Deciphering a Short Papal Cipher from 1721

George Lasry
The CrypTool and DECRYPT Projects
george.lasry@cryptool.org

Paolo Bonavoglia
Mathesis Venezia
c/o Convitto Liceo "Marco Foscarini"
paolo.bonavoglia@liceofoscarini.it

Abstract

As part of the DECRYPT project, hundreds of enciphered papal letters from the 16th, 17th, and 18th centuries were deciphered, and dozens of keys were recovered. Several ciphertexts remained unsolved, despite the use of sophisticated computerized algorithms, and were offered as public challenges in 2019. One of those is a short letter sent by the nuncio in Bruxelles to Rome, on October 9, 1721. It consists of groups of digit codes, separated by commas. After improving the algorithm, and with some manual work, the authors were able to recover most of the key and of the plaintext. In this article, they present the method they used to recover the key, and the decrypted message.

1 Introduction

In the 16th century, the ciphers used by the popes and their envoys (the nuncios) were among the most sophisticated and secure ciphers in Europe, together with those of Venice (Meister, 1906; Lasry et al., 2020; Bonavoglia, 2021). They were primarily homophonic ciphers, represented by digit code groups. To provide for additional security, the digits were written in continuous sequences, without separating the code groups. Furthermore, the code groups often contained a different number of digits, with either one, two, or three digits. As a codebreaker would first need to separate the sequence of digits into distinct code groups, long continuous sequences of variable-length code groups of digits made this part of codebreaking much more challenging. For example, if the letter *e* can be represented by the homophones *2*, *21*, or *212*, and the letter *t* is represented by the homophones *1*, *12*, and *121*, the sequence of digits *1212* may have ambiguous interpretations: (*1*=*t*, *212*=*e*), or (*1*=*t*, *2*=*e*, *12*=*t*), or (*1*=*t*, *21*=*e*, *2*=*e*), etc... (Meister, 1906; Lasry et al., 2020). However, from the 17th century, the papal cryptographic services underwent a gradual decline, employing simpler ciphers at the expense of cryptographic security (Lasry et al., 2020; Alvarez, 1996), and the code groups had most often the same number of digits.

The enciphered letter from the nuncio in Bruxelles to Rome, on October 9, 1721 is shown in Figure 1.[1] While the place and date are written in clear, the main contents are encrypted using groups of mostly two digits (e.g., *16*), with occasional groups of one digit (e.g., *1*) and of four digits (e.g., *9336*), that are separated by commas. One group, *222*, consists of three digits.

It is likely that those commas were added by the cipher secretary who deciphered the letter. When visually observing the sequences of digits, the spacing between two digits separated by a comma (e.g., *01,16*) is the same as for two digits not separated by a comma (e.g., *01*). If the commas had been added when enciphering the letter, the spacing when a comma is added would have been wider than between two digits not separated by a comma. Furthermore, such commas reduce the security of the cipher.

As this letter could not be deciphered as part of the work described in (Lasry et al., 2020; Megyesi et al., 2020), it was presented as a public challenge in 2019 (Lasry, 2019a), but no solutions have been offered to date for this challenge. Unlike other ciphers in the Vatican archives, for which multiple documents encrypted with the same key were found, for this specific cipher, only one letter was available, making its decipherment more challenging (Lasry et al., 2020).

[1] Source: ASV - Arch. Nunz. Colonia / 5.

2 Deciphering the Letter

The first step was to transcribe the ciphertext, which was found to consist of 198 comma-separated groups, with 47 unique code groups.

From the multitude of cipher examples in the references (Lasry et al., 2020; Meister, 1906), it was expected that the four-digit code would probably encode elements from a nomenclature (e.g., names, places, common words), and that the one-digit and two-digits groups would (mostly) represent homophones of letters of the Italian alphabet (with the letters *u* and *v* being interchangeable).

Since the code groups were separated by commas, it was first expected that the decipherment of this letter with the help of computerized algorithms would not be a difficult task, however, all attempts to recover the key with those algorithms did not succeed. While other cipher schemes (e.g., polyphonic ciphers) were considered, it became clear after running simulations with ciphertexts with a similar number of groups (less than 200) and about 50 distinct groups, that the existing algorithms were not powerful enough for short ciphertexts as the one in the letter.

