# Lost in Transcription of Graphic Signs in Ciphers

**Giacomo Magnifico and Beáta Megyesi**
Dept. of Linguistics and Philology
Uppsala University, Sweden

**Mohamed Ali Souibgui, Jialuo Chen and Alicia Fornés**
Computer Vision Center
Computer Science Department
Universitat Autònoma de Barcelona, Spain

## Abstract

Hand-written Text Recognition techniques with the aim to automatically identify and transcribe hand-written text have been applied to historical sources including ciphers. In this paper, we compare the performance of two machine learning architectures, an unsupervised method based on clustering and a deep learning method with few-shot learning. Both models are tested on seen and unseen data from historical ciphers with different symbol sets consisting of various types of graphic signs. We compare the models and highlight their differences in performance, with their advantages and shortcomings.

## 1 Introduction

Encrypted historical manuscripts contain a large variety of symbols taken from different symbol sets, often digits, characters of the Latin and Greek alphabets, or graphic signs such as the Zodiac or alchemical symbols. Diacritics and dots can be used systematically to make further distinction between the symbols. One of the main challenges that arises when transcribing ciphers is the transcription process which is normally the first necessary step to decrypt the manuscript at hand. Partly of fully automatizing the transcription would not only save time but also lead to more consistent transcriptions, even on the error side which would make correction easier and faster. Therefore, (semi-)automated transcription of historical ciphers would be of great help. Recently, Hand-Written Text Recognition (HTR) has made great progress. The main challenge with encrypted sources is the segmentation and recognition of the symbol set due to the variability of the written characters and the use of alphabets and symbol sets across ciphers.

HTR techniques (just as any field within Artificial Intelligence) are built upon unsupervised, semi-supervised or fully supervised methods, where the former does not need any annotated training data such as clustering, while the latter needs a rather big set to reach high(er) performance. Powerful architectures, magic black boxes, have recently become available through deep learning. These systems are known to be data- and energy hungry, requiring big data sets and huge computer power in terms of GPUs for training. Recently, supervised models have been developed where only a few transcribed examples of text segments is needed for training. The recent development in HTR makes transcription less expensive and more manageable.

The work presented in this paper investigates two approaches applied to the automatic transcription of ciphers with various types of symbol sets including a large variety of graphic signs. We compare unsupervised clustering that does not require any training data but needs a post-processing step for cleaning the output, and the supervised few-shot learning approach, which requires only a few examples of each symbol in a cipher for training but does not need any post-processing. We evaluate both methods on the same cipher sample and provide the Symbol Error Rate value for the models to measure their performance. We also give a time estimate for transcription of historical encrypted sources.

Next, in Section 2 we give an overview on previous studies on the automatic transcription of ciphers. In Section 3 we describe the clustering and the few-shot model architectures under evaluation. In Section 4 we present the performance of the models, in Section 5 we discuss the results, and in Section 6 we conclude the paper.

## 2 Transcribing Historical Ciphertexts

During the past years, several work have been published on the automated transcription of historical ciphers. Given the thorough monitoring of the factors that make or break the efficiency of an instrument for automated transcription — accuracy and time — previous studies have reported the advantages and efficiency of automated transcriptions based on Recurrent Neural Networks and compared to manual ones, once the accuracy of the instrument achiever results higher than 90% (Fornés et al., 2017).

The promising results presented in one of the more recent publications, involving the evaluation of an interactive online transcription tool (Baró et al., 2019) and (Johansson, 2019), confirmed the direction in which to proceed in the development of HTR tools — unsupervised models such as clustering developed in an image processing pipeline including cropping of the image, binarization, line- and character segmentation, and finally symbol recognition by clustering (Chen et al., 2018) and (Chen et al., 2021).

Other methods were recently proposed using powerful deep learning architectures. Such architectures used to require a big amount of training data with lots of manual effort for providing transcription of hundreds of images to be used for training for high performance. An example of such a study on the transcription of ciphers was the Seq2seq Attention model (Renfei, 2020). The need of reducing the expensive preparation step to produce training data and develop models that require only a few images of each alphabet symbol became emerging. (Souibgui et al., 2020) presented an architecture based on few-shot learning by detecting all symbols in a given alphabet in a textline image, and decoding the obtained similarity scores to the final sequence of transcribed symbols. The model was shown to be powerful on various types of ciphers.

