Cryptanalytic and historical challenges with unidentified encrypted documents from the early modern era

Camille Desenclos
Université de Picardie Jules Verne,
CHSSC / Université de Lorraine, CNRS,
Inria, LORIA
camille.desenclos@u-picardie.fr

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

Abstract

In most cases, historical encrypted documents include some parts in cleartext, such as headers, dates, signatures, or addresses, which allows the origin, date, and language of these documents to be established. An attached decrypted text, similar documents (same encryption, homogeneity of date or origin) in the same volume or box, or the catalog description may assist in that process. However, in a few cases, none of these are available, posing several challenges both from cryptanalytic and historical perspectives. Based on three 16th-century case studies, this paper aims to discuss a multidisciplinary method to proceed from an unidentified encrypted document to a workable transcription – decipherment and identification.

1 Introduction

Performing both a complete decipherment and an accurate historical analysis of a text requires basic identification information: who is writing? To whom? When? Such information can be missing in encrypted documents as well as in clear ones. With clear texts, the content can help determine the origin of the documents. It is the other way with encrypted texts: the origin is crucial to reconstruct the content (especially the names that are encrypted with a nomenclature).

Many encrypted documents from the early modern era are letters: the first words, courtesy formulas, dates, and/or signatures are thus cleartext, as is the last page, which typically contains the address and sometimes the sender's name, the sending and/or reception dates. Each of these cleartext segments offers the required identification information. In a few other cases, the encrypted document does not provide any identification information, but it is preserved alongside other encrypted documents (already decrypted and/or identified), allowing us to link all the documents together because of their similar ciphers. In both cases, not only are the sender, the recipient, and the language identified by this means, but also the date and the nature of the information (diplomatic, military, …) can be predicted. Those are valuable for cryptanalysis as they indicate the vocabulary and names we can reasonably expect.

Things become more challenging when the encrypted document does not provide any such information and cannot be matched with other encrypted documents. As some logic was probably involved in building up the volume or the box (same sender or recipient, same date range, same territorial area, same holder), it is tempting to assume that the documents (encrypted or not) are related, especially when they seem to be homogeneous (same topic, same date, …). However, it is possible that letters belonging to the same correspondence were dispersed and/or inserted into the wrong volume(s) or box(es). While dealing with unidentified encrypted letters, the global content of a volume or a box can serve as a hint, but not as a foundation for cryptanalysis.

These challenges, both for cryptanalysis and historical analysis, are common to every encrypted document, regardless of its origin, date, or content. This paper proposes a method, or at least some potential solutions, to overcome them, from three 16th-century case studies. As centralized archival practices were yet underdeveloped (especially in 16th-century France), Renaissance documents can present additional and specific challenges: they are often scattered between boxes or even archives, and the losses (missing letters within a correspondence, missing decrypted texts) are numerous. In addition, the content of the documents may be less intelligible, especially while cryptanalyzing: non-standardized spelling, titles rather than names for people's designation, …

Proceedings of the 8th International Conference on Historical Cryptology, pages 41-51
Poznań, Poland, 16-18 June, 2025

2 Presentation of the case studies

The three encrypted text sets have been chosen according to their representativeness for Renaissance encrypted documents: scattered letters, scattered pages, and an isolated letter. All three of them are preserved at the Bibliothèque nationale de France (BnF) and were produced during the 16th century. Each case also represents a different stage of the cryptanalysis process: cryptanalysis and identification are completed, cryptanalysis is completed but identification is incomplete, and ongoing cryptanalysis and identification.

2.1 Mary Stuart's scattered letters (1578-1584)

Several papers have recently been published on Mary Stuart's encrypted letters, which were found at the BnF (Lasry/Biermann/Tomokiyo 2023; Biermann/Tomokiyo/Lasry 2023). As the cryptanalytic work has been completed and published, we do not intend, in this paper, to uncover new elements but to examine the methodology that has led to the identification of the letters and see how much it can be replicated.

Before cryptanalysis, none of these letters had been identified except for one. It turned out, after cryptanalysis, that they had been written by Mary Stuart (Queen of Scots) between 1578 and 1584 and were addressed to the French ambassador to Elizabeth I (Queen of England), Michel de Castelnau, seigneur de La Mauvissière. In total, 57 letters are scattered within four volumes at the BnF: Français (fr.) 2988 (26 letters), fr. 3158 (1 letter), fr. 20506 (28 letters), Cinq-Cents de Colbert 470 (2 letters). Only the letter in volume fr. 3158, which has some parts in cipher and the rest in cleartext, had been identified as from Mary Stuart before the work of Lasry *et al.*: Castelnau's name and the date were written in cleartext, and an additional mention on the top margin identified Mary Stuart as the sender. For volume Cinq-Cents de Colbert 470, the connection with Castelnau and Mary Stuart could have been guessed: every document in the volume is related to him and mainly focuses on Mary Stuart; furthermore, numerous letters were written by her. For the letters within the two other volumes (fr. 2988, fr. 20506), as each letter is fully encrypted (even the address), nothing could help connect them with Mary Stuart or Castelnau.[1] Lastly, according to the catalog (the description has been updated since then), the encrypted letters in volume fr. 2988 (on which Lasry *et al.* started working) contained letters unrelated to Mary Stuart or England/Scotland, but rather related to Italy and written at the end of the 1520s.[2]

