

Inscription de l'Anse du Caro

Dossier réalisé du 26/07/19 au 28/08/19

François PARMENTIER

—

Ce dossier ne prétend pas comprendre précisément l'inscription laissée sur le rocher de l'Anse du Caro. Son objet est de saisir le problème avec la démarche la plus rationnelle possible, en prenant chaque élément pour ce qu'il est : une donnée sur l'inscription (partie I), le fruit d'une analyse automatisée de l'inscription (partie II) ou une théorie/tentative de décodage menant ou non à un texte doté de sens (partie III). Ainsi, ce dossier gagne en incertitude au fil des pages, pour terminer sur une partie différente (IV), constituée de perspectives et de conclusions provisoires.

Peut-être qu'une des tentatives effectuées percera le mystère de cette inscription ; peut-être pas. Peut-être, plus probablement, que les données laissées ici aideront d'autres à déterminer le sens de ce texte. Quoi qu'il en soit, je tiens à présenter ce dossier au jury, en espérant qu'il lui apporte des éléments de réflexion pertinents.

I. Données du problème

I.1. Contexte historico-géographique

Afin d'éviter quelques longueurs à ce dossier déjà consistant, le contexte sera résumé en quelques points :

- Géographie :
 - Coordonnées : 48° 34' 50" 20" N, 4° 44' 41" 06" O ;
 - Le rocher se trouve près de l'Anse du Caro, à quelques centaines de mètres d'Ilien-en-Traoñ, sur la presqu'île de Plougastel-Daoulas (Bretagne).
- Histoire :
 - **Date de création inconnue** : On lit usuellement 3 dates sur le rocher : 1787, 1786, 1771 (la logique voudrait qu'il faille sélectionner la plus tardive comme candidat probable de date de création) ;
 - **1920** : Inscription de l'année, paraît-il par un soldat russe¹ ;
 - **1979** : Signalement de l'inscription par Bernard Tanguy auprès de la Société Archéologique du Finistère² ;
 - **1984** : Article d'Yves-Pascal Castel dans le *Bulletin de la Société Archéologique*. Présentation d'une première transcription³⁻⁴ ;
 - **Mai 2019** : Lancement du concours de décryptage, après plusieurs recours à spécialistes tentés en vain⁵.

1 <https://www.ouest-france.fr/bretagne/brest-29200/qui-dechiffrera-les-mysterieuses-inscriptions-du-rocher-du-caro-plougastel-daoulas-6332545>

2 <https://societe-archeologique.du-finistere.org/inventaires/plougastel-daoulas-inscription-anse-caro.html>

3 Ibid.

4 https://actu.fr/bretagne/plougastel-daoulas_29189/linscription-mysterieuse-rocher-plougastel-daoulas-pres-brest_24707361.html

5 Ouest-France, ibid.

I.2. Transcription du texte

Comme indiqué plus haut, une première transcription de l'inscription a été effectuée par Yves-Pascal Castel, en 1984 ; une deuxième tentative a été fournie dans le document du concours de la mairie : elles se trouvent toutes deux, commentées, en Annexe 3. Au vu des clichés fournis par la mairie, trouvés sur Internet et aux recoupements faits avec ces deux tentatives précédentes, voici le résultat, avec lequel nous travaillerons tout au long de ce dossier.

La transcription est présentée de la manière suivante : à gauche se trouve l'identifiant du texte, composé d'une lettre désignant le rocher et d'un numéro de ligne. Plus d'informations sur le système de numérotation en Annexe 2. Les tirets bas indiquent une ou plusieurs lettres indéchiffrables ; l'alinéa montre que l'inscription ne commence pas tout à gauche du rocher, mais est (plus ou moins) déportée vers la droite.

- A1 GROCAR
- A2 dREAR DIO2EEVbIO
- A3 ARVREOIE2LA0hVEC
- A4 PEI AbEIEIE20I82E +
- A5 IbICEIG
- A6 _OAVI EKGE D
- A7 A2OMGAROPA γCDO'FET
- A8 dAR OA
- A9 O I E EM __ GEM E JAIEJ
- A10 IVEL AChEODCET DA-AOMA
- A11 CVLES ELdA RE IdIMEV2ME2
- A12 I __R ER
- A13 AR PRIGIL O d11(Cœur orné d'une croix)81
- A14 ObIIE bRI2bVILN EROIAL
- A15 ALVO4 ARbORSIV_T
- A16 CARCLO IVE PRE 2 T
- A17 V2OI RE2 E_____ I
- A18 VA_2 1920
- A19 2dARANdOC
- A20 AdREIRIO
- A21 I186 IEI2

B1 2h_
 B2 AND PIN
 B3 _A_AI
 B4 _A_VET
 B5 _AM_
 C1 ___A
 C2 _ PRET
 C3 OR
 C4 ONE_AI_ _
 D1 __E_
 D2 O
 D3 VET
 D4 I_T
 E1 __OIR
 E2 __RIC
 E3 __R_
 E4 __CE_

Il existe, bien sûr, de nombreux obstacles à une transcription indiscutable.

Tout d'abord, l'existence d'espaces et même de lignes distinctes est quelquefois discutable (voir l. A8). De même, certains signes, ici compris comme appartenant tous à la même catégorie, pourraient en fait être distincts : je ne marque pas de différence entre le o de petite taille centré en hauteur (l. A10) et un O prenant toute la ligne (l. A13) ; ni entre un l (trait vertical sans sérif, l. A2) et I (avec sérif aux extrémités, l. A13) ; ni entre un V bien net et droit (l. A2) et un autre spécimen plus courbé (début de l. A11). Certaines décisions ont dû être prises durant cette transcription, comme différence IV et N : pour ce faire, trois critères ont été utilisés – la jonction entre le I et le V, l'inclinaison des deux caractères, et leur taille⁶.

La question des dates observées est bien épineuse, car je ne retrouve pas les éléments cités dans les précédentes transcriptions. La date autour du cœur en A13 (si c'en est une), me semble être 1181, car tous les 1 disposent d'un sérif en haut tourné vers la gauche, et ne comporte pas de barre médiane. Je retrouve le même caractère en deuxième position l. A21, tandis que le premier n'a pas du tout de sérif : je l'ai donc transcrit I.

6 Ainsi, pour obtenir un N, la jonction entre I et V doit être faite, la première barre du V orientée à 45° et la hauteur du V et celle du I identiques : on peut alors supposer que I et V sont en fait une même lettre, c'est-à-dire un N. Dans le cas contraire (pas de jonction, mauvaise orientation, hauteurs différentes), on ne peut pas inférer la présence d'un N, mais bien d'un I et d'un V (ex. l. A10 et A16).

Ainsi, si les analyses menées sur cette transcription échouent, c'est que les choix que j'ai posés durant ce travail ont été mauvais. Il faudra donc fournir une transcription plus fine, qui prend en compte la taille et l'orientation des caractères, la présence ou non de sérifs. Pour ce faire, un déplacement sur le lieu-même devra sans doute être effectués, pour toucher (comme Saint Thomas) les signes et, ainsi, mieux les comprendre. On pourra notamment se pencher sur certains caractères particulièrement intrigants :

- Le C final de la ligne A3, très anguleux mais sans barre médiane (pour en faire un E) ;
- Le début de la ligne A4 ;
- La possible présence de caractères avant l'inscription de la ligne A5 ;
- L'apostrophe en A7 (hapax) pour simplement être un I appartenant à la ligne A6

I.3.Éléments symboliques de l'inscription

En dehors des chiffres, trois éléments non-alphabétiques sont inclus dans le texte :

- Une croix discrète et non-stylisée (l. A4) ;
- Ce qui est communément caractérisé comme un croissant de lune tourné vers la gauche (l. A6) ;
- Un cœur surmonté d'une croix (l. A13).

Pour ce qui touche les deux premiers symboles, on ne peut certes pas déduire grand-chose, étant donné qu'ils sont très génériques. La croix pourrait être un simple signe + ou suivre un texte religieux (comme, de temps en temps, les croix émaillent le texte dans les bréviaires). Le croissant, s'il s'agit vraiment d'un croissant, pourrait laisser à croire à un élément astrologique/astronomique, sans plus d'indice.

Le symbole le plus éloquent, donc, reste le cœur surmonté d'une croix, assimilé communément au Sacré Cœur de Jésus.

La dévotion envers le cœur du Christ est assez ancienne, mais elle a connu un essor considérable dans la France du XVII^{ème} et XVIII^{ème} siècle, avec des saints tel que Jean Eudes ou Marie Alacoque. La fête du Sacré Cœur, arrivant 19 jours après la Pentecôte, a été instituée en 1670 dans le diocèse de Rennes, puis étendue en France puis à l'Église entière. Dans l'inscription de Plougastel, la représentation du Sacré Cœur est donc congruente avec la lecture traditionnelle des dates 1787 (l. A13) et 1786 (l. A21).

Plus tard, le Sacré Cœur devient l'emblème du soulèvement royaliste Chouan, qui a notamment eu lieu en Bretagne ; ce symbole est par ailleurs resté dans l'emblème actuel de la Vendée. À titre personnel, je considère cette piste comme assez probable, car iconique, même si elle poserait question pour un adepte de la lecture traditionnelle des dates – 1787 et 1786, situées 7 et 8 ans avant les épisodes de Chouannerie. À noter que la lecture des dates sus-citées pose aussi question pour une autre raison : si l'on suppose que la dernière ligne (l. A21) donne la date d'écriture du texte, celle de la ligne A13 désigne donc une époque qui ne s'est pas encore produite. L'on aurait donc affaire à une prévision ou à une prédiction : à moins qu'il ne s'agisse pas là de dates.

Reste enfin la gravure sur le rocher E (si je ne m'abuse), désignée communément par le terme « voilier ». En vérité, la gravure est peu lisible, et montre surtout des lignes droites convergeant vers un même sommet. Pour ne rien faciliter, le rocher semble brisé à la base de la gravure. Ce détail, quoique possiblement critique, reste donc principalement inexploitable.

