The Upplandic Non-Lexical Rune Stones: Ciphers or Nonsense?

Viktor Wase
Stockholm, Sweden
viktorwase@gmail.com

Abstract

The so-called non-lexical rune stones use ordinary runes but contain nothing but nonsensical "words". It is not entirely uncommon for rune stones to contain hidden and enciphered messages, which is why this study investigates the possibility of the Upplandic non-lexical stones being ciphers. This is done using a graph clustering algorithm that sorts the stones into groups based on how similar their texts are.

The algorithm labeled all non-lexical stones as outliers (belonging to no group), with the exception of U1126 and U1128 that form a group on their own. As such it is deemed unlikely that any of the non-lexical stones (perhaps excluding U1126 and U1128) are ciphers.

1 Background

The occurrence of ciphers in and among runes are not at all uncommon. Even rune stones, placed out in the public for everyone to see, contain messages in the form of ciphers. It is therefore a natural conclusion that the purpose of the ciphers was not to convey a hidden meaning, but something else.

Take the Kareby baptismal font (signum Bo NIYR5;221B), for example. Its transcription, excluding a complicated bind rune, reads raþesaerkannamnorklaski (Bæksted, 1949). This gives: *raðe sa er kan namn orklaski*, which roughly translates to *Read those who can the name orklaski*. Orklaski is not a known name. However, if one replaces each rune with the one that precedes it in the younger futhark, then orklaski becomes þorbiarn, which is still a common name in Norway. This is a Caesar cipher and is simple to solve once you know how (Suetonius, 1914).

In a similar vein, the stone U 1165 ends with a series of long and short lines: ||′′′′||′′′|||′′′′′||′′′|||′′′′′′|||′′′′′ hiuk (Nordby, 2018, p. 392). This is a binary rune cipher, where each rune can be reduced to a pair of numbers. By pairing up the long lines with the short ones following to the right one gets 2/4, 2/3, 3/5, 2/3, 3/6, 3/5. The younger futhark is commonly divided into three parts (ætt): fuþork hnias tbmlR. The first number indicates which ætt, and the second number the index in the ætt. It should be noted that the ættir are numbered backwards as 3, 2, and 1. This gives airikr hiuk, which translates to Erik carved. Once again the name was the only part that was encrypted.

A third type of cipher can be found in DR 239, which contains the following inscription *þmk iii sss ttt iii lll*. This is called an istil-formula, since the runes can be shuffled into three words that end with istil: *þistil* (thistle), *mistil* (mistletoe) and *kistil* (box) (Nordby, 2018, p. 104).

1.1 Non-Lexical Stones

There are some rune stones, mostly in Uppland and Södermanland, Sweden, that have no apparent meaning. They look like regular rune stones and the runes are of standard runic form, but they do not form words and sentences. They are commonly referred to as non-lexical stones. The common belief is that they are produced by illiterate carvers (Bianchi, 2010, p. 165), but there are fringe theories about their actual message.

For example Stig Eliasson argues that these stones show some patterns that would not show up if it were pure and random gibberish (Eliasson, 2014). This could indicate, he argues, that they might be written in an unexpected language. This is concretized by suggesting that the Danish Sørup stone might be written in Basque (Eliasson, 2010).

Perhaps not surprisingly, people have considered the possibility of the non-lexical stones being ciphers. In 1923 Erik Brate wrote this about the stone U 466: "...designed with the intention to test the wit of the reader, which supersedes the abilities of our time" (from Swedish: utförd i avsikt att sätta läsarens skarpsinne på prov, som överstiga vår tids förmåga) (Wessén and Jansson, 1946, p. 279-281). Regarding U 298 he wrote that he believed it to be "hidden writing" (from Swedish: lönnskrift) (Wessén and Jansson, 1946, p. 6-7). Rikard Dybeck, the creator of the de-facto Swedish national anthem, wrote about U 427 in 1877: "the inscription, as of yet uninterpreted, will probably remain so for a long time to come." (from Swedish: inskriften, hittills otydd, lärer länge nog förblifva det.) (Wessén and Jansson, 1946, p. 214-216).

