounced.

Consider, for example, tokens appearing just before a drawing and those at the ends of lines. These two cohorts remain distinct from all others unless we reduce the p-value threshold below 0.005. This results in tokens at the ends of lines appearing similar to those immediately following a drawing. The p-value threshold must be lowered orders of magnitude further before one can attribute the differences between the tokens before a drawing or at the ends of lines (compared to any other cohort) to random chance alone.

In short, there is little ambiguity to be found in these results!

Tests of significance using p-values are based on solid statistical reasoning, and when properly applied and correctly interpreted, they provide the means to establish high confidence in analytical results. It should be noted, however, that the use of p-value as a measure of significance has raised some concerns among researchers in recent years. In fact, one highly viewed article in Nature magazine dealt specifically with this concern

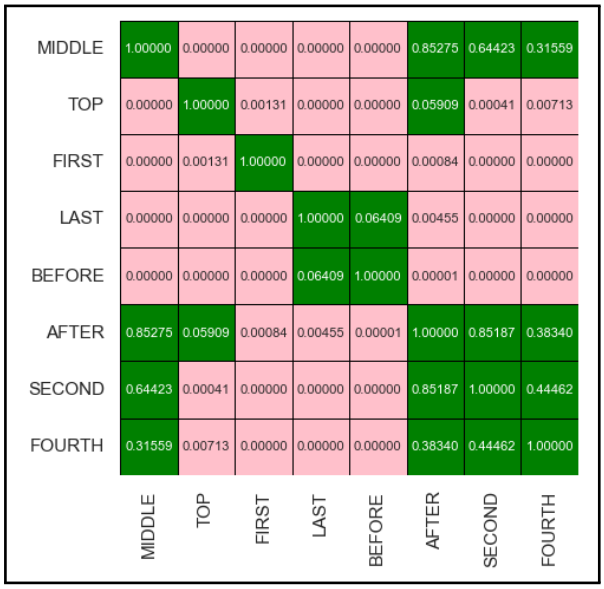


Figure 5: χ^2 Statistical Significance Matrix. *Lighter cells indicate that the pair of cohorts that are most probably governed by different causal phenomena resulting in observed differences between their distributions of token lengths.*

(Nuzzo, 2014), and the American Statistical Association subsequently issued a statement of caution in regard to the over-reliance on *p*-value testing (Wasserstein and Lazar, 2016).

In the present analysis of token lengths, we selected a threshold that was well in excess of that commonly considered sufficient, and ensured that the statistical requirements for the validity of these tests were well met. But, given that the implications of the results are both subtle and potentially controversial, it is worth repeating with precision what they do and do not tell us.

The tokens found in certain positions on a page, relative to paragraph layouts and drawing intrusions, exhibit a distribution of token length that differs from that found in the larger population of tokens spread throughout the middle of the paragraphs. The statistical tests indicate that it is extremely improbable that these differences can be explained by random variations in sampling. This implies they must be due to some underlying causal mechanism that is related to the token's position, although the analysis itself cannot reveal what that mechanism is.

7 Analysis of Token Positional Propensities

So far, we have looked at differences in the distribution of token lengths, but we have not yet looked at which tokens might account for the differences. We now look at particular tokens regardless of their lengths, and how the propensity of particular tokens differs between cohorts. Are the shifts in mean token lengths between cohorts due simply to the scribe choosing tokens based on their length alone? Or are the shifts an incidental consequence of choosing particular tokens more often, or less often?

To explore those questions further, we have taken the middle locations within the paragraphs of Voynichese as representing the "base" population of tokens. The assumption is that, if the scribe is influenced at all by position when choosing tokens, then regardless of whether that choice is driven more by semantic content, a cipher process, or some language simulation device, the positional consideration will be least when writing tokens in these middle locations. And so the MIDDLE cohort becomes our reference cohort. We have then looked at each unique token across the lexicon of the study corpus, and compared its frequency of occurrence within each of the subject cohorts —TOP, FIRST, LAST, BEFORE, and AFTER— to that within the MIDDLE cohort. We have then performed statistical significance tests to determine which tokens have a propensity to occur much more than expected ("affinitive"), or much less than expected ("aversive"), in each of the subject positions.

In the previous analysis, a *p*-value threshold was chosen to determine if the token length distribution at specific page positions differed significantly from that of the reference cohort. We saw that the basic result was unchanged over a wide range of potential *p*-value thresholds. Here, however, we are assessing a large number of unique tokens to determine for each, whether it indicates a propensity to be used by the scribe in various specific positions.

This 'propensity' should not be confused with 'prevalence'. Measures of the occurrence counts and relative frequencies of tokens, have often been reported by Voynich researchers. While that can give a useful quantitative measure of token prevalence in particular positions, by itself it is reductively simplistic —and potentially mis-

leading— for understanding usage propensity and whether the observed token usage is intentional due to behavioral or grammatical mechanisms, or an epiphenomenon due to sampling variations.

The better approach for our purpose is to consider the statistical significance and confidence level associated with a measurement of relative prevalence between positions before accepting it as an indication of propensity. We can do this by applying a binomial test, wherein we treat the occurrence of a particular token in the subject cohort as a binomial random variable whose probability is determined from the reference cohort. The binomial test, as an exact statistical procedure, does not presuppose the same conditions required by the χ^2 test, notably the need for the number of observations of the token to be of sufficient size.

We chose to use the same p-value threshold of 0.01, as used in the previous analysis, but to get a sense of the sensitivity of the results to that choice, we first performed a parametric analysis. Figure 6 shows, for each cohort, how the number of unique tokens considered to have a significant propensity would vary if we changed the choice of p-value threshold.

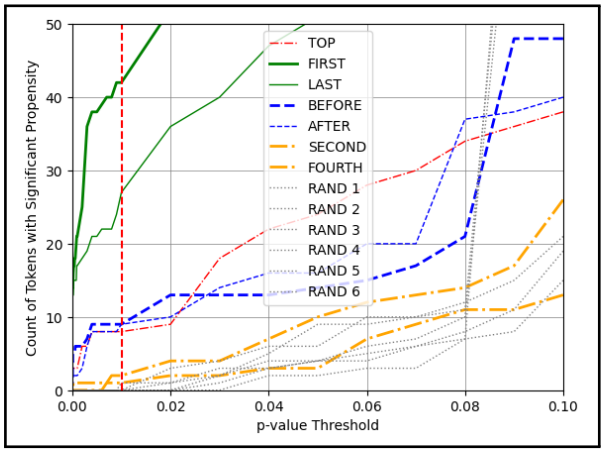


Figure 6: Count of Tokens with Significant Propensity: Variation with p-value Threshold

We can see in the figure that as we raise the threshold, more and more of the unique tokens pass the test of significance. This is not unexpected since no matter how different the observed count of any particular token is from its expected count, selecting a high enough threshold will render it explainable as random sampling error. But this trend also highlights the fact that the choosing p-value thresholds is always somewhat arbitrary.

An additional insight from Figure 6 is notable. The figure shows the validation cohorts, including six random and two positional cohorts (SECOND and FOURTH). Although similar in size to the smaller subject cohorts and comprising hundreds of unique tokens (see Table 1), they form a bundle of traces distinct from other traces in the plot, indicating few or no tokens with propensity at lower p-values. This reinforces the idea that the five subject cohorts significantly differ from the reference cohort in regards to usage of specific tokens with either higher (affinitive) or lower (aversive) propensity.

While the p-value tells us the probability of the observed data under the assumption of the null hypothesis (i.e. that the observed differences in the token counts in the subject and reference cohorts could happen by chance), it does not quantify the strength of evidence in favor of an alternative hypothesis (i.e. that the token's usage in a particular position, being governed by a different underlying mechanism, occurs with a different probability than in the reference cohort).

For this reason, we have included a Bayesian measure of significance, the Bayes Factor (B), which is an alternative to traditional frequentist methods that rely on p-values. The Bayes Factor quantifies the strength of evidence in favor of one hypothesis over another, and is defined by the ratio of the observed data's likelihood under a candidate hypothesis to its likelihood under the null hypothesis.

In our case, this Bayes Factor represents how much more likely it is that the scribe's token choices occur in the subject cohort according a different governing mechanism than that for his token choices in the reference (MIDDLE) cohort. The greater the Bayes Factor, the greater is our confidence in the hypothesis that the token truly has a highly affinitive, or highly aversive propensity to be used by the scribe in particular positions.

Categorical interpretations for the Bayes Factor as a measure of evidence strength have been suggested elsewhere (Wei et al., 2022) as follows:

$B < 3$	Worth only a bare mention
$3 \leq B < 20$	Positive
$20 \leq B < 150$	Strong
$B \geq 150$	Very Strong

While the p-value is bounded to the $[0,1]$ range, the Bayes Factor is unlimited with a range of $[0,\infty]$. So it is convenient to use the logarithm of B

instead. The above values of 3, 20, and 150 happen to correspond closely to $log(B)$ values of 1, 3, and 5 respectively (where $log(\dots)$ designates the natural logarithm).

To gauge whether a particular token has a positional propensity, we chose the very conservative approach of using *both* the binomial test and the Bayes Factor 'Very Strong' category. So for each cohort, we identified all tokens that resulted in both p-value≤ 0.01 and $log(B) \geq 5$.

Tilt	Token		Counts		Stats		
	Voynichese	Eva-	expected	observed	Propensity	*p*-value	*log(B)*
Affinitive	aiin	am	0	5	24.5	0.000002	>10
	chory	chory	0	4	19.6	0.000061	8.1
	oiin	om	0	4	19.6	0.000061	8.1
	y	g	0	4	19.6	0.000061	8.1
	sal	sal	0	5	12.2	0.000067	7.9
	chotaiin	chotaiin	0	4	9.8	0.000831	5.5
	dam	dam	1	14	9.8	0.000000	>10
	dy	dy	6	28	4.9	0.000000	>10
	dal	dal	3	14	4.3	0.000008	9.7
	dar	dar	5	15	3.2	0.000109	7.2
	daiin	daiin	36	59	1.6	0.000235	6.3
Aversive	chol	chol	37	1	0.0	0.000000	>10
	shol	shol	11	0	0.0	0.000019	>10
	sho	sho	9	0	0.0	0.000079	9.4
	chy	chy	15	2	0.1	0.000046	8.8
	chey	chey	9	1	0.1	0.000830	6.2
	chor	chor	22	2	0.1	0.000000	>10

Table 2: Tokens with Propensity for Last Position on a Line (Cohort: LAST)

All five of the positions represented by the subject cohorts were found to have certain tokens showing statistically significant propensity. This includes the positions that are intrinsic to the script itself (TOP, FIRST, and LAST), but surprisingly also the positions that are extrinsic in the sense that the position is defined by elements that are not part of the script (BEFORE and AFTER).

All of the identified tokens have been tabulated by position and by their tilt, either affinitive or aversive. Due to space limitations, only two example tables are included here (Tables 2 and 3).[6]

Note that the additional Bayes Factor criteria caused rejection of some of the tokens otherwise counted for in the traces in Figure 6, but ensured that the tables provide a catalog of only those tokens with very strong evidence of positional propensity that are also statistically significant.

[6]All of our study results, along with additional analyses, discussions, and the full set of tables, are available in the Supplemental Online Material at https://www.quantumlynxresearch.com/research.

We have also listed in the tables a quantified measure of propensity defined as the ratio of the best estimates of the probabilities of the token occurring in the subject vs the reference cohorts.[7]

Tilt	Token		Counts		Stats		
	Voynichese	Eva-	expected	observed	Propensity	*p*-value	*log(B)*
Affinitive	qotaiin	qotaiin	0	4	21.8	0.000040	8.5
	dy	dy	3	23	9.0	0.000000	>10
	dam	dam	1	5	7.8	0.000522	5.9
	s	s	7	19	2.8	0.000084	7.4
Aversive	chor	chor	10	0	0.0	0.000048	9.9
	chol	chol	17	3	0.2	0.000047	8.7

Table 3: Tokens with Propensity for Position Immediately Before a Drawing Intrusion (Cohort: BEFORE)

8 Conclusions

This study has produced several findings validated by formal tests of significance. They are not only significant in the formal statistical sense, but also in terms of their potential value toward understanding the nature and structure of the Voynichese script.

The overarching conclusion from this study can be summarized as:

> The distributions of the unique Voynichese tokens found in the Voynich Manuscript depend not only on their position within paragraphs and lines of script (intrinsic positioning), but also on position in relation to the hard boundaries imposed by the presence of drawings (extrinsic positioning).

More specific conclusions include:

1. Tokens immediately preceding the intrusion of a drawing on an otherwise continuous line of script, and tokens at the ends of lines, tend to be shorter than tokens located elsewhere.
2. Tokens in the top lines of paragraphs, at the beginnings of lines, and immediately following a drawing intrusion, tend to be longer.
3. Certain tokens have significant propensities to be either used or avoided by the scribe, depending on position.

[7]We derive estimates of these probabilities directly from a simple ratio of observation counts; no additive smoothing is used as it would distort true zeros which are legitimate for this sparse data.

4. Tokens exhibit positional propensity not only for positions dependent on the script itself, but for positions dependent on the drawings, which are extrinsic to the script.

A catalog of all tokens with significant positional propensity has been compiled for use in further research.

9 Implications

Of the several findings itemized in the previous section, the one regarding tokens adjacent to drawings is perhaps the most surprising. Anomalies in the tokens and glyphs found in the top lines of paragraphs and at the beginnings and endings of paragraphs have been reported before, although few of those efforts have considered the statistical significance of the reported observations. We are not aware, however, of any other efforts that have analyzed the Voynichese script in relation to positions that are dependent on elements extrinsic to the script itself.

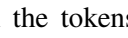
The finding that there exists significant propensity for particular tokens immediately before and after drawings is not only unprecedented, but may also have radical implications for understanding the nature of the manuscript, and particularly the Voynichese script.

A full discussion of potential implications is beyond the scope of this paper, but a couple of thought-provoking scenarios are worth mentioning.

