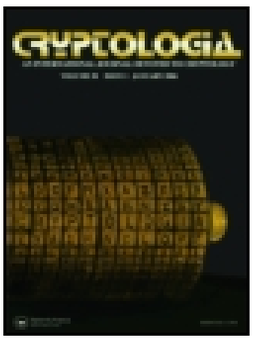

Cryptologia

ISSN: 0161-1194 (Print) 1558-1586 (Online) Journal homepage: https://www.tandfonline.com/loi/ucry20

Decryption of historical manuscripts: the DECRYPT project

Beáta Megyesi, Bernhard Esslinger, Alicia Fornés, Nils Kopal, Benedek Láng, George Lasry, Karl de Leeuw, Eva Pettersson, Arno Wacker & Michelle Waldispühl

To cite this article: Beáta Megyesi, Bernhard Esslinger, Alicia Fornés, Nils Kopal, Benedek Láng, George Lasry, Karl de Leeuw, Eva Pettersson, Arno Wacker & Michelle Waldispühl (2020): Decryption of historical manuscripts: the DECRYPT project, Cryptologia, DOI: 10.1080/01611194.2020.1716410

To link to this article: https://doi.org/10.1080/01611194.2020.1716410

CRYPTOLOGIA
https://doi.org/10.1080/01611194.2020.1716410

Decryption of historical manuscripts: the DECRYPT project

Beáta Megyesi[a], Bernhard Esslinger[b], Alicia Fornés[c], Nils Kopal[b], Benedek Láng[d], George Lasry[e], Karl de Leeuw[f], Eva Pettersson[a], Arno Wacker[g], and Michelle Waldispühl[h]

[a]Uppsala Universitet, Department of Linguistics and Philology, Uppsala, Sweden; [b]Faculty 3, University of Siegen, Siegen, Germany; [c]Computer Vision Center, Computer Science Department, Universitat Autònoma de Barcelona, Barcelona, Spain; [d]Budapest University of Technology and Economics, Department of Philosophy and History of Science, Budapest, Hungary; [e]External Researcher, DECRYPT/CrypTool projects, Givataim, Israel; [f]Informatics Institute, University of Amsterdam, Amsterdam, Netherlands; [g]Bundeswehr University Munich, Munich, Germany; [h]Department of Languages and Literatures, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

Many historians and linguists are working individually and in an uncoordinated fashion on the identification and decryption of historical ciphers. This is a time-consuming process as they often work without access to automatic methods and processes that can accelerate the decipherment. At the same time, computer scientists and cryptologists are developing algorithms to decrypt various cipher types without having access to a large number of original ciphertexts. In this paper, we describe the DECRYPT project aiming at the creation of resources and tools for historical cryptology by bringing the expertise of various disciplines together for collecting data, exchanging methods for faster progress to transcribe, decrypt and contextualize historical encrypted manuscripts. We present our goals and work-in progress of a general approach for analyzing historical encrypted manuscripts using standardized methods and a new set of state-of-the-art tools. We release the data and tools as open-source hoping that all mentioned disciplines would benefit and contribute to the research infrastructure of historical cryptology.

KEYWORDS

automatic decryption; cipher collection; historical cryptology; image transcription

1. Introduction

Ever since humans invented writing systems, there has been a need for hiding certain messages from others than the intended receivers. For this purpose, various types of techniques were developed throughout the centuries

CONTACT Beáta Megyesi beata.megyesi@lingfil.uu.se Uppsala Universitet, Department of Linguistics and Philology, Box 635, Uppsala, 751 26 Sweden.

from steganography to encryption. Thousands of encrypted manuscripts are found in archives, documents that are not yet available for historical research. Examples of such materials are diplomatic and military correspondence and intelligence reports, scientific writings, private letters and diaries, as well as manuscripts related to secret societies and magic. Many scholars and scientists are working on some of these documents in an uncoordinated fashion, in various scientific areas such as history, linguistics, philology, computer science, and computational linguistics, all with their own point of view, purpose and methods. They encounter the same or similar problems when confronted with encrypted sources.

Not much is known in detail about the encrypted documents throughout the centuries, especially during the Early Modern period from the late Middle Ages (c. 1500) to the Age of Revolutions (c. 1800). Whereas various algorithms and tools have been developed to decipher the most common forms of encryption, most of these are not suitable to deal with historical, hand-written encrypted documents that are written in nonstandardized languages, are often hybrid in nature (mix of ciphertext and cleartext, notes from different authors), contain mistakes, and are not available in digitized form. Large-scale studies have not been possible due to the lack of infrastructural resources and tools for historical cryptology.

In this paper, we present the DECRYPT project, aiming to build infrastructural support for historical cryptology by bringing the expertise of the various disciplines together, to transcribe, process and decrypt the historical encrypted sources. The resources and tools are released through a web-service with information about the provenance of the sources and other facts of relevance, and provide tools for (semi-)automatic decryption. We focus on the development of software tools for automatic analysis allowing users to decrypt various types of encrypted historical documents.