II. Analyses de l'inscription

II.1. Analyse fréquentielle

Avant de procéder à l'analyse, j'ai retiré deux symboles (lune et cœur, le + étant possiblement typographique) ainsi que la date 1920, très visiblement gravée par un autre auteur. Enfin, j'ai enlevé les tirets bas ajoutés lors de la transcription. Les analyses ont été menées sur la page <https://www.dcode.fr/analyse-frequences> avec ajouts manuels.

Je n'ai pas retiré d'autres signes, comme les symboles typographiques (apostrophe, tiret) ou les chiffres, qui peuvent toujours servir pour chiffrer un message. De fait, leur présence peut modifier la fréquence calculée, d'où les deux variantes corrigées : la première enlève les signes typographiques ; la seconde ces derniers ainsi que les chiffres.

L'inscription semble comporter 25 caractères alphabétiques, 4 chiffres et 3 symboles typographiques, apparaissant selon les fréquences suivantes :

Caractère	Nombre d'occurrence	Fréquence d'occurrences	Fréquence corrigée 1 (sans signes typo)	Fréquence corrigée 2 (sans signes ni chiffres)
E	43	14,38 %	14,53 %	15,03 %
A	33	11,04 %	11,15 %	11,54 %
R	28	9,36 %	9,46 %	9,79 %
I	28	9,36 %	9,46 %	9,79 %
O	25	8,36 %	8,45 %	8,74 %
2	15	5,02 %	5,07 %	5,24 %
V	14	4,68 %	4,73 %	4,90 %
C	12	4,01 %	4,05 %	4,20 %
И	12	4,01 %	4,05 %	4,20 %
L	9	3,01 %	3,04 %	3,15 %
d	8	2,68 %	2,70 %	2,80 %
T	8	2,68 %	2,70 %	2,80 %
M	7	2,34 %	2,36 %	2,45 %
b	7	2,34 %	2,36 %	2,45 %
G	6	2,01 %	2,03 %	2,10 %

P	6	2,01 %	2,03 %	2,10 %
D	5	1,67 %	1,69 %	1,75 %
N	5	1,67 %	1,69 %	1,75 %
1	5	1,67 %	1,69 %	-
h	3	1 %	1,01 %	1,05 %
8	3	1 %	1,01 %	-
Θ	2	0,67 %	0,68 %	0,70 %
J	2	0,67 %	0,68 %	0,70 %
S	2	0,67 %	0,68 %	0,70 %
-	1	0,33 %	-	-
K	1	0,33 %	0,34 %	0,35 %
F	1	0,33 %	0,34 %	0,35 %
+	1	0,33 %	-	-
'	1	0,33 %	-	-
γ	1	0,33 %	0,34 %	0,35 %
6	1	0,33 %	0,34 %	-
4	1	0,33 %	0,34 %	-

En bref

- 32 caractères : 25 alphabétiques, 4 chiffres et 3 symboles typographiques ;
- L'essentiel de l'alphabet est latin et en majuscule, mais présente certains autres caractères : 2, И, d, b, h, Θ et γ ;
- Les 10 caractères les plus fréquentes : E, A, R, I, O, 2, V, C, И, L (73,03 % de fréquence cumulée).

II.2. Algorithme de détermination de voyelles

Plusieurs algorithmes existent pour déterminer, dans un texte, quels caractères tiennent lieu de voyelles ou des consonnes : mais aucun n'est infallible. Le plus connu est sans doute celui de Sukhotine⁷ : j'utilise pour ma part l'algorithme de Mans Hulden OCPb (*Obligatory Contour Principle based*)⁸, mis en ligne par son auteur⁹. Cet algorithme est très efficace – seulement 7 erreurs lors de l'analyse d'un corpus de 503 langues.

⁷ Guy, J. B. M. (1991). Vowel identification: an old (but good) algorithm. *Cryptologia*, 15(3), 258–262.

⁸ Hulden, M. (2017). A phoneme clustering algorithm based on the obligatory contour principle. *Proceedings of The 21st SIGNLL Conference on Computational Natural Language Learning*.

⁹ Disponible à l'adresse <https://github.com/cvocp/cvocp>. Cependant, quelques modernisations ont dû être effectuées à la main pour convenir à la version de Python utilisée (37.4).

J'ai donc utilisé 3 corpus pour mener ces analyses :

1. Un dans lequel j'ai remplacé toutes les incertitudes (tiret bas dans la transcription) par des espaces, ainsi que la date 1920 et la mention des deux symboles (+ exclu) ;
2. Un dans lequel j'ai uniquement laissé les mots sans difficulté de lecture (sans tirets bas) ;
3. Une copie du premier dans laquelle j'ai supprimé les chiffres et symboles typographiques.

Il existe par ailleurs deux variantes de l'algorithme, basées sur des principes légèrement distincts : chaque corpus a donc été analysé selon ces deux variantes – la version originale est ici notée **a**, la version *tier-based* est dite **b**.

On aboutit donc à 6 analyses : 1a et 1b, 2a et 2b, 3a et 3b. J'ai enfin utilisé l'algorithme de Sukhotine (S.), à partir d'un tableau phonotactique (II.3.) à titre de complément.

La multiplication de ces analyses n'est pas anodine : loin de multiplier inutilement les résultats, elle les croise afin d'observer les lettres labiles et celles davantage stables – par ex. F apparaît quelquefois dans les voyelles, d'autres fois dans les consonnes. Cela permet donc de prendre avec plus de recul les résultats obtenus.

Les sorties de chaque analyse sont détaillées en annexe 4, mais un tableau synthétise ci-dessous les grandes lignes de ces résultats. L'algorithme *tier-based* (toutes les analyses b) permet également de différencier, globalement, les consonnes coronales de celles non-coronales, d'où leur indication dans le tableau.

Lettre	1a	1b	2a	2b	3a	3b	S.
E	V	V	V	V	V	V	V
A	V	V	V	V	V	V	V
R	C	C. non-coronale	C	C. coronale	C	C. non-coronale	C
I	V	V	V	V	V	V	V
O	V	V	V	V	V	V	V
ʒ	C	C. coronale	C	C. non-coronale	C	C. coronale	C
V	C	C. non-coronale	C	C. coronale	C	C. non-coronale	C
C	C	C. coronale	C	C. non-coronale	C	C. coronale	C
H	C	C. non-coronale	C	C. coronale	C	C. non-coronale	C
L	C	C. non-coronale	C	C. non-coronale	C	C. non-coronale	C
d	C	C. coronale	C	C. non-coronale	C	C. coronale	C
T	C	C. coronale	C	C. non-coronale	C	C. coronale	C
M	C	C. non-coronale	C	C. coronale	C	C. non-coronale	C

Lettre	1a	1b	2a	2b	3a	3b	S.
b	C	C. coronale	C	C. non-coronale	C	C. coronale	C
G	C	C. coronale	C	C. non-coronale	C	C. coronale	C
P	C	C. coronale	C	C. non-coronale	C	C. coronale	C
D	C	C. non-coronale	C	C. coronale	C	C. non-coronale	C
N	C	C. coronale	C	C. coronale	C	C. coronale	C
1	V	V	V	C. coronale	-	-	C
h	V	V	V	V	V	V	C
8	C	C. non-coronale	C	V	-	-	V
Θ	C	C. coronale	C	C. coronale	C	C. coronale	C
J	C	C. non-coronale	C	C. non-coronale	C	C. non-coronale	C
S	C	C. coronale	C	C. coronale	C	C. coronale	C
-	C	C. coronale	C	C. non-coronale	-	-	C
K	V	C. non-coronale	V	V	C	C. non-coronale	?
F	C	V	C	C. non-coronale	C	C. non-coronale	C
+	V	C. non-coronale	V	C. coronale	-	-	C
'	V	C. coronale	C	V	-	-	?
γ	V	V	V	V	V	V	?
6	V	V	V	C. non-coronale	-	-	?
4	C	C. coronale	C	C. non-coronale	-	-	C

Synthèse des analyses de détermination de voyelles. C : consonne ; V : voyelle. Le vert clair indique que les lettres concernées sont variablement classifiées en coronales ou non-coronales. Le vert foncé indique une variabilité du classement consonne/voyelle.

En bref

- Globalement, il semble que les lettres de voyelles désignent bien des voyelles, et les consonnes des consonnes ;
- Parmi les caractères particuliers, l'algorithme *OCPb* classe h et γ comme étant des voyelles ; Θ, 2, I, b et d comme des consonnes ;
- Les caractères 1, 8, 6, K, F, h, + et ' sont très instables selon l'algorithme, la variante et le corpus utilisés ;
- Au contraire les lettres stables sont : A, E, I, O, γ (voyelles) ; Θ, S, N (consonnes coronales) ; J et L (consonnes non-coronales).

II.3. Analyse phonotactique

L'analyse phonotactique décrit les probabilités que telle lettre soit suivie par telle autre, pour tous les caractères de l'alphabet utilisé.

L'analyse a ici été menée sur deux corpus différents : un comportant les espaces de la transcription proposée plus haut, l'autre n'interprétant comme espaces que les retours à la lignes de l'inscription. Dans tous les cas, les incertitudes ont été remplacées par des espaces. Les résultats obtenus sont en annexe 5 ; les points importants sont les suivants :

- 3 répétitions de lettres : 11, EE et II ;
- Les groupes de voyelles (identifiées en II.3) sont les suivants : EE, EO, EA, EI, EK, OA, OI, AO, (AA), AI, (AY), IE, IO, IA, II, hE ;
- Le bigramme « AR » est étrangement fréquent ;
- La lettre T se trouve toujours en fin de mot.

II.4. Analyse de l'entropie du texte

L'entropie d'un texte est la mesure de sa variabilité et, donc, de la quantité d'information qu'il contient. En effet, un message uniforme, peu variable, n'est que peu informatif ; au contraire, un message informatif est généralement peu uniforme. Un calculateur de la variabilité des lettres se trouve sur <https://planetcalc.com/2476/> : les espaces ont été retirées du calcul, car trop hypothétiques. Cependant, si l'on postule la présence d'espace, l'entropie devra être un peu augmentée.

L'algorithme donne une entropie de premier ordre de 4,19 à partir des données de la partie II.1 : l'entropie moyenne d'une langue naturelle se situe généralement entre 3,91 et 4,14¹⁰. La valeur semble donc normale pour un langage naturel.