2.2 Duprat's encrypted reports (1521)

As for Mary Stuart's letters, the second case consists of scattered letters but in three closer volumes. They are preserved at the BnF: fr. 2961 (24 pages, three text sets), fr. 3029 (42 pages, eight text sets), fr. 3092 (six pages, three text sets). These texts are currently being analyzed by the authors (publication forthcoming), but they can already be identified as reports from the French embassy (composed by Antoine Duprat, Jacques de Chabannes, Jean de Selve, and Robert Gedoin) at the French-English conference in Calais in 1521.[3] With one exception,[4] these texts are fully encrypted and are deprived of any address (See Appendix 1). In one encrypted document (fr. 3029, fol. 134), a mention has been added on the top of the first folio: "Lettre en chiffre adressée au roy" (Letter in cipher addressed to the King). However, while this suggests the encrypted text may also be in French, this does not help much, as we do not know which French king this is about. Regardless of this mention, the texts may be encrypted reports attached to letters. However, these reports are not preserved in a way that reflects their initial form. Following cryptanalysis, we have been able to locate historical decryptions (from the early 16th

[1] There are a few remaining seals that display the royal arms of Scotland, but their analysis would have required heraldic knowledge (which is not often mastered by codebreakers or even historians).

[2] Three reports, written in Italian, were mentioning Mary Stuart in volume fr. 2988. As the majority of documents and especially the other encrypted letters were related to Italy, the odds were in favor of 1520s Italy and not of 1570s Scotland.

[3] See about the Calais conference, Russell 1971. As Russell points out, the conference has been well documented by both sides (French/English). The journal of Jean Barillon, Duprat's secretary, in which some of the ambassador's letters were copied, is now preserved at the BnF (NAF 6777) and was published in the 19th century (Barillon 1897-1898).

[4] Fol. 105 (fr. 3029) contains cleartext segments. The Latin text does not offer any precise element for identification (no date, no name, no event). It is the only Latin text; the reports had all been written in French.

century) of some of the reports, allowing us to determine that these 14 texts do not form 14 reports. These plaintext copies helped to distinguish nine reports, one of which was scattered across two volumes, and two of which were incomplete (the missing parts of these two encrypted reports have not yet been found). As for the rest, no plaintext copies matching our decipherments have been found. It has nevertheless been possible to distinguish four other reports, presumably complete, while two other sets have yet to be investigated but presumably form at least three separate reports. Not only are the complete set of Duprat's encrypted reports scattered between three volumes, but their pages are filed in different volumes, which prevents a full understanding of the content when deciphering them. Finally, although the content of the three volumes suggested that the encrypted documents could be diplomatic letters written in the 1510s or the 1520s, no other documents in the volumes were related to the Calais conference. Before the advent of cryptanalysis, nothing could have led to identifying those reports as having been produced by the French embassy in Calais in 1521.

2.3 Fr. 6632 isolated document (1580s)

Although they are scattered in various volumes and sometimes fragmented, Mary Stuart's letters and Duprat's reports each offer numerous encrypted pages, which helps a lot for cryptanalysis. The third and last case study addresses the opposite challenge: a unique specimen. Volume fr. 6632 (BnF) contains a two-page text (possibly a letter) that does not present any identification element and is the only encrypted text of the volume (See Appendix 2). As in the previous examples, the other letters and the catalog description may offer some guidance. Volumes fr. 6628 to 6632 contain letters, mostly sent to King Henri III of France by Queen Mother Catherine de Medici and others. However, the many letter authors in these volumes do not help to formulate any hypotheses, except for the date (probably in the 1580s). At the current stage of the working process, this encrypted text has not been precisely identified.

3 Deciphering unidentified encrypted letters: challenges and methods

There are three different approaches to accessing the content of early modern encrypted documents (which have not been decrypted yet), whether or not they have already been identified:

- Finding the original cipher key and associating it with the encrypted documents
- Recovering the key based on available decrypted text
- Reconstructing the key via cryptanalytic techniques.

3.1 Finding the original cipher key in the archives

The first two approaches can be implemented without cryptanalysis expertise, but they require archival knowledge. Keys and encrypted letters are not systematically preserved in the same volume/box (or even in the same library/archives). If the archival situation varies from country to country, from archives to archives, researchers may face four situations:

1. Encrypted letters and cipher key are preserved in the same volume/box or group of volumes/boxes[5]
2. The cipher key is preserved along with other keys[6]
3. The cipher key is preserved alone in an unrelated volume or box[7]
4. The cipher key has been lost or destroyed.

Scenario 1 is the easiest one. As soon as a cipher key is preserved in a volume or box (or a homogeneous set of volumes or boxes), a comparison with the encrypted letter(s) can be performed. This is done first by comparing the ciphertext characters and secondly by attempting to decrypt the letter using the cipher key. We can know whether the cipher key matches the encrypted letters in a few minutes. In contrast, we cannot state the loss of a cipher key (Situation 4) before conducting extensive archival work.