2. Historical cryptology

Historical cryptology is the study of encrypted messages from our history aiming at their decryption and contextualization by analyzing the mathematical, linguistic and other coding patterns and their histories. There is no established discipline as yet, but the field has been sketched in broad outline by David Kahn in his Magnum Opus *The Codebreakers*, first published in 1967 (Kahn 1996). Since then some books and many case studies have been published (e.g. Bauer 2007; Schmeh 2015; Singh 2000), but no systematic exploration of this field has ever been undertaken. Algorithms and software for the analysis of historical ciphers are generally not in focus in current research on cryptology. This is because classical ciphers were

supplanted in the 1960–70s by the advanced number-theoretic methods that drive today's military and commercial encryption.

However, secrecy in history has received increasing attention in recent decades. Case studies have been published on particular aspects of the use, means and goals of cryptology (e.g. Bonavoglia and Pellegrino 2016; Lasry et al. 2017; Láng 2015; Patarin and Nachef 2010; Strasser 2012) but a systematic overview of relevant sources and their analysis is as yet missing. Focus has been on the inventors of enciphering methods (such as Battista Alberti, Trithemius, Della Porta, Selenus, Vigenère, Falconere), but their complicated techniques were rarely applied until the 18th century.

Neither the systematic collection of enciphered manuscripts, nor the production of software tools supporting the decryption of these received much research attention. Systematic studies have been difficult to perform since the encrypted research material lies scattered in archives and libraries, is poorly indexed and therefore difficult to find. And although (semi-)automatic decryption tools for some ciphers have been developed and some are publicly available (e.g. CrypTool 2), these are not directly applicable for systematic decipherment of authentic historic documents in large scale.

However, current science and technology enables a piecemeal achievement of the goals of the project: to provide a data collection of encrypted handwritten manuscripts, to semi-automatically convert the images to a machine readable format, to automatically detect various cipher types and the plaintext language, and to automatically decipher the encrypted documents.

To our knowledge, there are only two open-access research projects worldwide that deal with the automatic decryption of historical ciphers in larger-scale. One is the project called DECODE (DECRYPT, 2020): Automatic decoding of historical manuscripts, financed by the Swedish Research Council between 2015 and 2017, with partly the same objectives as the DECRYPT project, albeit on a much more limited scale without taking into account the historical aspects or advanced decryption algorithms. The DECODE project aimed at the development of computer-aided tools for semi-automatic decoding of historical source material. The project included the collection of several hundred ciphers and the development of a set of metadata to describe those. For decoding, historical texts from 14 languages were collected, and various language models were released. Transcription guidelines were developed, and preliminary studies of semi-automatic transcription by image processing were carried out (Fornés, Megyesi, and Mas 2017), i.e., tools that help automatic cryptanalysis and decoding.

Another project focusing on the implementation of various crypto algorithms, mostly for modern ciphers, is the CrypTool project (CrypTool, 2020). CrypTool is an open-source project since 1998 with more than 300 contributors from different universities worldwide. It became the standard for easy-access in teaching cryptology by building the most-widespread,

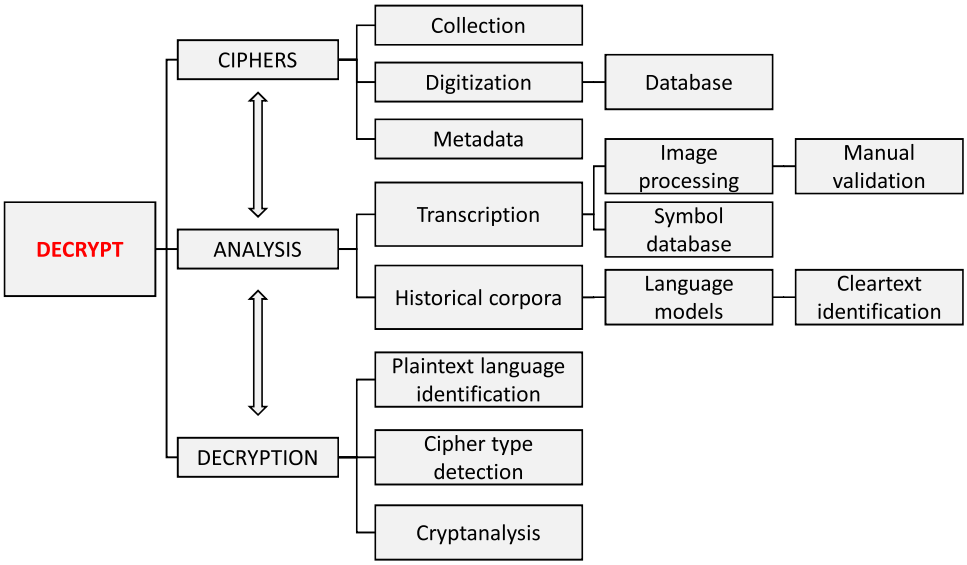

Figure 1. The DECRYPT project modules.

free e-learning programs for cryptography and cryptanalysis (Hick, Esslinger, and Wacker 2012; Kopal et al. 2014). In the area of classical ciphers, some of them are already implemented including their cryptanalysis. Another part of the CrypTool project is the international cipher contest *MysteryTwister C3* (MysteryTwister C3, 2009–2020) to solve ciphers.