À partir des données de la partie II.3., l'entropie de deuxième ordre a pu être également calculée. Étant donné qu'un algorithme adapté n'a pas été programmé, il a été exécuté sur une feuille Excel avec la formule donnée sur la page suivante : <https://www.mdpi.com/1099-4300/10/2/33/htm>. On obtient, avec le corpus avec espaces, une valeur de 2,02, à contraster avec l'intervalle de normalité : 3,01 à 3,37¹¹. Cependant, vu la lourdeur du processus, des erreurs humaines ont pu se glisser dans ce résultat, qui reste à prendre avec du recul – voir III.2.d. Aucune langue naturelle ne semble combiner une entropie de premier ordre forte avec un second ordre si faible¹².

En bref

- L'entropie de premier ordre vaut 4,19, ce qui situe l'inscription dans la normale haute des langages naturels ;
- L'entropie de deuxième ordre semble valoir 2,02, soit bien en-dessous de la valeur normale. Ce résultat doit, malgré tout, être considéré avec prudence, dû à de possibles erreurs de procédure (cf. *infra*).

10 Source : http://voynich.nu/a2_char.html, §2.4.

11 Ibid.

12 Voir graphiques *ibid*.

II.5. Analyse de l'indice de coïncidence

L'indice de coïncidence permet de calculer la probabilité de répétition d'une même lettre au sein d'un texte. Il permet notamment de faire la différence entre un chiffrement mono- et poly-alphabétique (voir infra).

L'indice a été calculé à partir de l'outil disponible sur : <https://www.dcode.fr/indice-coincidence>, pour divers corpus déjà utilisés :

- Corpus 1 (voir II.2), sans incertitudes, symboles, ni 1920 : 0,09037 ;
- Corpus 2, contenant les mots sans difficulté de lecture : 0,0862 ;
- Corpus 3, Corpus 1 d'où ont été retirés les symboles non-alphabétiques : 0,09037 ;
- Corpus d'analyse phonotactique (II.3) : 0,09012 (avec ou sans espaces).

En moyenne, dans une langue naturelle, l'indice de coïncidence se situe dans un intervalle allant de 0,053 à 0,085¹³. Les indices obtenus sont donc légèrement supérieurs à l'intervalle naturel.

En bref

- L'indice de coïncidence se situe entre 0,086 et 0,091 ;
- Cet indice est légèrement supérieur à l'intervalle naturel.

II.6. Autres pistes d'analyse

Plusieurs autres analyses auraient pu être poursuivies, telles que :

- L'analyse de bigrammes et trigrammes ;
- Des statistiques sur les mots utilisés (taille, fréquence d'utilisation – voir loi de Zipf) : cependant, la difficulté inhérente aux espaces incertaines empêche cette procédure ;
- Un calcul de corrélations à long-terme aurait pu être mené si le corpus présenté n'était pas si restreint ;
- Il serait possible, avec plus de moyens et de patience, de mettre au point un modèle caché de Markov, utile ou non dans le cas présent.

III. Tentatives de décodage

III.1. Détermination *a priori* de langues candidates

Sans rien regarder aux données obtenues lors de l'analyse, on peut déterminer *a priori* un panel de langues candidates, c'est-à-dire qui pourraient probablement être la langue de l'inscription :

- Géographiquement, le breton et le français sont deux candidats probables ;
- Attendu la symbolique religieuse du Sacré Cœur, on peut ajouter le latin ;
- Par recoupement d'alphabet, le grec ancien et le russe sont possibles, quoiqu'improbables ;
- Historiquement, l'on doit se situer vers le XVII^{ème} – XVIII^{ème} siècle (époque de vénération du Sacré Cœur), plus spécifiquement la fin du XVIII^{ème} siècle si l'on identifie le cœur orné d'une croix comme un symbole chouan.

13 Source : https://fr.wikipedia.org/wiki/Indice_de_co%C3%AFncidence#Autres_langues.

Afin de tester les possibilités de ces cinq langues, il est nécessaire d'établir un corpus pour chacune de ces dernières, correspondant globalement à l'époque concernée.

Le corpus le plus difficile à trouver a été celui en langue bretonne. En effet, il s'agissait de trouver un document de l'époque, qui soit si possible rédigé en dialecte cornouaillais, préservé autant que possible de l'influence du français. Mon choix, pour des raisons de congruence historique, s'est finalement arrêté sur « A beurs ar Roue », déclaration royaliste de 1795¹⁴. La numérisation de l'ouvrage ne valant rien, je l'ai retranscrit à la main. L'utilisation de la lettre Q, absente de l'inscription, a été supprimée en faveur de la lettre K ; les diacritiques des voyelles et la cédille du C ont été supprimés ; de même, les éléments récurrents (« Art. »), illisibles ou non-bretons (patronymes listées en fin de déclaration) ont été retirés. Le corpus final est donc constitué de 3 005 mots, 13 076 caractères (hors espaces), ce qui est amplement moins que les autres : je n'ai cependant pas pu rallonger le document en y rajoutant les mots d'un deuxième texte, la variabilité risquant d'être trop élevée. Il faut noter, pour terminer sur le sujet du breton, que la graphie de la langue bretonne est soumise à une grande variabilité, à la fois en fonction de l'époque et du lieu¹⁵ : il est même arrivé que, durant la retranscription du corpus, je n'aie pas trouvé la définition de certains mots contenus dans le texte (sans que cela soit dû à des mutations, par exemple).

Le corpus français est extrait des « Arrêts et règlements du Parlement de Bretagne » (chapitre 238 à 254)¹⁶, datant de 1737. Le tout a été retranscrit à la main, d'après une numérisation automatique par Google. J'en ai retiré les nombreuses références à d'autres ouvrages, aux dates de jugement et les parties en latin. Le corpus est donc long de 4 076 mots et contient 19 948 caractères (sans espaces)

Le corpus latin est constitué d'une encyclique de 1753, « Sollicita ac provida » du pape Benoît XIV¹⁷. L'idéal aurait été d'obtenir un texte en latin originaire de l'évêché de Cornouailles, où se trouve Plougastel, mais je n'en ai pas trouvé de numérisé. Les indications néerlandaises du corpus ont été retirées, et les U ont été changés en V, comme le veut l'usage. Ce corpus fait donc 4 060 mots et 26 201 caractères (espaces exclues).

Le corpus russe provient d'un écrit de Mikhaïl Vassilievitch Lomonosoff, « Informations historiques et topographiques sur l'antiquité de la Russie et spécialement de la ville de Voldoga et de son comté »¹⁸, plus spécifiquement les 3 984 premiers mots, dont ont été enlevées les numérotations bibliographiques citées dans le texte, constituant un ensemble de 22 475 caractères (sans espaces). L'intérêt de cet écrit, rédigé en 1777, est son orthographe de l'époque, avec l'utilisation de caractères archaïques supprimés lors de la réforme de 1918.

14 Sources : <https://books.google.fr/books?id=TKJYAAAACAAJ> et https://archive.org/details/bub_gb_TKJYAAAACAAJ.

15 Voir : https://fr.wikipedia.org/wiki/Orthographe_du_breton ; https://wikimonde.com/article/Orthographe_du_breton et <https://abp.bzh/histoire-de-la-langue-bretonne-evolution-de-la-graphie-31513>.

16 Source : <http://fontes-historiae-iuris.univ-lille2.fr/bibliotheque-numerique/recueils-darrets/du-parlement-de-bretagne/>

17 Source : http://www.documentacatholicaomnia.eu/01_01_1740-1758-_Benedictus_XIV.html.

18 [Source](#).

Faute d'avoir pu trouver mieux pour le corpus grec, j'ai sélectionné au jugé un auteur que l'on pourrait qualifier de « classique », étudié et traduit au XVIII^{ème} siècle en France : Hérodote. C'est le premier livre de l'« Histoire », du premier paragraphe au 32^{ème} inclus, qui a servi à constituer notre corpus de référence. Attendu que le texte de l'inscription ne dispose d'aucun diacritique, j'ai transposé l'extrait d'Hérodote en supprimant esprits, accents, et autres diacritiques présents dans la version classique du texte¹⁹ – le iota souscrit a été remplacé par la lettre iota en plein texte. Le corpus est donc finalement constitué de 4 088 mots, 21 455 caractères sans espaces.

Bien entendu, tous ces corpus ne sont qu'une approximation, une première tentative de comparaison guère parfaite : on pourrait citer, dans les œuvres thématiques, la trop grande récurrence de noms propres spécifiques (Jésus pour les écrits pieux, Solon pour Hérodote, etc.). Malgré tout, ce corpus pourra tout de même fournir quelques premiers éléments de réflexion.

III.2. Analyse des langues candidates

N.B. : tous les résultats ci-après sont résumés en un tableau synthétique (p. 18).

III.2.a. Analyse fréquentielle

Les analyses fréquentielles des langues candidates sont disponibles en annexe 6.

Il est tout d'abord possible de comparer la taille de l'alphabet de chaque langue, ou plus exactement le nombre de caractères différents utilisés :

- Inscription : 32 (dont 7 non-alphabétiques) ;
- Breton : 26 (dont 2 non-alphabétiques) ;
- Français : 30 (dont 5 non-alphabétiques) ;
- Latin : 31 (dont 7 non-alphabétiques) ;
- Russe : 48 (dont 11 non-alphabétiques) ;
- Grec : 25 (dont 1 non-alphabétique).

Par ailleurs, j'ai comparé la distribution fréquentielle au sein de l'alphabet de chaque corpus. Par exemple, dans l'inscription, la lettre la plus fréquente apparaît 14,38 % du temps, tandis qu'en russe, c'est 11,53 % : il y a une différence de 2,85 points. De même pour la deuxième lettre la plus fréquente, etc. Ainsi, plus la différence (absolue) cumulée entre ces pourcentages est grande, plus il est improbable que la langue de l'inscription soit effectivement celle de la langue candidate. Mes calculs mènent donc aux résultats suivants :

- Entre l'inscription et le corpus breton, il y a 17,31 points de différence en terme de répartition ;
- Entre l'inscription et le français, 30,69 points ;
- Le latin, 28,24 points ;
- Le russe, 31,22 ;
- Le grec, 20,87.

