

Encrypted epigraphy - the case of a mysterious inscription in the Neapolitan church of Santa Maria La Nova

Cosimo Palma

University of Naples "L'Orientale" / Via Duomo, 219, 80139 Napoli, Italy

University of Pisa / Lungarno Antonio Pacinotti, 43, 56126 Pisa, Italy

cosimo.palma@phd.unipi.it

Abstract

This paper contains all steps regarded as necessary and relevant towards the decryption of the epigraph placed in the "Turbolo Chapel" of the Neapolitan Church of Santa Maria La Nova. The inscription has been processed by means of Python-written procedures in combination with the decryption software *AZdecrypt*, thus displaying linguistic features converging unequivocally to the hypothesis that its clear-text be encrypted by monoalphabetic substitution from a natural language, possibly alongside with transposition and polyalphabeticism. A preliminary analysis on characters n-grams and vocals-consonants combinations has not succeeded yet in individuating any clear-text language among all the analysed corpora.

1 Introduction

Inside the Turbolo Chapel of Santa Maria la Nova (Naples, Italy) two epigraphs are to be found. The one on the observer's left side contains an indulgence statement from Pope Gregory XIII related to holy masses celebrated therein, in dedication to Miss Turbolo, the noblewoman Giovanna De Rosa. On the right side is placed an epigraph written in an unknown alphabet¹. This immediately recognizable feature does not let us understand that we behold an encrypted text yet: as it often happens in these cases, the only decisive proof that we are coping with an encryption would be finding out its decryption, or at least any other external

¹The inscription is constituted by approximately six hundred glyphs, against the one thousand seven hundred fifty of the papal indulgence. Beside the obvious space reasons, the insertion of the whole photography would have been superfluous because of the many deletions and erosions mostly present on the top and bottom areas. Refer to section "Historical alphabets and coeval codes review" for further details.



Figure 1: A snapshot of the encrypted epigraph of the Turbolo Chapel in the Neapolitan Church of Santa Maria La Nova (courtesy of the cultural association "Oltre il chiostro", leader manager of the monumental complex).

element which could let us pointing at it as such without further doubts.

2 Related Work

In few sparse newspaper articles, it emerges that serious attempts at decryption have been indeed undertaken, although related scientific publications seem impossible to find. The only two published works explicitly mentioning the inscription of the Turbolo chapel are either focused on the the artistic and architectural features of the whole church (Rocco, 1928), or on a specific set of historical events and aristocratic genealogy displayed in order to ascertain the reliability of the theory, according to which in the tomb placed outside, on the other side of the very same epigraph's wall, the count Vlad III of Wallachia (popularly known as Dracula) be buried (Miriello, 2021).

In the former work, the cultivated and generally

esteemed author assesses that the inscription is the Greek translation of the adjacent Latin inscription (and yet, even a profane would easily acknowledge that the first epigraph is too long, compared to the second, which in turns displays many glyphs obviously not belonging to the common Greek alphabet). In the latter, three conjectures about the inscription's origins are brought forth, among which I report only the most relevant one: the inscription shall be nothing else than a table used by the Franciscan monks for didactic purposes.

The church was considered in fact, at the end of the XVI. century, as one of the most flourishing university poles in Southern Italy where, among others, also oriental languages and calligraphy were taught. The same author reports that a radiocarbon-dating performed on the epigraph places it at around the XVI-XVII century and that a preliminary attempt for decryption, consisting in substituting each glyph by probable phonemes related to similar glyphs in other alphabets, has lead to no satisfactory results. It is not unusual to find mysterious inscriptions in western as well as eastern churches, although in these cases they are mostly present in the form of tetragrams (Moutafov, 2006), a text-length which cannot be compared to the one analysed in this seat.