Next, we will describe the DECRYPT project, aiming at a large-scale, cross-disciplinary approach for fast and reliable progress in the field of historical cryptology.

3. The DECRYPT project

DECRYPT focuses on the refinement and development of the tools involved in the automatic processing and decryption of encrypted historical sources. The aim of the development of these tools is to enable (semi-) automatic decoding of unknown and enciphered manuscripts. The different steps involved in this process are illustrated in Figure 1, and include 1) data collection and digitization, 2) analysis needed prior to decryption, and 3) decryption including cryptanalysis. These steps will be described in detail in the subsequent sections.

3.1. Collection and digitization

In order to study the development of historical cryptology in Europe, a large collection of encrypted sources of different types — keys, ciphertexts, encryption descriptions — in their original form are needed, collected from

various time periods, countries and regions, and written in many different languages. It is a time-consuming and cumbersome endeavor as encrypted sources are seldom indexed in libraries and archives, and these are therefore difficult to find. Once a historical encrypted source is found, it has to be digitized, if it is not available digitally. We can do so by ordering relevant images from the archives, often for a (quite expensive) fee and for personal use only. To collect ciphers, project participants, mainly historians specialized in particular time periods and areas, all with interest and expertise in historical cryptology dig the archives and libraries for ciphers. Currently, collections of historical sources are underway in the archives from Austria, France, Germany, Italy, Hungary and the UK. We welcome researchers to help us expand the collection.

The encrypted sources are described with a set of metadata with information about the current location, provenance, content and format of the manuscript to enable search, and research on historical contextualization and decryption. The encrypted sources with their metadata description are stored in a publicly available database, which will be described in Section 4.1.

3.2. Analysis

Ciphers normally use a key for an encryption algorithm to generate a *ciphertext* from the underlying *plaintext*. In archives, we usually do not find the keys nearby the ciphertext and the corresponding plaintext, as some of the documents have been destroyed or stored in various places without any links or information about their connection. To make systematic studies on cipher development and usage over time, we need to take care of various scenarios depending on the available document type:

(1) ciphertext only: Reproduce the key and the plaintext by applying cryptanalysis,
(2) ciphertext and plaintext: Reveal the key by segmenting the code sequence in the ciphertext and map those to the plaintext characters and n-grams.
(3) ciphertext and key: Generate the plaintext by applying a decoder, and
(4) plaintext and key: Produce a ciphertext by an encoder.

The first scenario (type 1), i.e. the cryptanalysis, is treated as the most difficult of the four types, where we have to make educated guesses about the plaintext language and the cipher type before decryption.

Mapping ciphertext and plaintext (type 2) has not been devoted much attention among cryptologists, but historians struggle with the mapping of

ciphertext symbols and codes to plaintext characters or words by manual, often time-consuming efforts. Here, we can help out by automatic methods where highly probable, alternative mappings of codes to plaintext entities are tested against historical language models. Allowing large scale studies where several ciphertexts can be tested against a plaintext, or vice versa is also a great advantage of using automatic ciphertext-plaintext mappings.

Cases where the key is available, either with a ciphertext (type 3) or a plaintext (type 4) are usually straightforward and complications occur only when the key contains ambiguities (e.g. in case of polyphonic keys), or if the manuscript is difficult to interpret because of damaged material (paper or parchment), or unclear handwriting style.

To process and analyze real-world ciphertexts, the first step is the digitization and processing of the historical source. We develop tools for (semi-) automatic transcription of the images, provide statistical analysis based on the ciphertext (e.g. n-gram statistics and clusters of symbols) and keys (e.g. key type recognition), develop and evaluate language models based on authentic historical language data that can be used for transcription, and in later steps for plaintext language identification and cryptanalysis.

To reduce the tedious, error-prone and time-consuming process of manual transcription of the images, we apply image processing using deep-learning architectures. We develop specific handwriting recognition methods to transcribe encrypted manuscripts. The transcription across different ciphers also applies to develop classifying symbol characters and to generate a symbol database across various types of symbol sets used in ciphers.