19 Ici : <http://remacle.org/bloodwolf/historiens/herodote/ungrechr.htm>.

Néanmoins, comparer la courbe de répartition de fréquence ne suffit pas : on peut également comparer la fréquence de répartition de chaque lettre en particulier (celle du E dans l'inscription contre celle dans le corpus russe, etc.). Là encore, il est possible d'aboutir à une différence cumulée. J'ai donc comparé chaque lettre se ressemblant graphiquement : ainsi, j'ai pu comparer le P à la lettre russe P (correspondant pourtant au son /r/). J'ai divisé la fréquence cumulée obtenue sur le nombre de caractères comparés (car elle augmente logiquement au fur et à mesure que l'on compare des paires de caractères). On obtient donc :

- Breton : 1,51 point par caractère (ppc) ;
- Français : 2,10 ppc ;
- Latin : 1,87 ppc ;
- Russe : 3,44 ppc ;
- Grec : 2,12 ppc.

III.2.b. Détermination de voyelles

Les mêmes méthodes que pour l'inscription ont été utilisées : l'algorithme OCPb et celui de Sukhotine. Les résultats sont précisés en annexe 7, avec quelques emphases pour les points importants.

Un indicateur de similarité reste le ratio voyelles/caractères, variant en fonction de l'alphabet, et donc des langues :

- Ratio de l'inscription : 6/32 ;
- Breton : 7/29 ;
- Français : 7/28 ;
- Latin : 6/27 ;
- Russe : 13/42 ;
- Grec : 8/27.

Au-delà de cet indicateur, on peut examiner de plus près les sorties afin de déterminer si chaque consonne produit bien le son que l'on attend d'elle. Ce que cela signifie, c'est que si une lettre X est classifiée en consonne non-coronale, tandis que toutes les autres langues classifient le son qu'elle produit comme coronal, il y a fort à parier que le son qu'elle évoque n'est en fait pas le son auquel on s'attend – ce pourrait être la marque d'une substitution monoalphabétique, par exemple.

Avant, donc, de passer en revue les correspondances coronales/non-coronales, rappelons que la signification de « coronal » pour l'algorithme OCPb n'est pas exacte au sens linguistique, mais que ce moyen de catégorisation peut tout de même de différencier deux groupes de consonnes, que l'on caractérise comme coronales ou non-coronales faute de trouver une autre appellation.

Ainsi donc :

- La lettre L est classifiée en non-coronal, comme toutes les lettres du corpus donnant le son /l/ ou apparenté ;

- La lettre N est classifiée en coronal pour l'inscription, tandis que les autres langues voient le son /n/ comme non-coronal ;
- La lettre Θ est dite coronale, comme le sont les phonèmes /f/ (russe), /t^h/ et /t/ (grec) ;
- Les classification de J (pour les sons /ʒ/ et /j/) et S sont compatibles avec certaines langues et incompatibles avec d'autres ;
- Globalement, le code C. 1. semble désigner un son coronal et C. 2 un son non-coronal – les lettres suivantes vont en ce sens : P, C, d (sauf en français), T (sauf en russe), b, G (sauf en latin), R. Trois lettres vont à l'encontre de cette conclusion : V, M et D (sauf en français).
- La lettre 2 est classifiée comme voyelle par l'algorithme de Sukhotine, ce qui me semble être une erreur due à la faiblesse de l'algorithme, attendu que cette classification ne se retrouve aucunement dans les sorties de l'algorithme OCPb. Rappelons que l'algorithme de Sukhotine identifie d'ailleurs souvent des consonnes comme voyelles dans les langues candidates, comme le N (français et latin), le V et le K (en français). On peut donc raisonnablement faire fi de cette incongruence et classifier 2 comme une consonne.

Cet examen des sorties mène donc globalement à une confirmation que les consonnes donnent bien les sons que l'on suppose être leurs, malgré une incongruence (lettre N, V, M et D) et de nombreuses incertitudes (lettre J, S).

En bref

- L'hypothèse selon laquelle les consonnes désignent bien les sons qui sont les leurs semble partiellement confirmée par les sorties OCPb.

III.2.c. Analyse phonotactique

Les tableaux phonotactiques de chaque corpus sont donnés en annexe 8. Ici, nous ne nous intéresserons qu'à quelques caractéristiques phonotactique, bien qu'il serait possible de mener de nombreuses autres analyses, plus profitables quoique plus complexes.

Étudions tout d'abord les paires de voyelles possibles dans chacune des langues étudiées. Voici d'abord un tableau récapitulant les paires de voyelles observées dans le texte du rocher (à partir du corpus avec espaces) :

Première lettre	Deuxièmes lettres possibles				
A	-	-	I	O	-
E	A	E	I	O	K
I	A	E	I	O	-
O	A	-	I	-	-
h	-	E	-	-	-

Paires de voyelles observées sur le rocher de l'anse du Caro. Le tableau se lit ligne par ligne, de la manière suivante : « À la suite de la lettre A, on peut trouver un A, un O, ou un y ; à la suite de E, un A, un E, etc. »

Suivent donc les tableaux des langues candidates, avec en vert les différences par rapport au tableau de l'inscription.

Corpus breton :

Première lettre	Deuxièmes lettres possibles				
A	A	E	I	O	U
E	A	E	I	O	U
I	A	E	-	O	U
O	A	E	I	-	U
U	A	E	I	O	U

→ 13 différences

Corpus français :

Première lettre	Deuxièmes lettres possibles				
A	-	-	I	O	U
E	A	E	I	O	U
I	A	E	I	O	U
O	-	E	I	-	U
U	A	E	I	O	-
K	A	E	-	-	-

→ 12 différences

Corpus latin :

Première lettre	Deuxièmes lettres possibles				
A	-	E	I	-	V
E	A	-	I	O	V
I	A	E	I	O	V
O	-	E	-	O	V
V	A	E	I	O	V

→ 15 différences

Corpus russe :

Première lettre	Deuxièmes lettres possibles									
А	А	Е	И	О	И	У	-	Й	-	-
Е	А	Е	-	О	И	-	-	Й	-	-
И	А	Е	И	О	И	У	Ъ	Й	-	-
О	А	Е	-	О	И	У	-	Й	-	-
И	-	Е	-	-	И	У	-	-	-	-
Ъ	-	-	-	-	И	-	-	-	-	-
У	-	Е	-	-	И	-	Ъ	Й	-	-
Ы	-	Е	-	-	-	-	-	Й	-	-
Ъ	-	Е	-	-	И	-	-	Й	-	-
Ь	-	Е	-	-	И	-	Ъ	-	-	-
Й	-	-	-	-	-	-	-	-	-	-
Э	-	-	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-	1	8
8	-	-	-	-	-	-	-	-	-	-

→ 36 différences

Corpus grec :

Première lettre	Deuxièmes lettres possibles						
Α	-	Ε	Ι	Ο	-	Υ	-
Ε	Α	Ε	Ι	Ο	Η	Υ	Ω
Ι	Α	Ε	Ι	Ο	Η	-	Ω
Ο	-	Ε	Ι	Ο	Η	Υ	Ω
Η	-	Ε	Ι	Ο	-	-	-
Υ	Α	Ε	-	Ο	Η	-	Ω
Ω	-	Ε	Ι	-	-	Υ	-

→ 23 différences

Cette comparaison n'est bien sûr qu'un moyen assez primitif d'observer les similarités et différences entre les textes étudiés. Il faudrait, pour bien faire, comparer les fréquences d'apparition des groupes de voyelles, et prendre en compte la difficulté à lire les espaces sur l'inscription. En vérité, même le comptage des erreurs effectué dans les tableaux ci-dessus est en soi débattable : mais cela reste un premier moyen de comparaison.

Par ailleurs, il nous faut examiner les lettres doubles existantes :

- L'inscription compte 1, A, E et I comme lettres répétables (soit 4/32) ;
- Le corpus breton présente les lettres comme répétables E, A, N, R, T, U, S, L, C, M, P et F (12/26), mais il s'agit essentiellement d'emprunts au français ;
- Le corpus français contient E, R, S, N, I, T, D, L, C, P, M et F doublées (12/29) ;
- Le corpus latin double les lettres I, O, V, R, C, L, D, T, M, P, N, S et F (13/32) ;
- Le corpus russe les lettres O, И, А, Е, І, Н, С, В, Т, Л, М, К, Д, П, Я, Ж, et les chiffres (17/37) ;
- Le corpus grec E, I, O, N, Σ, T, P, Λ, Π, K, M et Γ (12/25).

III.2.d. Entropie

L'entropie de premier ordre a été calculé pour tous les corpus :

- Pour l'inscription (rappel) : 4,19 ;
- Corpus breton : 3,99-4,09 ;
- Corpus français : 4,32-4,41 ;
- Latin : 4,12-4,14 ;
- Russe : 4,65-4,74 ;
- Grec : 4,14-4,17.

Idem pour l'entropie de second ordre :

- Inscription (rappel) : 2,02 ;
- Breton : 2,23 ;
- Français : 2,16 ;
- Latin : 2,25 ;
- Russe : 2,33 ;
- Grec : 2,22.

En bref

- L'entropie de deuxième ordre calculée est en fait seulement un peu plus basse dans l'inscription que dans les corpus étudiés, ce qui indique un défaut de calcul ou de référence. Malgré tout, sa valeur reste inférieure à celle des autres corpus.

III.2.e. Indice de coïncidence

Ce paramètre a également été calculé :

- Inscription (rappel) : 0,0862-0,0904 ;
- Breton : 0,0734 ;
- Français : 0,0802 ;
- Latin : 0,0731 ;
- Russe : 0,0514 ;
- Grec : 0,0710.