Meaningful	**Known Language**	Readable by modern scholars if not everyone else
	Unknown (Lost) Language	No longer spoken or known
	Constructed language	Never commonly used, but manufactured for use by a closed group
	Stenography	A system of shorthand for a known or lost language
	Encryption	An encrypted form of a known or lost language
	Delusional Rant	Expressions of a writer who believes he is recording lucid thoughts but which make no sense beyond short range syntax
Meaningless	**Language Simulation**	Manufactured to look like meaningful text
	Gibberish	Random writing, without any semantic content at all

Table 4: Possible Language Variants Underlying the Voynichese Script. *This is a comprehensive list of possibilities; however, they are not all mutually exclusive.*

Table 4 summarizes possible language variants that could explain a script like Voynichese. Regardless of which of these is true, having tokens with a propensity related to their adjacency to drawings implies an unnatural coupling of elements that are not part of the script itself, either to the semantic content of meaningful script or to the syntactic process that produced meaningless script.

For 'meaningful' script, the most plausible of the variants would seem to be selective use of stenography, where the scribe resorts to short-hand alternatives when choosing tokens next to a drawing. It is still difficult, however, to explain why the alternatives chosen would be biased to a small set of particular tokens unless perhaps they are a form of punctuation that is not tied directly to any prescribed semantic content. (And even this would not explain the affinitive  and aversive  and).

For 'meaningless' script, either of the variants —a language simulation or gibberish— is conceivable, providing the scribe simply decided to make atypical choices for tokens when encountering a drawing, and did so with a preference to a favored list.

In any case, our explanations for the observed extrinsic propensities are highly speculative at this stage, and so further research is underway.

Finally it is worth mentioning that, while we have applied considerable rigor in our analysis, we do not rule out the possibility of an overlooked systemic error or other plausible explanation for our results, and we welcome review of the calculations, or suggestions regarding other interpretations of the results.

References

Viv Bewick, Liz Cheek, and Jonathan Ball. 2003. Statistics review 8: Qualitative data – tests of association. *Critical Care*, 8(1):46, December.

Claire L. Bowern and Luke Lindemann. 2021. The Linguistics of the Voynich Manuscript. *Annual Review of Linguistics*, 7(1):285–308, January.

Julian Bunn. 2022. Word Positions on the Folios. https://voynichattacks.wordpress.com/2022/12/26/word-positions-on-the-folios/.

Lisa Fagin Davis. 2020. How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript. *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies*, 5(1):164–180.

Mary E. D'Imperio. 1978. *The Voynich Manuscript: An Elegant Enigma.* United States: National Security Agency/Central Security Service.

Daniel E Gaskell and Claire L Bowern. 2022. Gibberish after all? Voynichese is statistically similar to

human- produced samples of meaningless text. In *CEUR Workshop Proceedings*, University of Malta, November.

William Romaine Newbold. 1928. *The Cipher of Roger Bacon*. University of Pennsylvania Press.

Regina Nuzzo. 2014. Scientific method: Statistical errors. *Nature*, 506(7487):150–152, February.

Gordon Rugg. 2004. An Elegant Hoax? A Possible Solutionto the Voynich Manuscript. *Cryptologia*, 28(1):31–46, January.

Jorge Stolfi. 2000. On the VMS Word Length Distribution.

Ronald L. Wasserstein and Nicole A. Lazar. 2016. The ASA Statement on *p* -Values: Context, Process, and Purpose. *The American Statistician*, 70(2):129–133, April.

Zhengxiao Wei, Aijun Yang, Leno Rocha, Michelle F. Miranda, and Farouk S. Nathoo. 2022. A Review of Bayesian Hypothesis Testing and Its Practical Implementations. *Entropy*, 24(2):161, January.

René Zandbergen. 2021. The Cardan grille approach to the Voynich MS taken to the next level, 10.48550/arXiv.2104.12548.

René Zandbergen. 2022. Analysis of the Illustrations. https://voynich.nu/illustr.html.

René Zandbergen. 2023. Text Analysis - Transliteration of the Text. https://voynich.nu/transcr.html.

Development of the Block Cipher LAMBDA1 in 1990

The Block Ciphers DES, GOST and LAMBDA1

Winfried Stephan
Mathematician, Retired
wstephan@mein.gmx

Abstract

In 1990, it became apparent that the German Democratic Republic (GDR) would leave the socialist community of states. This involved the gradual reduction of cooperation between the cipher services of these countries and the separation of cipher connections.

LAMBDA1 is a block cipher developed in East Germany in 1990. It was designed for a cipher device for which a Soviet algorithm was originally intended. The plan was to use a predecessor of the Soviet block cipher algorithm, called GOST. This now had to be replaced. The aim was to provide a cipher algorithm that could not be easily decrypted by either the Warsaw Treaty countries states or the NATO countries.

The background to these considerations was the assumption that the GDR would confirm to exist as an independent state for an extended period in a kind of transitional phase.

The article describes the circumstances under which the LAMBDA1 algorithm was developed in just one month. It was based on the results of previous projects and was then intensively analyzed.

The project was only abandoned when it became clear that the unification of the two German countries would take place at short notice and was imminent.

The algorithm below is described only to the extent necessary to understand the development process.

1 Publications on LAMBDA1 and T-316 after 1990

Like all state secrets of the German Democratic Republic, information about cipher algorithms and cipher machines was kept secret until 1992. With the dissolution of the GDR, the Stasi Record Archive (BStU) received all existing documents on the cipher service, with a few exceptions. They were registered there and only gradually made available to the public. Since then, this office has been integrated into the Federal Archives (Bundesarchiv).

The publication of the LAMBDA1 algorithm and the T-316 GO cipher device was made possible in particular by Jörg Drobick's many years of research into the cipher services of the GDR.

Programs to implement the algorithm are publically available, as well as a video tutorial showing encryption and decryption using the T-316 GO cipher device (Drobick 1989). Additionally, two more implementations are available: (CrypTool 2) and (github.com).

A bachelor thesis was written based on this material (Altenhuber 2018).

So while the technical details of LAMBDA1 have already been published, this paper explains the reasons and historical circumstances under which the algorithm was developed.

2 The Situation in 1990

Elections to the People's Chamber of the GDR were held on March 18, 1990. The majority of the newly elected parliamentarians voted in favour of unification of the GDR with the German Federal Republic (FRG). This marked the beginning of the unification process, although the timeframe was initially open. A

transition period of about two to four years was generally assumed.

It was therefore clear that the GDR would withdraw from international relations, in particular from the Council for Mutual Economic Assistance and the Warsaw Treaty.

It should also be noted that according to the regulations in force in the FRG, communications could not be secured using NATO procedures, as the GDR was still a member of the Warsaw Treaty. NATO technology was not allowed to be used in these countries.

3 Tasks of the Cipher Services

In the GDR, the Central Cipher Authority (ZCO) was responsible for the technical management and control of the ciphering facilities (CW). Until 1989, this was a department of the Ministry for State Security (MfS). It was then assigned to the GDR Ministry of the Interior in January 1990. After reunification, work continued on its dissolution. The ZCO remained under the control of the Ministry of the Interior of the Federal Republic of Germany until its final dissolution on December 31, 1990.

The tasks of the ZCO included the development, production and provision of new ciphering techniques and other cryptological procedures, means and methods for encrypting messages, in addition to providing guidance to the operational services.

Let's take a closer look at a development project that has been underway since the mid-1980s:

At that time the hardware basis for cryptographic machines was changing. Mainframe computers were used in computer centers and machines with CPU-based cipher implementations were developed. In addition, a corresponding algorithm was needed for commercial applications. New algorithms were needed for this and an analogue to the American block cipher algorithm (FIPS77) was already under development in the Soviet Union (GOST89).

As a general rule, the methods and algorithms used at the level of state secrets should not be used on a large scale in the commercial sector. As in the USA there should be a separation between encryption methods used in the government or military sector and those used in the commercial sector.

The predecessor of the new cipher was developed by the Russian State Security Committee (KGB) in the 1970s. As early as spring 1984, Soviet cryptologists informed the GDR about their work on this block cipher algorithm in two lectures: "Properties of the Data Cipher Algorithm (DCA)" (1986) and (1987) (Killmann W., Stephan W. 2024 Appendix A). It is the forerunner of the algorithm which is now known as GOST No. 28147-89 (GOST89)[1].

It should be used as a block cipher algorithm in the socialist countries. In the GDR, it was therefore assumed that in the future the Soviet standard would be used if necessary. The algorithm was still classified as "Top Secret" until 1990.

The use of DCA or GOST would have required the permission of the Soviet Union. However, this was no longer to be expected due to the political situation.

The process of unbundling between the socialist states also included the disentangling of relations between the encryption services of these countries.

For example, all cipher machines used by the armies of the Warsaw Treaty had to be handed over by the GDR army to the Soviet army. This also concerned confidential documents on cooperation with the Soviet cipher services, encryption algorithms and encryption devices that were in use or under development at the time. This also included all existing documents relating to the DCA and GOST.

This situation is described below using the development of the T-316 data cipher machine as an example. For this device the use of GOST was intended.

The development of the devices was already well advanced. The T-316 was one of the first encryption devices in the GDR to use a programmable microprocessor system. This made it possible to implement various encryption algorithms, taking into account the technical framework. The T-316 was designed and built

[1] From this point on, GOST will be used as an abbreviation for GOST No. 28147-89.

for both civilian and military use. The civilian version, called the T-316 GO (Figure 1), was fitted in an attaché case and the military version T-316 M in a metal housing.

The development of the T-316 devices must have begun before 1987, because the BStU already possesses a document from 19 August 1987, entitled *"Demand planning for the T-316 and T-314 cipher devices under development"* (ZCO1 1987).

Figure 1: The civilian variant T-316 GO
Source http://scz.bplaced.net/316.html

By 1990, ten T-316 GO units had already been produced. It was also planned to produce another 50 units in 1990 and 850 units in 1991.

In addition, the Czechoslovakia was to receive a loaner device for testing, which indicates that it was intended to provide T-316 to other socialist countries (Drobick 1990).

In this situation, an encryption algorithm other than DCA or GOST had to be found for the T-316 GO in a very short period of time. (Drobick 1989). This explains the time pressure under which the LAMBDA1 algorithm was developed.

It is noteworthy that in the BStU documents analyzed so far, GOST or its predecessor DCA do not appear as the algorithm actually intended for the T-316 device. It seems that the secrecy of these facts has been maintained until today. Therefore, this information is new and comes from the developers and the author of this paper.

4 DES as Candidate for Replacement

The first candidate for replacement was the internationally renowned American DES.

The American data encryption standard DES (FIPS 77) was the most widely used standard for non-classified data encryption at the time.

However, since its publication, there have been concerns about its security. The ZCO has been monitoring and analyzing publications on DES since at least the early 1980s.

Let us recall some of the design features of this block cipher.

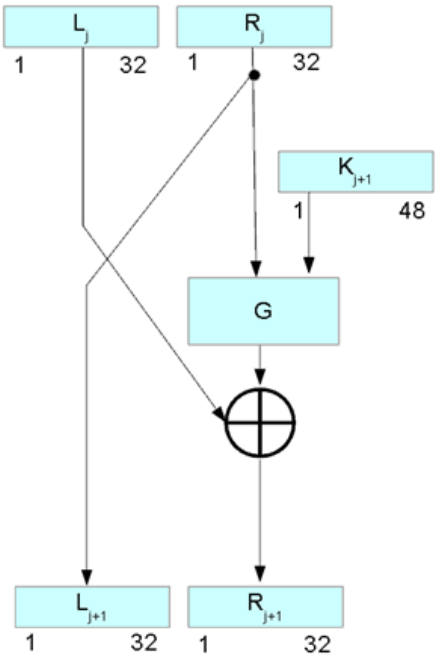


Figure 2: The round function of a Feistel cipher

Figure 2 outlines the basic principle of a round of a Feistel cipher. The DES algorithm performs 16 of these rounds.

An important component of the round function is the so-called Feistel function *G*. The construction of this function corresponds to the principle of diffusion and confusion established by Claude Shannon (Figure 3).

In an assessment *Possibilities and dangers of using the DES* (ZCO2.1990) the following three potential weaknesses were highlighted:

• 56 bits for the key are already in 1990 not enough. It is conceivable that the most powerful decryption services, using all the possibilities of science and technology (special hardware, parallelization, etc.), will realize the TPM (total trial method - brute force) with an effort that goes to the limits of their capacity.

• It is unknown whether decryption services are aware of the laws of DES (trapdoors), which – under any real assumptions – require much less effort than trying out 2^{56} variants.

• In the DES literature in 1990, it was speculated that the S-boxes and key scheduling might contain trapdoors.

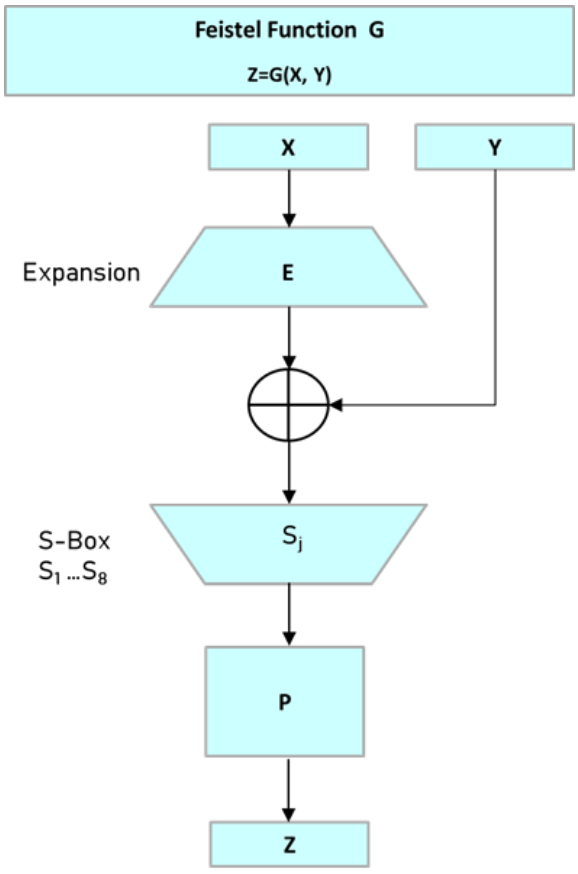


Figure 3: The Feistel function *G* for the DES

The following conclusions were drawn from this:

The possibility of using the DES as an alternative to GOST was considered for applications with lower security requirements. For state secrets in particular, it was ruled out on the basis of the assessments made, in particular because of the small size of the key. A potential decryption – by whomever – must be ruled out.