[5] See for instance, BnF, fr. 4030. The volume contains both the original cipher key and the encrypted letters from Villeroy to Hotman (1609-1612).

[6] Volume fr. 3995 (BnF) contains 68 cipher keys (1580-1595) that were, mainly, used by the Duke of Nevers.

[7] See for instance, BnF, fr. 3349. The volume contains an unidentified cipher key (probably from the 1580s), along with letters from Henri III and Catherine de Medici but none of them is encrypted.

As for Scenarios 2 and 3, Renaissance cipher keys were not systematically labeled by their users. The name of the users (mainly the recipient's name) and/or the date may have been written on the verso.[8] However, many cipher keys remain unidentified or, at the very least, are not sufficiently identified (i.e., lacking a date or name). As a result, it may be impossible to determine the origin of cipher keys without matching them to encrypted letters. This can be a vicious circle: encrypted letters are crucial to identify cipher keys, but cipher keys are crucial to decrypt letters. Moreover, the catalogs often do not list all the documents in the volume/box, regardless of whether they can be identified. Although the precision of catalog descriptions varies from archives to archives and even from collection to collection, we often have to deal with descriptions such as "several cipher keys",[9] with incorrect descriptions,[10] or with descriptions that do not mention the cryptographic nature of the documents.[11]

Linking a cipher key to its encrypted letters would require reviewing all the collections of several libraries and archives to ensure that we have browsed every box and volume that may contain a cipher key, regardless of the description provided by the catalogs.[12] Secondly, as the encrypted letters are not identified, it is impossible to exclude any cipher key based on criteria such as the date, the user, the territorial area, or the language: we would need to compare each cipher key to the encrypted letters manually. It is both time-consuming and ineffective. In contrast, cryptanalysis can help find the original cipher key. Matching letters and original cipher keys is much easier when the key, or part of it, has been reconstructed: it no longer requires trying to decrypt the first lines of the encrypted document with each key whose ciphertext characters seem to match and, because of the newly deciphered content, we can focus on cipher keys (and other encrypted texts) relevant to a specific date and use. A copy of the cipher key that was used by the French ambassadors in Calais, for instance, still exists: Jean Barillon, Duprat's secretary, has copied the key in his "journal" (BnF, NAF 6777), which is a record of the conference (copies of the diplomatic instructions, of the key, of some letters). Because of its nature, this journal is not preserved alongside or near the encrypted reports (which are not preserved with the other letters from the ambassadors). Matching the encrypted reports with the copy of the cipher key was possible only after cryptanalysis, as this copy (not listed in the library catalog) had already been identified and mentioned in Desenclos 2021 as a copy of the key used by the French ambassadors in Calais.[13] Similarly, Kopal and Waldispühl, who worked on a previously unidentified letter from Emperor Maximilian II (1575), were able to find the original cipher key (and two other encrypted letters) in the DECODE database, only after having cryptanalyzed the first letter (Kopal/Waldispühl 2022).

3.2 Recovering the cipher key from available decrypted text

The same issue arises with the second approach (recovering the key based on available decrypted text) as several configurations can be observed, depending on whether the decrypted text has been copied on the original document (margin, between lines) or another folio:

- Encrypted letters are preserved in several but related volumes (groups of manuscripts or boxes dedicated to the same negotiation, the same person, or the same time range), and the decrypted text has been copied on the original letters or on separate folios that can be linked to the

[8] See for example, BnF, fr. 6204, fol. 8: "chiffre avec M. d'Usson, 14 septembre 1701" (cipher with Mr d'Usson, 14 September 1701).

[9] The BnF catalog suggests that volume NAF 8431 contains "diplomatic ciphers". Without inspecting the volume, we cannot know how many cipher keys and whose key the volume really contains.

[10] According to the catalog, volume fr. 3362 (fol. 57) would contain "a cipher". However, it is not a cipher key but part of an encrypted document.

[11] Volume fr. 7129 contains 3 cipher keys; none is mentioned in the catalog.

[12] For French early modern encrypted documents, the keys can be found mainly at the BnF, the Archives nationales (national archives), the Archives diplomatiques (diplomatic archives) or the Service historique de la Défense (military archives), but one can also find isolated keys in local or private archives or in another library.

[13] Another cipher key has been found in volume fr. 2961 (BnF). It does not contain the nomenclature, but only the alphabet symbols. This key could have been easily matched to the encrypted reports: it is held before one encrypted report. However, encrypted letters in volume fr. 2961 have been found only after the first stages of cryptanalysis have been performed on the reports in volumes fr. 3029 and 3092.

encrypted letters (mentions on top or on the back folio, decrypted text which immediately follows the encrypted letter, …).
- Encrypted letters are scattered between volumes of different origins, and the decrypted text has been copied on the original letters
- The decrypted text has been copied onto another folio and preserved in a different volume from the letters, or lost or destroyed over the centuries.