Au terme de cette longue partie d'analyse, il est possible de résumer les résultats obtenus en un tableau :

En bref					
Critère/Langue candidate	Breton	Français	Latin	Russe	Grec
Nombre de caractères	3	2	1	4	5
Nombre de caractères (corrigé)	3	1	3	4	3
Fréquence des lettres 1 à 1	1	2	3	5	4
Distribution des fréquences	1	4	3	5	2
Ratio voyelles/caractères	3	2	1	5	4
Couples de voyelles	2	1	3	5	4
Répétabilité des lettres observées	3	3	5	1	3
Propension à la répétition	4	2	1	3	5
Entropie de premier ordre	3	4	2	5	1
Entropie de deuxième ordre	3	1	4	5	2
Indice de coïncidence	2	1	4	5	3
MOYENNE	2,55	2,09	2,72	4,27	3,27

Rang de ressemblance des langues candidates au texte de l'inscription en fonction des divers paramètres étudiés. Plus le rang est proche de 1, plus la langue désignée ressemble à celle de l'inscription. Les demis rangs correspondent à des cas d'ex æquo.

Aucune des langues candidates n'apparaît, au terme de cette comparaison, comme avantagée par rapport aux autres : il semble cependant que le grec et, surtout, le russe soient de piètres prétendants. Quelle que soit la langue candidate, reste à expliquer la distribution fréquentielle, le faible nombre de lettres doubles, la faible entropie de second ordre, et l'indice de coïncidence très élevé. Est-ce un effet de la taille du corpus ?

III.3. Diverses tentatives sans modification

Cette partie est dédiée aux tentatives d'exploration sans modification de lettres : il s'agit en fait de prendre les mots en alphabet latin sans difficulté de lecture ni anomalie (caractère spécial, minuscule, etc.) et d'en chercher sur le net (DuckDuckGo, Wikipedia, Wiktionary, entre autres) les significations possibles – en privilégiant, bien entendu, les langues candidates. Les recherches en alphabet cyrillique et grec, faisant intervenir des modifications d'alphabet, sont de fait catégorisées dans la partie suivante.

Voici donc les résultats de mes recherches :

- GROCAR (A1) : rien
- EKGE (A6) : rien.
- AOMA (A10) : langue édoïde du Sud-Est du Nigeria (aussi appelée « ivbiosakon ») ; prénom féminin.
- CVLES (A11) : deuxième personne du singulier, présent, du verbe latin cvlo (faire accoupler).
- AR (A13) : préposition ou article défini en breton (remplace « an » dès 1650) ; particule présente en de nombreuses langues, comme le basque, l'irlandais, le vieil anglais, etc.
- PRIGIL (A13) : prénom indien.
- EROIAL (A14) : rien.
- CARCLO (A16) : rien.
- AND (B2) : mot présent en anglais, vieil anglais, irlandais, etc.

En bref

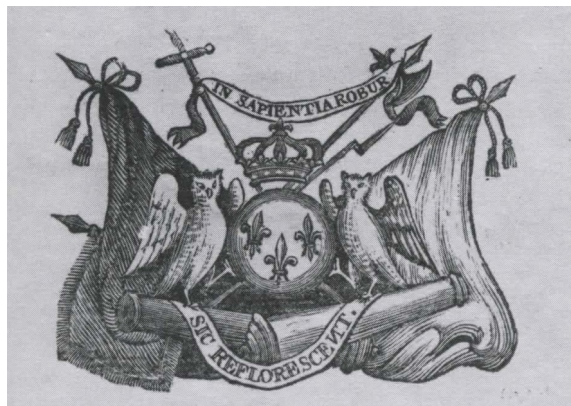
- Les recherches à partir du texte en clair n'aboutissent à aucun résultat net, allant en faveur de l'hypothèse du chiffrement ou de l'usage d'une langue peu connue.

III.4. Chiffrement par simple modification alphabétique

III.4.a. Simple renversement de lettres

Attendu la présence du s renversé (Z) dans les caractères de l'inscription du rocher, j'ai tendance à interpréter de même *И* comme un N renversé, et non un caractère cyrillique.

En m'intéressant à la symbolique du Sacré Cœur et donc à la Chaounerie, j'ai trouvé l'image, au demeurant banale, des armoiries royalistes dans un Almanach de 1795 :



Intéressé par cette curieuse répétition, j'ai consulté un archiviste par rapport à l'existence de caractères inversés, spécifiquement à ce N inversé de l'Almanach. Il m'a dit qu'il s'agissait d'une erreur banale et anodine, probablement due à un facteur humain (inversion d'un caractère mobile par oubli, négligence ou analphabétisme). Ledit archiviste m'a indiqué spontanément que cela existait également sur support gravé, là encore dû à un facteur humain – généralement l'analphabétisme du graveur qui suivait distraitement son modèle. Je l'ai interrogé sur la fréquence de ce genre de phénomènes, qu'il m'a affirmé très basses – il n'a pas pu me citer d'exemple connu où l'on ait pu trouver une telle inversion.

Je restais tout de même curieux de trouver une telle répétition lors de mes recherches. Se pouvait-il que ce N inversé ait une quelconque signification dans le contexte des révoltes royalistes de la fin du XVIII^{ème} ? Cette question a abouti lors de ma transcription du texte breton « A beurs ar Roue », qui comportait les mêmes armoiries, mais sans cette inversion du N. Il semble donc qu'il faille m'en tenir aux dires de mon archiviste et considérer, probablement, les lettres inversées comme des éléments anodins et non-significatifs. Ce fut d'ailleurs le choix de Yves-Pascal Castel, auteur de la première transcription du rocher de l'Anse.

Pour évaluer la probabilité de ce phénomène, l'on peut se pencher sur les caractéristiques phonotactiques des lettres renversées, et les comparer à leurs équivalents non-inversés :

- Les placements phonotactiques de N et И diffèrent de 130,91 points ;
- S et 2 de 201,65 points ;
- b et d : 182,27 points ;
- À titre de comparaison, certains couples de lettres non-apparentées ont été comparés, et l'écart moyen vaut 326,48 points. Cependant, certains couples peuvent avoir une différence assez faible (e.g. L & M : 119,52).

Si cet indice phonotactique reste donc à raffiner, le comportement des lettres inversées et droites est assez comparable d'un point de vue phonotactique, ce qui ne va pas à l'encontre de l'hypothèse du renversement.

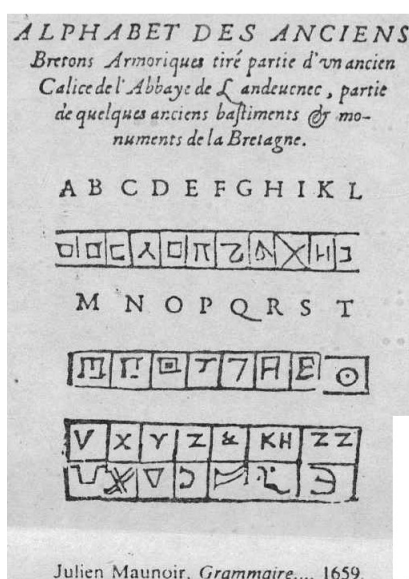
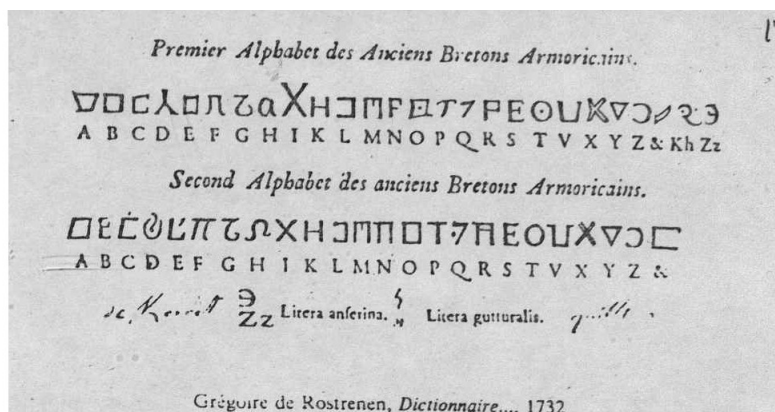
Cependant, le renversement de telles lettres ne peut suffire à rendre intelligible des mots comme V2OИ (A17) qui, recherchés sous la forme VSON (voire USON) ne donnent rien. De même, si le renversement de lettres suffisait, comment expliquer l'échec des recherches en clair ?

En bref

- Le renversement des lettres est une hypothèse possible mais qui ne peut suffire à décrypter l'inscription.

III.4.b. Alphabet armoricain

Dans le communiqué de la Société Archéologie du Finistère²⁰, l'inscription de Guengat est mentionnée, sans plus de précision. Il s'agit en fait d'un article d'Yves-Pascal Castel sur une inscription de l'église de Guengat²¹, qui mêle des caractères de l'alphabet latin à d'autres, plus méconnus. Cet article, intitulé « Ésotérisme ? L'Alphabet Armoricain, une curiosité peu connue », présente donc une inscription de Guengat, en alphabet armoricain, accompagnées les alphabets bretons détaillés par Le Pelletier, Rostrenen et Maunoir dans leurs Dictionnaire et Grammaire respectifs. Les dites alphabets sont les suivants :



Alphabets Armoricains cités par Y.-P. Castel dans l'article éponyme

Servant possiblement de source les uns aux autres, ces alphabets se recourent de manière satisfaisante. Ce qui pose plutôt question, c'est que de tels symboles sont peu fréquents dans l'inscription de Plougastel : il doit alors s'agir d'une substitution partielle.

20 <https://societe-archeologique.du-finistere.org/inventaires/plougastel-daoulas-inscription-anse-caro.html>.

21 Il est question d'une autre inscription, du même auteur, au même lieu, sur cette page : <http://www.guengat.com/8/eglise04.html>.

Il nous faut donc travailler à partir de l'inscription et y identifier des caractères armoricains afin de les rendre latins :

- Les lettres C et M ne changent pas.
- Θ pourrait être un O ou un T, mais l'algorithme OCPb classe ce caractère comme une consonne : il faut donc privilégier la substitution pour T ;
- E donnerait S et B, mais ce n'est pas compatible avec la sortie de l'algorithme OCPb ;
- T pourrait être un P ;
- A pourrait être un R, mais là encore, c'est incompatible avec les résultats précédents ;
- Idem pour O, pouvant donner T ;
- Notre J pourrait fort bien être un L ;
- En B3, l'on pourrait, avec un peu d'imagination voir une sorte de Π, pouvant symboliser un N ou F.