Another article (almost homonym to the present one) first formulates the epistemological conundrum of an encrypted object situated in plain sight (Rosenmeyer, 2019). According to the author, the magical or mystic power attributed to the letters of the alphabet within oriental and Egyptian culture explains the use of acrostics in religious contexts in Egypt. Furthermore, she states that the encoder may want to tease the observer with a riddle, as an intellectual game, and not with the real purpose to conceal a message. A further scenario, as subtle as important, is the case where something is by law or enforcement compelled to be written, implicitly assuming its understandability, without expressing it, thus opening the possibility for the reluctant engraver to perform a sort of Isolt's oath, whereby the requirement is fulfilled in its form, but not in its substance. Rosenmeyer's study bears also the merit for describing a rare encryption method, based on the art of *isopsephy*. As in the rabbinic *gematria*, it associates to every letter of the Greek alphabet a numerical value, ranging from 1 to 900. By the most common encryption strategy, every sign translates in the one

resulting from the difference between the numeral order immediately superior, and the numeric value itself. For instance, if Ψ equals to 700, it shall be read as 1000 minus 700, hence 300, the value of the letter τ . In accordance to this calculation, all glyphs corresponding to numbers with a 5, such as 50 and 500, do not change. Despite its interestingness, I have regarded the inclusion of this approach to the present paper not relevant, since it still falls under the umbrella of a monoalphabetic substitution, which could be eventually easily investigated to assess whether it has been applied by following isopsephic principles.

3 Problem Statement

Plenty of superficial observations, such as the amount of different characters, in the same range of standard alphabets, are sufficient to assume that the epigraph is most likely the encryption of a text written in a natural language. However, the fact itself that the artefact is located in plain sight, allows us to imagine that whatever information contained therein must not be exceptionally secret, and is required to be unlocked by a group of people possessing the suitable key or a sufficient amount of time. The dense conglomerate of history and mystery the epigraph is entangled with induces first at taking a step back and deciding from which angle the decryption shall be approached. The very same is undoubtedly, at the same time, a source of motivation for undertaking this challenge.

4 Methodology

Because of the abundant literature available for the historical aspect, and of the lack of a serious quantitative analysis of the object, I decided to focus only on the tasks strictly related to decryption, in the hope that my leads, once joined with the others from different domains, could finally bring to a holistic solution of the puzzle. The major steps I undertook to attempt the epigraph decryption are:

- Review of historical alphabets and coeval codes;
- Statistical analysis on single characters;
 - Candidate-languages corpora harvesting and pre-processing;
 - Transposition of the inscription's glyphs into Latin characters;

- Preliminary analysis on vowels-consonants intertwining;
- Index of Coincidence calculation for text-snippets extracted from each corpus;
- Shannon Information Entropy calculation for text-snippets extracted from each corpus;
- Friedman's test calculation for every candidate language;

- Statistical analysis on N-grams;
 - Generation of a N-grams file suitable for the *AZdecrypt* software;
 - Creation of a ".ini" file for initialization of the software calculation;
 - Copying the main epigraph's transliteration into the software's input window;
 - Running the solver for every relevant decryption mode;
 - Output files storage and analysis.

In the following, only the non-trivial steps among the above listed ones will be described in deeper detail.

5 Review of historical alphabets and coeval codes

Before diving into the canonical decryption procedure, for which I followed the path laid down in (Knight et al., 2011; Hauer and Kondrak, 2016) on the basis of (Pommerening, 2021), I first pursued a review of all existing alphabets, which substantially confirmed the already mentioned range of alphabets participating in the epigraph's composition, such as old Slavonic, Greek, Latin and Coptic (Miriello, 2021). In addition to those, the Carian alphabet² was found to be the alphabet which, singularly, contains the highest amount of the inscription glyphs.

Not only existing alphabets, but also invented ones have been taken in duly consideration³, as listed in (King, 2001; Della Porta, 1563; Trithemius, 1518; Somogyi, 1906; Schöning, 2014; Meister, 1902; Kranz and Oberschelp, 2009;

²Carian is an extinct language spoken until the first century B.C. in Caria, a region of western Anatolia, and in Egypt (Adeigo, 2006).