There was also no other acceptable, rapidly deployable method for this area, especially as it was difficult to assess whether other services could decrypt such an alternative algorithm. The FEAL and MASSEY algorithms were discussed. They were still quite new at the time and there were only a few publications on their cryptanalysis.

The way out was to develop them in-house, officially based on DES, unofficially using the knowledge of the GOST. The starting point was a Feistel cipher, which was the common basis for both block algorithms.

Only the structure of the round function, the S-boxes, the expansion function and the number of 16 rounds were taken over from the DES.

5 GOST as Source for LAMBDA1

As early as 1984, East German cryptologists performed a control evaluation of the Soviet block cipher algorithm. It was based on a comparison with the published articles on DES. The knowledge gained was very useful in 1990 when an own algorithm had to be constructed.

Official sources on the DCA have not yet been found. However, there are no significant cryptographically relevant differences between the GOST algorithm and its predecessor.

Only features of the GOST that were used in the design of LAMBDA1 are listed here. The full description of GOST can be found on the website cited in the source (ZCO3).

The cipher standardized under GOST 28147-89 is a Feistel cipher with a key length of 256 bits, which corresponds to 32 8-bit characters. It is the Soviet counterpart to the Data Encryption Standard (DES).

The block length was also equal to 64 bits. At 32, the number of rounds was twice that of DES.

The key space is much larger and therefore the use of TPM (total trial method - brute force) is practically impossible.

Key scheduling is solved differently for GOST than for DES:

The first eight 32-bit round keys K_i are obtained from the key K dividing them into eight blocks, i.e. $K = (K_8, K_7, K_6, K_5, K_4, K_3, K_2, K_1)$, the round keys K_9 to K_{16} and K_{17} to K_{24} correspond again to the keys K_1 to K_8, the last eight round keys K_{25} to K_{32} are the first round keys in reverse order. All in all, we get the key order K_1, ... , K_8, K_1, ... , K_8, K_1, ... , K_8, K_8, ... , K_1.

The analogues to the S-boxes in the DES were eight interchangeable permutations P, each of which was used to realize four bit substitutions. They could also be used as secret key elements (long-term keys).

Instead of the bitwise addition of the binary vectors in the individual rounds in the DES, the addition *mod* 2^{32} was used.

6 LAMBDA1 Key Space and used Round Keys

The exact definition of the LAMBDA1 algorithm can be found in (ZCO3). A brief overview of this description is given in the chapters 6 to 9.

The key space for LAMBDA1 and the round key generation is borrowed from DCA/GOST. This was obvious as they also use 256 bits and the T-316 was already prepared for this key size. The key consists of 32 characters (S) of eight bits each (B):

$$S=(S1,S2,\ldots,S32)=(B_1,B_2,\ldots,B_{256})$$

The function T^{11} is a cyclic shift of a 48-bit vector. In the 16 round keys, only 192 bits are used. Accordingly, they are cyclically shifted by 11 bits each (see equations below). The keys K_{17} and K_{18} are used only once in the middle of the calculation in round 8. They are only 32 bits long.

The bits are assigned to the rounds according to the following rule:

$$K_1 := (B_1, \ldots B_{48})$$
$$K_2 := (B_{49}; \ldots B_{96})$$
$$K_3 := (B_{97}, \ldots B_{144})$$
$$K_4 := (B_{145}, \ldots B_{192})$$
$$\forall j \in 5,12: \quad K_j := T^{11}(K_{j-4}) \ldots$$
$$\forall j \in 13,16: \quad K_j := T^{11}(K_{25-j}) \ldots$$
$$K_{17} := (B_{193}, \ldots B_{224})$$
$$K_{18} := (B_{225}, \ldots B_{256})$$

The key S should be used as a time key and be valid for 7 days.

7 LAMBDA1 The modified Feistel Function

Until 1990, it was always assumed that trapdoors were hidden in the DES due to the design of the round function in conjunction with key scheduling. These would have led to a reduction in decryption effort. For this reason, the Feistel function for LAMBDA1 was modified. Only the expansion function and S-boxes are adopted from the Feistel function of the DES.

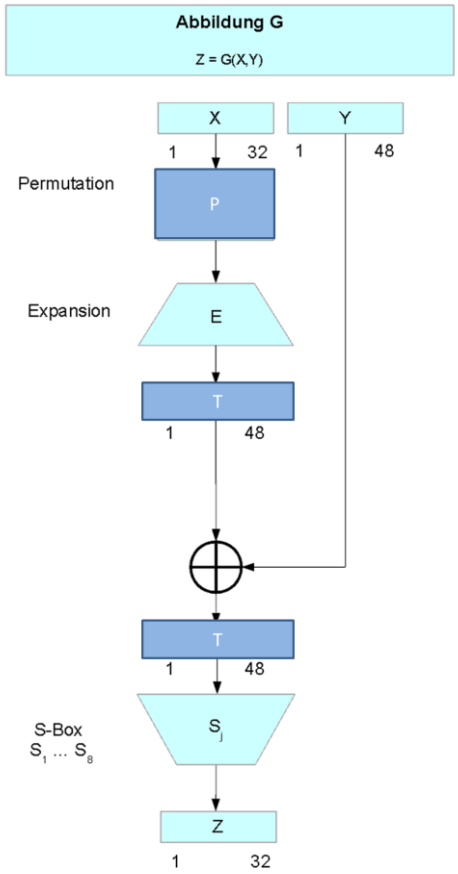


Figure 4: The modified Feistel function LAMBDA1

The idea of integrating a permutation P is retained. It is now positioned before the expansion function E. In addition, two cyclic shifts T of length 48 have been added. The additions are highlighted in darker color in Figure 4.

There was a well-founded hope that this change would eliminate all potential built-in weaknesses.

In addition, care was taken to ensure that the findings from the analysis of the round function of DES and GOST could also be applied to the modified version of LAMBDA1 (Chapter 10).

8 LAMBDA1 The Special Role of the 8th Round

The eighth round occupies a special position (Figure 5). It is therefore modified.

Here, in addition to the usual described round function, an additional round key (K_{17},K_{18}) with a length of 32 bits each is added using two additions *mod*2^{32} (also marked in darker color in Figure 5). This effects the mathematical description of the algorithm. The dependencies of the variables are no longer just binary. The idea for this came from the knowledge about the GOST. The idea was to add another option to neutralize possible trapdoors in the S-boxes of the DES. Due to the additions of this other addition, the K_{17} and K_{18} affect different bits in their sum depending on the contents of the 32-bit vectors to be summed.

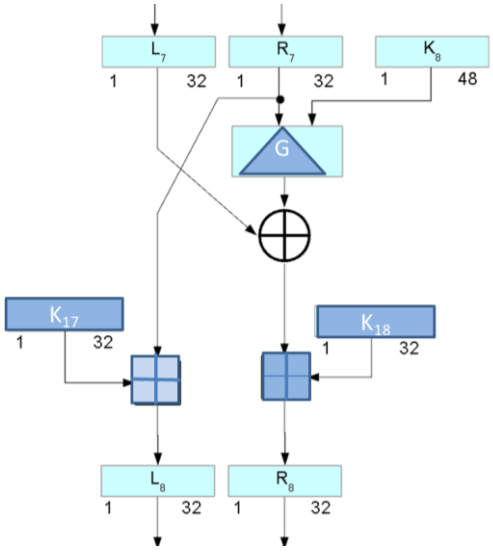


Figure 5: The eighth round of LAMBDA1

If a crypto attack has progressed to the middle round and information about the round keys is already available, then a barrier should be set up here, which would then also have to be overcome. For example, consider a scenario such as a meet-in-the-middle attack.

9 The Development Outcome

The algorithm was handed over to the developers of the T-316 device for technical implementation on March 12th, 1990.

Figure 6 illustrates the basic structure of the algorithm.

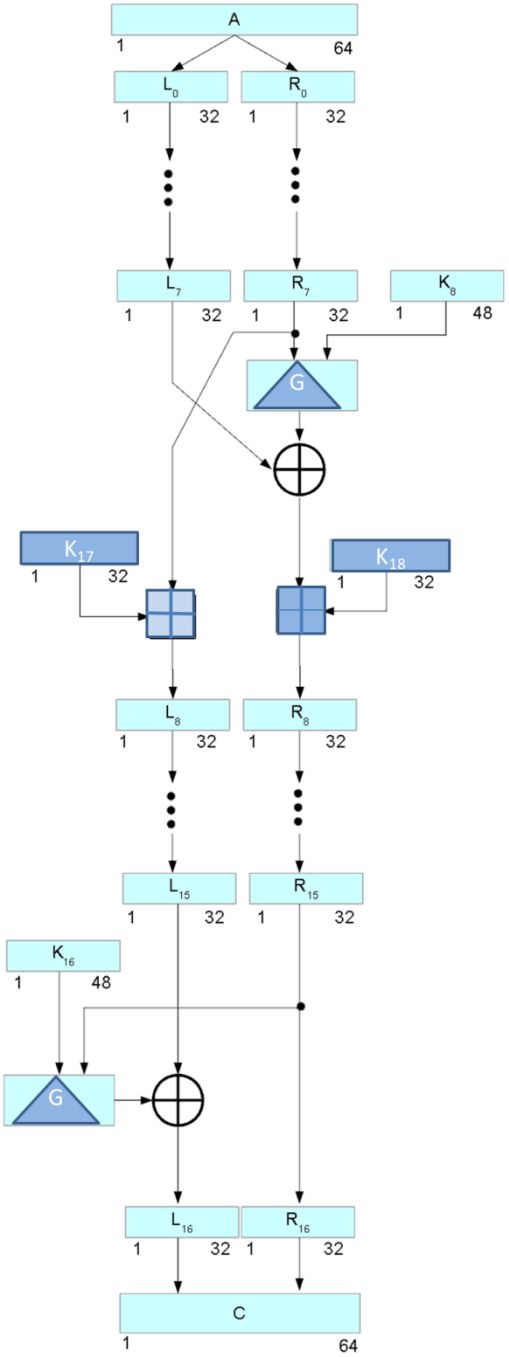


Figure 6: The Algorithm LAMBDA1[2]

[2] All figures are based on the sketches in, *DESCRIPTION LAMBDA1* (ZCO3 1990).

The algorithm was developed within about three weeks in February and March 1990. The following changes were made to the DES:

• The effective key length was increased from 56 bits to 256 bits.

• The way in which the key sequence is generated for the 16 round keys with a length of 32 bits each has been changed (section 6).

• The number of rounds is the same as the DES with 16 rounds.

• The 8th round has been modified by inserting two additional key vectors and adding them *mod* 2^{32}.

• Omission of the initial and final permutations of the DES. It is well known from the literature that the initial and final permutations in DES have no cryptographic relevance. That's why they were simply omitted.

• There is a built-in cryptographic reserve that can be activated if the permutation is kept secret. In this context, the term "long-term key" (commutator/permutation *P*) was introduced.

• If weaknesses or trapdoors were to appear, it would be conceivable to counteract them with this permutation.

• Different permutations P can also be used to define algorithms for different application areas and to separate them from each other.

• Obviously, the basics of GOST were sufficiently camouflaged. This was a desirable side effect at the time. The design results in the desired difference to both DES and GOST.

• Cryptographic attacks should be prevented, or at least made more difficult, for the intelligence services, which may be well versed in DES.

A thorough and comprehensive development analysis was not possible due to time constraints. To minimize the risk of use, tried and tested basic structures (Feistel ciphers) were used. Once development was complete, a control analysis was carried out with the resources available.

10 Analysis of the LAMBDA1 Algorithm

The cryptologic analysis began in mid-March and was interrupted after an inventory report in June 1990. Nine graduate mathematicians were involved, but they were only able to devote about 30% of their working time to the task. This means that about 180 hours of analysis work were invested.

Due to the chosen design, it was possible to use all known publications on the DES in the GDR in 1990 for the evaluation of LAMBDA1. This gave some assurance that the cryptographic properties of the algorithm were of good quality. The cryptology group's experience from years of development and analysis activities could also be successfully applied.

According to the report *Assessment Report LAMBDA1* (ZCO4 1990), in the short time until 22nd June, about 70 sources were examined and evaluated in terms of LAMBDA1 by these employees. From a cryptological point of view, the following results are worth mentioning, which are reproduced here in abbreviated form:

1. Following the DES-like functions examined in Even S., Goldreich O. (1983), studies of the group of LAMBDA1-like functions have been shown that the alternating group is present.

2. All weak and semi-weak keys with palindrome properties (Simmons G. J., Moore J. H. 1987) have been clearly identified. However, in relation to the key size of 2^{256}, their occurrence is orders of magnitude lower than in the DES.

3. Statements have been proven that significantly restrict the number and type of automorphisms in a basic automaton model of the LAMBDA1 algorithm. This excludes several classes of algebraic structures that could possibly be used cryptologically.

4. Double transitivity is an essential algebraic property of finite groups that should be proved for block ciphers. The verification of this property is also of interest for the analysis of LAMBDA1. So far, the theoretical basis for a computerized test has been worked out.

5. There are a number of different statistical studies in the literature that attempt to discover undesirable regularities in the cipher algorithm (Leung A. K., Tavares S. E. 1984). Experiments on the avalanche effect and the strict avalanche effect and on the testing of special bit dependencies have been planned and partly started. Due to the low performance of the

available computers, only a few tests could be carried out, a final evaluation is currently not possible. However, the tests so far do not show any abnormalities.

The results show that the specialists of ZCO were able to apply the experience gained from other analyses here and to achieve comparable results.

In 1990, the report concluded that there were no objections to use LAMBDA1 to protect classified information up to Top Secret (GVS) until Q1/91. The authorization was granted for a limited period of time, as a control analysis was planned in the near future.

The algorithm was proposed for implementation in a T-316 cipher despite the very short development and analysis period.