Les autres lettres seraient de l'alphabet latin.

Les modifications seraient donc les suivantes :

- Θ → T :
 - ARVREOIE2LAΘhVEC (A3) devient ARVREOIE2LATHVEC ;
 - AbEIEIE2ΘI82E (A4) devient AbEIEIE2TI82E ;Ces deux mots, même avec retournement de 2 et I, ne fournissent aucun résultat, sans doute à cause de la difficulté à déterminer la présence d'espaces.
- T → P : cela pourrait expliquer pourquoi T n'existe qu'en fin de ligne, puisqu'il est en fait une variante orthographique de P.
 - AChEODCET (A10) donne donc AchEODCEP ;
 - PRET (C2) donne PREP, mot moyen irlandais signifiant « début, rebond » ;
- J → L : JAIEJ (A9) donne LAIEL (prénom utilisé aux États-Unis).
- Le mot en A3, pouvant être lu ΑΠΑΙ, pourrait donner :
 - ΑFAΙ : « afañ » signifie « framboise » en gallois, « afañ » embrasser en breton ; « afán » mot espagnol pour « désir » ; c'est aussi le nom de divers éléments (saint et rivière gallois, village camerounais ou iranien) ;
 - ΑΡΑΙ : « apan » (ou « apán » en hongrois) est également présent dans de nombreuses langues et désigne également un village mexicain.

Dans tous les cas, ces substitutions ne règlent pas les problèmes rencontrés dans le texte en clair (voir III. 3).

À noter, également, que certaines inscriptions semblent relever partiellement de l'alphabet armoricain mais ne pas pouvoir être totalement lues par les grilles de lecture fournies plus haut. C'est notamment le cas de l'inscription du pilier de Saint Tromeur, situé dans la paroisse de Sainte Triphine²².

22 La source la plus complète à ce sujet se trouve sur : <http://marikavel.org/bretagne/sainte-trephine/accueil.htm>.

En bref

- L'alphabet armoricain pourrait être utilisé dans l'inscription du rocher, mais certaines de ses équivalences sont contradictoires avec l'algorithme OCPb (O→T, e.g.). Plus globalement, cet alphabet ne suffit pas à décrypter le texte.

III.4.c. Alphabet grec

Comme annoncé plus haut, nous nous sommes basés pour l'heure sur l'alphabet latin. Mais deux langues candidates utilisent d'autres alphabets : le grec et le russe. Ainsi, cette partie examine la possibilité d'une transposition du texte en alphabet latin vers l'alphabet grec – la partie suivante fera de même avec le cyrillique russe.

Pour obtenir l'équivalence la plus probable possible, nous nous basons sur plusieurs indicateurs (classés ici par ordre d'importance) :

- La ressemblance graphique, prenant en compte les variantes épigraphiques ;
- Le classement par l'algorithme OCPb ;
- La fréquence d'apparition de chaque lettre.

Suit donc le tableau d'équivalence puis le corpus transcrit en alphabet grec. Les éléments difficilement conciliables avec les données et les points d'hésitation sont indiqués avec un point d'interrogation. D'abondantes notes de pied de page permettent de mieux expliquer les choix effectués.

Caractère	Équivalent grec	Caractère	Équivalent grec
E	E ε	b	Δ δ (d renversé) ²³ ?
A	A α	G	Γ γ ²⁴
R	P ϱ ²⁵ ?	P	Π π ²⁶ ?
I	I ι (et/ou Z ζ ²⁷ en fonction des sérifs ?)	D	P ϱ ²⁸ ?
O	O ο	N	N ν
ç	Σ σ/ς ²⁹	h	H η

23 Préféré à B β (lettre peu fréquente). Pour raison de divergence par rapport à l'algorithme OCPb, on a également supprimé A α (https://en.wikipedia.org/wiki/File:Greek_minuscule_Alpha.svg).

24 Préféré à Σ σ/ς à cause de sa fréquence.

25 Certaines variantes régionales ajoutaient déjà le pied du R : <https://fr.wikipedia.org/wiki/Rh%C3%B4>.

26 P et R ne peuvent symboliser tous deux la lettre P, puisque l'algorithme ne les classe pas dans la même catégorie.

27 <https://fr.wikipedia.org/wiki/Z%C3%A0ta#D%C3%A9riv%C3%A9s>.

28 Privilégié à Δ δ pour des raisons fréquentielles, mais on peut débattre de ce choix.

V	N v ³⁰ ?	Θ	Θ θ
C	Σ σ/ς ³¹⁻³² ?	J	?
И	N v ?	S	Σ σ/ς
L	Λ λ ³³⁻³⁴	K	Κ κ
d	Δ δ ³⁵ ?	F	Φ φ/ϕ ³⁶ ?
T	T τ	γ	Υ υ ³⁷ ?
M	M μ ³⁸	8	Ω ω ³⁹
4	P ρ ⁴⁰⁻⁴¹	+	Ξ ξ ⁴²

Cette équivalence considère donc les caractères suivants comme extra-alphabétiques : ‘, 6, 1, -. Il faut noter que le tiret et les chiffres arabes ne sont guère utilisés en grec classique.

Les lettres suivantes de l’alphabet grec ne figurent pas dans le tableau d’équivalence : B, Z, X et Ψ. Cependant, leur fréquence d’occurrence dans le corpus grec est inférieure à 1 %.

Globalement, on appréciera la difficulté de dresser un tableau d’équivalence, et le caractère très éclectiques des variantes épigraphiques utilisées. Comment expliquer une telle variabilité et l’usage de ces variantes archaïques et régionales en Bretagne, à une époque où le Sacré-Cœur était connu ? Contextuellement, ces éléments posent des difficultés : mais les données priment. Cela signifie que si l’équivalence proposée est concluante et donne un texte doté de sens, il faudra bien se résoudre à accepter cette hypothèse. Le texte décrypté pourrait d’ailleurs contenir des explications.

29 Préféré à Z ζ à cause de sa fréquence d’occurrence élevée. I ι (voir https://commons.wikimedia.org/wiki/Greek_epigraphic_letters) a été écarté, conformément à l’algorithme OCPb.

30 Υ υ a été écarté car V est classifié comme consonne.

31 Voir la forme lunaire, apparue au Moyen-Âge : <https://fr.wikipedia.org/wiki/Sigma#Typographie>.

32 Privilégié à K κ et B β (https://fr.wikipedia.org/wiki/Alphabets_grecs_archa%C3%AFques#Tableau_r%C3%A9capitulatif) pour raisons fréquentielles.

33 [https://fr.wikipedia.org/wiki/Nu_\(lettre_grecque\)#D%C3%A9riv%C3%A9s](https://fr.wikipedia.org/wiki/Nu_(lettre_grecque)#D%C3%A9riv%C3%A9s).

34 Préféré à F f (https://fr.wikipedia.org/wiki/Alphabets_grecs_archa%C3%AFques#Digamma_pamphylien), caractère plus régional.

35 Privilégié à A α (https://en.wikipedia.org/wiki/File:Greek_minuscule_Alpha.svg), incongruent avec la classification OCPb.

36 Préféré à F f (<https://fr.wikipedia.org/wiki/Digamma>) caractère plus archaïque et régional.

37 Privilégié par rapport à Γ γ et X χ (<https://fr.wikipedia.org/wiki/Chi#%C3%89volution>) pour sa classification en tant que voyelle.

38 Préféré à M μ ([https://fr.wikipedia.org/wiki/San_\(lettre_grecque\)](https://fr.wikipedia.org/wiki/San_(lettre_grecque))), caractère moins classique.

39 <https://fi.wikipedia.org/wiki/Omega#Variantteja>.

40 Voir graphie étrusque : <https://fr.wikipedia.org/wiki/Rh%C3%B4>.

41 Préféré à A α à cause d’OCPb.

42 https://fr.wikipedia.org/wiki/Xi?uselang=fr#Alphabets_archa%C3%AFques

Voici donc la transcription transposée suivant le tableau ci-dessus :

- A1 ΓΡΟΣΑΡ
- A2 ΔΡΕΑΡ ΡΙΟΣΕΕΝΔΙΟ
- A3 ΑΡΝΡΕΟΝΕΣΛΑΘΗΝΕΣ
- A4 ΠΕΝ ΑΔΕΝΕΝΕΣΘΙ8ΣΕ Ξ
- A5 ΝΔΙΣΕΝΓ
- A6 _ΟΑΝΙ ΕΚΤΕ Ɔ
- A7 ΑΣΟΜΓΑΡΟΠΑ ΥΣΡΟΨΕΤ
- A8 ΔΑΡ ΟΑ
- A9 Ο Ι Ε ΕΜ __ ΓΕΜ Ε ?ΑΙΕ?
- A10 ΙΝΕΛ ΑΣΗΕΟΡΣΕΤ ΡΑ-ΑΟΜΑ
- A11 ΣΝΛΕΣ ΕΛΔΑ ΡΕ ΙΔΙΜΕΝΣΜΕΣ
- A12 Ι __Ρ ΕΡ
- A13 ΑΡ ΠΡΙΓΙΛ Ο Δ11Ω1
- A14 ΟΔΠΕ ΔΡΙΣΔΝΙΑΝ ΕΡΟΙΑΛ
- A15 ΑΛΝΟΡ ΑΡΔΟΡΣΙΝ_Τ
- A16 ΣΑΡΣΛΟ ΙΝΕ ΠΡΕ Σ Τ
- A17 ΝΣΟΝ ΡΕΣ Ε_____ Ι
- A18 ΝΑ_Σ
- A19 ΣΔΑΡΑΝΔΟΣ
- A20 ΑΔΡΕΙΡΙΟ
- A21 Ι1Ω6 ΝΕΙΣ
- B1 ΣΗ_
- B2 ΑΝΡ ΠΙΝ
- B3 _Α_ΑΝ
- B4 _Α_ΝΕΤ
- B5 _ΑΜ_
- C1 ___Α
- C2 _ ΠΡΕΤ
- C3 ΟΡ
- C4 ΟΝΕ_ΑΝ_ _
- D1 __Ε_

D2 O
D3 NET
D4 I_T
E1 __ONP
E2 __PIΣ
E3 __P_
E4 __ΣE_

Par conséquent, voici quelques essais de traduction pour certains mots isolés – ceux contenant des caractères avec incertitude d'équivalence sont indiqués en italique :

- ΓΡΟΣΑΡ (A1) : rien.
- ΕΚΤΕ (A6) : rien. La préposition εκ marque un mouvement vers l'extérieur, comme le latin *ex*.
- ΑΟΜΑ (A10) : rien.
- ΣΝΛΕΣ (A11) : rien (et difficilement prononçable).
- ΑΡ (A13) : αϝ' peut-être vu comme l'élision de ἄϝα, mot de liaison très polysémique (or, donc, puis, etc.).
- ΠΡΙΓΙΑ (A13) : rien.
- Ο (A13) : ό, article défini masculin nominatif.
- ΟΔΙΕ (A14) : rien.
- ΕΡΟΙΑΛ (A14) : rien.
- ΣΑΡΣΛΟ (A16) : rien.
- ΝΕΙΣ (A21) : peut-être rapproché de νέος, « nouveau ».
- ΑΝΡ (B2) : rien.
- ΠΙΝ (B2) : rien.
- La transcription, proposée à Google Traduction, ne donne rien d'intelligible, même par fragments.