More recently Craig P. Bauer argued similarly in his book Unsolved!. He concludes with the following remark: “A statistical study needs to be conducted on groups of related stones, such as those from Uppland, Sweden, with currently unreadable runic inscriptions to see if they might have been enciphered in the same manner.” (Bauer, 2017, p. 115-126).

1.2 Classification of Runic Cryptology

K. Jonas Nordby created a classification of runic cryptology (Nordby, 2018, p. 76). The two top classes are permutation and substitution ciphers. Permutation means that the runes are sorted in some unusual order, and substitution means that a specific symbol represents a specific rune. Permutation ciphers are simple to detect, since the frequencies of the runes are the same as in non-encrypted texts. Substitution ciphers are a bit trickier. One has to differentiate between mono-alphabetic substitutions ciphers (commonly abbreviated MASC) and homophonic substitution ciphers. The former being a cipher in which one symbol represents one rune, and the latter several symbols can represent one rune (Dooley, 2018, p. 9). The homophonic substitution ciphers can be excluded from this study since they require more symbols than the used alphabet, and the non-lexical stones only use the symbols from the futharks.

However, one of the sub-classes of substitution in Norby’s classification is neither mono-alphabetic nor poly-alphabetic. It is called jǫtunvillur, and in it each rune is replaced by the last rune in its name (Nordby, 2018, p. 135). In English this would entail that B is enciphered to E, since the letter is pronounced bee. Likewise F would be enciphered to F, since it is pronounced eff. The problem is that C would also be enciphered to E. This makes it a very inpractical cipher that is very hard to read. Nordby argues that it might have been a tool for learning the names of the runes (Nordby, 2018, p. 149).

There exists ciphers that are dependent on the position of the letter as well. For example, A might be encoded as B if it is the first letter of a text but encoded as C if it is the second letter. These ciphers tend to be highly complex and nothing of the sort has been found in the Viking era Scandinavia. The earliest examples found are from the 16:th century (Bonavoglia, 2020, p. 46). These are therefore excluded from the search, and the algorithm is not expected to be able to find any such ciphers.

1.3 Aim

The aim of this study is to develop an algorithm that takes a collection of short texts, from the stones, and divides them into groups. Each group will contain stones that are similar, in the sense that the frequencies of the runes are similar. If a stone is dissimilar to all the other stones then it will be classified as a singleton. This algorithm will then be applied to a collection of stones with both ordinary texts and non-lexical texts.

There are two foreseeable outcomes. Either, only one large group is formed with most of the regular stones and all of the non-lexical stones are filtered away as singletons. Or, a large group is formed with most of the regular stones, and a second group is formed with a portion of the non-lexical stones. Note that all non-lexical stones do not need to be in this second group, since it is possible that some of them are ciphers while others are not. The second outcome would indicate that the ciphers are distinct from the regular stones, but similar to each other. This means that there is some underlying pattern that could indicate the existence of a cipher. If neither of these are the true outcome, then the algorithm will not have been successful in separating the regular stones from the non-lexical stones, and a new algorithm will have to be developed.

The goal of the algorithm should be to be able to find stones that use mono-alphabetic or jǫtunvillur-like substitution ciphers, without being

Baseline	Non-Lexical
U 32, U 46, U 56, U 69, U 91	U 298
U 96, U 99, U 109, U 124	U 370
U 132, U 135, U 144, U 147	U 427
U 151, U 155, U 164, U 165	U 466
U 166, U 175, U 184, U 186	U 468
U 189, U 192, U 193, U 217	U 483
U 224, U 227, U 240, U 244	U 522
U 257, U 259, U 261, U 276	U 811
U 292, U 305, U 327, U 328	U 902
U 342, U 345, U 365, U 368	U 983
U 372, U 373, U 390, U 397	U 1126
U 423, U 431, U 435, U 441	U 1128
U 442, U 486, U 494, U 495	
U 528, U 530, U 574, U 577	
U 580, U 582, U 585, U 594	
U 606, U 620, U 660, U 662	
U 683, U 732, U 750, U 768	
U 814, U 826, U 856, U 866	
U 875, U 903, U 911, U 941	
U 943, U 949, U 960, U 961	
U 967, U 969, U 972, U 978	
U 994, U 1003, U 1028, U 1037	
U 1045, U 1060, U 1070, U 1127	
U 1129, U 1131, U 1146, U 1148	
U 1151, U 1157, U 1172	