³The two most remarkable examples, are the *Lingua Chaldeorum* of Rudolph IV Duke of Austria, and the *Lingua Ignota* of Saint Hildegard of Bingen. to which even a glossary of more than one thousand words is linked.

Figure 2: A diachronic display of Carian alphabets. Glyphs resembling the epigraph's ones are red-circled.

Rous and Mulsow, 2015). The quoted works also contain the state of the art of XVI century cryptology, which at the time was still in its dawn. If the inscription's radio-carbon dating were to be considered reliable (which is highly probable, since it matches the dates engraved on the Turbolo's tomb), the hypothesis for polyalphabetic or homophonic substitution ciphers cannot be safely discarded yet (see section "Statistical Analysis on single characters" for an empirical assessment of this question), because they first originated exactly in those years (Della Porta, 1563; Trithemius, 1518). The survey in this domain can be regarded as accomplished only if yet another set of symbols is taken in consideration, namely those which could be mapped to concepts, instead of usual alphabet characters, as in the case of alchemical signs⁴.

Despite the remarkable similarity with some of the symbols occurring in the epigraph, an unambiguous way to match them with single characters could not be found. For instance, taking only

⁴The combination alchemy-cryptology characterises most of the Middle Ages- and Renaissance occultism.

A Table of Chymicall & Philosophicall Characters wth their significations as they are usually found in Chymicall Authors both printed & manuscript.

Figure 3: Table of alchemical symbols taken from "The last will and testament of Basil Valentine, monke of the Order of St. Bennet" (1671). Symbols resembling the epigraph's ones are red-circled. Source: *Wikimedia*.

the initial letter of the represented element, would result in a heavy homophony, which once practically transposed would yield no meaningful word or sentence.

6 Statistical analysis on single characters

The epigraph clearly underwent major deletions, mostly in the lower part: the entirety of text at our disposal is therefore not suited for a satisfactory statistical analysis, which usually delivers best results when performed over larger samples. However, the preliminary manual attack have been unsuccessful, thus leading to a computer-aided one⁵. The first step towards the decryption was to establish an arbitrary letter-glyph mapping in order

⁵The project's repository can be accessed at <https://github.com/Glottocrisis/MariaLaNova>.

Epigraph's glyph (EG)	Related Latin letter (CL)	EG	CL
Δ	A	⊙	O
B	B	Π	P
C	C	∇	Q
Δ	D	ℙ	R
E	E	∇	S
Z	F	+	T
∇	G	∇	U
N	H	∇	V
T	I	∇	W
Δ	L	F	X
I	M	∇	Y
N	N	Δ	Z

Figure 4: Mapping between Latin and epigraph's glyphs.

to produce a machine-readable document. Afterwards, a handful of candidate languages have been selected according to various factors, such as their prestige at the period to which the inscription itself belongs, as well as their diversity⁶, and harvested online, mostly through the online corpus database HistCorp (HistCorp, 2023)⁷.

6.1 Index of Coincidence

The uncertainty drawn by these preliminary observations has induced the necessity of a statistical analysis. The related literature reports about different methods for clear-text language assessment for an encrypted text, among which the Index of Coincidence and the Shannon's Information Entropy represent the most known and efficient ones. The Index of Coincidence (IC) is a measure which reflects the probability in a given text that two randomly selected letters coincide. It is used in cryptology for the identification of the clear-text language, since every language has a relatively constant IC. In the equation of the IC, N represents

⁶Considering different linguistic families knowingly accelerates the process of the clear-text language identification.

⁷The corpus for Old Luxembourgish has been obtained manually, by scanning a copy of the *Codex Mariendalensis*, a manuscript from the beginning of the XIV century about the life of Yolanda of Vianden consisting of 5,963 lines of rhyming couplets in the Moselle-Franconian German dialect.

