A new attack on the mysterious inscription of Santa Maria La Nova

Cosimo Palma
University of Naples "L'Orientale"
University of Pisa, Italy
cosimo.palma@phd.unipi.it

Yll Rugova
University of Prishtina
Kosovo
yll.rugova@trembelat.com

Paolo Bonavoglia
Mathesis Venezia c/o Liceo Foscarini
Italy
paolo.bonavoglia@liceofoscarini.it

Abstract

Expanding upon the established hypothesis of monoalphabetic substitution with potential transposition and polyalphabetic elements, this analysis of the Santa Maria la Nova epigraph incorporates Ancient Greek, Old Church Slavonic, Old Romanian and Old Albanian, thus exploring the possibility that the cipher's plaintext derives from historically under-examined languages, particularly those with cultural and historical ties to medieval Naples and its Eastern Mediterranean networks. Special attention is given to aligning the analyzed corpora with the epigraph's actual textual rendering and to the evaluation of multilingualism.

1 Introduction

After publishing the preliminary cryptoanalysis of the ciphered epigraph in the Turbolo Chapel at Santa Maria la Nova (Palma, 2023), the public interest in this fascinating artifact has been further stirred by the production of a novel documentary that revolves around the same places of the inscription (Repubblica, 2024). Despite the failed decryption attempt, the attention reclaimed from the scientific community, whose main outcome is represented by this contribution, has been positively responded to.

In the "Future Work" section of (Palma, 2023), checking for magical/esoteric words or conlangs (such as Hildegard's *Lingua Ignota*, magical papyri, gnostic libraries and other books about occultism and alchemy) was proposed as a possible approach to carry on the investigation. Format constraints necessitate postponing this approach in favor of a more promising research direction. The present work focuses on monoalphabetic substitution attacks in relevant languages, prioritizing initial decryption stages:

1. Obtaining the clearest possible transcription based on the most recognizable inscription segments;
2. Selecting historically grounded candidate languages;
3. Performing attacks on smaller epigraph portions before attempting alternative methods, considering potential multilingualism.

2 Data Preprocessing and preliminary evaluation

To comply with the first listed principle, a manual transcription has been newly performed, as verifiable in Figure 2.

In addition, the CTTS transcription tool[1] has been used to provide an aided transcription (see Figure 1).

One of its output, Figure 3, helped us notice that some glyphs cluster overwhelmingly in one side (top or bottom) of the inscription. For instance, the sixth stripe from right, representing the frequency and position of the glyph  in the epigraph, clearly shows it only appears in the first half of the inscription.

This observation strengthens the hypotesis that some glyphs can be homophones, an hypotesis that is difficult to reconcile with the character's frequency distribution, which fully aligns with a pure monoalphabetic substitution cipher.

An analysis of digrams composed of most frequent characters (as shown in Figure 4) opens a new avenue for investigation, as every language must display a costant behavior for digrams as well as for their opposite (obtained by switching the letters).

In the following sections, four languages have been more thoroughly investigated because of their promising features.

[1] https://github.com/CrypToolProject/CTTS/tree/main.

Proceedings of the 8th International Conference on Historical Cryptology, pages 106-110
Poznań, Poland, 16-18 June, 2025

Figure 1: A snapshot of the encrypted epigraph of the Turbolo Chapel in the Neapolitan Church of Santa Maria La Nova from the transcription environment CTTS.

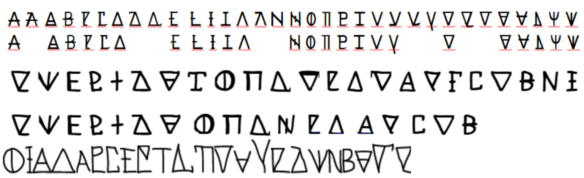

Figure 2: Comparison of characters transcription from the epigraph performed independently by the authors (the first two providing a minimal and extended character's interpretation; the latter a single one.)

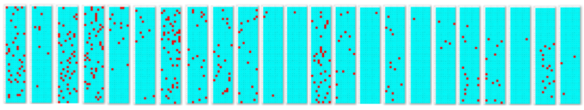

Figure 3: A visualization capturing both position and frequency of each inscription's glyph.