Table 1: The stones in the baseline group and the non-lexical group.

Figure 1: U 99 from the baseline group. Picture taken 1931.

confused by permutations.

2 Dataset

The dataset used is the offline version of the Scandinavian Runic-text Database (samnordisk runtextdatabas). The scope of the study will be limited to the Upplandic stones. There are over 1100 such stones. We form two groups based on these stones. The first is the baseline group and the second is the non-lexical group. See table 1.

The baseline stones are 100 stones that are longer than 10 total runes and contain only sixteen-rune younger futhark without extensions. They also had to have a translation in the Scandinavian Runic-text Database, to ensure that they do indeed have a lexical meaning. The stones are chosen randomly. To reduce the scope of the study we will only focus on stones with young futhark. The non-lexical group is based on the separation made by Marco Bianchi in his doctoral thesis (Bianchi, 2010, p.170-199) with some removals. The four stones U 523, U 835, U 1170 and U 1175 were removed since they contain rune-like signs, but no actual runes. U 888 and U 1179 were removed since they contain very little of the original message. U 493 and U 1180 were removed because they contained the letter e, which is not part of the younger futhark. U 529 was removed since the runes are very shallow and hard to read, to the point that Scandinavian Runic-text Database did not have any runes in its entry. U 1061 and U 596 did not have entry either, so it was removed. Finally, U 1078 was removed since it only had four symbols on it. This leaves 12 stones.

All uncertain runes, guesses and non-futhark signs were removed from the dataset. Old sources (indicated by [ ] in the Scandinavian Runic-text Database) were used. All stones with ciphers (indicated by < >) and variants of words (indicated by /) in the transcription, were excluded from the baseline group. This gives 499 runes in the non-lexical group and 4998 in the baseline group.

3 Algorithm

3.1 Similarity Measure

The goal of the algorithm is to cluster the stones based on how similar they are. But similarity has yet to be defined in this context. Each stone is converted to a list of 16 numbers, one for each rune in the younger futhark. This number represents the frequency of the rune in the stone. For example if a carving were to have 60 % i-runes and 40 % l-runes, then its list would contain zeros except for the numbers representing i and l which would be 0.6 and 0.4. The distance metric chosen is the common Pythagorean metric, but in 16 dimensions.

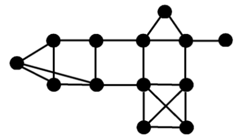
A clustering algorithm based on this similarity measure won't be confused by permutations since the frequencies of the runes remain unchanged if the order of the runes is changed. As a matter of fact, this means that the algorithm will not be able to differentiate pure permutations from non-enciphered stones. Substitution ciphers should be detectable since their distribution of frequencies will change.

3.2 Clustering

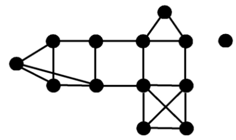
Before the data is passed to the algorithm it has to be converted into a graph of points connected by edges. Each point represents a stone in the dataset, and it is connected to all other stones that are similar to it.

First the Pythagorean distance of all pairs of stones are calculated. The median of these numbers is set as a threshold; if the distance is lower than the median then the pair is connected by an edge (similar) otherwise they are not. This choice of threshold is arbitrary and two other threshold values are used as well to ensure robust results. These values are the 40:th and 30:th percentile of the distance of all pairs of stones.