The algorithm and the improvements used to recover the key for this cipher are described in Section 3. Following those improvements, fragments of meaningful words could be recovered, with some initial assignments of the homophones to letters. With additional manual work, and after realizing that some code groups were likely to represent short prepositions or words (e.g. *il*, *ne*, *per*), additional parts of the ciphertext could be deciphered, and a more complete key recovered.

The key and the decryption are shown in Figure 2 and Figure 3. The digit *2* is used exclusively as a null symbol (either *2*, *22*, or *222*). Several codegroups likely represent short words or prepositions (e.g., *35=che*, *47=de*). Some homophones could not be reliably assigned to letters (e.g., *67*, *45*), and some of them might also represent short words or prepositions. The nomenclature elements (*9336*, *9485*, *9356*) could not be identified, as additional material (e.g., additional or longer ciphertexts) is required for that task.

The deciphered text can be approximately transcribed and as follows:

Sopra il curato 45? 9441? 43? te diuinis 67? rats(?) di sapere qualche cosa di 9336? 9485? di Tournai. Sopra l affare di Colonia mae 55? a quel che uedo ne(?) è informato. Se? stima a proposito che io 76? scriua per 9356? ad un suo corrispondente lo farò uolentieri ma attendo i suoi ordini per questo.

This can be loosely translated as follows:

About the parish priest 45? 9441? 43? (te diuinis?) 67? rats(?) to know something 9336? by Tournai. About the Cologne affair mae55(?) as far as I see you are informed. If you think for this purpose that I write for 9356? to a correspondent of yours, I will gladly do so but I await your orders for this.

3 Computerized Decipherment

As part of the DECRYPT project, various tools have been developed for the recovery of the homophonic cipher keys, from ciphertexts (Lasry et al., 2020; Megyesi et al., 2020).

The primary codebreaking tool requires a reference corpus, composed of texts in the target language. The tool uses the corpus to compute the frequencies $R_{i,j,k}$ of all possible trigrams of consecutive letters i, j, and k (e.g., *uni*, *ent*, etc...), and uses them to search for an optimal key solution, using a simulated annealing algorithm and a fitness score $Score(K)$ that is computed for a candidate key K, as follows:

- Decode the ciphertext using the candidate key K.
- Compute the frequencies of all the trigrams of letters - $F_{i,j,k}$ in the resulting decrypted text.
- Compute $S(K)$, the fitness score for K, as follows: $S(K) = \Sigma_{i,j,k}(F_{i,j,k} \cdot \log(R_{i,j,k}))$

During the simulated-annealing search, the tool performs transformations (or changes) in keys, looking to improve the fitness score. The following transformations are tested at each iteration:

- Swap the assignments of any two homophones. For example, if $x \rightarrow T$ and $y \rightarrow E$ (ciphertext symbol x represents T, and ciphertext symbol y represents E) before the transformation), then after the transformation: $x \rightarrow E$ and $y \rightarrow T$.[2]
- Change the assignment of a single homophone, e.g., instead of $a \rightarrow N$ (before the transforma-

[2]Note that this operation does not change the number of homophones mapped to E or T.

tion), we will have $a \rightarrow R$ after the transformation.[3]

Given a long enough ciphertext, and a suitable language corpus, this algorithm is likely to correctly recover most of the key mappings between the homophones and the alphabet letters. More details on the technique may be found in (Lasry et al., 2020). For more details on simulated annealing algorithms for codebreaking, see (Lasry, 2018).

For short ciphertexts as the one here, two improvements were required, in order to obtain an initial decipherment.

- Using a corpus of Old Italian books, instead of a generic Italian corpus.
- Changing the scoring function to use the frequencies of 5-grams of letters (such as *menti*, *dente*), instead of 3-grams.

With the improved algorithm, and some additional manual work, most of the text and the key could be successfully recovered.

4 Conclusion

This works illustrates the challenge of codebreaking of short homophonic ciphertexts, even in the case the digit code groups are separated with commas. Without the commas, the process would have been evern more challenging.

Unfortunately, without the meaning of the nomenclature elements (*9336*, *9485*, *9356*), it is difficult to interpret the historical meaning of the message. More work is required with the assistance of a linguist (expert in Old Italian) to complete the decipherment and the key recovery.