However, the LAMBDA1 evaluation report of June 1990 quoted above became the final report due to the political events.

The three literature sources cited here are examples of the state of knowledge at the time.

11 LAMBDA1 is not used

On 24 July 1990, the President of the Central Office for Information Security (Zentralstelle für Sicherheit in der Informationstechnik, ZSI) and future President of the Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik, BSI), Dr. Otto Leiberich, made a business trip to Dahlwitz-Hoppegarten to visit the ZCO of the (still) GDR.

There is an internal protocol of the ZCO in which the conclusions from this visit are summarized. An excerpt from this *"Protocol of the ZCO on the Conclusions from this Business Trip of August 1, 1990"* (Killmann, Stephan. 2024 Appendix F) documents main objectives of this trip:

• Specification of the tasks of the ZCO for the territory of the GDR and, after unification, for the territory of the former GDR in a transitional period.

• Continuation of scientific and scientific-technical work, profiling of specialists to carry out analysis work for the evaluation and certification of information technology systems.

In this context, measures to continue work on T-316 and LAMBDA1 are also listed:

• T-316 Completion of the production introduction, preparation of the acceptance of the equipment by the manufacturer, preparation of the use of the equipment by the user, deadline: 11/90,

• Implementation of the framework agreement with Steremat-GmbH for the commercial use of the device, deadline: 8/90,

• Continuation of work on the development and analysis of the LAMBDA1 and DELTA algorithms and inclusion of the FEAL and MASSEY algorithms in the studies, deadline: 11/90.

The points quoted from the minute's show that the ZCO intended to continue working on the algorithm and the devices, also in coordination with or at least with the knowledge of the ZSI.

It is quite clear, that if the unification had not taken place so quickly, LAMBDA1 would probably have been used by the GDR to secure its communication; especially since ready-to-use T-316 GO devices were already available.

However, political developments in the GDR led to a rapid unification of the two German states. In the July and August, a government crisis developed in the GDR. This was accompanied by a rapid decline in the GDR economy. There was no longer any talk of a transitional period of about two to four years, as had been discussed in political circles. The system changed rapidly.

The two German states were unified on 3 October 1990. The LAMBDA1 block cipher algorithm was not used after that, nor was the T-316 device.

Acknowledgments

The author would like to thank Wolfgang Killmann and Franz-Peter Heider for fruitful discussions and support. The author would like to thank Jörg Drobick for publishing interesting information about the GDR cipher service on his website.

References

Michael Altenhuber. 2018. *Analyse und Implementierung der DDR-Chiffriermaschinen T-310/50 und T-316* Bachelorarbeit Nr. 1510239014
Analysis and implementation of the GDR cipher machines T-310/50 and T-316

CrypTool. 2.1 (Stable Build 9589.1) Programm https://www.cryptool.org/en/ct2/

Github.com. *https://github.com/tassadarius/LAMBDA1/tree/master/docs* (visited on 2023-11-20)

Jörg Drobick. *1990. Extensive documentation on the LAMBDA1 algorithm*
http://scz.bplaced.net/des.html#lambda
(visited on 2023-11-08)

Jörg Drobick. 1989. Informationen zu *T-316 GO*. *http://scz.bplaced.net/316.html*
(visited on 2023-11-08)

Shimon Even and Oded Goldreich. 1983. *DES-like function can generate the alternating group* - IEEE Trans. Inf. Theory, 1983, Vol. IT-29, No. 6, pp. 863-865

A. K. Leung and Safford E. 1984. *Sequence complexity as a test for cryptographic systems* - Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings.

G. J. Simmons and J. H. Moore 1987. *Cycle Structure of the DES for keys having palindromic (or anti-palindromic) sequences of rounds keys.* - IEEE Trans. Software Eng., 1987, Vol. SE-13, No. 2, pp. 262-273

Wolfgang Killmann and Winfried Stephan. 2024. *Das DDR-Chiffriergerät T-310: Kryptographie und Geschichte*. Springer Verlag, ISBN 978-3-662-67584-7
The GDR cipher device T-310: Cryptography and history

FIPS77. Federal Information Processing Standard Publication No. 46, January 1977. *DES Data Encryption Standard*
International Standard ISO 8372. 1987-08-15 Information processing - *Modes of operation for a 64-bit block cipher algorithm*

GOST89. *Encryption, Decryption and Message Authentication Code (MAC)*
https://www.rfc-editor.org/rfc/rfc5830
(visited on 2023-12-11)
Russian Federal standard for electronic encryption, decryption, and message authentication algorithms GOST 28147-89

ZCO1. 1987. *Bedarfsplanung zu den in der Entwicklung befindlichen Chiffriergeräten T-316 und T-314*, 19.08.1987 in BStU-ZAIG 25862
Demand planning for the T-316 and T-314 cipher devices under development

ZCO2. 1990. *Möglichkeiten und Gefahren der Nutzung des DES*, 20.02.1990
http://scz.bplaced.net/des.html#lambda1,
(visited on 2023-11-20)
Possibilities and dangers of using the DES

ZCO3. 1990. *LAMBDA1 64bit Blockchiffrierung, BESCHREIBUNG LAMBDA1*, 04.04.1990
http://scz.bplaced.net/des.html#lambda1,
(visited on 2023-11-20)
LAMBDA1 64bit block cipher, description

ZCO4. 1990. Referat 21. *Sachstandsbericht LAMBDA1* 22.06.1990;
http://scz.bplaced.net/des.html#lambda,
(visited on 2023-11-20)
LAMBDA1 Inventory Report

Cryptology and redaction – a strange symbiosis

Dermot Turing
Kellogg College
60-62 Banbury Road
Oxford OX2 6PN UK
dermotturing@btinternet.com

Abstract

This paper explores the relationship between cryptology and redaction. Redaction can be a frustration to historical cryptology research. Examples of redactions of historical papers relevant to cryptology are presented. It is concluded that the practice of redaction is often ineffective and the policy rationale behind redactions difficult to understand.

1 Introduction

Cryptography is concealment: concealment of content of a communication, while the communication itself is overt. For historians of cryptology, the concealment can go further than the content of communications. The processes of cryptography and cryptanalysis may themselves be secret, requiring a further layer of obfuscation, created through security laws, non-disclosure contracts and censorship. When disclosures are allowed, they are frequently partial, with documentation released into the public domain only with redactions. This paper examines the interplay between redaction and cryptology.

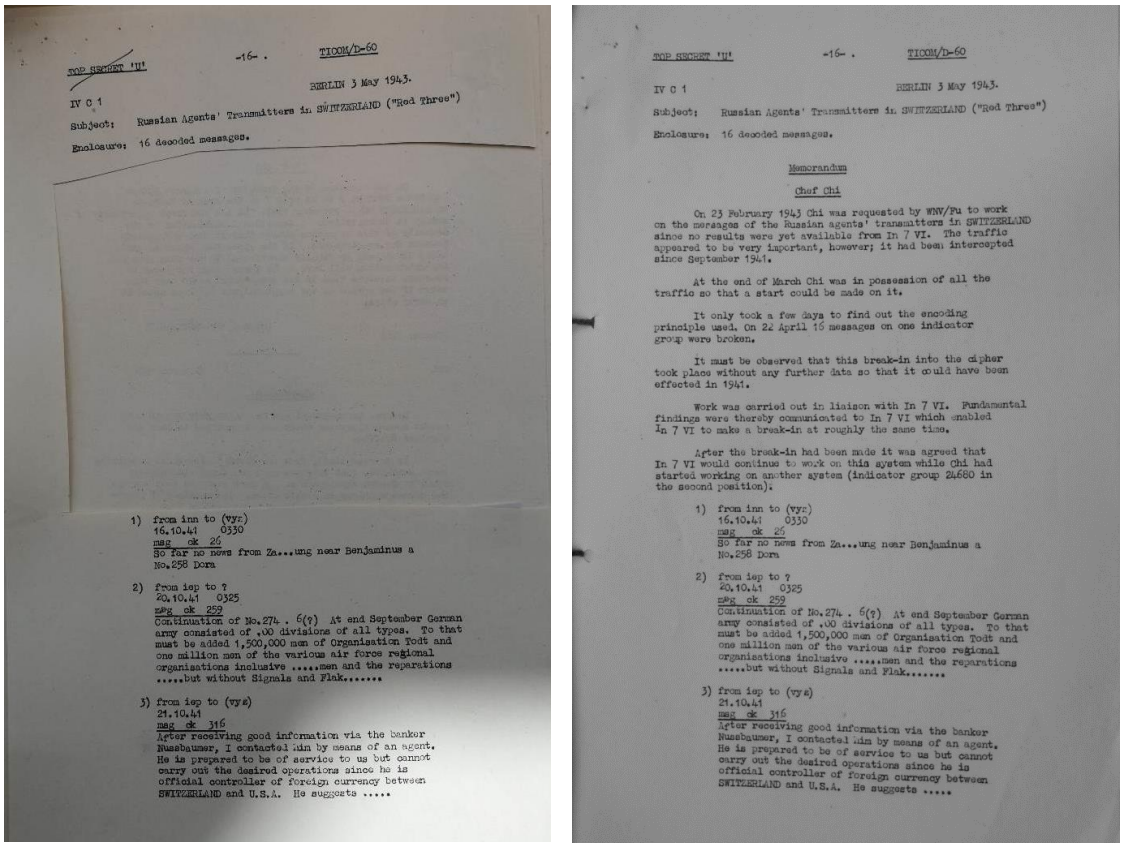
State-imposed secrecy concerning the cryptologist's art is probably as old as the art itself. For the last hundred years it has become increasingly difficult for state authorities to deny or obscure the existence of official cryptanalysis. Public demand can encourage disclosure of historical documents, but disclosure is rarely comprehensive. In different countries, there are different policy objectives and different standards. Some documents are withheld, others redacted. The UK's Public Records Act 1958 requires that public records be transferred to the National Archives within 20 years of their creation, unless they 'are required for administrative purposes or ought to be retained for any other special reason'.[1] 'Retained' is an expression broad enough to include redaction. In the United States, redaction may be justified under section 3605 of the National Security Agency Act of 1959 (50 USC 3605), which states: '… nothing in this chapter or any other law … shall be construed to require the disclosure of the organization or any function of the National Security Agency … or of the names, titles, salaries, or number of the persons employed by such agency.'

Redaction is, evidently, integral to the process of declassification of historical records. Declassification has previously been covered extensively in the literature (summaries can be found in Bennett, 2002, and on the NSA's 'Declassification and Transparency Initiatives' webpage[2]). However, this paper is not about declassification per se: its aim is to consider redactions in the field of cryptology, and to highlight certain curious relations between the two subjects.

2 Redaction of Cryptological Papers

To understand the theory and practice of redaction as applied to cryptology, one may consider a few examples, chosen from among the many instances which researchers encounter. The first is a file seized at the end of World War 2 by the Allied TICOM squads sent to Germany to obtain materials and information relating to German cryptanalytic capabilities (Rezabek, 2016). This was numbered T-1650 by the TICOM registry and, many years later, returned to Germany as part of a collection now in the

[1] Section 3(4).
[2] https://www.nsa.gov/Helpful-Links/NSA-FOIA/Declassification-Transparency-Initiatives/ (accessed 3 April 2024), which has links not only to declassified material but also to policy memoranda.



TOP SECRET 'U' -16- TICOM/D-60

IV C 1 BERLIN 3 May 1943.

Subject: Russian Agents' Transmitters in SWITZERLAND ("Red Three")

Enclosure: 16 decoded messages.

1) from inn to (vyz)
16.10.41 0330
msg ck 26
So far no news from Za...ung near Benjaminus a
No.258 Dora

2) from isp to ?
20.10.41 0325
msg ck 259
Continuation of No.274 . 6(?) At end September German army consisted of .00 divisions of all types. To that must be added 1,500,000 men of Organisation Todt and one million men of the various air force regional organisations inclusive .....men and the reparations .....but without Signals and Flak.....

3) from isp to (vyz)
21.10.41
msg ck 316
After receiving good information via the banker Nussbaumer, I contacted him by means of an agent. He is prepared to be of service to us but cannot carry out the desired operations since he is official controller of foreign currency between SWITZERLAND and U.S.A. He suggests .....

TOP SECRET 'U' -16- TICOM/D-60

IV C 1 BERLIN 3 May 1943.

Subject: Russian Agents' Transmitters in SWITZERLAND ("Red Three")

Enclosure: 16 decoded messages.

Memorandum

Chef Chi

On 23 February 1943 Chi was requested by WNV/Fu to work on the messages of the Russian agents' transmitters in SWITZERLAND since no results were yet available from In 7 VI. The traffic appeared to be very important, however; it had been intercepted since September 1941.

At the end of March Chi was in possession of all the traffic so that a start could be made on it.

It only took a few days to find out the encoding principle used. On 22 April 16 messages on one indicator group were broken.

It must be observed that this break-in into the cipher took place without any further data so that it could have been effected in 1941.

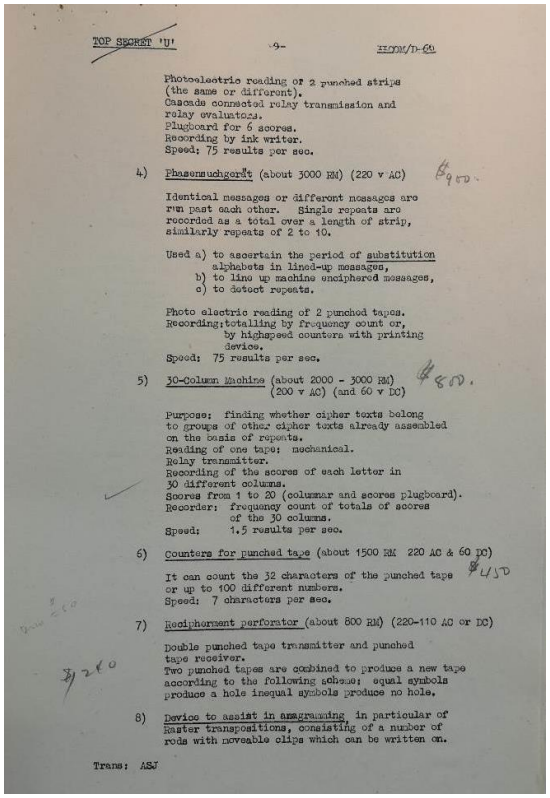
Work was carried out in liaison with In 7 VI. Fundamental findings were thereby communicated to In 7 VI which enabled In 7 VI to make a break-in at roughly the same time.