Given the promising results of clustering and the few-shot architectures, we choose to evaluate each and compare them on the same set of ciphers. The evaluation was conducted in order to provide more specific directions on how to implement the best performing pieces of architectures, as well as to provide an analysis of the strength and shortcomings of the two models.

## 3 Automatic Transcription

Before we describe the involved models to be evaluated, let us introduce the datasets used in the experiments.

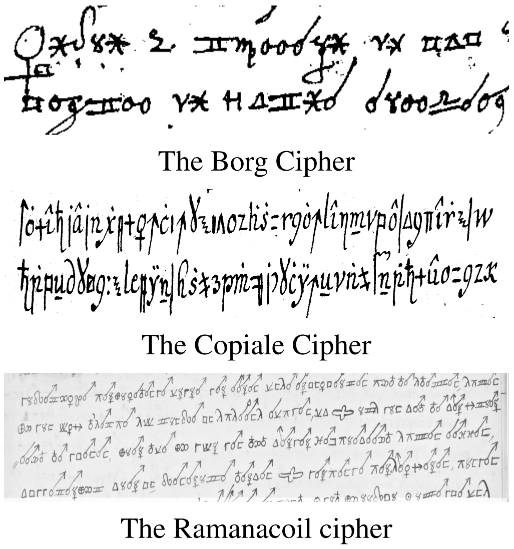
### 3.1 Data

We selected three decrypted ciphers with transcriptions freely available. The ciphers contain various symbols sets of different size and different hand-writing styles. The three ciphers are exemplified in Figure 1.

The Borg cipher (Aldarrab et al., 2017) is a historical encrypted source originating from the 17th century. It consists of 408 pages with 34 eclectic symbols with space between the code sequences. The symbol set ranges from Latin letters and diacritics to Zodiac and alchemical symbols. The hand-writing is greatly varied and sometimes rather difficult to interpret, since the symbols are connected not only horizontally but oftentimes also vertically across lines leading to many touching symbols. 16 transcribed pages containing $\sim$ 17 lines and $\sim$ 280 characters per page on average were used for the test set, with a total number of characters around $\sim$ 4 480.

The Copiale cipher (Knight et al., 2011) originates from the 18th century. It consists of 100 different symbols including digits, Latin and Greek letters, diacritics, punctuation marks, and a big variety of graphic signs. The cipher is meticulously written with clearly segmented symbols and straight lines. The total amount of pages used as test set for model evaluation is 24, with $\sim$ 18 lines and $\sim$ 720 characters per page on average; the total amount of characters was $\sim$ 17 280.

The Ramanacoil cipher (Dinnissen and Kopal, 2021) reveals two Dutch East India Company letters from 1674. The cipher consists of 55 different graphic signs. The symbols are clearly separated and spaces are used to mark symbol sequences. The cipher is meticulously written but the lines are not necessarily straight. The symbols are tiny and the pages contain the highest number of lines and characters per page on average among the three manuscript, reaching $\sim$ 40 lines and $\sim$ 2 240 glyphs per page. Therefore, we only chose 8 pages for test set, reaching a total amount of characters around $\sim$ 13 440. Noteworthy that no pages from Ramana-



The Borg Cipher

The Copiale Cipher

The Ramanacoil cipher

Figure 1: Example of the three ciphers.

coil were used for training — the cipher has been unseen — as opposed to the two other ciphers.

### 3.2 Unsupervised Clustering

The architecture we choose to evaluate is based on the work by (Baró et al., 2019) and (Chen et al., 2021). The transcription process of the tool follows five macro-steps including binarization, segmentation, clustering, label propagation, and transcription, which are presented below in operational order and illustrated in Figure 2.

**Binarization**. The conversion of the manuscript image into a black and white picture with a stark contrast between glyphs and background. The threshold that produced the results is based on the experiments by (Sauvola and Pietikäinen, 2000).

**Segmentation**. A two-fold process that involves line segmentation followed by character segmentation, where the tool divides the image into single lines and the single lines into smaller rectangles to isolate the glyphs. The user can choose a preset of measures from known ciphers (e.g. space between the lines, avg. surface of the symbols, space between symbols) or input a custom set to fine-tune the model to specific characteristics of a particular manuscript.

**Clustering**. The process recognizes and groups characters together through *k-mean clustering* with the use of a hierarchical algorithm (Almazán et al., 2014). After clustering, the user can set the number of clusters that the tool has to output — ideally the total number of unique symbols plus one for uncertain cases — or let the tool automatically decide the total number of clusters.