Indeed, image processing of historical sources in general and of encrypted sources in particular is challenging because the handwriting style and material degradation highly distort the shape of characters, which provoke errors in the transcription. Since transcription and deciphering are usually separated subsequent tasks, errors in transcription are propagated, affecting heavily the decryption. Therefore, we investigate the integration of image processing and automatic decryption into one single step (as was proposed for the first time by Kevin Knight and presented in a pilot study (Yin et al. 2019)) or into an iterated pipeline (feedback). This joint architecture will hopefully fasten the time-consuming transcription, minimize the errors and create synergy effects as both image processing and automatic decryption tools rely on statistical language models and clustering of symbols, which could be commonly used.

Language models for historical language data are needed in order to identify the plaintext language, and help reduce the search space during decryption. As historical texts contain a large variation in spelling and as languages change over time, we need to develop models to handle this variation during decryption. The user will be provided with educated guesses

on the basis of language models and can choose the correct solution on the basis of several solutions found by the cryptanalysis. Many documents found include ciphertext and cleartext, i.e. non-encrypted text written in a particular language. While language identification is widely used and successful for longer texts, there is no trivial solution for short text sequences consisting of only a few words. Code switching is also common in historical texts, where short cleartext sequences may occur in several languages (e.g. Italian and Latin) just as we can find it in modern, nonstandard texts in social media (e.g. English in Swedish texts). So we develop algorithms to 1) identify cleartext sequences and their language considering spelling variation and language change and 2) develop various language models to optimize decryption, and improve the guesses of the analyzer.

3.3. Decryption

The cryptanalysis part deals with different cipher types like monoalphabetic substitution, homophonic substitution, polyalphabetic and polygraphic substitution, various types of transposition, nomenclatures, and mixtures of those. The various cryptanalysis algorithms will be implemented along with the language models. Also, original and generated keys will be mapped to the ciphers in the database. The user is able to interact with the tool by choosing among various algorithms including the detection of null elements, code structure and decryption algorithms. To handle erroneous cases, the tool will also include an interactive part, where the user will be able to correct the output on the basis of language models to improve the result, and the system will be re-trained to learn better models. The project will result in an interactive tool for the decryption of ciphers, or other historical manuscripts written in an unknown language or writing system.

4. Resources

Next, we describe the already available resources and tools that we have developed so far.

4.1. The DECODE database

To store and share encrypted sources, we developed a database, called DECODE (see DECRYPT, 2020), aiming at the systematic collection and description of ciphers, keys and related documents (Megyesi, Blomqvist, and Pettersson 2019) to create infrastructural support for historical research in general, and historical cryptology in particular. The collected cipher records — ciphertexts and keys — are annotated with metadata schema developed specifically for historical ciphers. Information includes the

current location of the manuscript, its provenance, computer-readable transcription, possible decryption(s) and translation(s) of the ciphertext, images, cleartext found in connection to the encrypted source, and any additional materials of relevance to the particular manuscript.

At the time of writing, the database contains over 1,100 ciphertexts and keys from Early Modern Times collected from various archives and libraries in Europe: Austria, Belgium, Germany, Hungary, Italy, the Netherlands, UK, and the Vatican City. The earliest records originate from the 15th century, and the latest from 1809. About 33% of the material consists of original keys. Out of 634 ciphertexts, appr. 30% have been decrypted, and 30% are transcribed in text format (utf-8) allowing further processing for cryptanalysis. The according plaintext languages found yet are Dutch, English, French, German, Hungarian, Italian, Latin, Spanish, or a combination of these (e.g. English-Latin, Hungarian-Latin, Italian-Spanish). Most of the records are short, one or two-page images, but we also find longer ciphertexts up to 410 pages. The great majority of the ciphertexts are encrypted using digits, but ciphertexts with alphabetic characters and graphic symbols (like Zodiac and alchemical signs) are also present. The known cipher types in the database are mostly based on simple or homophonic substitution, with or without nomenclatures, but polyphonic substitutions also appear.

The database allows for search in the existing collection for all users. Registered users with an account (i.e. professionals in historical cryptology) may also edit existing, or upload new encrypted manuscripts. Our hope is that knowledgeable users will contribute to enlarge the database by uploading new material for a growing collection.

4.2. Historical corpora and language models

In order to identify the underlying language behind a ciphertext (i.e. the plaintext language) and correct various hypotheses about particular plaintext character sequences or words, language models describing character- and word-based patterns for many language varieties from different time periods are necessary resources. Within the DECODE project, the HistCorp collection (Pettersson and Megyesi 2018) was created, which is a freely available open platform (HistCorp, 2018) aiming at the distribution of a wide range of historical corpora and other useful resources and tools for researchers interested in the study of historical texts. The platform currently contains a monitoring corpus of historical texts from various time periods and genres for 14 European languages: Czech, Dutch, English, French, German, Greek, Hungarian, Icelandic, Italian, Latin, Portuguese, Slovene, Spanish, and

Swedish. The collection is taken from well-documented historical corpora, and distributed in a uniform, standardized format. The texts are downloadable as plaintext, normalized with regard to spelling, and some are annotated with part-of-speech and syntactic structure. In addition, preconfigured language models and spelling normalization tools are provided to allow the study of historical languages. In addition, the user may upload historical texts to create his/her own language models using the online tools.