After the break-in had been made it was agreed that In 7 VI would continue to work on this system while Chi had started working on another system (indicator group 24680 in the second position).

1) from inn to (vyz)
16.10.41 0330
msg ck 26
So far no news from Za...ung near Benjaminus a
No.258 Dora

2) from isp to ?
20.10.41 0325
msg ck 259
Continuation of No.274 . 6(?) At end September German army consisted of .00 divisions of all types. To that must be added 1,500,000 men of Organisation Todt and one million men of the various air force regional organisations inclusive .....men and the reparations .....but without Signals and Flak.....

3) from isp to (vyz)
21.10.41
msg ck 316
After receiving good information via the banker Nussbaumer, I contacted him by means of an agent. He is prepared to be of service to us but cannot carry out the desired operations since he is official controller of foreign currency between SWITZERLAND and U.S.A. He suggests .....



TOP SECRET 'U' -9- TICOM/D-60

Photoelectric reading of 2 punched strips (the same or different).
Cascade connected relay transmission and relay evaluators.
Plugboard for 6 scores.
Recording by ink writer.
Speed: 75 results per sec.

4) Phasenuchgerät (about 3000 RM) (220 v AC)

Identical messages or different messages are run past each other. Single repeats are recorded as a total over a length of strip, similarly repeats of 2 to 10.

Used a) to ascertain the period of substitution alphabets in lined-up messages,
b) to line up machine enciphered messages,
c) to detect repeats.

Photo electric reading of 2 punched tapes.
Recording: totalling by frequency count or, by highspeed counters with printing device.
Speed: 75 results per sec.

5) 30-Column Machine (about 2000 - 3000 RM) (200 v AC) (and 60 v DC)

Purpose: finding whether cipher texts belong to groups of other cipher texts already assembled on the basis of repeats.
Reading of one tape: mechanical.
Relay transmitter.
Recording of the scores of each letter in 30 different columns.
Scores from 1 to 20 (columnar and scores plugboard).
Recorder: frequency count of totals of scores of the 30 columns.
Speed: 1.5 results per sec.

6) Counters for punched tape (about 1500 RM 220 AC & 60 DC)

It can count the 32 characters of the punched tape or up to 100 different numbers.
Speed: 7 characters per sec.

7) Reciphorment perforator (about 800 RM) (220-110 AC or DC)

Double punched tape transmitter and punched tape receiver.
Two punched tapes are combined to produce a new tape according to the following scheme: equal symbols produce a hole unequal symbols produce no hole.

8) Device to assist in anagramming in particular of Raster transpositions, consisting of a number of rods with moveable clips which can be written on.

Trans: ASJ

Political Archive in Berlin.[3] TICOM translated the file's contents into English and circulated the translation as document D-60 among the American and British signals intelligence services. Copies of D-60 are available at the US National Archives Records Administration and the UK National Archives.[4] In the British copy, pages 8 and 9 have been excised completely, retained under the Public Records Act. These pages are, however, open to view in the American copy, but page 16 is heavily cut from the American version notwithstanding that the page is completely open to view in the British copy (see Exhibit 1). The differences are, perhaps, surprising, given the oft-celebrated cooperation between the US and UK signals intelligence services (Smith, 2022).

Such inconsistencies allow us to consider the content of the excised passages. The pages cut from the British copy of D-60 in 2004 concern eight types of machinery invented by German

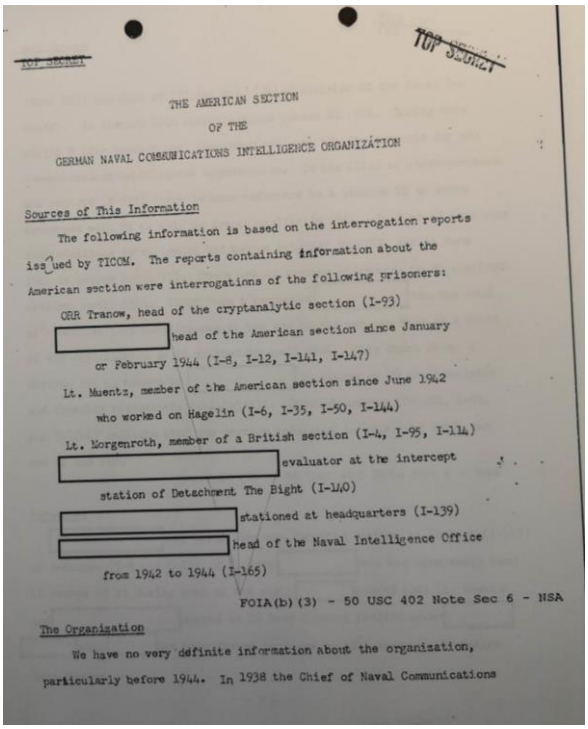
[3] Auswärtiges Amt, Berlin, Collection S8.

[4] NARA RG 457 Entry P4 Box 8; UK HW 40/174.

codebreakers to simplify attacks on superenciphered codes. Little technical detail is included in these pages. The heavily cut page 16 of D-60, cut out when the document was declassified in 2009, concerned the breaking of messages of a group of agents in Switzerland feeding intelligence on Germany to the Soviet Union (but not the decoded messages themselves). Why these subjects appeared to be sensitive to the different authorities, well after the discontinuance of code systems vulnerable to the machine methods outlined in D-60 and public knowledge of the German success against the agents' messages (Flicke, 1957) is hard to comprehend.

The next example relates to lesser redactions, where only individual words or phrases are covered up. One comes from another TICOM-related paper where people's names have been concealed (see Exhibit 2). Keeping people's names confidential might indicate a good rationale for redaction. But enough information remains to allow the researcher to fill in the blank spaces: the source documents (I-8, I-12, etc) are cited in the document, and these are publicly available, enabling the missing names to be reconstructed (here Schulze, Biege, Holtermann, and von Baumbach). It is unclear why these individuals' names were to be obscured, when that of Wilhelm Tranow, the German Navy's premier cryptanalyst who personally broke many Allied naval codes in World War 2 to devastating effect, was not. Perhaps the answer is that Tranow's achievements had been in the public domain for over thirty years by 2011 when this document was declassified (Kahn, 1978).

Similar examples can be found in the papers of the celebrated American cryptologist, William F. Friedman, which were declassified by the NSA in November 2014 and are available to view on its website in redacted form. Here, uniform length of each letter or number in the redacted telexed original facilitates a letter-count as a clue to the missing element. The potential solution can be checked by referring to other declassified documents, which freely give the names of the relevant personnel. Furthermore, so far from concealing information, the redaction has actually supplemented it. The redaction note references 50 USC 3605, implying that the persons whose names were concealed were linked to the NSA (see Exhibit 3).



TOP SECRET

TOP SECRET

THE AMERICAN SECTION
OF THE
GERMAN NAVAL COMMUNICATIONS INTELLIGENCE ORGANIZATION

Sources of This Information

The following information is based on the interrogation reports issued by TICOM. The reports containing information about the American section were interrogations of the following prisoners:

ORR Tranow, head of the cryptanalytic section (I-93)

[redacted] head of the American section since January or February 1944 (I-8, I-12, I-141, I-147)

Lt. Muentz, member of the American section since June 1942 who worked on Hagelin (I-6, I-35, I-50, I-144)

Lt. Morgenroth, member of a British section (I-4, I-95, I-114)

[redacted] evaluator at the intercept station of Detachment The Bight (I-140)

[redacted] stationed at headquarters (I-139)

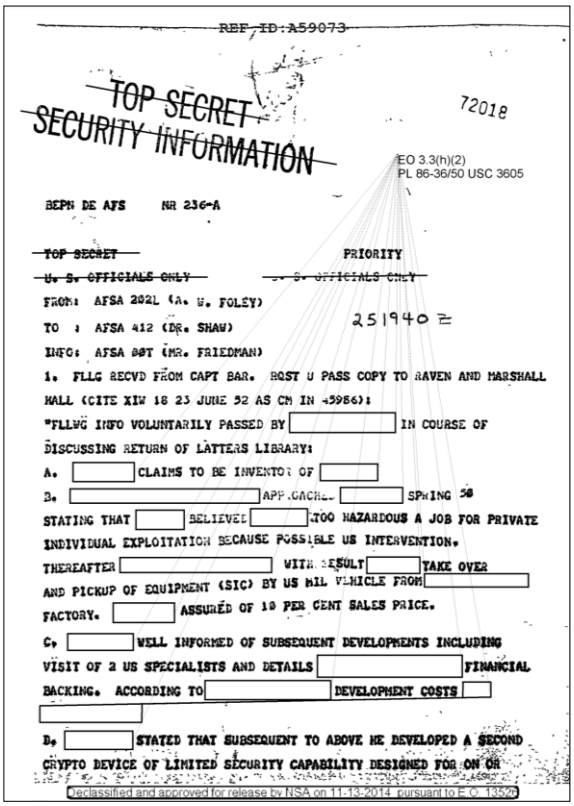
[redacted] head of the Naval Intelligence Office from 1942 to 1944 (I-165)

FOIA(b)(3) - 50 USC 402 Note Sec 6 - NSA

The Organization

We have no very definite information about the organization, particularly before 1944. In 1938 the Chief of Naval Communications

Exhibit 2. Page from NARA RG 38 Entry A1-1030 Box 74 Folder 3640/10. US declassification authority 003003.



REF ID:A59073

~~TOP SECRET~~
~~SECURITY INFORMATION~~

72018

EO 3.3(h)(2)
PL 86-36/50 USC 3605

BEPN DE AFS NR 236-A

~~TOP SECRET~~ PRIORITY
~~U. S. OFFICIALS ONLY~~ ~~U. S. OFFICIALS ONLY~~

FROM: AFSA 282L (A. W. FOLEY)
TO : AFSA 412 (DR. SHAW) 251940Z
INFO: AFSA 88T (MR. FRIEDMAN)

1. FLLG RECVD FROM CAPT BAR. RQST U PASS COPY TO RAVEN AND MARSHALL HALL (CITE XIW 18 23 JUNE 52 AS CM IN 45986):

"FLLWG INFO VOLUNTARILY PASSED BY [redacted] IN COURSE OF DISCUSSING RETURN OF LATTERS LIBRARY:

A. [redacted] CLAIMS TO BE INVENTOR OF [redacted]

B. [redacted] APP CACHE [redacted] SPRING 50 STATING THAT [redacted] BELIEVED [redacted] TOO HAZARDOUS A JOB FOR PRIVATE INDIVIDUAL EXPLOITATION BECAUSE POSSIBLE US INTERVENTION. THEREAFTER [redacted] WITH RESULT [redacted] TAKE OVER AND PICKUP OF EQUIPMENT (SIC) BY US MIL VEHICLE FROM [redacted] FACTORY. [redacted] ASSURED OF 10 PER CENT SALES PRICE.

C. [redacted] WELL INFORMED OF SUBSEQUENT DEVELOPMENTS INCLUDING VISIT OF 2 US SPECIALISTS AND DETAILS [redacted] FINANCIAL BACKING. ACCORDING TO [redacted] DEVELOPMENT COSTS [redacted]

D. [redacted] STATED THAT SUBSEQUENT TO ABOVE HE DEVELOPED A SECOND CRYPTO DEVICE OF LIMITED SECURITY CAPABILITY DESIGNED FOR ON OR

Declassified and approved for release by NSA on 11-13-2014 pursuant to E.O. 13526

Exhibit 3. NSA Friedman collection, Folder 395, doc A59073. Papers enabling reconstruction are in NARA RG 457 Entry P4 Box 20 ("Professor Vierling's library"), previously declassified in March 2012. Note the redaction authority cited top right (50 USC 3605).

3 Discussion and Conclusions

What can we learn from these rather unsatisfactory redactions? First, it seems that redaction is a weak way to protect secrets. Redaction techniques are, in general, susceptible to textual analysis or even to simple copy-paste removal of superimposed blanking (Bland et al., 2023; Ingram, 2019). Redactions applied to historical cryptological papers are vulnerable to techniques which would be recognised by codebreakers of the period from which they originate: contextual analysis allowing linguistic interpolation; parallel availability of the same text in a different communication; word length analysis; and fingerprinting. The fact that these documents are likely to be of interest to students of cryptanalysis – the very techniques which the authorities wish to obscure – adds a spice of irony to the redaction exercise.

One may then ask why the redactions were made in the first place. In a democracy, there is a tug-of-war between the security imperative of protecting the state and the expectation of openness and accountability of state agencies. One category of legitimate non-disclosure arises from the need to protect vulnerable persons from reprisals or breach of privacy, a particular concern where the individuals concerned are or were agency members. Another is where the papers reveal a cryptanalytical technique, or a pathway towards a technique, which could expose current national secrets.

But other, more dubious rationales may be at play: protection of official or national reputation, predilection for secrecy ('if in doubt, leave it out'), the power of mystique, and so forth. It is odd that many mid-twentieth century papers on codebreaking remain classified: surely security-related reasons for concealment have now lapsed. It is, unfortunately for historians, possible that they may remain under wraps indefinitely, since political priorities and budgets do not lend themselves readily to reviews of previous classification decisions.

Acknowledgments

The author would like to thank the anonymous reviewers for helpful feedback. Thanks are also due to the staff of NARA for friendly advice during the conduct of research, and to the President of Kellogg College for a Visiting Fellowship.

References

Gill Bennett. 2002. Declassification and Release Policies of the UK's Intelligence Agencies. *Intelligence and National Security*, 17(1): 21-32.

Maxwell Bland, Anushya Iver and Kirill Levchenko. 2023. Story Beyond the Eye: Glyph Positions Break PDF Text Redaction. *Proceedings on Privacy Enhancing Technologies*, 2023(3): 43-61.

Wilhelm Flicke. 1957. *Agenten Funken nach Moskau*. Verlag Welsermühl, Wels, Austria.