To date, only one public challenge about a papal cipher remains unsolved (Lasry, 2019b).

Acknowledgments

This work has been supported by the Swedish Research Council, Grant 2018-06074, DECRYPT – Decryption of historical manuscripts.

[3]Note that this transformation increases the number of homophones assigned to R, and decreases the number of homophones assigned to N. To ensure that the key is well balanced in terms of distribution of homophone assignments, a certain maximum number of homophones per regular element is specified when running the algorithm.

References

David Alvarez. 1996. Faded lustre: Vatican cryptography, 1815–1920. *Cryptologia*, 20(2):97–131.

Paolo Bonavoglia. 2021. The ciphers of the Republic of Venice an overview. *Cryptologia*, pages 1–24.

George Lasry, Beáta Megyesi, and Nils Kopal. 2020. Deciphering papal ciphers from the 16th to the 18th Century. *Cryptologia*, pages 1–62.

George Lasry. 2018. *A Methodology for the Cryptanalysis of Classical Ciphers with Search Metaheuristics, Ph.D. thesis, University of Kassel, Germany*. Kassel University Press. `http://www.upress.uni-kassel.de/katalog/abstract.php?978-3-7376-0458-1`.

George Lasry. 2019a. MysteryTwisterC3 - Vatican Challenge Part 3. `https://mysterytwister.org/challenges/level-x/the-vatican-challenge-part-3`, [Accessed: January, 14, 2022].

George Lasry. 2019b. MysteryTwisterC3 - Vatican Challenge Part 5. `https://mysterytwister.org/challenges/level-x/the-vatican-challenge-part-5`, [Accessed: January, 14, 2022].

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.

Aloys Meister. 1906. *Die Geheimschrift im Dienste der päpstlichen kurie von ihren Anfängen bis zum Ende des XVI Jahrunderts*, volume 11. F. Schöningh.