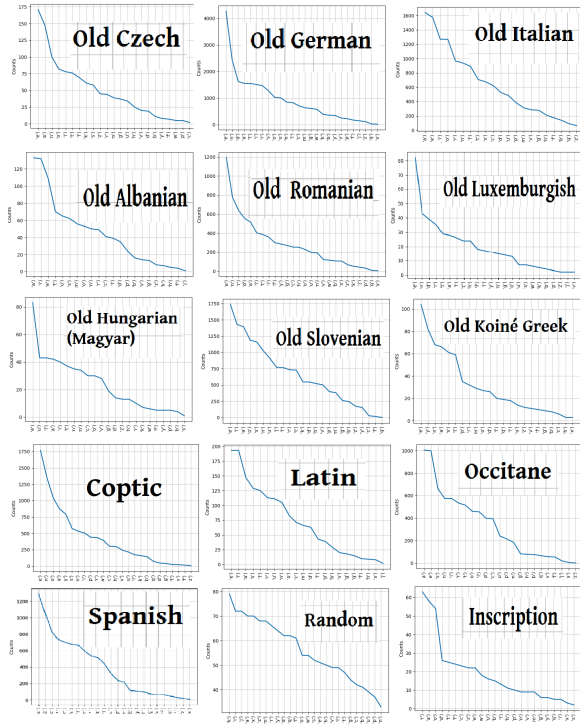


Figure 5: A synoptic view of character frequency distribution for all candidate languages, performed on random snippets with similar length to the inscription’s one, extracted by related corpora.

the length of the text, and n_1 through n_c are the frequencies (as integers) of the c letters of the alphabet.

$$IC = \frac{\sum_{i=1}^c n_i(n_i - 1)}{N(N - 1)/c}$$

Nonetheless, it shall be considered, as observable from the adjacent epigraph as well, that the writing style and conventions for epigraphs is very different from the normal ones: the suppression of all doubles, as well as the use of abbreviations, automatically result in a lower IC. On the other side, the *scriptio continua*⁸ enhances the same probability, since the final letter of a word can match the initial letter of the following ones. In order to parametrize these discrepancies in the IC equation, it would be needed to analyse these changes on a larger corpus of epigraphs, written in both styles, a task which is not necessary to undertake for this work. The IC has been automatically calcu-

⁸From Latin, literally: continuous writing, i.e. space between words is omitted. This writing style is typical of Greek epigraphy as well.

lated by means of the project’s ”getIOC” function, which implements the above mentioned mathematical formula.

Candidate language	IC
Inscription	1.72
Random text	1.03
Magyar (Old Hungarian)	1.61
Old Luxemburgish	1.87
Old Romanian	1.71
Magyar (Old Hungarian) 2	1.75
Coptic	1.87
Koiné Greek	1.88
Latin	1.94
Old Albanian	1.73
Old Slovenian	1.52
Old Spanish	1.81
Old Italian	1.53
Old German	1.95
Old Czech	1.30

Table 1: Normalized Index of Coincidence values for every candidate language.

6.2 Other statistical tests

Another measure through which a whole language can be captured is the Shannon Information Entropy (SIE), defined as follows (Shannon, 1951):

$$H(x) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i).$$

The formula calculates the average level of *information*, inherent to the variable’s possible outcomes (in our case, the characters), whereby a more *informative* outcome is intended to be the less expected one.

Other methods for the assessment of the clear-text language are statistical similarity tests, such as the χ^2 , the Kolmogorov-Smirnov or the Kullback-Leibler divergence tests, conceived to measure how similar two samples are, i.e. how probable it is that they are generated by the same distribution function. Once again, the limited length of the inscription does not allow to take this method in serious consideration. Condition for the comparison is that the samples, algorithmically rendered as two vectors or arrays, be of the same length. Moreover, their values shall be normalized, since the character frequency distributions