En bref

- L'équivalence proposée avec l'alphabet grec, pourtant optimisée au niveau paramétrique, ne semble pas convenir, malgré quelques résultats encourageants. Même les mots sans grand difficulté de transcription (ΕΚΤΕ, ΑΟΜΑ, ΝΕΙΣ) restent inintelligibles.

III.4.d. Alphabet cyrillique ancien

Il s'agit ici de procéder de même que dans la partie précédente. Pourtant, *a priori*, plusieurs éléments rendent cette hypothèse improbable : les faibles scores du russe lors de la comparaison de langues candidates ; le caractère И unanimement catégorisé comme

consonne (Й, lettre trop spécifique, n'est pas un bon substitut) ; l'absence d'apostrophe en cyrillique ; l'impossibilité à faire de b et d des знаки, i.e. des signes spéciaux modifiant la prononciation des consonnes précédentes – b et d étant parfois en début de mot.

Suit donc la table d'équivalence, ajustée en fonction des trois critères sus-cités :

Caractère	Équivalent cyrillique	Caractère	Équivalent cyrillique
E	Е е (et Э э ?)	b	В в ⁴³ ? Б б ?
A	А а (et Я я ?)	G	Г г
R	Р р	P	П п
I	И и (et И и ?)	D	Д д
O	О о	N	Н н
ʒ	З з	h	Ч ч ⁴⁴
V	В в	Θ	Θ θ
C	С с	J	Ж ж
И	И и ⁴⁵	S	С с ?
L	Л л	K	К к
d	Б б ⁴⁶ ? В в ?	F	Ф ф
T	Т т	ʎ	ʎ ⁴⁷⁻⁴⁸
M	М м	bɪ	Ы ы ?
4	Ц, ц	Ю	Ю ю ?

Les caractères 1, 6, 8, -, ' et + de l'inscription sont considérés comme extra-alphabétiques. Les caractères У у, Ш ш, Щ щ, Ю ю, Х х, Ъ ъ, Ь ь, et Ъ ѡ manquent à l'appel dans le tableau de correspondance – voir aussi tous les caractères archaïques⁴⁹.

On remarque déjà l'improbabilité d'une telle équivalence, avec l'absence totale de signes permettant de distinguer les deux modes de prononciation propres au russe : le premier dit « mou » (c'est-à-dire palatalisé, prononcé à l'avant de la bouche et en collant la langue au palais), marqué par le signe ь et certaines voyelles comme е, я, и, ю (et ë depuis 1793) ; au contraire, le second est dit « dur » (c'est-à-dire vélarisé, prononcé au fond de la bouche), marqué par le ъ et les voyelles équivalentes aux premières : э, а, ы, у, о. Ici, on n'a aucun marqueur de vélarisation ou palatalisation.

43 Ъ ъ, Ь ь, et Ъ ѡ ont été écartés, car classifiés comme voyelles par ОСПб.

44 Privilégié à Н н en raison de la fréquence d'apparition.

45 C'est un choix extrêmement discutable, mais cette équivalence a été préférée à Н н pour des raisons graphiques, bien que le symbole de l'inscription soit classifié comme une consonne. On pourrait arguer

46 Préféré à В в en raison de la forme cursive de la lettre б. Ъ ъ, Ь ь, et Ъ ѡ ont été écartés, car classifiés comme voyelles par ОСПб.

47 [https://fr.wikipedia.org/wiki/Ouk_\(lettre_cyrillique\)](https://fr.wikipedia.org/wiki/Ouk_(lettre_cyrillique)).

48 Préféré à Х х, classifié comme une consonne.

49 https://fr.wikipedia.org/wiki/Alphabet_cyrillique#Alphabet_originel.

En l'absence, donc, de ce double système, il me semble compromis de fournir une transcription complète comme cela a été le cas avec l'alphabet grec. Il est plus adapté de focaliser les essais sur quelques mots – ceux déjà recherchés.

Un dictionnaire d'ancien orthographe⁵⁰ a été utilisé pour les mots suivants :

- GROCAR (A1) : aucun mot ne débutant par грo... et faisant cette taille ou moins n'est disponible sur le dictionnaire.
- EKGE (A6) : idem pour эк... et ек... (spécifiquement экре et екре).
- АОМА (A10) : la combinaison ао... n'existe pas en début de mot.
- АR (A13) : le mot ар seul, ou avec un signe modificateur (знак) derrière.
- PRIGIL (A13) : пригі..., приги..., пригі... et приги... ne donnent rien.
- О (A13) : о seul n'existe pas.
- ОбИЕ (A14) : rien pour обии..., обие..., оби..., овии..., овие..., et овii...
- EROIALE (A14) : rien pour ерои... ou ероi...
- CARCLO (A16) : rien pour сарсл...
- ИЕИ2 (A21) : иеи... ou иеi...
- AND (B2) : pas de mot анд ou аид.
- PIN (B2) : pas de mot рин ni pi...

En bref

- L'équivalence proposée avec l'alphabet russe est lamentable : il est difficile d'établir une correspondance nette, notamment avec les fréquences (souvent très divergentes), et les essais de transcription n'aboutissent pas.

III.4.e. Autres pistes alphabétiques

Nous n'avons essayé ici qu'une poignée d'hypothèses, mais d'autres sont toujours possibles, faisant intervenir d'autres alphabets et/ou langues :

- En Europe :
 - Θ est une abréviation latine utilisée au Moyen-Âge pour désigner certains mots débutant par un o, comme *obitus* ou *obiit* (cf. A14)⁵¹. On pourrait donc s'intéresser aux abréviations possiblement utilisées, bien qu'à ma connaissance, les abréviations paléographiques utilisent de nombreux diacritiques et signes spéciaux, ici absents.
 - On peut trouver quelques ressemblances avec l'alphabet étrusque, comme l'usage du signe 4, 8 ou d'un proche parent de Θ.
- En Amérique :
 - Le d pourrait être vu comme un P culbuté, utilisé pour la transcription de langues siouanes⁵². Mais c'est le seul détail qui va dans ce sens.
 - Le 8 est utilisé en Algonquin.

50 <http://www.dorev.ru/>.

51 Voir <https://fr.wikipedia.org/wiki/%EA%9D%8A>.

52 https://fr.wikipedia.org/wiki/P_culbut%C3%A9.

- En Afrique :
 - L'alphabet copte pourrait constituer un bon candidat, puisqu'il utilise les lettres ϩ, Ⲗ, ϫ, et ressemble de près à l'alphabet grec.
 - Le tfinagh, fonctionnant avec un faible de nombre de voyelles, comporte quelques caractères ressemblants. Mais son O désigne le son /r/, pas une voyelle.
- En Asie :
 - La transcription romaine de la langue zhuang utilise les signes ɹ et b pour marquer les tons. Cependant, elle utilise d'autres caractères absents de l'inscription – voir ci-après le début de la déclaration universelle des droits de l'homme, « Bouч bouч ma dəŋz lazɓuŋ coub miɹ cuyouz, cinbyenɹ cəuɹ genli bouчbouч Biŋzɹdəŋz ».
 - Les lettres Θ et h peuvent évoquer des symboles cyrilliques (Oe et Chha) utilisés pour écrire des langues non-slaves en Asie : cependant, la plupart de ces alphabets datent du XX^{ème} siècle.
- Θ pourrait également être une lettre additionnelle latine (O rayé), mais là encore, il s'agit d'un ajout trop récent.

III.5. Hypothèses de chiffrement

III.5.a. Chiffrement mono-alphabétique

Considérons dans un premier temps les conditions d'un chiffrement mono-alphabétique dans le cadre de l'inscription du Caro. Il ne peut s'agir du cas le plus classique, c'est-à-dire de rotation (ROT-n, César, etc.) où chaque lettre est remplacée par sa n^{ème} suivante (e.g. A par K, B par J, etc.) : en effet, dans ce cas, ce qui semble être de voyelles (A, E, I, O) ne seraient pas identifiés comme telles par les algorithmes de détermination de voyelles. Un décalage simple ne peut donc suffire. Autre hypothèse improbable : celle d'un chiffrement semblable à celui du code Navajo, où chaque lettre est remplacée par une syllabe. Dans ce cas, oui, les voyelles sont préservées, mais on observe également une grande récurrence de polygrammes (groupes de plusieurs lettres), ce qui n'est pas le cas dans l'inscription.

L'hypothèse la plus vraisemblable est donc celle d'un chiffrement semblable à celui de l'argot Malespín⁵³, où certaines voyelles sont interchangées entre elles, et de même pour les consonnes. Ainsi, les voyelles restent des voyelles, mais le mot est méconnaissable : « malespin » devient « pelasmon ». Bien sûr, l'utilisation spécifique de cette méthode dans l'inscription est improbable, puisqu'elle est originaire de l'Amérique centrale et date du XIX^{ème} siècle.