5. Tools

To analyze ciphers, we are developing methods and tools for various types of input as described in Section 3.2. In case the ciphers have been digitized but not yet transcribed, image processing techniques can be used for easing their transcription. With the aim to provide generic transcription tools that do not require annotated data, we first focused on developing an unsupervised method for semi-automatically transcribing ciphers of any kind (Baró et al. 2019). Thus, the system is able to detect, cluster and transcribe symbols in document images with a low user intervention, and without the need of having annotated data to train the system for each specific manuscript.

Given a bunch of ciphertexts and keys, we automatically map a ciphertext to a key and return the plaintext if the cipher type is known. We measure the output of the mapping by applying historical language models based on HistCorp to make educated guesses about the correct decryption of ciphertexts (Pettersson and Megyesi 2019). The method is implemented in a publicly available online user interface where users can upload a transcribed key and a ciphertext and the tool returns the plaintext output along with a probability measure of how well the decrypted plaintext matches historical language models for 14 European languages.

For cryptanalysis, we focus on commonly used historical cipher types in Early Modern times based on substitution, including homophonic and polyphonic substitution ciphers with or without nomenclatures, of fixed or variable length codes, with or without nulls. The set of tools which we already developed consists of ciphertext clustering tools to group similar ciphers belonging to the same cipher type, a broad set of parsers to tokenize the transcriptions of the ciphertexts into their core elements, and key recovering tools from known-plaintext and ciphertext-only.

Up to the time of writing, we implemented 12 different parsers for different substitution ciphers from the Vatican in CrypTool 2 (CT2), allowing us to tokenize nearly all Vatican ciphers stored in the DECODE database, decrypt them (if a key is available) and make further statistical analyses.

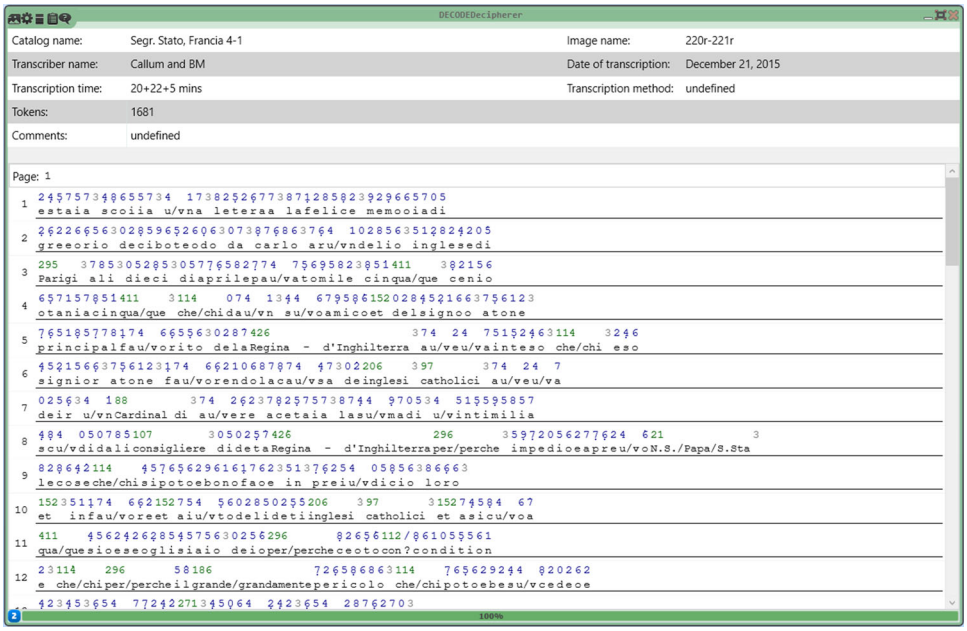

Figure 2. *DECODE Decipherer* component in CrypTool 2 – showing the parsing and decryption of a Vatican cipher.