3 In the footsteps of Kostantin Castrioti: the Old Albanian lead

After the death of Scanderberg, many Albanians migrated to Italy; part of them, to Naples (Schmitt, 2014). The church of Santa Maria la Nova in Naples is important for Albanological studies precisely for this reason. Within its premises, for example, is located the grave of Konstantin Kastrioti (d. 1500), nephew of Gjon Kastrioti (Skenderberg) — Albanian national hero (Ciarlanti, 1644; Padiglione, 1879). The dedication below the sculpture on the sarcophagus clearly states that it was raised by Donika Kastrioti, Skenderbeg's wife, in 1500, for commemorating the untimely death of her nephew (Amarelli, 2012). When in the summer of 2024 a delegation of the Academy of Science of Kosovo visited Napoli, they were informed that the grave of Konstantine Kastrioti was at one point somewhere in the Turbolo chapel, where the undeciphered inscription takes up a corner of a wall. Such changes within the church seem to have been a common occurrence (Forgione, 2023). In any scenario, this would mean that the mysterious inscription could belong to a period before the Turbolo family restructured the chapel (1582), maybe even to the year 1500 when Konstantine's grave was placed there.[2]

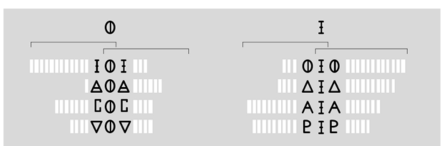

Figure 4: A diagrammatic sample of most frequent digrams.

Until the late 17th century, Albanian was mainly written using the Latin script. The oldest Albanian written text is an inscription with Latin letters, found on a manuscript dated 1462. It is a well known short (Elsie, 2017) word baptism formula originally from Durrës in Albania. The oldest book written in Albanian is from 1555, known as *Meshari* (Mass book), a book printed using Venice typographic materials surviving in a single copy kept in the Vatican Library in Rome. This book was also written in Latin script, with 4 additional symbols mainly borrowed from cyrillic *bosancica* script. Though a manuscript from 1400 mentions an original alphabet for Albanian texts (Schmitt, 2014), no traces of such a script were witnessed until the end of the 17th century (Rugova, 2022). A certain manuscript called *Anonimi i Elbasanit*, dated to be from the end of the 17th, or the middle of the 18th century is the first known text in Albanian with an original alphabet (Elsie, 2017). In subsequent decades several other such original scripts emerged, mainly in what is known as the orthodox milieu of the Albanian speaking communities [ibid.]. One of them seems to have been shared by the Arumanian speaker of Albania as well (Demiraj, 2019). No similarities can be spotted between these original scripts and the one of the Napoli inscription. However, given the

[2] A work from Simone Papa (Il Vecchio) in 1488 is one example of works finished there before the Turbolo family made their changes in 1580s (Bryan, 1849).

plurality of different Albanian scripts mentioned above, another one from 1500 would indeed be a delightful discovery for Albanian studies, but not a surprising one. That, together with statistical cues–as shown in (Palma, 2024), was the reason we used an Albanian corpus to try to decipher the Napoli inscription. For this purpose, given the closeness to Konstantin's period, the Meshari of Gjon Buzuku was used as the main Albanian corpus.

It has been harvested from the Titus-Project (Gippert, 1996)[3] and processed by means of the cryptology suite *Malanova*[4] previously developed as codebase for (Palma, 2023). Enlarging the corpus with further documents has been momentarily avoided, given the exploratory nature of the analysis. Moreover, every corpus in Old Albanian presents slightly different scripts and even vocabularies, which would have posed an issue of alignment that cannot be tackled in the limited scope of this survey. Using the software *AZDecrypt*[5], loaded with the 5-grams generated from the preprocessed corpus, an attack has been run using the *Substitution* as well as the *Substitution Skips and Nulls* functions, respectively loading all different epigraphy transcriptions as described in section 2. Every test has been run setting different levels of entropy: 1, 5 and 10, each time for the whole epigraph, then the upper, and then the lower half.

4 Tracking down the Balkan way: failed attempts with Old Romanian and Old Church Slavonic

Previous statistical analysis have also shown a good fit between the character distribution of the inscription with Old Church Slavonic and Old Romanian. This correlation has been further confirmed by calculating the *Entropy* and the *Index of Coincidence* (Friedman, 1922). The brevity of the text does not allow to be overly optimistic about such metrics, that can be reliable only on larger portions of text. However, it cannot be neglected that these encouraging values have been measured only for these languages, which by chance corresponds to the language whose alphabets feature in the inscription, and are in the closest geographical area surrounding the one where Old Albanian

Carattere	Frequenza Assoluta	Frequenza Relativa.	lettere	
Φ	49	13.0 %	A	11.94 %
Δ	45	11.9 %	O	9.87 %
Ⱥ	41	10.8 %	I	9.61 %
Ɨ	39	10.3 %	T	8.77 %
Ᵽ	28	7.4 %	E	8.58 %
⊏	25	6.6 %	Σ	7.84 %
Ꝟ	18	4.8 %	N	5.90 %
Ɲ	17	4.5 %	H	4.96 %
∇	16	4.2 %	P	4.82 %

Figure 5: Comparison of the frequencies of the cipher and those of ancient Greek

was spoken. The very same approach as described in section 3 has been applied for these languages. The related corpora have been harvested from the Language Bank of Finland *Kielipankki*[6] and from (Niculescu and Dimitrescu, 2000) for Old Romanian, by means of OCR digitization[7].