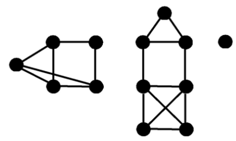
The clustering part of the algorithm is based on an algorithm called Highly Connected Subgraphs (HCS). It is a rather simple algorithm that takes a graph and looks for the smallest set of edges without which the graph will become disconnected - a so-called minimum cut. It then repeats this procedure on the two new separated graphs. It stops dividing a graph when its minimum cut contains n / 2 or more edges, where n is the number of points in the graph (Erez Hartuv, 2000). Consider the graph below, for example.



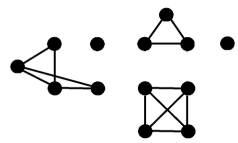
The minimum cut contains only one edge, the rightmost one. If that edge is removed, then the graph is divided into two disconnected sub-graphs, thus creating the graph below.



The minimum cut of this graph contains two edges, since one cannot divide the graph into two disconnected sub-graphs by removing only one edge. It is worth noting that the algorithm has two choices here: either it removes the triangle at the top (creating a singleton) or the horizontal edges of the square in the middle. In these cases the outcome is random. Let us say that it chooses the square.



With two sub-graphs (ignoring the singleton) the algorithm will process them separately. In both cases the minimum cut is two, and removing the edges gives the following graphs.

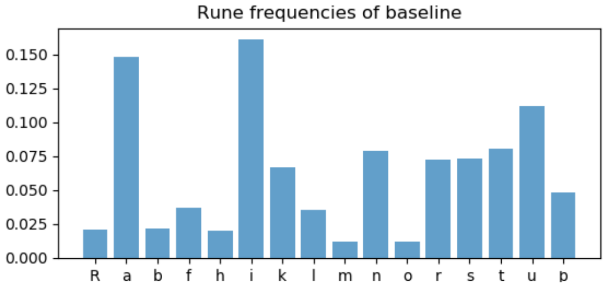


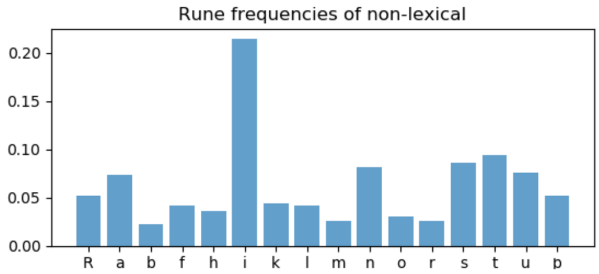
At this point the minimum cut of each sub-graph is larger than or equal to half of the number of points in each sub-graph, which means that the algorithm stops.

This paper used the Python implementation found att *github.com/53RT/Highly-Connected-Subgraphs-Clustering-HCS*.

4 Results

Before we get to the result of the clustering algorithm, let's quickly examine the rune frequencies of the baseline group and the non-lexical group, as seen in the figures below.





The frequency distributions are clearly different, as seen by the huge spike in *i* in the non-lexical group and the lack of such a spike in *a*. However, it is clear that the non-lexical distribution is not simply a reordered version of the baseline. This indicates that there is no widespread use of a substitution cipher in the non-lexical stones.

4.1 Graph Algorithm Result

The results from the algorithm are quite interesting. The overall behaviour was the same no matter if the threshold value was the median, or the 40:th or 30:th percentile. The result was: one large sub-graph containing the majority of the stones, one tiny sub-graph containing the pair U 1126 and U 1128, and then a lot of singletons. The remarkable part is that the non-lexical stones were almost always filtered out from the large sub-graph - the only exception being U 298 when the threshold was the median. See table 2 for the full results for the 40:th percentile. The non-lexical group is marked in bold font.

The algorithm can clearly filter out the non-lexical stones from the bulk of the regular stones. The fact that it did not group the non-lexical stones together means that it seems unlikely that there is any widespread use of substitution ciphers, permutation ciphers or a combination of the two. Roughly a third of the baseline was excluded from the large group. It should be noted that the outlier group contains stones with common forms. For example

U 135 (translated): *Ingifastr and Eysteinn and Sveinn had these stones raised in memory of Eysteinn, their father, and made this bridge and*

Figure 2: The non-lexical stone U 1126. From *Upplands runinskrifter*, part 2 (1946).