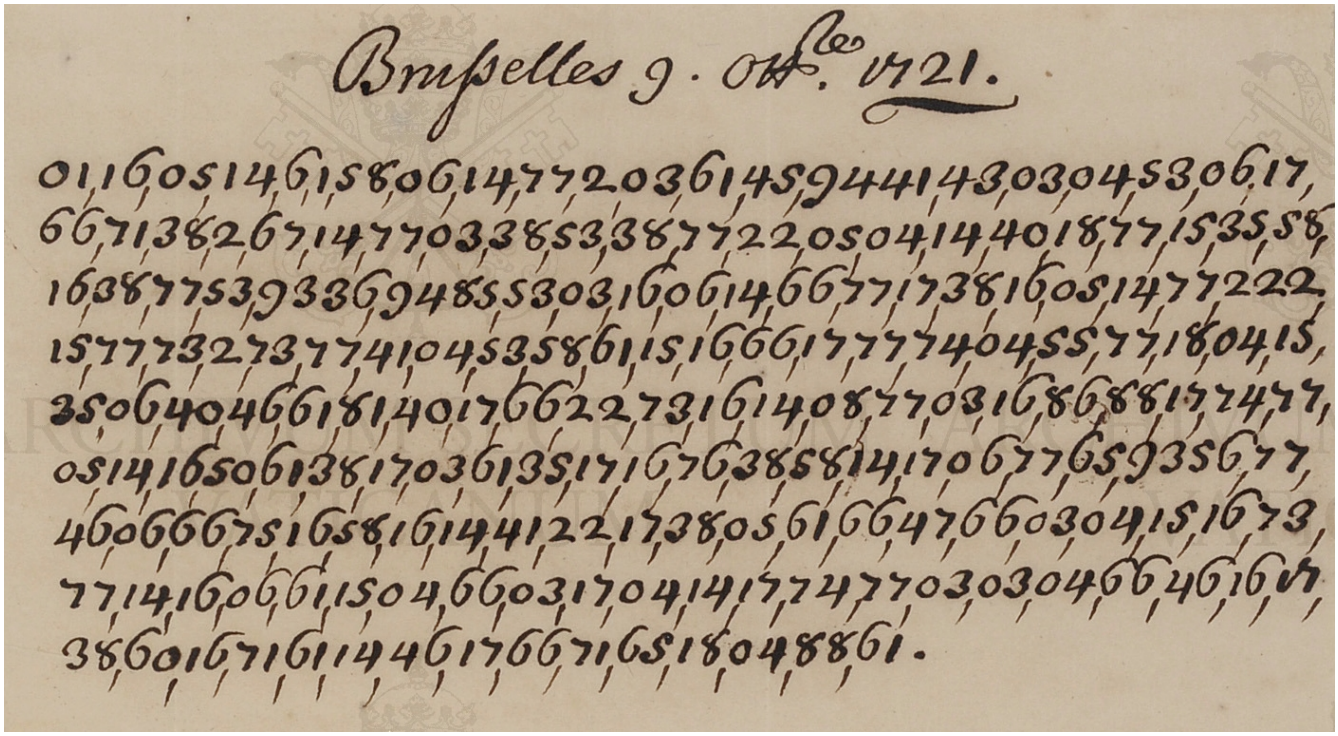

Figure 1: The Ciphertext - Source: ASV - Arch. Nunz. Colonia / 5

Alfabeto

a	*c*	*d*	*e*	*f*	*i*	*l*	*m*	*n*	*o*	*p*	*r*	*s*	*t*	*u*
6	58	46	04	73	17	15	08	66	16	05	14	01	03	06
77			40		71				61	50	41	38		60

Nulle

2	22	222

Dizionario

che	35	*de*	47	*di*	53	*il*	1	*ma*	74
ne?	81	*per*	65	*qu*	18	*se?*	86	*st*	88

Figure 2: Reconstructed Key

01	16	05	14	6	1	58	06	14	77	2	03	61	45	9441	43	03	04	53	06	17		
s	*o*	*p*	*r*	*a*	*il*	*c*	*u*	*r*	*a*	–	*t*	*o*	45	9441	43	*t*	*e*	*di*	*u*	*i*		
66	71	38	2	67	14	77	03	38	53	38	77	22	05	04	14	40	18	77	15	35	58	
n	*i*	*s*	–	67	*r*	*a*	*t*	*s*	*di*	*s*	*a*	–	*p*	*e*	*r*	*e*	*qu*	*a*	*l*	*che*	*c*	
16	38	77	53	9336	9485	53	03	16	06	14	66	77	17	38	16	05	14	77	222			
o	*s*	*a*	*di*	9336	9485	*di*	*t*	*o*	*u*	*r*	*n*	*a*	*i*	*s*	*o*	*p*	*r*	*a*	–			
15	77	73	2	73	77	41	04	53	58	61	15	16	66	17	77	74	04	55	77	18	04	15
l	*a*	*f*	–	*f*	*a*	*r*	*e*	*di*	*c*	*o*	*l*	*o*	*n*	*i*	*a*	*ma*	*e*	55	*a*	*qu*	*e*	*l*
35	06	40	46	61	81	40	17	66	22	73	16	14	08	77	03	16	86	88	17	74	77	
che	*u*	*e*	*d*	*o*	*ne?*	*e*	*i*	*n*	–	*f*	*o*	*r*	*m*	*a*	*t*	*o*	*se?*	*st*	*i*	*ma*	*a*	
05	14	16	50	61	38	17	03	61	35	17	16	76	38	58	14	17	06	77	65	9356	77	
p	*r*	*o*	*p*	*o*	*s*	*i*	*t*	*o*	*che*	*i*	*o*	76	*s*	*c*	*r*	*i*	*u*	*a*	*per*	9356	*a*	
46	06	66	75	16	58	16	14	41	22	17	38	05	61	66	47	66	03	04	15	16	73	
d	*u*	*n*	*su*	*o*	*c*	*o*	*r*	*r*	–	*i*	*s*	*p*	*o*	*n*	*de*	*n*	*t*	*e*	*l*	*o*	*f*	
77	14	16	06	61	15	04	66	03	17	04	14	17	74	77	03	03	04	66	46	16	17	
a	*r*	*o*	*u*	*o*	*l*	*e*	*n*	*t*	*i*	*e*	*r*	*i*	*ma*	*a*	*t*	*t*	*e*	*n*	*d*	*o*	*i*	
38	60	16	71	61	14	46	17	66	71	65	18	04	88	61								
s	*u*	*o*	*i*	*o*	*r*	*d*	*i*	*n*	*i*	*per*	*qu*	*e*	*st*	*o*								

Figure 3: Deciphered Plaintext