Candidate language	SIE
Inscription	~ 4.1
Random text	~ 4.7
Magyar (Old Hungarian)	~ 4.1
Old Luxemburgish	~ 4.2
Old Romanian	~ 4.3
Coptic	~ 4.2
Koiné Greek	~ 5.0
Latin	~ 4.1
Old Czech	~ 4.3
Old Albanian	~ 4.4
Old Slovenian	~ 4.4
Old Spanish	~ 4.2
Old Italian	~ 4.2
Old German	~ 4.1

Table 2: Shannon Information Entropy-value for every candidate language.

may be originated from corpora of different size⁹. After proceeding with the visualization of the letter frequency distribution for every candidate language, a Friedman’s Test has been performed by means of the *CrypTool* software (CrypTool, 2023) to assess the probability of monoalphabetic substitution, which as expected yielded the same result for every language (except for English, where a slight possibility for homophony is still contemplated). It has to be highlighted that homophonic ciphers usually display more than twenty-four letters, while polyalphabetism evens out characters frequency towards distributions typical of random texts. Magyar, Old Albanian and Old Romanian are the languages whose IC more resemble the inscription’s.

The same calculation performed over text samples of different length, for any language, shows that the IC may vary even significantly, which compels us to consider this result as an indication, more than a prediction. The most important takeaway of this kind of analysis is that a randomly generated text shows a completely different value. The same point can not be made for the SIE, whereby the value related to the Greek language results even higher than a random text. Moreover, the lack of significant discrepancies among the entropy values suggests that this kind of measure, at least in our case, can not be effectively

⁹The procedures yielding these statistical tests’ results are to be found in the project *MariaLaNova*, under the file “Graphs.py”.

exploited alone in discerning between randomness and meaningfulness of a clear-text language.

7 Statistical analysis on N-grams: a brief excursus on the AZdecrypt software

The cryptologic attack based on N-grams was supported by *AZdecrypt* (AZdecrypt, 2023), the same software used to decrypt the famed Zodiac cypher¹⁰. Although it is not the most user-friendly and advanced software at disposal in terms of supporting community and releases periodicity, it is extremely flexible to incorporate new corpora.

AZdecrypt is conceived for modern cryptanalysis, especially for homophone ciphers, nevertheless it is easily adaptable to historical ciphers¹¹. The n-grams vocabularies included in *AZdecrypt* are formatted in binary, yet it is possible to include them through a textual file. In the project *MariaLaNova* the function which generates a N-grams file suitable to be processed by *AZdecrypt* can be run by the command:

```
ngramsAZ(file, 5, case = "lower")
```

, whereby the first parameter is a corpus, the second one the desired N , and the third, optional, for rendering the output in lower- or uppercase. The output file will enlist each N-gram immediately followed by its log value, a number between 0 and 255 obtained by:

$$\log_{10}(ngramfrequency_{corpus}) * 10$$

All N-grams followed by “000” could be removed. The GUI library used for *AZdecrypt* does not support Unicode. Hence, only languages that can be represented in ASCII are visually supported. A workaround for this problem is substituting Unicode with ASCII and then providing a ASCII to Unicode mapping table in the n-gram *.ini* file. The *.ini* format is used for simple text files containing initialization parameters. In *AZdecrypt*, it accompains in the “Ngrams” folder every N-gram file. Its appearance for Persian, a language not supported by Unicode, is:

¹⁰<https://www.iflscience.com/fbi-confirms-zodiac-killers-infamous-340-cipher-has-been-decoded-and-his-message-finally-revealed-62044> .

¹¹The repository containing all the outputs, sorted by language and decryption mode can be accessed at <https://github.com/Glottocrisio/AZDecryptMariaLaNova> .

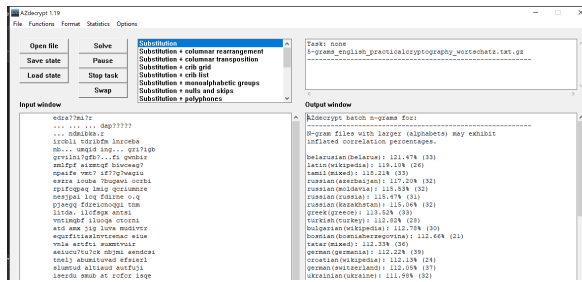


Figure 6: The AZdecrypt "Languages" function.