Mathew Ingram. 2019. Thank you to everyone who can't redact documents properly. *Columbia Journalism Review*, 10 January 2019.

David Kahn. 1978. *Hitler's Spies*. Macmillan, New York, USA.

Randy Rezabek. 2016. *TICOM: the Hunt for Hitler's Codebreakers.* Rochester, NY, USA.

Michael Smith. 2022. *The Real Special Relationship*. Simon & Schuster, London, UK. p 401.

Post-quantum trails: an educational board game about post-quantum cryptography.

Jelizaveta Vakarjuk
Cybernetica AS
Tallinn University of Technology
Tallinn, Estonia
jelizaveta.vakarjuk@cyber.ee

Nikita Snetkov
Cybernetica AS
Tallinn University of Technology
Tallinn, Estonia
nikita.snetkov@cyber.ee

Abstract

Post-quantum cryptography has gained more and more attention with the recent developments in quantum technology. There are already standard drafts for the novel post-quantum cryptosystems and organisations are starting the process of migration to post-quantum cryptography. However, the migration process has many challenges that need to be taken into account. Moreover, the algorithms themselves have become more complicated, making it more difficult to educate people about post-quantum cryptography. We propose to use gamification to make it easier to explain the main challenges and obstacles as well as the main steps of the migration process to the non-cryptographic community. We propose a board game that is built using the gamification taxonomy of Toda et al. to ensure a smooth learning process.

1 Introduction

There are big changes coming to the cryptography world with the new standards of post-quantum cryptography (PQC). PQC refers to cryptographic schemes that work on classical computers, but are resistant to both classical and quantum computer attacks (Alagic et al., 2022). An exact estimation of time when a large enough quantum computer will be available is considered an ambiguous task. However, we can use results from the survey conducted by evolutionQ Inc. (Michele Mosca, 2023), where 37 international quantum computing experts were asked a series of questions about the developments in the field. One of the questions was to identify the likelihood of having a quantum computer that can factorize a 2048-bit number in less than 24 hours in upcoming decades. A majority (20/37) of the respondents answered that it is about 50% likely or more likely to have such quantum computer in next 15 years. Organisations should start thinking of their migration strategies to make the process of switching systems from one class of cryptographic schemes to the other more smooth (Attema et al., 2023). However, with new algorithms, new challenges appear. The fact that new algorithms are more complicated to understand and have different limitations that make it challenging to fit PQC into existing protocols, is among the most prominent challenges. Additionally, the migration process is expected to be more challenging and resource-consuming compared to migration from DES to AES or from SHA1 to SHA2 (Banerjee et al., 2023). Therefore, the challenges that might be encountered during the process are novel and might be unique to each system. Considering increased complexity of the algorithms and challenges, it has become more difficult to educate people on this topic. In this paper, we propose to use gamification to explain the main challenges and steps in the migration process to people with different knowledge backgrounds. We develop a board game, where throughout the game process, players learn different families of PQC algorithms, the main obstacles in the PQC research process and the process of migrating to PQC, how the research process works in general.

1.1 Gamification

Active learning strategies are widely employed in learning environments to better engage learners with the course material, encourage critical thinking and discussions. Studies show that the learners' examination grades improve and the course failure rate decreases, if some of the active learning techniques are used throughout the course (Freeman et al., 2014). One of the examples of active learning is usage of gamification within the educational environment. Gamification has been extensively used for education with

the goal to increase learners' motivation and engagement with the material (Dichev and Dicheva, 2017). It is a supporting mechanism that incorporates various game elements to the educational activities. To add more formalism to the gamification methodology, a gamification taxonomy was proposed by Toda et al. (Toda et al., 2019). The aim of the taxonomy is to support the design of learning environments that are using gamification elements. The taxonomy is split into five dimensions – performance/measurement, ecological, social, personal, fictional. Each of the dimensions has various elements with examples on how those can be implemented in the educational environment. The *measurement dimension* is important for providing learners with feedback on their progress and actions that have been taken during the course of the game. The *ecological dimension* is responsible for the learner's interaction with the game environment. The *social dimension* is connected with the social side of interactions with the environment and the other learners. The *personal dimension* is related to the learner using the game environment. The *fictional dimension* is related to the learner's experience when interacting with the game though narrative and storytelling.

2 Post-quantum trails game

Post-quantum trails is a competitive game, where all the players have the same goal to achieve. The goal of the players is to finish their migration process before the quantum computer is built. Player, who finishes their migration process first, wins the game. If the quantum computer is built before one of the players finished their migration, the game wins. Throughout the game, players are not only focusing on the research actives, but also spend resources on marketing to advertise their research. There are several types of cards in the game – event cards, action cards, and team member cards. Example cards are presented on Figure 1. All the example cards contain pictures generated by the Adobe Firefly [1]. *Event cards* correspond to the main events that are happening in the world and that lead to some consequences in the research process. *Action cards* correspond to the activities that support the research, development and marketing processes. *Team member cards* correspond to the team members that can be hired to move the research and development process forward and improve it.

During the game, players can gain different resources – science points, influence points and money. Science points are needed to advance with research activities, influence points are connected to marketing activities and money is needed to hire new team members and be able to respond to the events coming from the event cards. Some of the event cards speed up the development of the quantum computer.

Event cards The research process can be influenced by different events happening in the field, findings by other research teams and difficulties in funding. For the event cards, we have collected different events from the sources like pqforum, PQC conferences (NIST, PKI, ETSI), Twitter threads of researchers from the field of PQC, papers connected with PQC. The main idea behind these cards is to illustrate what are the challenges of the research process in novel areas, how unexpected findings can destroy some of the development strategies of players, and simulate consequences of some bad events. Event cards contain news that influence their actions for this turn. Additionally, there are cards connected to the development of quantum computing, which are adding quantum tokens to the timeline.

Action cards There are different activities that can support the research process financially and also offer moral support. The content of the action cards is mostly based on the authors' research experience in the field of cryptography. The goal of these cards is to illustrate positive events that occur in the research process. Action cards help players to gain resources and hire new team members.

Migration path On the migration path, the players need to advance their player figures. To advance to the next level, players need to satisfy certain requirements or spend resources. It is possible to skip some of the levels, but if the player decides to complete an optional level it gives them protection against certain events from the event deck. The migration path has the following steps:

1. Security proof (optional) – gives additional science points to the player.
2. Publication – research should be submitted to journals/conferences to be evaluated by the other experts in this area. This step aims to

[1] `https://www.adobe.com/products/firefly.html`

Figure 1: Examples of cards of different type

simulate that before standardisation process researched scheme undergoes initial cryptanalysis.

3. Standard – step inspired by the NIST PQC standardisation process.
4. Hybrid mode (optional) – to make sure that the system is secure even if a novel attack appears on an implemented PQC scheme, it is advised by many organisations (ANSSI, 2023; BSI, 2023; H, 2023) to use PQC together with classical cryptography.
5. Real-life implementation – once the scheme is analysed and selected to be one of the future standards, it should be tested within the real systems. It is essential to perform testing and analysis as early as possible to be able to spot challenges and limitation at early stages and have enough time to fix them.
6. Worldwide usage – this is the final card of the migration path that the player should complete to win the game. It illustrates that the scheme was adopted and is being actively used.

Team member cards In order to be more successful in research and marketing activities, players can hire new team members. There are two main types of team members – researchers and marketing specialists. Researchers award players science points and there are three types of researchers – junior researcher, researcher and senior researcher. Marketing specialist awards influence points that are required to complete the final step of the migration process (worldwide usage).

2.1 Game development

To develop a board game idea and all the game elements that would supplement education process, we used the gamification taxonomy of Toda et al. (Toda et al., 2019). From the measurement dimension, we applied progression, points and stats. *Progression* – during the game, players advance in their migration process which helps them to identify how they move forward to the end goal (building last worldwide usage card of the migration process) Additionally, there is a quantum computer development progression bar, which indicates the approximate time remaining for completing the goal. *Points* – players gain science points and influence points for successfully performing different actions. *Stats* are connected with the progression, when a player moves their figure on the migration path and that shows which tasks the player has successfully completed. All the measurement dimension elements ensure that the players get enough feedback on their progress throughout the game and the game does not end unexpectedly for the players. Therefore, making sure that the players do not get frustrated or disoriented (Toda et al., 2019).

From the ecological dimension, we applied chance, imposed choice, economy, rarity and time pressure. *Chance* is used in different places throughout the game. There are action cards that allow players to gain resources, where the amount depends on the dice roll. There are event cards that are randomly drawn at the beginning of a player's turn that result in different consequences. Finally, players draw action cards at random from the corresponding deck. *Imposed choice* – there are certain migration steps that are not necessary for the successful completion of the end goal, but which bring additional points or provide protection against certain event cards. Players can choose whether to complete these steps or not,

which in turn, influences how the players will advance in the future. *Economy* is directly connected to resources that the players gain during the game. Players are spending gained resources to hire new team members and to participate in the research activities. *Rarity* – there is a limited number cards that give more resources to the players. *Time pressure* – there is the quantum computer development timeline that has a limited number of places that are filled with tokens during the game. Once all the fields are covered by tokens, the game ends. The presence of these ecological dimension elements ensures that the player does not get the feeling that their actions are aimless and do not impact the outcome of the game. Players are presented with the choice over the course of the game, limited resources of certain cards ensure that there will be enough interest among the players to find those rare cards. The addition of some chance elements makes it more interactive and interesting, but the amount of cards related to a random dice roll is limited to mitigate frustration of the players from bad luck and lessen the randomness.

From the social dimension, we used competition and social pressure. *Competition* is present since the players are competing among each other to finish their migration process first and win the game. *Social pressure* is partially achieved through the event deck. Even though, the events are appearing randomly, blame for the events with bad consequences (e.g., adding a quantum token) is put not on the deck itself but on the player who has drawn this card. Through the usage of the social dimension elements, we ensure that the players are motivated by trying to overcome their co-players and accomplish the goal faster. However, some players may be discouraged if they are not doing so well as their co-players. Therefore, our future plans include introducing a cooperative mode where the players will be competing with the game and not with the other players. Another option is to divide players into teams, so they will be competing in teams, not individually, which is less discouraging.

For the personal dimension, we have added the following elements – novelty, objectives, renovation. To achieve *novelty*, we added variability to all the card decks, so there will be enough different events and actions to keep the players engaged. *Objectives* – players have the same end goal, i.e., to complete the migration process. Additionally, there are smaller objectives along the path, for example, to publish a paper in a journal or a conference, a player needs to collect a certain number of science points. *Renovation* – there is a limited number of action cards that can be played together with the other action cards that require rolling a die to decide the outcome. Those cards allow the player to re-roll a die if the initial outcome is not satisfactory for the player. These personal dimension elements ensure that the players are not misguided throughout the game and have a clear goal to accomplish, variability of the cards ensures that the players do not lose interest during the game. Additionally, the renovation element allows players to re-do their unsuccessful actions.

From the fictional dimension, we added the *narrative element*. This is achieved through the introduction to the game, where the players will receive an explanation of the game world and the players' role in this world. Moreover, players draw event cards that are drawn each turn, add details to the game flow that may influence the player's actions during their turn.

3 Future work

The next phase of the research work is to try out the game with different audiences to collect their feedback and improve, clarify and modify the game where needed. Additionally, using the players' feedback, we will analyse if their understanding of the PQC migration process has improved after playing the game. If the players' feedback confirms our concerns about the single-player competitive mode, we will additionally develop a cooperative mode for the game to ensure that the educational environment is not discouraging or frustrating to the learners.

Acknowledgments

This work was funded by the Estonian Research Council under the grant number PRG1780 and the European Union under Grant Agreement No. 101087529. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Gorjan Alagic, Daniel Apon, David Cooper, Quynh Dang, Thinh Dang, John Kelsey, Jacob Lichtinger, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, and Daniel Smith-Tone. 2022. Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process. https://doi.org/10.6028/NIST.IR.8413-upd1, Jul.

ANSSI. 2023. ANSSI views on the Post-Quantum Cryptography transition (2023 follow up). https://cyber.gouv.fr/en/publications/follow-position-paper-post-quantum-cryptography, Aug.

Thomas Attema, João Diogo Duarte, Vincent Dunning, Matthieu Lequesne, Ward van der Schoot, and Marc Stevens. 2023. The PQC Migration Handbook. GUIDELINES FOR MIGRATING TO POST-QUANTUM CRYPTOGRAPHY. Technical report, Applied Cryptography and Quantum Algorithms (TNO) and Cryptology Group (CWI) and Netherlands National Communications Security Agency (AIVD).

Aritra Banerjee, Tirumaleswar Reddy.K, Dimitrios Schoinianakis, and Tim Hollebeek. 2023. Post-Quantum Cryptography for Engineers. Internet-Draft draft-ietf-pquip-pqc-engineers-02, Internet Engineering Task Force, October. Work in Progress.

BSI. 2023. Cryptographic Mechanisms: Recommendations and Key Lengths, Jan. BSI – Technical Guideline TR-02102-1, https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG02102/BSI-TR-02102-1.html.

Christo Dichev and Darina Dicheva. 2017. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education*, 14(1):9, February.

Scott Freeman, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415.

John H. 2023. Migrating to post-quantum cryptography, Nov. National Cyber Security Centre blog post, https://www.ncsc.gov.uk/blog-post/migrating-to-post-quantum-cryptography-pqc.

Marco Piani Michele Mosca. 2023. Quantum Threat Timeline Report 2022. https://globalriskinstitute.org/publication/2023-quantum-threat-timeline-report/, December. Accessed: 2023-11-28.

Armando M Toda, Ana C T Klock, Wilk Oliveira, Paula T Palomino, Luiz Rodrigues, Lei Shi, Ig Bittencourt, Isabela Gasparini, Seiji Isotani, and Alexandra I Cristea. 2019. Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, 6(1):16, December.