Decrypted texts were preserved, or at least have been identified, only for Duprat's reports (BnF, fr. 2966-2967). However, the connection with the encrypted reports could not have been made as there were no elements available to match a report to its relevant decrypted text: the decrypted texts are contained in another volume, and even if the decrypted texts are partially identified (sender and, sometimes, recipient), the encrypted reports were not. Only recovering the content via cryptanalysis has made this connection possible.

Although it is one of the hardest configurations (unidentified encrypted letters and a separate decrypted text), Duprat's case enhances the two main requirements for recovering a cipher key (without cryptanalysis): a decrypted text that exists and that can be linked to the unidentified encrypted document(s). Matching encrypted and decrypted texts together (as matching encrypted texts and cipher keys), whether to find relevant decrypted texts or enlarge the set of texts encrypted with the same cipher, can only be based on spot checks or an exhaustive inventory of all encrypted texts. The first method cannot be performed when the texts are not identified, and the effort needed for the second method might be prohibitive: matching documents requires both finding other similarly-looking encrypted documents and reconstructing each cipher key manually from these documents.

A third approach could be to look at close volumes or boxes (same territorial area, same period, same writers), but it can be hard to perform without any slight identification element. To overcome this, we can try to determine the origin of the documents and then investigate nearby volumes from a similar origin, rather than similar themes. All three volumes containing Duprat's reports belong to the former collection of Philippe de Béthune, part of which was organized by reigns (Solente 2001). As the content of volume fr. 3029 belongs to the sub-collection related to Francis I's reign, we could simply look into Béthune's volumes that belong to the same sub-collection. However, in the sub-collection related to Francis I's reign, there are more than 170 volumes. Other encrypted reports have certainly been found in volumes that belong to this sub-collection (fr. 2961 and fr. 3092). However, we cannot be absolutely sure that Duprat's reports had not already been scattered before the constitution of Béthune's collection and preserved in distinct collections or even libraries. In fact, one of the decrypted texts does not match any encrypted reports and suggests that at least one report is missing. Even by investigating the volumes only based on their origin, the archival investigation work remains highly time-consuming and represents a significant investment for a few unidentified texts.

In Mary Stuart's case, the pre-existing identification of the letter in volume fr. 3158, the credible identification of Castelnau as the recipient or intermediary of the letters in volume Cinq-Cents de Colbert 470, and the similarities between all the encrypted letters (e.g., "K" at the beginning and in the addresses) could, in theory, have led to identifying all these letters as Mary Stuart's. That would have required prior knowledge of the existence of these four text sets with similar graphical ciphertext characters. Yet, they are scattered in very distinct volumes: two of them were not supposed to contain Mary Stuart's letters, and all four belong to different collections. Volumes fr. 2988 and fr. 3158 come from Béthune's collection,[14] volume fr. 20506 from Gaignières's collection and volume Cinq Cents de Colbert 470 from Colbert's collection. Even if work similar to the work done for Duprat's case (going through all the letters from the same collection) had been done, it would have been unsuccessful. Matching all four volumes was made possible only after successful cryptanalysis. According to Lasry/Biermann/Tomokiyo 2023, the familiarity with the ciphertext characters in

[14] Volume fr. 2988 was part of the ensemble related to Francis I's reign, while volume fr. 3158 was part of the ensemble related to Francis II's reign. Although Mary Stuart was Francis II's wife, the encrypted letters in fr. 3158 were not written during Francis II's reign (nor Francis I's).

volume fr. 2988 made possible a match with volume fr. 20506, on which Tomokiyo was working for another purpose (to add new collections to Cryptiana).[15] While this was partially by luck (looking at this volume at this particular time), this highlights the benefits of systematic surveys but also their limitation: Tomokiyo reconstructs keys but works only on documents that are digitized and put online by the archives and libraries (part of the collections only); the DECODE database is not limited to digitized documents but depends on the contributors intakes (Heder/Megyesi 2022); Desenclos is conducting a systematic survey at the BnF (Desenclos 2021), but it has not been yet published. For Mary Stuart's letters, identifying Castelnau as a possible recipient of the encrypted letters made the search feasible by inspecting only volumes that may contain letters from and to Castelnau (Lasry, Biermann, and Tomokiyo 2023). While some documents' history can be retraced, this is not always the case. As Mary Stuart's letters were addressed to Castelnau, they should have been (at least at first) preserved along with his papers. The fate of Castelnau's papers is unclear, and although a part of them has been preserved as an ensemble (fr. 3158), the other part has been scattered throughout history. Castelnau's papers may have been integrated into Montmorency's papers (through marriage and patronage), from which Philippe de Béthune, François-Roger de Gaignières, and Jean-Baptiste Colbert have built up their own collection. No matter which thread is pulled, there is no way (except for a tedious, exhaustive inventory of thousands of volumes) to match documents, whose consistency and identification are often lost for centuries. Analyzing a volume composition cannot be considered a solid way to identify a document, but rather a hint.