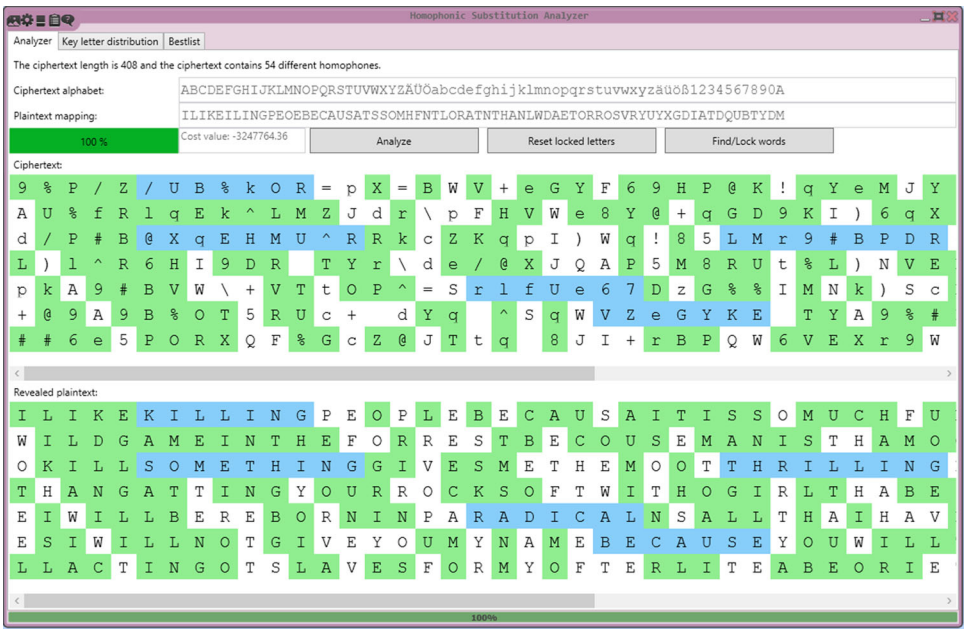

Figure 3. *Homophonic Substitution Analyzer* in CT2 – Analyzing the infamous Zodiac-408 letter.

Figure 2 shows how such a Vatican cipher can be parsed and decrypted using CT2.

We also worked on automated homophonic key recovery (known-plaintext as well as ciphertext-only). We implemented cryptanalysis algorithms

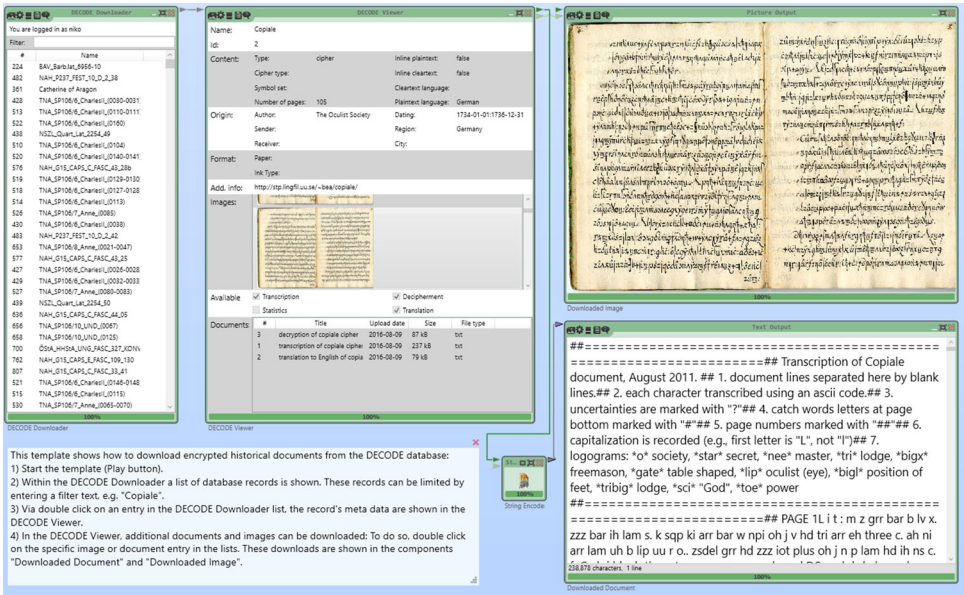

Figure 4. The integration of the DECODE database with the open-source program CrypTool 2.

based on hill climbing and simulated annealing, allowing the user to break homophonic substitution ciphers in CT2 (Kopal 2019). Figure 3 shows the *Homophonic Substitution Analyzer* component in CT2 analyzing the infamous Zodiac-408 letter. The tool allows the user to automatically and semi-automatically analyze homophonically encrypted texts. Already revealed words can be marked and "locked" for further analysis steps. In Figure 3, locked letters are marked green and automatically found words are marked blue. In future work, we plan to adapt the analyzer in such a way that it is able to analyze homophones of different lengths, e.g. different number of digits per homophone, as well as to work with ciphertexts containing nulls.

Prototypes are developed first as console-based applications without sophisticated user interfaces. After improving the new algorithms, more user-friendly versions are implemented in the open-source program CT2 as well as in the aforementioned web service for transcription and cipher analysis.

The DECODE database and CT2 are connected through a JSON-based API which allows CT2 to query the database and download transcribed texts, and a transcribed ciphertext can be analyzed by a pre-trained CT2 module. Figure 4 shows the CT2 application with the DECODE Downloader and Viewer components.

The algorithms and tools are currently under development and will be released throughout the project in CT2 and on the website of the DECRYPT project (DECRYPT, 2020). Unsolved ciphers from the DECODE database have been also published as a challenge on MysteryTwister C3 and two were successfully solved by crypto enthusiast for the first time.