Decipherment of a German encrypted letter sent from Sigismund Heusner von Wandersleben to Axel Oxenstierna in 1637

Michelle Waldispühl
University of Oslo
Norway
michelle.waldispuhl@ilos.uio.no

Nils Kopal
University of Siegen
Germany
nils.kopal@uni-siegen.de

Abstract

We present our work on an encrypted letter from the Thirty Years' War written by the ally of the Swedish Empire, Sigismund Heusner von Wandersleben in 1637 and sent from Kassel to the Swedish High Lord Chancellor Axel Oxenstierna. We describe our analysis of the ciphertext including information on the cipher type, the process of cryptanalysis and challenges for the decipherment. We include the edition of the letter in the current state of decipherment and summarize its content.

1 Introduction

It is not unusual that encrypted sources are stored in archives without having been deciphered. It is a laborious process to decrypt historical ciphers and oftentimes, historians and archivists working with these documents do not have the resources to perform a cryptanalysis of unknown ciphers. This fact may lead to sensational finds such as the recent discovery of unknown letters by Mary Stuart in the Bibliothèque Nationale de France (Lasry et al., 2023). For cryptanalysts interested in historical ciphers, searching archives systematically for undeciphered material is not always straightforward. However, with the help of specific search entries, such as "undeciphered", "unknown writing", and more effectively, by talking to experienced archivists, such documents can be found (Megyesi et al., 2024). Assisted by computer-based tools such as those provided by the DECRYPT project[1] undeciphered documents can be cryptanalyzed and deciphered on the own computer in a (semi-)automatic way.

In this brief paper, we present the decipherment and cryptanalysis of an encrypted letter from the Swedish National Archives, which has not been deciphered before. It is a letter sent by Sigismund Heusner von Wandersleben, an ally of the Swedish Empire, to the Swedish High Lord Chancellor, Axel Oxenstierna, in 1637.

[1] www.de-crypt.org (Megyesi et al., 2020).

2 The letter

The letter is three pages long and includes cleartext in German and ciphertext. Additionally, there is an attachment to the letter also including ciphertext passages. We show the first page in Figure 1; for the other pages see the entry in the DECODE database (Heusner von Wandersleben, Sigismund, 1637).

The document is stored at The Swedish National Archives in the Oxenstiernska samlingen, volume E 622 A. It is included in a collection containing 14 letters written by Sigismund Heusner von Wandersleben in the years 1632–1638. Only this one letter is encrypted. The ciphertext was collected in fieldwork by Beáta Megyesi and uploaded to the DECODE database. We ordered digitations of all 14 letters from the Swedish National Archives.

3 The cipher

The encrypted passages are written in a homophonic substitution cipher using digits as ciphertext code elements. We have identified 85 homophones for the plaintext alphabet letters. The homophonicity is uneven with a maximum of eight homophones for the letter 'e' and a minimum of one code element for the letters 'k' and 'p'. The digits used as code elements for the alphabet letters range between 4 and 202. In the ciphertext, each code element is separated using dots.

In addition to the encoding of the alphabet letters, there is also a nomenclature in use where three-digit code elements encode lexical plaintext elements. From several syntactic contexts we can deduce that we here mainly have to do with place

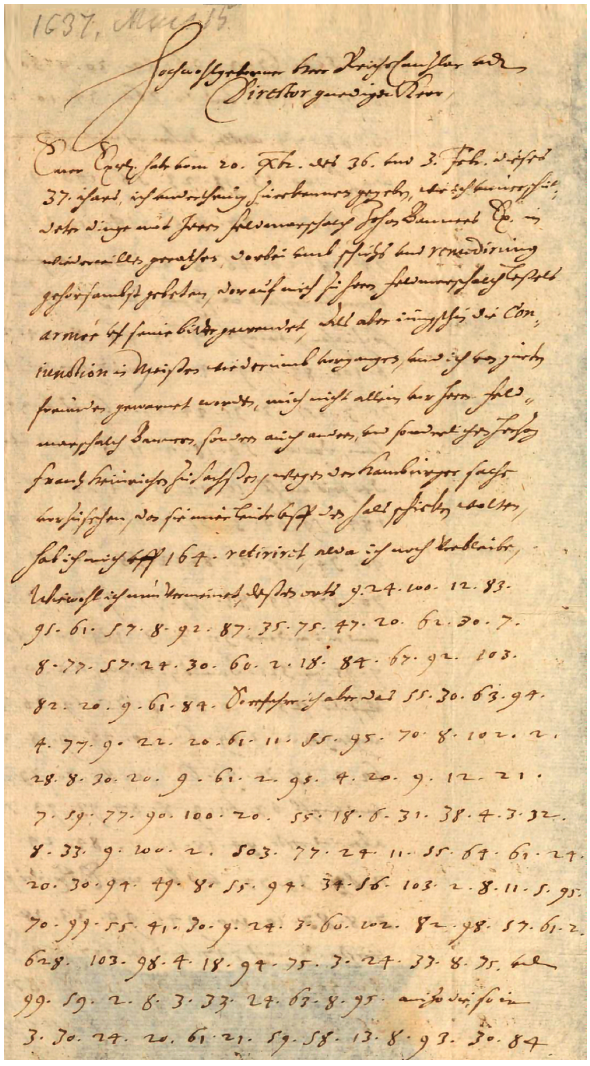


Figure 1: Page 1 of Heusner von Wandersleben's letter from 1637. (Heusner von Wandersleben, Sigismund, 1637)

names and personal names, e.g., the passage 'hab ich mich vff 164. retiriret' (*I have withdrawn myself to* 164.) on page 1. The complete key can be found in the entry for the record in the DECODE database (Heusner von Wandersleben, Sigismund, 1637).

4 Transcription

We transcribed the collection of 14 letters using the automatic text recognition tool *Transkribus* (https://readcoop.eu/transkribus/, accessed 25 September 2023). The decision for this tool lied in the high amount of cleartext in German "Kurrentschrift" for which Trankribus' transcription models are well trained. Furthermore, the cipher alphabet contains only digits which is a symbol set also covered by Transkribus' models — in contrast to other rare scripts used in e.g. the Copiale cipher. For these symbol sets, *Transcript Tool* has been developed (Szigeti and Héder, 2022) and is a more suitable choice.

The transcription output provided by the Transkribus tool was manually validated, and special care was taken to correct any mistranscribed digits in the encrypted passages. Subsequently, the ciphertext was extracted from the first letter to further process for cryptanalysis.

5 Cryptanalysis

Due to the high number of distinct code elements, we hypothesized that the cipher is homophonic. Since the cleartext passages were written in German, we assumed that the plaintext language was also German, a hypothesis that was confirmed during the cryptanalysis. We cryptanalyzed the letter using two components from the cryptanalysis tool CrypTool 2 (CT2) (Kopal, 2018): (1) the Homophonic Substitution Analyzer (HSA) (Kopal, 2019), which enables to perform (semi-)automatic cryptanalysis, and (2) the Substitution component, which allows for the decryption of a given ciphertext when the key is (partially) known. Initially, we utilized the HSA's automatic cryptanalysis algorithm, which, after several restarts, yielded partially correct words. Subsequently, we iteratively refined the generated output and restarted the algorithm to further enhance the automatic discovery of more plaintext segments. Additionally, we performed a frequency analysis in CT2 on the code elements. This allowed to identify the frequency distribution of the code elements, e.g., the two most frequent represent the letter 'e' (despite the high amount of homophones) and code elements occurring only once are candidates for nomenclature elements.

5.1 Close-reading and writer-specific dictionary

After the first round of cryptanalysis, we reintegrated the decrypted passages into the cleartext passages and applied a manual close-reading method. First, obvious false letters were corrected, i.e., in the passage HERRUNDKNECIT *Herr und Knecht* 'master and servant' where the last I must be H. After several rounds of further improving the key applying this method, the homophonic cipher could be broken entirely.

Our plan to incorporate a writer-specific dictio-

nary based on Heusner Wandersleben's other 13 letters in order to find his specific spelling patterns turned out not to be useful in our analysis because the material consisting of 4,067 tokens was too limited.

5.2 Search for the original cipher key

Upon deciphering significant parts of the homophones and subsequently revealing portions of the key, we searched for the original key. Since the cipher has similarities with other keys used in Swedish correspondence in the 30 Years' War, such as the Camerarius key or the Beaumont key (Stålhane, 1934; Waldispühl, in press), we examined the documentation of seventeenth-century keys from the Swedish National Archives in the DECODE database. We also reviewed keys from the 1600s in the Hesse State Archives based on the fact that the letter was sent from Kassel, Germany. Sometimes, one is fortunate enough to find the original key (Kopal and Waldispühl, 2022), but in this case, we were not. That is why we have not been able to decipher the nomenclature elements of the cipher.

5.3 Challenges during the decipherment

The ciphertext contained a high number of homophones relative to its length, complicating the cryptanalysis. For instance, the letter 'e' is represented by seven homophones. Some homophones occurred only once or twice in the whole text, which was problematic for automatic decryption. However, this challenge could be met by manual cryptanalysis. An open problem is the absence of a key for the nomenclature elements which keeps them indecipherable. The lack of punctuation and word separation in the ciphertext obscures syntactic structures, complicating sentence delineation and interpretation initially. Furthermore, in historical alphabets, I/J and U/V are not separated and share identical code elements. Since CT2 uses the modern alphabet as a starting point in the automatic analysis, the analyzer has to decide either to use e.g. U or V, and based on the decision, half of the words are decrypted in a "wrong" way (Waldispühl et al., in press). Lastly, the key seems to show a certain pattern in how the code elements are distributed for the plaintext letters, i.e., 12-A, 11-C, 10-B, 9-D. However, we did not fully understand the system to exploit it for cryptanalysis.

6 Decrypted letter

In the following, the plaintext is rendered in capital letters and remaining undeciphered nomenclature code elements are given as they appear in the original text. To facilitate reading, word separators are introduced in the plaintext. Words followed by a question mark are cleartext passages difficult to read.

Page 1: 1637, Mais-15
Hochwohlgeborner Herr Reichchanzlar
vndt Director gnediger Herr
Eurer Excell. habe vom 20. Xmbr.
des 36. vnd 3. Febr. dieses 37.
ihars ich vnderthenig zuerkennen
gegeben, wie ich vnuerrichtetdeter
dinge mit Herrn Feldmarschalch Johan
Banners Ex. in wiederwillen gerathen,
darbei vmb schutzs vnd remedirung
gehorsambst gebeten, darauf mich zu
herrn Feldmarschalch Leßels armée
vf seine bitte gewendet, Als aber
iüngsthin die Coniunction in Meißen
wiederumb vorgangen, sond ich von
gueten freunden gewarnet worden, mich
nicht allein vor herrn Feldmarschalch
Bännern, sondern auch anderen, vnd
sonderlichen Hern frantz heinrichen
zu sachßen, wegen der Hamburger sache
vorzusehen, da sie mier leute vff den
hals schicken wolten, hab ich mich
vff 164. retiriret, alda ich noch
verbleibe, Wiewohl ich nun vermeinet,
deßen orts DIE ALTE DEUOTION GEGEN DIE
CRON ZU FINDEN. So erfahre ich aber
das HERR UND KNECHT SEHR GEENDERT UND
AM GANZEN HOFE AUSSER DER 503. NICTT
EINER MEHR UFFRECHT SCHWEDISCH IA DER
628. FAUORISIREI vnd CARESRIRET aizo
die so in SEINEM ABWESEN

Page 2: dem 503. ZUWIEDER GEWESEN
vnd die 764. UBERGEBEN wollen, daher
ich mich nicht wenig Verwundert wie
L.EWOLFF DERORTEN NEGOCIIREN SONNEN
deme es furwahr an nottieftigen
vnderhalt, in? sogar das ich darfur
erschrocken, höchlichen gebracht?, UON
EINEM FELDZUG wirdt geredet, WOHIN IST
STILL. es kan aber nichts großes sein,
dan die FORCE nicht alda, vnd man der

Zeit nicht BASTANT 808. 385. 184. AUS DEM LANDT ZU TREIBEN da 503. SOLL IN DAS FELD vnd ein anderer AUS DEM LANDT in 321. damit ist es ein KUCHEN Vff die 255. BESTALLUNG ingleichen die 289. GELDER machet man GROSSES HERANTZ vnd viehl REDENS UON. Es ist aber das erste NOCG NICHT CLAR vnd kan was E. Ex. die sache mit SELBER CRON dero hohen verstandt nach 747. vnd den 617. RECHT IM DIRECTORIO FASSEN schon alles dergestalt ge-

Page 3: machet werden, das man DOCH DIE CRON 708. SUCHEN vnd von derselben DEPENDIREN UND BITTEN muss von deren MAN SICH SONSTEN AUSZUHALFETERN GEMEINET Wegen des andern ist kein vberflues, vnd erfolget sparsam genug, das sich also die sachen wohl geben, Euer Ex. habe meiner schuldigkeit nach ich dieses mit wenigen gehorsamlichen anfuegen wollen, die ich des allerhöchsten schvzs? vnd dero zu beharrlich gnaden mich vnderthenig empfehle vnd dero resolution mit sachßen erwarte, Vol. CASSEL den 15. May 1637. Eurer Excelz. Werthiger gehorsamer diener 384.
ich habe mich des EWOLFREN seines Comptorii? et?

Page 4: Man hat hier in der aller größen geheimb etlich 385. CENTNER METALLISCHE SPEISSE ZUSAMMEN GESCHMELZET vnd solche verdecket auf 147 gefüret, daselbst etliche schon hierzu gemachte STUCKE ZU TAUSCHEN vnd werden mit dem UFFBRUCH wie ich in vertrauen vernommen den nechsten sich eihlen vndt ALLE ABWERTS GEHEN. Alhier wie ich Vermercke wirdt IOHAN UON UFFELN COMMENDANT 571. GUNTEROT aber GEHET MIT ANDEREN ZU FELDT varleßet sich mercken, ob seye etwas UON 703. ALHIER wie auch ANDERE vnd haben GROSSE HOFFNUNG
dieses wird gleich bey schließung meines schreibens ahn? EWOLFFEN bey einem gueten ort geschrieben.

7 Some historical context and summary of the content

The author of the letter, Sigismund Heusner von Wandersleben, served as a Swedish counselor and general war commissioner 1631–1638. Born 1592 in Coburg, he entered a diplomatic career after his studies first as a counselor of Herzog Wilhelm IV. of Sachsen-Weimar and then as a Swedish General War Commissioner (Warlich, 2011).