3.3 Reconstructing the key via cryptanalytic techniques

For unidentified encrypted texts, their decipherment through cryptanalysis is often the best, or even the only way to identify the text, to enable finding other encrypted texts (with the same key) and, of course, to access their content. Deciphering historical encrypted documents typically consists of four stages:

1. Computerized cryptanalysis to identify the meaning of most ciphertext characters representing individual letters of the alphabet – the homophones in the common case of homophonic ciphers.
2. Identifying the meaning of the remaining ciphertext characters representing individual letters of the alphabet and those representing doubled letters, nulls, or requiring an action (e.g., to delete the previous character). This phase is typically performed manually, based on the partial results obtained in Stage 1, by analyzing all instances of the given ciphertext characters. The more ciphertext material is available, the higher the chance those characters will be identified in their majority and accurately. Also, by examining which deciphered sequences are plausible and which are not, it is possible to determine which ciphertext characters are likely part of the nomenclature.[16]
3. Identifying the meaning of nomenclature elements representing common words, prepositions, or parts of words. This is also performed by examining the various instances of those elements in their textual and sometimes historical context.
4. Identifying nomenclature elements that represent proper names. Identifying the text and its historical context is critical at this stage (see Section 4).

These four stages are the same, regardless of whether the document has been identified. However, the absence of any identification element (date, nature, context) makes cryptanalysis harder, even impossible in some cases. Identifying the meaning of some nomenclature elements, either names (Stage 4) or infrequently used words (Stage 3), relies on prior knowledge of the context; without it, cryptanalysis will remain incomplete. Identifying, in Stages 1 and 2, the ciphertext characters for the letters of the alphabet (homophones) and frequent nomenclature words (Stage 3), in contrast, will generally still be feasible given enough ciphertext, but may require testing more hypotheses, notably for two reasons: both the language and the cipher

[15] Cryptiana is a website that aims to identify and reconstruct historical ciphers from archives documents in (https://cryptiana.web.fc2.com/code/crypto.htm).

[16] This also allows for determining the structure of nomenclature symbols, such as the addition of certain diacritics, or specific ranges of numerical codes.

features are completely unknown.[17] Cryptanalysis software and algorithms rely on language statistics computed from a corpus of the original language. Even for encrypted letters with the slightest identification, language can be determined or, at the very least, narrowed down to a few possibilities. Assuming the wrong language will cause cryptanalysis algorithms to fail. For Mary Stuart's case, as shown in Lasry/ Biermann/Tomokiyo 2023, a wrong initial assumption based on the catalog and the content of the first studied volume (fr. 2988) – that the letters were originally in Italian – resulted in the project being significantly delayed. If the language is unknown and cannot be inferred from other clues or if the first assumption does not give any conclusive result, all possible languages (and their territorial variations) but also all possible types of ciphers (transposition, homophonic substitution, polyalphabetic substitution, …) and all possible structures of ciphers must be tested for successful cryptanalysis while they are usually guessed from the date, the nature and the author of the document. A French cipher from the 1580s is more likely to be homophonic, with 2-4 homophones per alphabet letter, to utilize nulls, and to feature a nomenclature that incorporates both names and common words. In contrast, a French cipher from the 1520s may be monoalphabetic or have a low number of homophones and a small nomenclature (mostly names). Both structures can be observed in the fr. 6632 letter (homophonic cipher with at two to four homophones per letter of the alphabet) and in Duprat's reports (monoalphabetic substitution, twelve nomenclature elements, and five nulls).

French Renaissance ciphers were less complex and had fewer features, so the cryptanalysis of earlier encrypted texts was usually easier. With Duprat's reports, given the small number of distinct ciphertext characters (about 35), it was hypothesized that the cipher could be a simple monoalphabetic substitution cipher with a few additional elements, such as names (small nomenclature) and nulls. Based on the mention "Lettre en chiffre adressée au roy",[18] in the top margin of one of the encrypted texts (fr. 3029, fol. 134), we rightfully expected the original language to be Middle French. Computerized cryptanalysis easily recovered the ciphertext characters assigned to individual letters of the alphabet; then, six characters representing nulls and four nomenclature elements were also recovered with manual work.[19] In contrast, when the cipher is homophonic and/or uses an important nomenclature, more hypotheses must be formulated, making cryptanalysis harder as these stages are seldom sequential, and the process is mainly iterative. When deciphering Mary Stuart's letters, as shown in Lasry/Biermann/Tomokiyo 2023, the special ciphertext characters (to delete or repeat the last ciphertext character) and the structure of the nomenclature elements – five different meanings for a single ciphertext character based on adding diacritics – were hard to identify. Some of the diacritics were first assumed to be individual ciphertext characters. Only after they identified the origin of the letters were Lasry *et al.* able to locate examples of related cipher keys in archives, similar in structure and contents to the one used in the collection they analyzed,[20] and then to confirm several cryptanalytic hypotheses. Even at this stage, the knowledge of the cryptographic practices (in a specific area, at a specific time) is fundamental for cryptanalysis, and unidentified texts deprive the codebreaker(s) of essential elements for success.