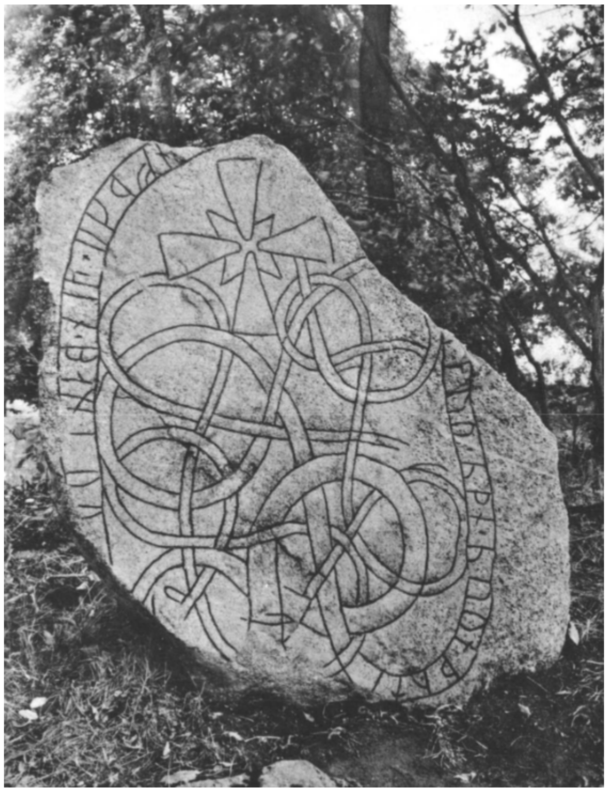


Figure 3: The non-lexical stone U 1128. From *Upplands runinskrifter*, part 2 (1946).

Large Group	Small Group	Singletons
U 32, U 56, U 69	**U 1126**	U 46
U 91, U 96, U 99	**U 1128**	U 132
U 109, U 124, U 144		U 135, U 155
U 147, U 151, U 164		U 184, U 192
U 165, U 166, U 175		U 217, U 244
U 186, U 189, U 193		U 257, U 292
U 224, U 227, U 240		**U 298**, U 305
U 259, U 261, U 276		U 345, U 365
U 327, U 328, U 342		U 368, **U 370**
U 372, U 390, U 423		U 373, U 397
U 431, U 435, U 486		**U 427**, U 441
U 494, U 530, U 580		U 442, **U 466**
U 582, U 585, U 606		**U 468**, **U 483**
U 660, U 662, U 732		U 495, **U 522**
U 750, U 768, U 814		U 528, U 574
U 826, U 866, U 875		U 577, U 594
U 903, U 911, U 941		U 620, U 683
U 949, U 960, U 961		**U 811**, U 856
U 969, U 972, U 978		**U 902**, U 943
U 1003, U 1028		U 967, **U 983**
U 1060, U 1070		U 994
U 1127, U 1129		U 1037
U 1146, U 1151		U 1045
U 1157, U 1172		U 1131
		U 1148

Table 2: The results from the algorithm. Most stones belong to one large group, many do not belong to any group (singletons) and two form a small group. The non-lexical stones are in boldface.

this mound.

U 244 (translated): *Fasti had the stone cut in memory of Fastulfr, his son.*

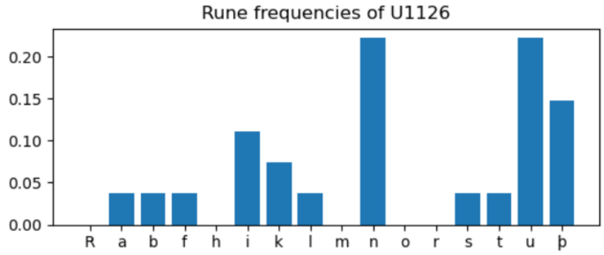
Both of these have common forms and it would be expected that a similar ones would exist in the baseline group, but even so they have been marked as singletons. This indicates that the methodology is not perfect at the stone-level, even if the algorithm manages to catch the overall larger trends.