6. Conclusion

In this paper, we presented the DECRYPT project: Decryption of historical manuscripts aiming at the creation of an infrastructure for historical cryptology including a large collection of ciphertexts and keys, and tools for the automatic processing and analysis of various types of ciphers from Early Modern times. We also presented our planned future, long-term goals along with our on-going work of publicly available resources and tools for historical cryptology (DECRYPT, 2020) developed so far.

Among the resources we presented the DECODE database that provides a large number of encrypted sources with metadata and transcriptions, and the HistCorp collection consisting of authentic historical texts for 14 European languages with character- and word-based language models that can be a helpful resource for cryptanalysis.

Among the tools, we described on-going work on (semi-)automatic transcription, and presented algorithms and tools for cryptanalysis of historical ciphers from Early Modern times, including ciphertext clustering tools aiming at the grouping of ciphertexts on the basis of their cipher type, key recovering tools from known-plaintext and ciphertext-only, and a broad set of parsers to identify the internal code structure of the transcribed ciphertext for the segmentation of the core elements, such as the homophones, nomenclatures, and nulls. We welcome interested and knowledgeable people in historical cryptology to contribute to the DECRYPT project.

Acknowledgements

The authors are grateful to their colleagues in the HICRYPT network for valuable discussions and suggestions concerning the need for infrastructural support for historical cryptology, especially and in alphabetical order: Camille Desenclos, Kevin Knight, Anne-Simone Rous, and Gerhard Strasser.

Funding

This work has been supported by the Swedish Research Council [grant 2018-06074]. Alicia Fornés also acknowledges the Ramon y Cajal Fellowship [RYC-2014-16831] and the Spanish project [RTI2018-095645-B-C21].

About the authors

Beáta Megyesi is a senior lecturer in computational linguistics at the Department of Linguistics and Philology, Uppsala University, Sweden. She is specialized in digital philology and natural language processing of non-standard language data. She serves as the PI of the DECODE and DECRYPT projects, aiming at the development of infrastructural resources and tools for historical cryptology.

Bernhard Esslinger is a professor for IT security and cryptology at the University of Siegen, Germany. Before, he was head IT security at Deutsche Bank and CISO at SAP. He is specialized in asymmetric cryptography and in the didactical aspects of the overall area of cryptology. He is the head of the CrypTool project and serves as a cryptanalysis expert of the DECRYPT project.

Alicia Fornés is a senior Research Fellow at the Computer Vision Center and the Universitat Autònoma de Barcelona, Spain. She has a broad expertise in the analysis and recognition of images of documents. Her research interests include historical document image analysis, handwritten text recognition and graphics recognition.

Nils Kopal is a computer scientist and cryptanalyst working as a postdoc at the University of Siegen, Germany. He is specialized in cryptanalysis of classical ciphers and distributed cryptanalysis. He is leading the development of the open-source software CrypTool 2. In the DECRYPT project he is responsible for developing tools for cryptanalysis of historical and classical ciphers and integrating these in the DECRYPT pipeline and CT2.

Benedek Láng is a historian and a medievalist (PhD, Central European University, Medieval Studies Department, 2003). At present, he works as a professor and head of department at the Budapest University of Technology and Economics (Dept of Philosophy and History of Science). He is a historian of science, specialized on two major topics: late medieval manuscripts of learned magic, and early modern secret communication (artificial languages and cipher systems).

George Lasry is a computer scientist in the high-tech industry in Israel. He obtained his PhD in 2017 with the research group "Applied Information Security" (AIS) at the University of Kassel, Germany. Prior to this, he worked for many years in the development of communications systems, and also managed R&D and sales organizations. His primary interest in cryptographic research is the application of specialized optimization techniques for the computerized cryptanalysis of classical ciphers and cipher machines. Using such a technique, he solved in November 2013 the Double Transposition (Doppelwürfel) cipher challenge which was published by Klaus Schmeh in 2007.

Karl de Leeuw is an intelligence historian who published extensively about the history of cryptology in the Netherlands. He is chief editor of *The History of Information Security: A Comprehensive Handbook* (2007) and of a book series for Springer in the same field.

Eva Pettersson is a researcher in computational linguistics at the Department of Linguistics and Philology, Uppsala University. Her research interests include digital humanities and natural language processing of historical text. She has developed several methods for spelling normalization of historical text, and for information extraction from historical sources.

Professor Arno Wacker is the head of the research group "Privacy and Compliance" with the Bundeswehr University Munich in Germany. He has been part of the leading group of the open source project CrypTool 2 and a member of the steering group of MysteryTwister C3. His main research interests are modern security protocols for decentralized distributed systems, e.g. for volunteer computing scenarios, and computerized cryptanalysis of classical ciphers. Additionally, he has taught classical and modern cryptology in classes at the University of Kassel and at special workshops for pupils at local schools.