Those challenges are compounded when the cipher is complex (e.g., a large number of homophones) and/or if the ciphertext is short. For the third case study (fr. 6632), we have only obtained partial and tentative results, given the short length of the text (1,100 ciphertext characters) and the high number of distinct ciphertext characters (120). Based on this and the dates of the other letters in the volume (France, 1580s), we hypothesized that this cipher was homophonic with many homophones per letter and included at least a few tens of nomenclature elements. Computerized cryptanalysis (Stage 1)

[17] In the only other documented example of unidentified deciphered documents we have found (Kopal/Waldispühl 2022), the authors had also to base on assumptions about language, date and cipher type to process cryptanalysis.

[18] Letter in cipher, addressed to the King.

[19] Only six of the twelve nomenclature elements are used in the encrypted reports we have identified so far.

[20] Lasry *et al.* have found two ciphers, one used between Mary Stuart and Guillaume de L'Aubespine, baron de Châteauneuf, Castelnau's successor as French ambassador in England, and another preserved along Castelnau's papers. Both share similar features (diacritics, similar nomenclature vocabularies). Their finding helped Lasry *et al.* to confirm several cryptanalysis hypotheses (Lasry/Biermann/Tomokiyo 2023).

was challenging due to the cipher complexity and the short ciphertext, and it required multiple tests with various parameters, leaving many illegible gaps, as expected, because several ciphertext characters were wrongly assumed to be homophones. In Stage 2, all the homophones could be interpreted, corrected, and distinguished from nomenclature elements (which could not be interpreted at this stage), resulting in additional legible segments. Here, the transcription of the first six lines after Stage 2 (the homophones corrected at this stage are marked in bold, and nomenclature codes with a question mark):

"*Madame* ***dair*** *aiant eu* ***? ?****modité en une* ***f****este* ***?*** *se* ***f****eit ché****s*** *elle, entra* ***? ?*** *en propos* ***? ? ?*** *tres p****r****egnantes raisons* ***? ?*** *persuad****? ? ?*** *veoir* ***? ? ?****ntendre au ma****rr****iage sur* ***? ?*** *ses plaintes* ***? ?*** *gran?s protestations de sa"*

Stage 3 was tedious and yielded incomplete results, while Stage 4 could not be performed due to the short ciphertext and the lack of historical context. About ten ciphertext characters could not be identified, including some likely to represent proper names:

"*Madame, Dair aiant eu* ***la com****modité en une feste* ***que*** *se feit ches elle, entra* ***bien ?*** *en propos* ***avec ? avec*** *tres pregnantes raisons* ***pour luy*** *persua****de****r et de veoir* ***?*** *et* ***d'e****ntendre au mariage sur* ***? ?*** *ses plaintes* ***et les*** *gran****de****s protestations de sa"*

4 Identifying unidentified encrypted letters: challenges and methods

For Stages 3 and 4, in particular, knowing the historical context and identifying the document are crucial. Although that depends on the size of the nomenclature and the length of the encrypted document, cryptanalysis can fuel a virtuous circle: reconstructing the plaintext, even without names, can enable the identification of the encrypted document; this identification can, in turn, help identify the words and names (nomenclature); a more complete decipherment can improve the identification of the document, and so on.

Even if the names have not been deciphered yet, the first stages of decipherment may offer several levers for identifying the document, such as names spelled and encrypted in full, or references to an event or a person. As described in Lasry/Biermann/Tomokiyo 2023, Mary Stuart's letters were identified as such based on clues in the deciphered text. There were mentions of a son ("mon fils"), of being in captivity ("ma liberté"), as well as feminine adjectives, and names spelled in full, such as Walsingham (Elizabeth I's principal secretary and her spymaster) or Spain. Similarly, with Duprat's reports, the main clues came from names spelled out in full before encryption. Due to the small size of nomenclature, most of the names were spelled and encrypted in full (Fitzwilliam, bishop of Durham, Duke of Albany, …) and helped a lot to identify the documents as French reports produced during negotiations with England. This has been confirmed by mentions of discussions between "le chancelier" (the chancellor, here Antoine Duprat) and "le cardinal" (the cardinal, here Cardinal Thomas Wolsey) in the context of concurrent discussions with ambassadors from various places (Hungary, Venice, ...). All those pointed toward the negotiations that took place between July and November 1521 in Calais.

It may come as a surprise to see so many names not being encrypted with nomenclature elements. In Duprat's reports, the reason was the simplicity of the cipher: French ciphers in the early 1520s were not as advanced as Italian or Spanish ones, and here, the cipher was designed for a time-limited use in a not-too-far-away embassy. The needs, in terms of security, were not the same. In other cases, such as Mary Stuart's or even the letter from volume fr. 6632, that is not surprising either, and these are no exception. Not all names were intended to be encrypted through a nomenclature (infrequently-mentioned names, people who burst into the political scene after the cipher key was created, etc.). Moreover, encrypting names using homophone elements after fully spelling them out limited the frequency of nomenclature elements assigned to important names. For 21st-century codebreakers, having names encrypted in full facilitates the identification of the origin and context of the text, further assisting in the decipherment process. Certainly, such encryption practices also helped early modern codebreakers, but only in recovering names. When the letters were intercepted and deciphered, the (approximate) date of the letter and its probable origin were already known: the codebreaker knew when and where the letter had been intercepted. He could easily guess the language, nature, and features of the cipher, as well as some frequently used words and names. The decipherment of unidentified documents through codebreaking highlights the gap between early modern codebreakers and 21st-century cryptanalysis. The latter is disadvantaged because