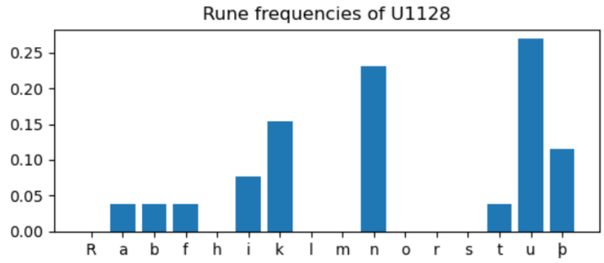
This brings us to U 1126 and U 1128. The fact that these stones are grouped together is rather intriguing. These stones are currently placed next to each other at the Alunda church This was not known to the algorithm, and yet it managed to pair them together. This does, of course, not mean that they are ciphers. It only means that they are similar. The inscriptions of the two stones are:

U 1126 *uluiuþnis-... ...-þnf]a · nnu · ub ' tnþk · uþnki*

U 1128 *...nfþku* × *–in · ban-iuu ...-nuu ' kþn ' kuunþkt-*

See the figures below for the runic frequencies of U 1126 and 1128.





Both U1126 and U 1128 differ greatly from the baseline frequencies. They might be monoalphabetic substitution ciphers. Or perhaps they are encoded with a cipher in which multiple runes are enciphered to the same symbol? They are unlikely to be jotunvillur since that only has 6 unique runes (Nordby, 2018, p. 137) and U 1126 and U 1128 has 11 unique runes together.

It seems Rikard Dybeck was right. The inscriptions will remain uninterpreted for at least a while more.

5 Acknowledgements

A big thank you goes out to Dr. James Knirk for vast and valid feedback. Any and all shortcomings and faults are of course my own.

References

Craig P. Bauer. 2017. *Unsolved!: The History and Mystery of the World's Greatest Ciphers from Ancient Egypt to Online Secret Societies*. Princeton University Press.

Marco Bianchi. 2010. *Viking age written culture in Uppland and Södermanland*. Department of Nordic Languages. Uppsala University.

Paolo Bonavoglia. 2020. Trithemius, Bellaso, Vigenère - origins of the polyalphabetic ciphers. *Proceedings* *of the 3rd International Conference on Historical Cryptology*, pages 46–51.

Anders Bæksted. 1949. Kareby-fontens runeindskrift. *Fornvännen - Journal of Swedish Antiquarian Research*, pages 49–53.

John F. Dooley. 2018. *History of Cryptography and Cryptanalysis - Codes, Ciphers and Their Algorithms*. Springer International Publishing.

Stig Eliasson. 2010. Change resemblances or true correspondences? on identifying the language of an 'unintelligible' scandinavian runic inscription. In *Transeurasian verbal morphology in a comparative perspective: Genealogy, contact, chance*, pages 43–79. Otto Harrassowitz Verlag.

Stig Eliasson. 2014. Runic inscriptions in an unrecognized foreign tongue? methodological preliminaries to language identification. *8:th International Symposium on runes and runic inscriptions*.

Ron Shamir Erez Hartuv. 2000. A clustering algorithm based on graph connectivity. *Information Processing Letters*, 76:175–181.

K. Jonas Nordby. 2018. Lønnruner - kryptografi i runeinnskrifter fra vikingtid og middelalder.

Suetonius. 1914. *Lives of the Caesars, Volume I: Julius. Augustus. Tiberius. Gaius. Caligula.* Harvard University Press.

Elias Wessén and Sven B. F. Jansson. 1946. *Upplands runinskrifter, part 2*. Kungliga Vitterhets Historie Och Antikvitets Akademin.