The letter is addressed to Axel Oxenstierna who served as the High Lord Chancellor of Sweden before, during, and after the Thirty Years' War, from 1612 until his death in 1654.

The text discusses various diplomatic and military matters in spring 1637, i.e., the time when the Swedish army had to flee North to Pommern after they had defeated the Saxon army in Wittstock in autumn 1636 (Murdoch et al., 2012). Heusner von Wandersleben mentions personal conflicts with leaders such as Johan Banér, Franz Heinrich of Saxony, and others. There are concerns about potential threats. He also talks about his own actions, including a retreat, and he seeks guidance and support of Oxenstierna. Towards the end, there are discussions about military forces, territories, and financial matters.

8 Conclusion

We have presented the methods and results of our partly successful decipherment of a hitherto undecrypted letter written by Heusner von Wandersleben to Oxenstierna in 1637. Our work lays the ground for historians to further analyze the letter's content and contribute contextual knowledge in order to address the undeciphered nomenclature elements. The fact that the nomenclature elements remained obscure to us means in turn that they make the cipher secure. However, the switching between cleartext and ciphertext had an opposite effect and contributed significantly to our successful cryptanalysis. Another aspect strengthening the cipher's security is its high number of homophones that were varied extensively in the ciphertext. Further research into a possible systematic ways of the homophone variance as well as the patterns of cleartext-ciphertext switching would reveal more knowledge about how keys were applied in practice.

Acknowledgments

We are grateful to our colleagues in the DECRYPT project for valuable feedback on our ongoing work. We want to thank especially Beáta Megyesi for collecting the ciphertext letter and sharing it with us. We are also indebted to the archivists at the Swedish National Archives for digitizing and openly sharing their materials. This work has been supported by the Swedish Research Council, grant 2018-06074, DECRYPT – Decryption of Historical Manuscripts.

References

Heusner von Wandersleben, Sigismund. 1637. The National Archives of Sweden, Oxenstiernska samlingen, volym E 622 A, fols. 1-4, DECODE ID 4332, link: https://de-crypt.org/R/4332.

Nils Kopal and Michelle Waldispühl. 2022. Deciphering three diplomatic letters sent by Maximilian II in 1575. *Cryptologia*, 46(2):103–127.

Nils Kopal. 2018. Solving Classical Ciphers with CrypTool 2. In *Proceedings of the 1st International Conference on Historical Cryptology HistoCrypt 2018*, number 149, pages 29–38.

Nils Kopal. 2019. Cryptanalysis of Homophonic Substitution Ciphers using Simulated Annealing with Fixed Temperature. In *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt*, pages 107–16.

George Lasry, Norbert Biermann, and Satoshi Tomokiyo. 2023. Deciphering Mary Stuart's lost letters from 1578-1584. *Cryptologia*, 47(2):101–202.

Beáta Megyesi, Bernhard Esslinger, Alicia Fornés, Nils Kopal, Benedek Láng, George Lasry, Karl de Leeuw, Eva Pettersson, Arno Wacker, and Michelle Waldispühl. 2020. Decryption of historical manuscripts: the DECRYPT project. *Cryptologia*, 44(6):545–559.

Beáta Megyesi, Alicia Fornés, Nils Kopal, Benedek Láng, Michelle Waldispühl, Vasily Mikhalev, and Bernhard Esslinger. 2024. Historical Cryptology. In Bernhard Esslinger, editor, *Learning and Experiencing Cryptography with CrypTool and SageMath*, pages 97–138. Artech House, Norwood.

Steve Murdoch, Kathrin Margarete Gertrud Zickermann, and Richard Adam Marks. 2012. The Battle of Wittstock 1636: Conflicting Reports on a Swedish Victory in Germany. *Northern Studies*, 43:71–109.

Henning Stålhane. 1934. *Hemlig skrift. Coder och chiffrermaskiner*. Lindfors, Stockholm.

Ferenc Szigeti and Mihály Héder. 2022. The TRANSCRIPT tool for Historical Ciphers by the DECRYPT project. In *Proceedings of the 5th International Conference on Historical Cryptology, HistoCrypt22*, pages 208–211.

Michelle Waldispühl, Beáta Megyesi, Nils Kopal, and Alicia Fornés. in press. Grapholinguistic features of historical ciphers and challenges for computer-based transcription and cryptanalysis. In Kerstin Kazzazi, Michael Schulte, and Gaby Waxenberger, editors, *From the Maya Script to the Germanic Runes – Case Studies on the Typology of Scripts and Research on Writing Systems*. Reichert, Wiesbaden.

Michelle Waldispühl. in press. Verschlüsselte Briefe im Schwedischen Reich: Mehrsprachigkeit und Geheimschrift während des Dreißigjährigen Kriegs. In Dessislava Stoeva-Holm and Michael Prinz, editors, *Praktiken der Mehrsprachigkeit im Schwedischen Reich (1611–1721)*. Harrossowitz Verlag.

Bernd Warlich. 2011. Der Dreißigjährige Krieg in Selbstzeugnissen, Chroniken und Berichten. Heusner von Wandersleben, Sigismund, link: https://www.30jaehrigerkrieg.de/heusner-von-wandersleben-sigismund-2/.

Poster abstracts

From Myths to Methods: Teaching Cryptography with the Enigma Machine

Tobias Baumeister and **Felix Schmutterer** and **Dietmar Fey**
Chair of Computer Science 3
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany

Abstract

Computer Science, like most engineering disciplines, usually pays little respect to its own history when it comes to teaching. This case study explores the integration of historical perspectives into computer science education, particularly focusing on the history of cryptography, exemplified most importantly via the Enigma machine. A novel teaching approach was designed and tested, which includes unplugged activities (i.e. no-technology activities) aimed to introduce and motivate students to both the technical complexities as well as the historical origins of different cryptographic methods. Students were challenged with progressively complex decryption assignments: Using substitution principles, they first decoded a simple Morse sequence. The second task involved solving a Bacon cipher to apply prior knowledge on a more intricate level. The third task involved deciphering a Caesar cipher, requiring deeper understanding of monoalphabetic substitution and historical research into cryptography's extensive history. Lastly, students were challenged with the decryption of an Enigma encoded weather report (encrypted by using only one Enigma rotor), which seemed to be an almost impossible challenge to most.

Through these assignments, students were guided to grasp encryption concepts and historical contexts, enhancing both their motivation and subject-specific understanding. Following the assignments, the lecturer provided historical background on the ciphers and further motivated the topic with background knowledge on the Enigma machine and Bletchley Park.

Evaluation results from two German university classes (a seminar on the "History of Computing" (Hoc) (n = 14) and a lecture "Computer Science for Engineers" (n = 33)) indicate mostly positive effects on student motivation.

Table 1: Average (and standard deviation) student rating regarding motivational aspects

Item	HoC Seminar	CSE Lecture
(a) Crypto	2.00 (σ =.88)	1.82 (σ =.64)
(b) Enigma	2.43 (σ =.85)	2.42 (σ =.75)
(c) WWII	2.43 (σ =.85)	2.09 (σ =.84)

After the teaching unit, students were asked to rate 18 statements on a Likert scale from 1 (Strongly agree) to 4 (Strongly disagree). In three items, we assessed whether students felt motivated by the teaching unit to learn more about (a) cryptography, (b) the Enigma machine and (c) World War II. 12-21% of the students (depending on the course) indicated that they felt less motivated regarding cryptography (c.f. table 1). Further exploration of the collected data shows that these respondents also struggled more with solving the decryption tasks, had the least prior knowledge of crytptography and had less interest in history in general compared to the students who were motivated. These findings emphasise the need for heterogeneous activities aimed at different student preferences and backgrounds. However, the overwhelming majority did respond positively to this interdisciplinary approach. These findings offer opportunities for future research and curricula design, in which the teaching of engineering topics alongside their rich history should be explored further.

Polyalphabetic cipher decryption function learning with LSTM networks

Oriol Closa
Independent Scholar
oriolcm@kth.se

Abstract

While Recurrent Neural Networks have been applied to a wide range of problems, from language modelling to time series forecasting among many others, the possibility of approximating a decryption function from a machine producing pseudorandom sequences seems intuitively not something they would be good at. However, we show how LSTM networks are indeed capable of not only learning but also extracting external key information from known-plaintexts of only 15 characters in length. In order to do that, we model and simulate simpler ciphers such as the Vigenère and the Playfair along with more complex machines like the Siemens and Halske T52d (without KTF) and the Hagelin C-38 in which we train our networks with. Furthermore, we also analyse the effects of different input types such as randomized data, German literature and war intercepts decrypted by the FRA in Sweden. As a result, this approach has proven to be effective for the Vigenère and the C-38 as well as partially for the T52d while giving negative results for the Playfair. Although this is without doubt not better than preexisting techniques, the intention is not to describe it as a better method to extract the key of a given ciphertext but rather to demonstrate the potential of other non-standard approaches to accomplishing such tasks which can be said of being completely possible.

Origins of NSA's Communications Security Mission

Evan Rea
Center for Cryptologic History
National Security Agency
earea@nsa.gov

Abstract

Today the National Security Agency has a dual mission; to collect foreign signals intelligence and protect National Security Systems. Today, this protective mission is known as Cybersecurity. But from the time of World War II into the 1980s this was known as communications security, or COMSEC. Though less well known and described in the historical record, the US COMSEC effort continued unbroken from the end of World War II, through NSA's predecessors, and, ultimately, into the new organization.

Early government efforts to organize it looked quite different than those for Communications Intelligence (COMINT), and led to a balanced arrangement between NSA and the armed services which had previously owned the COMSEC mission. In this new dynamic, NSA took on the role of a COMSEC service provider, setting technical standards and developing the cryptography and the hardware to apply to it. The military became a customer, providing requirements and the needed funding. NSA forged ahead, developing codes and encryption devices and putting them into service throughout the world.

The utility and durability of some of these machines lead to a long service life. For one offline encryption device known as the KL-7, it's ubiquity lead to a starring role in two espionage cases and a major international incident. The first espionage case was that of US Army Warrant Officer Joseph G. Helmich. Under pressure from his supervisor to settle outstanding debts, Helmich sold key lists and an operations manual to the Soviet Union in 1963 and 1964. However, he wasn't identified as a suspect until 1974, and not arrested until 1981. The second, and better-known espionage case involving the KL-7 was that of the John Walker spy ring. Unlike Helmich, who passed a limited amount of material on a few occasions, US Navy Warrant Officer John Walker and his accomplices supplied the Soviet Union with key lists and settings for several cryptographic machines, including the KL-7, for 18 years between 1967 and 1985. Finally, when North Korea seized the USS Pueblo and its crew in January 1968, they found a KL-47 machine, rotors and operating manuals on board, among other cryptographic prizes.

Project "Postmaschine" (1921-1925) ⇨ Ruin of Chima AG 1925 ⇨ Reichswehr + Chima AG

Claus Taaks
claus_taaks@web.de

At HistoCrypt 2023 in Munich, I presented the early development of the cipher machine, which was called "Enigma" since the end of 1923.

The result is summarised on the left side of the poster:
Since 1920, "Scherbius & Ritter", in competition with Damm, had been working at great expense on the "Postmaschine"[1], the only model in which there was any interest.
This project had already failed by the end of 1922, and Damm's "Cryptograph" failed in 1923. Nevertheless, Chima AG was founded in 1923, and the "capitalists' group" wanted to push through the postal machine business after all with intensive advertising.
Finally, the group used illegal means to introduce the last model of the “Postmaschine”.

On 15 July 1925, the Reichstag's Barmat Committee of Inquiry also dealt with the postal business of Chima AG. It became known that the Reichspost's decision in favour of the "Postmaschine" was based on fraud. Chima AG was de facto bankrupt. How could a company continue to operate that had only produced few prototypes and very small series?

The right-hand side of the poster shows how this business nevertheless continued:
The obligations from the previous deals, mainly the contracts with the Schiele & Bruchsaler Group (order of 1,000 machines), were settled, at a heavy loss.
The balance sheet for 1925 was based on very creative accounting. (For example, the 60% stake in N.V. Ingenieurs-bureau Securitas was valued higher than the share capital of Chima AG.)
The creditors of Chima AG kept quiet, someone had seen to that.

The delivery of glow-lamp machines to the navy had already begun in 1925. The army was still waiting for the "writing" machines that already had been announced in 1919.

Chima AG had to be restructured as quickly as possible:

- The "inventors' group" around Scherbius and Willi Korn redesigned the machines.
- Lawyer Rudolf Heimsoeth organised relations with the major banks and with politicians.
- The remaining shareholders, above all Deichsel and Rinke, placed the energetic Elsbeth Rinke in the management.
- The manufacturer was the Berlin company Körnig & Kassner.
- Companies called "Securitas" were dissolved or wound up. In 1927 Chima AG gave up its 60% stake in N.V. Securitas and in return got its patents back.

And the Reichswehr, which had been interested in the machine from the very beginning, since the end of the war?
It remained in the background and had the final say in everything from then on. This also meant that it determined the operability and simplification of the design and changed the criteria of possible use in wartime. Arthur Scherbius, who already had nothing more to say as the "capitalist group" was at the helm, could only make suggestions to the Reichswehr's cipher bureau.

(The subsequent phase of development, from 1928 onwards, was intensively researched by Dermot Turing and others: it consisted of the introduction of the Enigma as a glow-lamp machine throughout the Reichswehr in the early 1930s and the reaction to it in Poland, France and Great Britain.)

[1] The German terms "Postmaschine" (postal machine), "Kapitalistengruppe" (capitalists’ group) and "Erfindergrupppe" (inventors’ group) are used since 1923 in the commercial register of Chima AG. Previously, the cipher machines were simply called "Scherbius-Maschinen".

Published by:

NEALT Proceedings Series 53
ISSN 1736-6305 (online)
ISBN 978-99-1683-384-1
https://hdl.handle.net/10062/98422
https://doi.org/10.58009/aere-perennius0084
D-Space at Tartu University Library
ISSN 1736-8197, eISSN: 1736-6305