the knowledge of the writing context has been lost, making identifying common words and names more complex, if not impossible. For example, with our third case study (fr. 6632), it is quite possible that an early modern codebreaker would have been able to identify the parties of the marriage mentioned in the text, whether or not the marriage negotiations were public (rumors, intelligence, etc.). Conversely, it is impossible, at our times, to identify those names without any other supporting element. Cryptanalysis may initiate a virtuous circle, but can also lead to a dead end. With the fr. 6632 letter, the challenge is increased by the frequent use of distinct nomenclature elements for names. The reconstructed text segments are short and insufficiently explicit about the people or events mentioned. Here lies a second challenge that 21st-century researchers face: many elements are implicit in historical documents, whether or not they are encrypted. People are never mentioned by their full names but only by shortened names, which can apply to several people from the same bloodline (e.g., the Duke of Albany), or by one of their titles (the cardinal, the Grand Squire, etc.). These made immediate sense to early modern people and can make sense to 21st-century historians when the date is available. Without additional information, these shortened names and titles can remain insufficient to enable the identification of a document.

In many cases, nevertheless, it is possible to identify, at least partially, the sender and recipient and the date. In Mary Stuart's letters (Lasry/Biermann/Tomokiyo 2023), the date and the recipient's name were encrypted and have been recovered via cryptanalysis. In Duprat's reports, no identification elements were written, even in cipher. As a first step, only assumptions can be made about the sender, recipient, or date. Although the sender is the most essential element, two difficulties may arise: the letter's content may not be sufficient to identify the source; there may be several senders. The letters of the French embassy in Calais were written and sent either on behalf of the four ambassadors (joint signature) or by only one ambassador (single signature). The phrase 'moy chancellier' (I, chancellor) might suggest that only Duprat wrote some letters; however, the same phrase is also found in unencrypted letters sent by the whole embassy.

Identifying the addressee can sometimes be easier when the encrypted document appears to be a letter. Letters begin with the addressee's name, or at least with an adapted term ('Mr. Chancellor'), which allows either identifying the addressee or formulating some hypotheses. These can be confirmed by the document's content and/or by additional documents. The cleartext letters from the French embassy in Calais, for example, were mostly addressed to Francis I or Florimont Robertet (who was then in charge of Foreign Affairs); it is easy to assume that reports were also addressed to one of them; but nothing in the report gives a clue about the precise addressee (no words such as "Your Majesty" or "Sir"). Except when the addressee's name is written at the beginning of the letter, identifying the addressee can be particularly complex if the sender is not identified. In the case of fr. 6632 letter, the first word of the letter ("Madame") could indicate that the sender was addressing a woman (possibly Catherine de Medici, given the other documents in the volume), but the text could also begin without addressing anyone. "Madame Dair" – the first two words of the text – could be one of the text's subjects rather than the addressee.

Even if it is often easy to determine the time range of a document, it can be harder to determine a more precise date (day, month, and year). Given the nature of the cipher (number of homophones, type of ciphertext characters, ...) and the content of the document, it may be possible to identify, at least, the decade in which it was written. Although we have been able to date Duprat's reports to 1521 (the date of the embassy), no precise date (day and month) has yet been established for the encrypted reports. Doing so will require reading every unencrypted letter in related volumes to establish a daily or, at least, weekly chronology of the embassy. This work may be successful, but, for now, we have only been able to provide approximate dates (e.g., between 2 October and 10 October) for some reports.

Even a partial or hypothetical identification may enable further work to be carried out: research in the archives can be conducted, as the new identification elements may help limit the investigation's scope. Nevertheless, challenges similar to those outlined in sections 3.1 and 3.2 may persist, depending on the varying accuracy of catalogs and on the information elements that remain missing. Although researchers may get stuck at this stage due to a lack of new identification elements, the deciphering, then the identification, even partial, of the document and,

possibly, the discovery of additional sources (letters, encrypted or not, produced by the same sender and recipient, cipher key, decrypted texts) can lead to valuable results, particularly with recovering nomenclature elements. In the same way that the linguistic context enables the identification of certain common words (by reconstructing the meaning of the sentence), the historical context enables the identification of remaining ciphertext characters, whether they represent common words ("army," "peace," etc.) or proper names (people and places). Even before linking it to the cipher key copied by Jean Barillon, the identification of Duprat's letters had enabled the identification of four nomenclature elements. Only two causes can prevent the nomenclature from being completely reconstructed: the frequency of the ciphertext characters is too low, and they cannot be attributed with certainty to a word or name; the encrypted text is too short and contains numerous nomenclature elements.