**Label propagation**. Two parameters are required from the user, the alpha value (default set to 0.2) and a confidence threshold, as a method of soft-assignment of a cluster-related label to each glyph that resembles the specific cluster. The values used for the experiments conducted in this study were 0.2 and 0.8.

**Transcription**. Outputs a transcription as a single-file or page by page, as the user prefers.

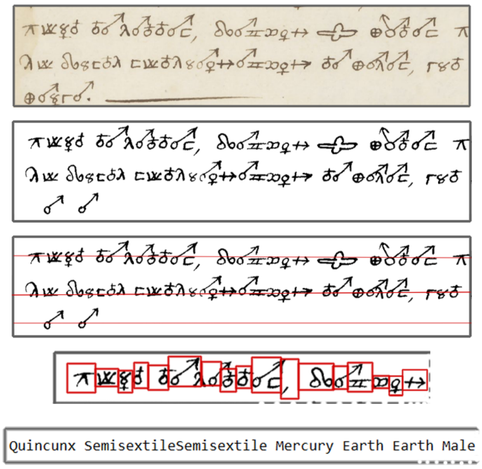


Figure 2: Processing steps of clustering of the Ramanacoil cipher.

### 3.3 Few-Shot Modeling

The few-shot architecture is based on the work presented in (Souibgui et al., 2020). Quite differently from the previous model, the main steps for user interaction are only two: the data *preprocessing* and the model run in the *command line*. Given that the architecture was used as an unsupervised model, the trained weights were generated from the Omniglot dataset (Lake and Tenenbaum, 2015) which served as baseline (Souibgui et al., 2020) for our experiments.

**Preprocessing**. Since the transcription architecture does not have an automated line segmentation process, the document had to be manually cropped and divided into single lines. Then, the lines were resized vertically to a height of 105 px before running the model using the same size as during training. The architecture-specific feature called *alphabet support* requires the user to provide 10 examples of each character in the cipher alphabet. Having more (and different) samples usually boosts the performance. However, we can also choose to copy the same character multiple times to obtain 10 samples to save time and minimize the human effort.

**Running the model**. Starting the model is done through a single command line, with which the user can alter the number of ”shots” of the architecture (i.e. the iterations of the model, max 5) and input a threshold for the transcription akin to the clustering architecture. The main characteristic of the architecture is the multi-layered structure used for extracting features from both the *alphabet support* images and the input document, presented in (Souibgui et al., 2020). This model currently requires great computer power in terms of GPU.

## 4 Results

Upon obtaining the output transcriptions for each of the three ciphers, the accuracy of the two models was evaluated through the use of Levenshtein's distance to calculate Symbol Error Rate (SER) in order to provide a standardised result. The formula for Levenshtein's distance SER is given below, with $N$ representing the total number of symbols in a line, and the total number of operations that is necessary to carry out to turn the transcription output into the gold standard output measured as three types of operations: $S$ symbol substitution, $D$ symbol deletion, and $I$ symbol insertion.

$$SER = \frac{S+D+I}{N} \tag{1}$$

The performance of the two architectures is shown in Table 1. The results measured as symbol error rate clearly show the overall better performance of the few-shot architecture both on seen *[S]* and unseen data *[U]*, respectively, compared to the clustering architecture.

Where it is expected for the few-shot model to show a larger performance gap when performing on seen data [S], there is the presence of a small gap when the few-shot model transcribes unseen data [U]; therefore, the model performs better than the clustering not only when working as intended (as a supervised architecture), but when underperforming (as an unsupervised architecture) as well.

| **Dataset** | **Clustering** | **Few-Shot** |
|---|---:|---:|
| Borg [S] | 0.638 | 0.150 |
| Copiale [S] | 0.350 | 0.104 |
| Ramanacoil [U] | 0.800 | 0.754 |
| *Avg. SER* | *0.596* | *0.336* |

Table 1: Symbol error rate (SER) and average value (*bottom line*) for each model performing on the three ciphers.

The meticulously written Copiale cipher leads to lowest transcription errors by both systems, while the unseen Ramanacoil — not surprisingly — contains the largest number of transcription errors by both methods.