5 It may be Greek, after all

As mentioned in (Rocco, 1928), the author assesses that the inscription is the Greek translation of the adjacent Latin inscription. Although even a profane would easily acknowledge that the first epigraph is too long, compared to the second one, there are indeed many signs similar to Greek letters (Α Β Δ Ε Ι Λ Ν Ο Π Ρ Τ Υ ψ Φ) appearing in the epigraph. If we also consider that father Rocco was esteemed as a cultivated literate, this kind of mistake would sound excessively superficial. What if he knew something he could not openly disclose? Another element worth considering is the final sign of each line. Of the 20 counted lines, 9 ended with Δ, 6 with A and 5 with I. This is quite unexpected; most languages have words ending with a large number of different letters. Conversely, in Italian almost all words end with the four vowels A O I E. And in Greek they end in a very large part with A, H or Σ.

In classical cryptanalysis, the first step after frequency analysis, is identifying the vowels. In addition to the individual letters, the frequency of digrams is very useful. The vowels agree with almost all consonants, while the consonants are more quarrelsome; calculating the entropy[8] of the

[3]https://titus.uni-frankfurt.de/texte/etcs/alban/buzuku/buzuk.htm

[4]https://github.com/Glottocrisio/MaLaNova

[5]https://github.com/doranchak/azdecrypt

[6]https://www.kielipankki.fi/corpora/ccmh/

[7]We thankfully acknowledge Francesco Pastore for providing this resource.

[8]About entropy and cryptanalysis see (Bauer, 2007)

digrams of each letter, the vowels have a value of about 4, while the consonants have values under 3.5 except for R (P in Greek) which has values slightly lower than the vowels. Hence, it turns out that the signs similar to O, I, A, are almost certainly vowels. The Δ sign also seems to be a vowel. However, even not knowing the original language, these data may be useful, because most frequencies of the same letter are not so different: this case is a good example. Assuming Latin as the language and the highest frequencies, the word BASILEIA, which does not exist in Latin, turned up in the first complete line, while ΒΑΣΙΛΕΙΑ exists in ancient and modern Greek. Putting aside the Latin language and moving on to Greek, something very interesting emerges: a whole line can be deciphered according to this hypothesis . After ΒΑΣΙΛΕΙΑ there are four signs ;A;Σ and at the end of the line, there are four more characters: E;IA , resulting in this partially decrypted line:

ΒΑΣΙΛΕΙΑ ΝΑΥΣ..Α Ι ..ΕΡΙΑ

Later, a name from the most remote school memories (the Odyssey) emerges: ΝΑΥΣΙΚΑ Nausicaa the daughter of the king of the Phaeacians who welcomes the shipwrecked Ulysses. In reality between Σ and A there is a practically empty space, two or three letters erased. If we assume that they are IK it comes out precisely . Later we found that the island of the Phaeacians, today identified with Corfu, was called in ancient times ΣΧΕΡΙΑ![9] giving:

ΒΑΣΙΛΕΙΑ ΝΑΥΣΙΚΑ.ΣΧΕΡΙΑ

Three semantically related concepts appearing in the same row strongly suggest a pattern. However, this hypothesis faces these challenges:

1. Up to now trials based on this same substitution pattern to the remaining text, did not yield meaningful texts;
2. Another row contains a seemingly meaningful Greek phrase, but with inconsistent character mappings for at least two glyphs.

A plausible explication of these problems is the use of multiple languages. While this would significantly complicate decryption, it's supported by the emergence of plausible Greek syllables throughout the text, consistent with typical Greek letter distribution patterns.

[9]An homophone was used here for Σ; not so strange, because Σ has also the ς variant and the sum of frequencies is almost the same as in Greek.

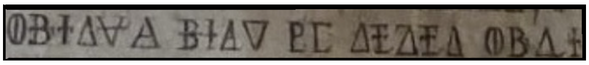

Figure 6: A detail of row 22, showing the only palindrome word of the whole inscription.