5. Conclusion

Regardless of their date or origin, unidentified encrypted documents are less prevalent than identified ones. Researchers often do not address them, and these documents pose multiple historical, archival, and cryptanalytic challenges. Being deprived of any identification element complicates every step of the document analysis, especially for early encrypted documents. When facing an encrypted document, the first reaction is to search for the original cipher key, or encrypted and possibly decrypted texts. This is most often not feasible with unidentified documents, requiring exhaustive and time-consuming work in archives and libraries or a fair amount of luck. Our three case studies suggest that cryptanalysis may be the most effective way to achieve, quicker and better, the identification (and analysis) of early modern documents, even though cryptanalysis itself faces additional challenges as neither the language, the historical context, nor the cipher type is known a priori. However, cryptanalysis can enable the identification of the document thanks to the deciphered content.

While the three cases have varying success (depending not only on the absence of identification but also on the size of the encrypted texts), they all emphasize the dual nature of codebreaking for early modern documents: it does not rely only on cryptanalytic techniques but also on historical understanding. Reconstructing the nomenclature, for instance, depends on the document's identification and historical knowledge of the context. With unidentified documents, even if it is easier to describe the cryptanalysis process as distinct stages, it looks more like an iterative process: multiple tests are made based on hypotheses, cryptanalysis leads to better identification, which leads to better cryptanalysis… This representation of the process as iterative highlights the benefits of continuous collaboration between codebreakers and historians: historical and cryptanalytic inputs are relevant and essential throughout the entire process, not just at the beginning (cryptanalytic techniques) or at the end (historical knowledge).

Funding

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

6. References

Barillon, J., 1897-1899. *Journal de Jean Barillon, secrétaire du chancelier Duprat, 1515-1521*, ed. Pierre de Vaissière, 2 vol.

Biermann, N., Tomokiyo, S., and Lasry, G., 2023. *What encryption errors can reveal: cross-cipher errors in Mary Queen of Scots' letters*. Proceedings of the 7th International Conference on Historical Cryptology HistoCrypt2024, 17-26.

Desenclos, C., 2021. *Écrire le secret quotidien. Pratiques de la cryptographie au sein de la diplomatie française (XVIe-premier XVIIe siècle).* Spies, Espionage, and Secret Diplomacy in the Early Modern Period, 85–103.

Lasry, G., Biermann, N., and Tomokiyo, S., 2023. *Deciphering Mary Stuart's lost letters from 1578-1584*. Crypotologia, 47, 101-202.

Héder, M., Megyesi, B., 2022. *The DECODE Database of Historical Ciphers and Keys: Version 2.* Proceedings of the 5th International Conference on Historical Cryptology HistoCrypt2022, 111-114.

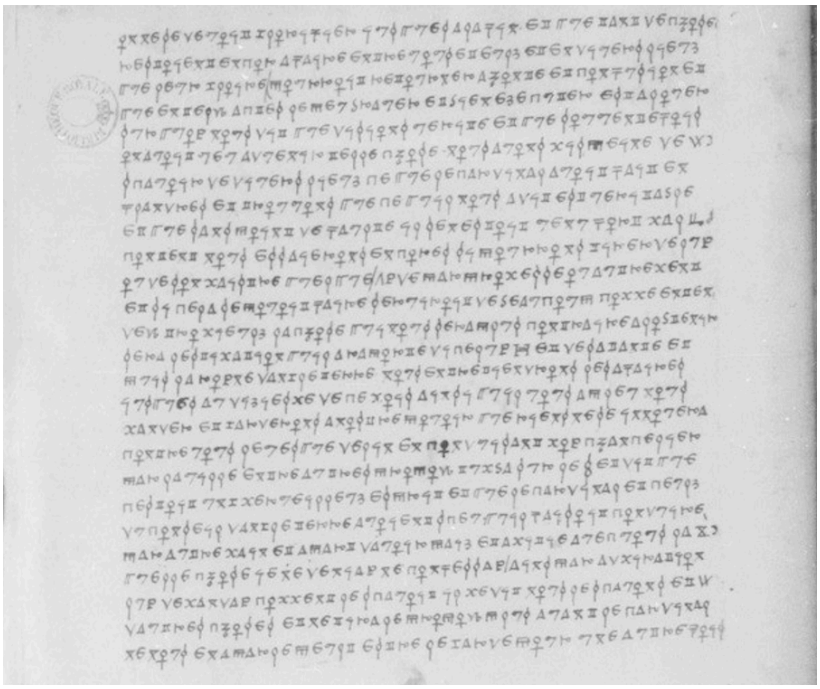
Kopal, N., Waldispühl, M., 2022. *Deciphering three diplomatic letters sent by Maximilian II in 1575.* Cryptologia, 46, 103-127.

Russel, J. G., 1971. *The Search for Universal Peace: the Conferences at Calais and Bruges in 1521*. Historical Research, 44, 162.

Solente, S., 1980-2001. *Les manuscrits des Béthune à la Bibliothèque nationale*.

7. Appendix

7.1. End of a scattered Duprat's report (BnF, fr. 3092, fol. 101 ©Gallica-BnF)



7.2. First page of the fr. 6632 letter (BnF, fr. 6632, fol. 39 ©BnF)