### 4.1 Transcription Time

We present the average time required to transcribe a hand-written cipher of 10 pages with 250 lines in total, as shown in Table 2. We look at the preprocessing, parameter setting and post-processing steps for each method and count the exact time it takes to finalize each step. It should be noted that the data reported in Table 2 is only representative of the time required to go from image to model output, the kind used for the performance evaluation shown in Table 1; time for further corrections that would be required on the resulting automated transcription is not taken into account.

The values in Table 2 are derived from an estimate of the time required for each page of the three datasets, respectively made of 16 pages with ∼17 lines per page for Borg, 24 pages with ∼18 lines per page for Copiale, and 8 pages with ∼40 lines per page for Ramanacoil.

The time required to assist the systems to produce the transcription output differ greatly from each other. While clustering has a fast pre-processing due to the automatized process of segmentation, the post-processing phase is time-consuming in order to reach high(er) performance in transcription. On the other hand, the few-shot model requires preprocessing in terms of time-consuming manual line segmentation and a segmentation of a few examples of each symbol in the document, called *Alphabet support*.

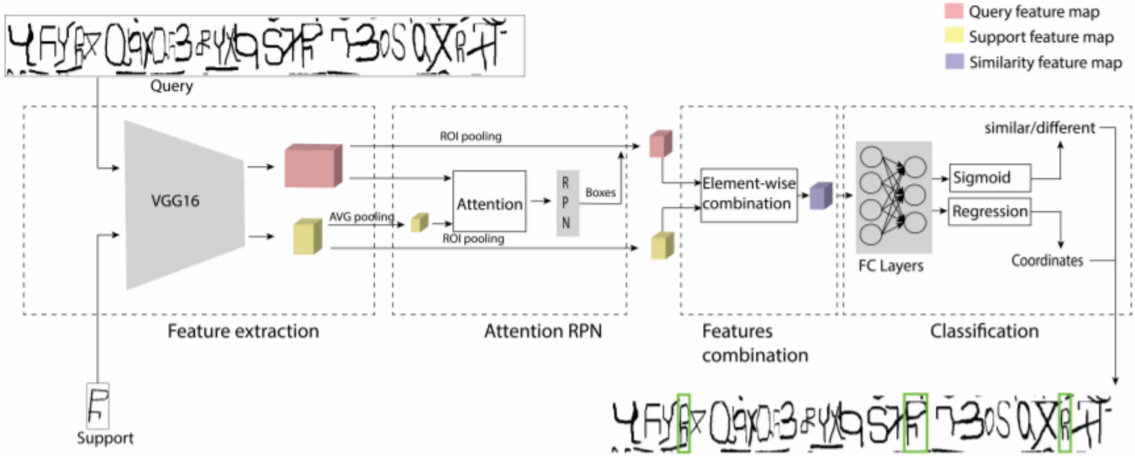


Figure 3: The few-shot model architecture.

| Process | Clustering | Few-shot |
|---|---|---|
| *Line preprocessing* | × | *∼8h 20' (2' / line)* |
| *Alphabet support* | × | *∼1h 15' (3' / char.)* |
| Parameter setting | ∼0h 15' | ∼0h 01' |
| Cluster clean & label | ∼1h 00' | × |
| Automated processes | ∼2h 10' | ∼0h 50' |
| *Avg. Time Elapsed* | *∼2h 55'* | *∼0h 51' (+ prep.)* |

Table 2: Estimate of time required for a full transcription of 10 pages with 250 lines. All user-independent processes have been unified under *Automated processes*.

## 5 Discussion

Above all, we would like to emphasize that the two architectures presented in the paper are not fully comparable since they do not use the same set of pre- and postprocessing steps which makes comparisons a bit unfair. However, the results indicate some major advantages and drawbacks with each method.

The few-shot model architecture achieves highest performance shown as lowest symbol error rate both on seen and unseen datasets. However, the model requires manual segmentation of lines, and at least five examples of each symbol in each cipher document. On the other side, while the current implementation of the clustering architecture performs lower in the transcription output, it allows for automatic segmentation of lines and symbols. One major weak point of the clustering approach that we identified is the failure of symbol segmentation in cursive writing style. Implementing and improving the segmentation algorithms on line and character levels would greatly increase performance for faster and more accurate transcription for both approaches.

A tool including both methods and allowing end-to-end transcription from image upload to transcription output would be beneficial for a more adequate and systematic evaluation of the methods. Since the advantage of the two architectures can be said to be complementary, they could be combined for higher performance and less user effort allowing automatic segmentation and postprocessing equally in both architectures.