The twenty-second row of the inscription contains the only palindrome in the text (see Figure 6). Among all possible five-letter palindromes in Koiné Greek, only ρητηρ (rhetor) is consistent with the character frequencies shown in Figure 5. This line is particularly notable for its repeated glyphs. Manual decryption through character substitution yields: **ἀγορεύ(-) γονυ δέ ῥητήρ ἀγνό**, meaning "... to harangue on one's knees, but rather as a venerable rhetor".

6 Final remarks and Future Work

In cryptanalysis, the successful decryption of consecutive meaningful words typically enables complete reconstruction of the cipher through systematic steps. However, this case presents unusual challenges. Several critical questions emerge: is the current decryption attempt entirely incorrect? Is this particular line accurately deciphered while other sections follow different patterns? Does the inscription employ multiple languages? Automated analyses using Old Romanian and Old Albanian corpora have not yielded meaningful results[10]. Given the epigraph's manageable size, manual cryptanalysis appears to be a more promising approach. While definitive solutions remain elusive, several avenues for future investigation present themselves. The most promising directions include:

1. Systematic testing of Old Albanian dialectal variations;
2. Deepening the Greek lead, trying to exploit the "crib grid" function of AZDecrypt to produce constrained output based on partial decryptions achieved from manual attempts;
3. Involving in the quest experts in the candidate languages.

[10]The output files of the automatic attacks using AZDecrypt are stored in the codebase: `https://github.com/Glottocrisio/AZDecryptMariaLaNova/tree/main/main/Output`. Results from the present work are stored in the folders "gre", "rom", "alb" and "slav".

References

Marco Amarelli. 2012. Costantino e La Casa Castriota. Nuovi Contributi Sulla Biografia e Gli Scritti Di «Filonico Alicarnasseo». *Critica Letteraria*, 40(154):121.

Friedrich Ludwig Bauer. 2007. *Decrypted secrets: Methods and Maxims of Cryptology*. Springer, Berlin.

Michael Bryan. 1849. *A Biographical and Critical Dictionary of Painters and Engravers, from the Revival of the Art... to the Present Time...* H. G. Bohn, London.

Giovanni Vincenzo Ciarlanti. 1644. *Memorie Historiche Del Sannio Chiamato Hoggi Principato Vltra, Contado Di Molisi, e Parte Di Terra Di Lauoro, Prouincie Del Regno Di Napoli*. Camillo Cauallo, Isernia.

Bardhyl Demiraj. 2019. Enigma e Gravurës Së Ardenicës. *Shejzat-Pleiades*, (1–2):121–125.

Robert Elsie. 2017. *Albanian Alphabets: Borrowed and Invented*. Centre for Albanian Studies, London.

Alessandro Forgione. 2023. Dracula é sepolto a Napoli? *Fenix*.

William F. Friedman. 1922. The index of coincidence and its applications in cryptanalysis. *Riverbank Publication*, 22.

Jost Gippert. 1996. Titus - alte und neue perspektiven eines indogermanistischen thesaurus. *Studia Iranica, Mesopotamica et Anatolica*, 2:49–76. Published in 1997.

A. Niculescu and F. Dimitrescu, editors. 2000. *Testi romeni antichi (secoli XVI-XVIII)*. Vulgares eloquentes. Antenore.

Carlo Padiglione. 1879. *Di Giorgio Castriota Scanderbech e de' Suoi Discendenti: Narrazione*. Stab. Tipografico del Cav. Francesco Giannini, Napoli.

Cosimo Palma. 2023. Encrypted epigraphy - the case of a mysterious inscription in the neapolitan church of santa maria la nova. In *International Conference on Historical Cryptology*.

C. Palma. 2024. Epigrafi cifrate nelle chiese antiche. In *ERUDITORUM ACTA 2024*, volume 3 of *De Cifris Koine*, pages 35–51. De Cifris Press, September. Cosimo Palma, 0000-0002-8161-9782.

La Repubblica. 2024. Vampiri per caso, la vera storia di dracula a napoli: la prima del film di syusy blady. *la Repubblica*, October. Online newspaper article.

G. Rocco. 1928. *Il convento e la chiesa di Santa Maria La Nova*. Tipografia Pontificia degli Artigianelli, Napoli.

Yll Rugova. 2022. Dëshmi Mbi Një Alfabet Origjinal Të Shqipes Nga Shekulli XVII: Kozmai i Durrësit Si Shpikës i Mundshëm i Alfabetit Të Parë Origjinal Të Shqipes. *Shejzat-Pleiades*, 7(3–4):21–37.

O. J. Schmitt. 2014. *Skënderbeu, Rishtypje e zgjeruar*. Fjala Publishing, Tiranë.