Michelle Waldispühl is a senior lecturer in German linguistics and language education at the Department of Languages and Literatures, University of Gothenburg, Sweden. She is a historical linguist specialized in philology, Germanic language history and the linguistics of

writing. She has worked on historical spelling variation, especially in runic and onomastic sources.

References

Baró, A., J. Chen, A. Fornés, and B. Megyesi. 2019. Towards a generic unsupervised method for transcription of encoded manuscripts. Paper presented at the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATECH). doi:10.1145/3322905.3322920.

Bauer, F. 2007. *Decrypted secrets — Methods and maxims of cryptology*, 4th ed. Berlin, Heidelberg: Springer.

Bonavoglia, P., and C. Pellegrino. 2016. The last poem of Pietro Giannoneâ — finally decrypted. *Cryptologia* 40 (5):411–27. doi:10.1080/01611194.2015.1087072.

CrypTool. 2020. The CrypTool project develops the world's most-widespread free e-learning programs in the area of cryptography and cryptanalysis. Accessed October 21, 2019. http://www.cryptool.org.

DECRYPT. 2020. The DECRYPT website contains resources and tools for historical cryptology, developed in two projects financed by the Swedish Research Council: The DECODE project (Automatic Decoding of Historical Manuscripts, 2015-2017) and the DECRYPT project (Decryption of Historical Manuscripts, 2019-2024). Accessed October 21, 2019. https://cl.lingfil.uu.se/decode/.

Fornés, A., B. Megyesi, and J. Mas. 2017. Transcription of encoded manuscripts with image processing techniques. In *Digital humanities*. New York, NY: Scribner.

Hick, S., B. Esslinger, and A. Wacker. 2012. Reducing the complexity of understanding cryptology using CrypTool. Paper presented at the 10th International Conference on Education and Information Systems, Technologies and Applications (EISTA 2012), Orlando, Florida, USA.

HistCorp. 2018. HistCorp is an open-source collection of historical texts and language models for 14 European languages. Accessed October 21, 2019. https://cl.lingfil.uu.se/histcorp/.

Kahn, D. 1996. *The codebreakers: The comprehensive history of secret communication from ancient times to the internet*. New York.

Kopal, N. 2019. Cryptanalysis of homophonic substitution ciphers using simulated annealing with fixed temperature. Paper presented at the Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt 2019, 107–16, Mons, Belgium: Linköping University Electronic Press, June 23–26.

Kopal, N., O. Kieselmann, A. Wacker, and B. Esslinger. 2014. CrypTool 2.0. Open-Source Kryptologie für Jedermann. In Datenschutz und Datensicherheit (DuD) 5/2014.

Lasry, G., I. Niebel, N. Kopal, and A. Wacker. 2017. Deciphering ADFGVX messages from World War I Eastern Front. *Cryptologia* 41 (2):101–36. doi:10.1080/01611194.2016.1169461.

Láng, B. 2015. Shame, love, and alcohol: Private ciphers in early modern Hungary. *Cryptologia* 39 (3):276–87. doi:10.1080/01611194.2014.915270.

Megyesi, B., N. Blomqvist, and E. Pettersson. 2019. The DECODE database: Collection of ciphers and keys. Paper presented at the Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt19, Mons, Belgium, June.

MysteryTwister C3. 2009–2020. MysteryTwister C3 (MTC3) is an international cryptography competition, offering more than 200 riddles (challenges) at four levels of difficulty. Accessed October 21, 2019. https://www.mysterytwisterc3.org.

Patarin, J., and V. Nachef. 2010. I shall love you until death" (Marie-Antoinette to Axel von Fersen). *Cryptologia* 34 (2):104–14. doi:10.1080/01611191003621212.

Pettersson, E., and B. Megyesi. 2018. The HistCorp collection of historical corpora and resources. Paper presented at the Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, Helsinki, Finland, March.

Pettersson, E., and B. Megyesi. 2019. Matching keys and encrypted manuscript. Paper presented at the Proceedings of the 22nd Nordic Conference on Computational Linguistics, 253–61, Turku, Finland: Linköping University Electronic Press, September 30–October 2.

Schmeh, K. 2015. Encrypted books: Mysteries that fill hundreds of pages. *Cryptologia* 39 (4):342–61. doi:10.1080/01611194.2014.988369.

Singh, S. 2000. *The code book: The science of secrecy from ancient Egypt to quantum cryptography*. Anchor Books.

Strasser, G. F. 2012. Late 18th-century French encrypted diplomatic "Letters of Recommendation" — or, how to unwittingly carry your own warrant. *Cryptologia* 36 (3): 230–9. doi:10.1080/01611194.2012.688694.

Yin, X., N. Aldarrab, B. Megyesi, and K. Knight. 2019. Decipherment of historical manuscript images. Paper presented at the Proceedings of ICDAR.