N-gram size=b5
 N-gram factor=90.11
 Entropy weight=1
 Alphabet=#<*)576%4\$
 ,3:-+?1;0(2&'!8'/.>=9=
 Temperature=700

, whereby in the first line the "b" stands for "binary"¹². It should be deleted for all non-binary formatted N-grams files. The alphabet line shall contain all characters present in the related N-grams file. The *temperature* variable refers to the probability of accepting a modification with a lower fitness. It continuously decreases, emulating the process of annealing in metallurgy, therefore the name.

The strategy adopted in my study to avoid unsupported characters is transposing the corpus into Latin characters *before* generating the N-grams file. This is achieved by the functions contained into the file "Replace.py", and at the state of the art are available for Greek, Coptic and Cyrillic. Other alphabets can be mapped easily following the same model used for the other "Replace" functions.

Before attempting at the decryption via all modes available in the software sound protocol expects to use the function "Languages" to determine which in which clear-text language a given input cipher could be by selecting "File", then "Batch n-grams (substitution)", and by opening "Languages.azd" under "Languages". Nevertheless, it has been my previous statistical analysis to suggest me which languages should have been prioritized in my software-supported attack, namely Latin, Hungarian, Czech, Romanian, Albanian and Church Slavonic.

By clicking on "File" then "Load N-grams" the

¹²For a binary 4-gram file, for example, the first byte would represent the 0 to 255 (log) value of the AAAA n-gram, the second byte would be AAAB... etc. up to ZZZZ. All possible n-grams have to be included in this order.

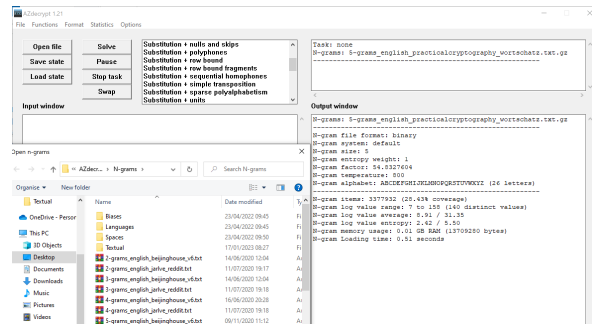


Figure 7: A typical view of AZdecrypt's working environment.

folder with all N-grams is accessed. Before running the decryption in one of the modes selected in the list above¹³, on the window's right side the ".ini" file content, as well as statistical observations on the uploaded N-grams file are displayed.

The N-grams analysis has embraced also an experiment, for which the software was not required. Moving from the observation that all analysed languages display all their vowels within their first eight letters ranked over frequency, a vocals-consonants intertwining can be investigated even if we are still unaware of the clear-text. By replacing in the epigraph *all possible* vowels (all first eight ranked glyphs) with a *V*, we can observe if the behaviour of possible vowels with consonants reflects the one of the other languages. This experiment is based on the assumption that while not all *Vs* are surely vowels, all not-*Vs* are consonants needs. After performing this operation, plenty of all-consonants 3-, 4 and 5-grams are to be found. Usually, three consonants in sequence are already pretty seldom, five almost impossible, in almost any language, which would constitute a strong case either for transposition, or even for randomness.