All in all, using the few-shot model is recommended — in case of access to high-end GPUs —- for datasets with regular line orientation (such as the Copiale cipher) and datasets with an alphabet smaller than 30 symbols (such as the Borg cipher). In cases when the line segmentation as part of the preprocessing has to be done manually, it might be unsuited for datasets longer than 10 pages due to the length of the manual segmentation required being equal to the speed of the full manual transcription of the dataset. In case of CPU-only machines, datasets with highly irregular line orientation, and datasets with alphabets larger than 30 symbols, the use of the clustering architecture is suggested in its current state.

## 6 Conclusion

We presented an evaluation and comparison between two structurally different transcription tools developed for hand-written text recognition and applied to historical ciphertexts. We showed how the use of few-shot architecture with supervised learning with a few examples used for training achieves the highest performance. Further development of the tool is the implementation of an automatic line and symbol segmentation in order to make the process of creating training data and select examples of glyphs faster and easier for the user.

We also showed that the clustering tool including its fully automated pre-processing of segmentation and its unsupervised nature are promising and could be used for ciphers and other scripts with unknown symbol systems.

It is worth exploring a combination of the two approaches; where the automatic segmentation is used along with the examples from already selected and corrected clusters to be used with the Few-Shot model. We believe that the combination of the two architectures would be ideal to take as the next steps.

## Acknowledgments

This work has been supported by the Swedish Research Council, grant 2018-06074: DECRYPT - Decryption of historical manuscripts, the Spanish project RTI2018-095645-B-C21 and the CERCA Program / Generalitat de Catalunya.

## References

Nada Aldarrab, Kevin Knight, and Beáta Megyesi. 2017. The Borg Cipher. https://cl.lingfil.uu.se/ bea/borg. Accessed: 2020-01-31.

Jon Almazán, Albert Gordo, Alicia Fornés, and Ernest Valveny. 2014. Word spotting and recognition with embedded attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(12):2552–2566.

Arnau Baró, Jialuo Chen, Alicia Fornés, and Beáta Megyesi. 2019. Towards a Generic Unsupervised Method for Transcription of Encoded Manuscripts. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATECH)*, pages 73–78.

Jialuo Chen, Pau Riba, Alicia Fornés, Joan Mas, Josep Lladós, and Joana Maria Pujadas-Mora. 2018. Word-Hunter: A Gamesourcing Experience to Validate the Transcription of Historical Manuscripts. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 528–533. IEEE.

Jialuo Chen, Mohamed Ali Souibgui, Alicia Fornés, and Beáta Megyesi. 2021. Unsupervised Alphabet Matching in Historical Encrypted Manuscript Images. In *Proceedings of the 4th International Conference on Historical Cryptology (HistoCrypt 2021)*.

Jörgen Dinnissen and Nils Kopal. 2021. Island ramanacoil a bridge too far. a dutch ciphertext from 1674. In *Proceedings of the 4th International Conference on Historical Cryptology (HistoCrypt 2021)*.

Alicia Fornés, Beáta Megyesi, and Joan Mas. 2017. Transcription of Encoded Manuscripts with Image Processing Techniques. In *Digital Humanities*.

Kajsa Johansson. 2019. Transcription of Historical Encrypted Manuscripts — Evaluation of an Automatic Interactive Transcription Tool. Bachelor thesis in Language Technology, Uppsala University, Sweden.

Kevin Knight, Beáta Megyesi, and Christiane Schaefer. 2011. The Copiale Cipher. In *Invited talk at ACL Workshop on Building and Using Comparable Corpora (BUCC)*. Association for Computational Linguistics.

Salakhutdinov R. Lake, B. M. and J. B. Tenenbaum. 2015. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.

Han Renfei. 2020. Using attention-based sequence-to-sequence neural networks for transcription of historical cipher documents.

Jaakko Sauvola and Matti Pietikäinen. 2000. Adaptive dovument image binarization. *Pattern Recognition*, 33, Issue 2:225–236.

Mohamed Ali Souibgui, Alicia Fornés, Yousri Kessentini, and Crina Tudor. 2020. A few-shot learning approach for historical ciphered manuscript recognition. In *Proceedings of the Internaitonal Conference on Pattern Recognition (ICPR 2020)*.