8 Results

By following the above mentioned steps, plenty of text files have been generated for every considered plain-text language, currently still under examination. Feeding *Google Translate* with some of the outputs' snippets from Latin or Hungarian, we are faced with fascinating translations of outstanding coherence. In spite of this, they cannot be correct at the same time; furthermore, *Google Translate*

¹³Among all, I have used only the functions for "Substitution", "Substitution + Nulls & Skips", "Substitution + simple transposition", "Substitution + sparse polyalphabetism".

showed to be extremely unreliable because meaningless words are therein often rendered with their most similar meaningful term, and translated accordingly. This experience has let me understand that a real awareness of the candidate languages is an indispensable condition for consistent improvements towards decryption. I have done my best with the languages I sufficiently know, such as Latin, Greek, Italian, Spanish and German, although even they cannot be confidently sorted out from our quest. For all the other ones, it is unavoidable gathering help from the scientific community, because only a synergistic effort can finally break this centuries-lasting riddle.

The statistical approach seems to fail at a first attempt, but there are plenty of strategies which can still be taken in account. Besides transposition and substitution, there are indeed other gimmicks which could allow to the ciphertext to maintain its statistical features, as occurring in the seventh challenge of Bellaso (Bellaso, 1553), where the characters frequency distribution resembles our inscription's, although it has been encrypted mainly by the method of *scrambled alphabets*.

9 Concluding remarks and future work

Differently from other case studies, where the encrypted text is usually far longer and cleaner, we have been put against a cryptological challenge for which probabilistic solving heuristics may not be enough, eliciting the need for semantics-aware decryption. Nevertheless, there are still some options to be taken into consideration, still coherent with the monoalphabetic substitution, like the conjecture according to which the epigraph may be edited in more than one language. This path could be walked by creating many bilingual corpora, from which to extract the N-grams. Yet another possible scenario could be the absence of vowels, i.e. the inscription may contain only consonants.

An option which cannot be explored realistically is the use of *steganography* (Trithemius, 1518): even in that case we would fall back in the great obstacle of the used esoteric glyphs, which is even more aggravated by the established practice among coeval cryptographers to willingly insert mistakes or abbreviation in the clear text, in order to impede decryption. In case that insights from collateral elements, maybe discovered inside the church, will suggest that the inscription may

contain one or more words, the *AZdecrypt* function *Substitution + Crib grid* can be used to check whether in the epigraph be contained combination of glyphs allowing that very word. This way has been already tried with terms such as "Holy Mary" and "Jesus Christ", translated into the main candidate languages.

A further idea which shall not be discarded, may be the creation of a corpus entirely constituted of magical/esoteric words, harvested from grimoires, glossaries of conlangs such as Hildegard's *Lingua Ignota*, magical papyri, gnostic libraries and other books about occultism and alchemy: if the alphabet was invented, no one forbids that also the related vocabulary was.

10 Acknowledgments

This work has benefited from the support of many scholars, providing me various suggestions and material for improvement. Therefore I thankfully acknowledge my supervisors Dr. Maria Pia Di Buono and Prof. Dr. Johanna Monti, as well as by Prof. Dr. Beata Megyesi and Dr. Carola Dahlke, for providing me precious critics related to the first draft's flaws in structure and content;

Dr. Richard Bean and Dr. Ivan Parisi, who have bestowed me valuable insights to substantially change the first version of the paper, in particular by shifting my attention to a more thorough statistical analysis;

Francesco Pastore and Francesco Afro Di Falco, for having shared with me very interesting considerations regarding the possible sense of the epigraph, particularly from historical and the esoteric standpoint;

Prof. Dr. Giuseppe Reale, head of the association "Oltre il Chiostro", which manages the museum complex of Santa Maria La Nova, for presenting me very detailed descriptions of the church (in particular concerning the architectonic elements strictly related to the analysed artefact), for granting me free entrance in the museum as well as high quality digital pictures of the inscription;

Dr. Jarl Van Eycke and David Oranchak have been very patient and generous in showing me the approach to integrate the *AZdecrypt* software with the multilingual historic corpus I needed for my inquiry.

References

- I. J. Adeigo. 2006. *The Carian Language*. Brill.
- AZdecrypt. 2023. Azdecrypt. <https://github.com/doranchak/azdecrypt>. Accessed: 2022-11-30.
- G.B. Bellaso. 1553. *Il vero modo di scrivere in cifra*.
- CrypTool. 2023. Cryptool portal - cryptography for everybody. <https://www.cryptool.org/de/>. Accessed: 2022-11-30.
- G. B. Della Porta. 1563. *De Furtivis Literarum Notis*.
- B. Hauer and G. Kondrak. 2016. Decoding anagrammed texts written in an unknown language and script. *Transactions of the Association for Computational Linguistics* 4, page 75–86.
- HistCorp. 2023. Histcorp-historical corpora. <https://cl.lingfil.uu.se/histcorp/index.html>. Accessed: 2022-10-23.
- D. A. King. 2001. *The Ciphers of the Monks: A Forgotten Number-notation of the Middle Ages*. Franz Steiner Verlag, Wiesbaden.
- K. Knight, B. Megyesi, and C. Schaefer. 2011. The copiale cipher. In *4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web BUCC*, pages 12–19.
- H. Kranz and W. Oberschelp. 2009. *Mechanisches Memorieren und Chiffrieren um 1430, Johannes Fontanas "Tractatus de instrumentis artis memorie"*. Boethius Band 59. Franz Steiner Verlag.
- A. Meister. 1902. *Die Anfänge der moderne diplomatische Geheimschriften- Beiträge zur Geschichte der italienischen kryptographie des XV Jahrhunderts*.
- L. Miriello. 2021. *Sulla presunta tomba di Dracula a Napoli*. I Polifemi. Stamperia del Valentino.
- E. Moutafov. 2006. Translating encrypted messages: greek and slavonic tetragrams as a mixture of languages or as a universal code.
- K. Pommerening. 2021. *Cryptology Part I: Classic Ciphers (Mathematical Version)*.
- G. Rocco. 1928. *Il convento e la chiesa di Santa Maria La Nova*. Tipografia Pontificia degli Artigianelli, Napoli.
- Patricia A. Rosenmeyer, 2019. *Encrypted Inscriptions: a Paradoxical Practice*, pages 373–392. Brill.
- Anne-Simone (Ed.) Rous and Martin (Ed.) Mulsow. 2015. *Geheime Post-Kryptologie und Steganographie der diplomatischen Korrespondenz europäischer Höfe während der Frühen Neuzeit*, volume 106 of *Historische Forschungen (HF)*. Duncker & Humblot, Berlin.
- F. Schöning. 2014. *Geheimschrift im deinste der päpstlichen Kurie von ihren Anfängen bis zum Ende des XVI Jahrhunderts*. Nabu Press.
- C. E. Shannon. 1951. Prediction and entropy of printed english. *Bell System Technical Journal*, 30(1):50–64.
- J.W. Somogyi, 1906. *Caratteristiche strutturali di cifrari monoalfabetici italiani nei secoli XIV e XV*, pages 195–213.
- J. Trithemius. 1518. *Polygraphiae libri sex*.

11 Appendix

List of relevant functions from *MariaLaNova*

Freqvoc: this function takes a ".txt" file as input and returns a frequency dictionary of the words contained therein.

letterngramsfilerank: this function generates a n-grams file of a given input file ranked by frequency.

generateioccorpus: the input corpus of "ngramsaz" is created by selecting files whose Index of Coincidence is most similar to the epigraph's one by means of this function.

ngramsaz: this function takes a corpus as input (as the one generated in the previous function) and generates a ".txt" file containing the logarithmic values for all n-grams in the format requested by azdecrypt (5-grams): "ex-ampl1123xampl009ample007".

patternsearch: this function loops through a particular word list seeking for a pattern. It takes the number of characters and the pattern as parameters. 'k' means consonant, 'y' vowel and 'u' all different letters.

matchwordscrco: this function loops through a ".txt" file to find a word matching with the input word as parameter i.e.: *word="polo"* and *file="*bebe; marmelade; mouse; poster; fata", the result will be "fata". In this version, we assume that the script contains no space (see **matchword** otherwise).