En somme, seule une interversion au sein des voyelles et également au sein des consonnes est compatible avec les données observées. Gardant cette conclusion à l'esprit, nous pouvons procéder à une tentative de décodage par analyse fréquentielle. La fréquence d'apparition des caractères variant selon les langues, nous ferons un essai par langue candidate.

53 Voir : <https://www.dcode.fr/argot-malespin>.

La méthodologie sera la suivante : remplacer la voyelle la plus fréquente dans l'inscription par celle la plus fréquente dans la langue candidate, et faire de même pour les consonnes, et ainsi de suite. Après chaque remplacement, le dictionnaire de la langue concerné sera consulté afin de vérifier la possibilité de tels mots.

La substitution a été tenté sur deux transcriptions : l'une telle que présentée en I.2, l'autre avec inversion des caractères (cf. III.4.a).

III.5.a.1. Breton

Dans la transcription originale, on observe un problème dès la troisième substitution : en effet, la lettre I, dans l'inscription est la troisième voyelle la plus fréquente : en breton, la voyelle occupant ce rang-là est la lettre O. Or cette substitution mène à un mot défiant les lois phonotactiques du breton : ObIIE (A14) devient « i_ooe ». Ajustant ce paramètre (I reste I, O reste O), la substitution est poursuivie malgré tout. L'étape suivante consiste à faire du R un N, consonne la plus fréquente dans le corpus : AR devient alors « an », article défini en breton⁵⁴. Ensuite, O reste O ; 2 devient R ; V, T. Le début d'A3, ARVREOIE2, devient donc « antneo_er », lequel ne donne aucun résultat (même en supprimant le « an », voir le « ant » initial). Mais peut-être s'agit-il de « ant ne ». C devient d. CVLES (a11) donne alors « dt_e_ » (les tirets représentent des consonnes), alors que la combinaison dt en début de mot n'est pas admise en breton. La transformation totale du corpus, donnée en annexe 10, a ensuite été effectuée sans supervision.

La transcription renversée a été établie en opérant les modifications suivantes : 2 → S, I → N et d → b. L'analyse fréquentielle a été à nouveau opérée pour ce corpus, ainsi que la détermination de voyelles par algorithme OCPb (en supprimant les incertitudes). Les résultats détaillés de ces deux analyses sont donnés en annexe 10 : toujours est-il que A, E, I, O, h, γ, ' , 1 et 6 sont classifiés comme des voyelles. La substitution commence de la même manière que le corpus non-modifié, notamment avec la question du I et O. Une première difficulté peut être identifiée en A14 : EROIAL donne « enoia_ », mot introuvable. Plus tard, ObIIE (A14) devient « odiie » et dREAR « dnean », mots également introuvables. À partir de ce moment, la substitution n'est plus supervisée.

III.5.a.2. Français

La même méthode a été ci-après appliquée : tous les résultats sont précisés en annexe.

Dans le corpus brut, A restant A et de même pour R, on obtient le mot « ar », qui ne figure pas dans les dictionnaires. De même, RE2 donne « res », également inconnu. I restant I et O devenant U, EROIAL (A14) devient « eruia_ », mot inconnu. Suite à ces échecs, le reste de la substitution n'est pas supervisé. On notera le « icimensmes » (A11).

Dans le corpus renversé, nous avons le même problème avec ar. En remplaçant b par T, on obtient « trear » pour dREAR (A2) : ce mot n'est pas français. dAROA (ou dAR OA, A8) donne « tarua » ou « tar ua », qui ne sont pas des mots français. Idem pour la supervision.

54 Voir le dictionnaire ici utilisé : <http://devri.bzh> en complément du Favereau.

III.5.a.3. Latin

Le corpus brut interchange la plupart des voyelles. A devenant E et R devenant T, AR devient « et », que l'on reconnaît être une conjonction. Cependant, au fur des substitutions, le résultat devient davantage improbable, par exemple avec « etrtia_ » en début de A3, difficilement compréhensible⁵⁵. En A17, V2OИ devient « rsam », également inconnu. De même, en A15, ALVO4 donne « ecra_ », non-identifié en latin. La substitution devient non-supervisée et ne conduit à rien.

Le corpus corrigé, dès la 5^{ème} substitution, donne lieu à « _tiet » en A2 (pour dREAR), « a_vvi » (pour ObIIE, A14) et « e_tivtva » en A20 (pour AdREIRIO).

III.5.a.4. Russe

La problématique de substitution des voyelles s'est déjà posée en III.4.d. : il y a trop peu de voyelles dans l'inscription par rapport à l'alphabet cyrillique russe. Nous gardons cela à l'esprit pour les substitutions suivantes :

Dans le corpus brut, la substitution des 4 premières voyelles et de la première consonne donne « но » pour RE (A11), qui peut aussi s'écrire (dans sa version molle) « нio », « нio » ou éventuellement « нё », dans le cas d'une inscription postérieure à 1793. Aucun de ces éléments ne donne de résultat⁵⁶. Le début de la ligne A3 pourrait donner « ирное », possible déclinaison de « ирный », lui-même déformable en « ирный », variante orthographique du mot « aérien ». Ce n'est guère convaincant, mais ça l'est déjà davantage que les précédents essais. En A10, IVEL donne « apo_ » ou « apë_ » (également recherché « яро » et « ярë »), ne donnant aucun résultat. Plus loin, en A11, EldA donne « оам_ », « ioам_ » ou « ëам_ », sans résultat (même en intercalant des signes durs ou mous). À partir de là, la substitution n'est plus supervisée.

Dans le corpus modifié, le début est similaire : « но » (et similaires) pour RE (A11), AR (A13) donne « ин » ou « ын », mots introuvables. EROIAL (A14) donne « онеи_ » (et variantes) qui est introuvable. Le résultat entier est en annexe 10.

III.5.a.5. Grec

La substitution sur corpus brut donc « av » (pour AR, A13) signifiant « si » en grec moderne ; « oa » semble, lui, désigner le sorbier en grec ancien ; « ve » (RE, A11) peut évoquer une variante orthographique de « vη- », préfixe de négation. Pourtant, une fois les 2^{ème} et 3^{ème} consonnes les plus fréquentes remplacées, le début d'A3 donne « avтvεo_ » (ou « av тvεo_ »/« avт vεo_ »), inconnu⁵⁷ ; IVEL (A10), « ιτε_ », (_ étant une consonne) ; CARCLO (A16), « ρανρ_o » – également introuvables. La transcription n'est alors plus supervisée, mais manque de consonnes grecques pour remplacer celles du rocher.

55 Le dictionnaire utilisé prend en compte les flexions des mots : <https://www.grand-dictionnaire-latin.com/>.

56 Sauf en nénéste : <https://ru.wiktionary.org/wiki/%D0%BD%D1%91>.

57 Voir l'outil de recherche Wiktionnaire avec joker :

https://tools.wmflabs.org/anagrimes/chercher_graphie.php?langue=&type=&genre=&liste=table&flex=oui&loc=oui&gent=oui&nom_propre=oui&no_diac=OK&transcrit=OK&liste=table#liste .

Au même stade que dans le corpus brut, le corpus renversé donne « *_vocav* » (GROCAR, A1) et « *_vεav* » (dREAR, A2), tous deux inconnus. De même, RE (A17) conduit à « *νες* », inconnu quelles que soient les diacritiques – il signifie Nescafé en grec moderne, signification difficile à prêter à l'inscription du rocher... De même, AOMA (A10) « *αο_α* », n'aboutit à rien. La substitution complétée est en annexe 10, comme toutes les autres.

En bref

- Un chiffrement mono-alphabétique est possible, mais doit attribuer aux voyelles une valeur de voyelles, et de même pour les consonnes.
- Des tentatives d'analyse fréquentielle prenant en compte ces contraintes n'aboutissent à rien, quelle que soit la langue candidate employée. L'échec apparaît généralement à partir de 4 ou 5 substitutions.

III.5.b. Chiffrement poly-alphabétique

Au XVIII^{ème} siècle, plusieurs méthodes de chiffrement poly-alphabétique existent déjà, comme le chiffre de Vigenère ou le cadran d'Alberti. Il est donc possible, *a priori*, que cette inscription soit le fruit d'un chiffrement poly-alphabétique. Malgré tout, cette méthode diminue de beaucoup l'indice de coïncidence qui, ici, est particulièrement élevé. Ce point est si caractéristique du chiffrement poly-alphabétique que l'on peut écarter d'entrée cette hypothèse.

En bref

- Un chiffrement poly-alphabétique est hautement incompatible avec l'indice de coïncidence observé.

III.5.c. Chiffrement par transposition

Comme pour l'hypothèse du chiffrement mono-alphabétique, il nous faut examiner les conditions selon lesquels l'inscription du rocher peut-être interprétée comme le fruit de cette méthode en particulier.

Le chiffrement par transposition consiste à déplacer les lettres d'un message sans affecter leur teneur. Par exemple, utiliser des anagrammes (« anagramme » devient « manga amer »), changer le placement de lettres au sein d'un mot ou d'une phrase (« ceci est un exemple » devient « ciec set nu meplxee ») sont des chiffrements par transposition. Ces méthodes sont généralement reconnues comme plus robustes que celles de substitution mono- et poly-alphabétiques. Contrairement à ces dernières méthodes, elles ne modifient ni les fréquences d'apparition, ni l'indice de coïncidence.

Cependant, pour qu'une méthode de chiffrement par transposition ait été utilisée, il faut qu'elle place les voyelles de telle sorte à ce qu'elles soient reconnues comme des voyelles par l'algorithme OCPb (puisque c'est le cas pour l'inscription du rocher).

Le deuxième élément à prendre en compte est la récurrence de bigrammes comme AR (9 fois), RE (7), ET (5), VE (5), RI (4), etc. : il peut s'agir autant de simples hasards ou d'une véritable caractéristique – reste à savoir s'il s'agit de l'un ou de l'autre.

Le dernier élément qui reste à expliquer est finalement la position de T, toujours en fin de mot, souvent en fin de ligne.

La première conséquence de cette hypothèse de chiffrement par transposition, c'est que, la fréquence d'apparition des lettres étant conservée, il faut supprimer les langues candidates qui montrent des fréquences d'apparition trop différentes, c'est-à-dire toutes sauf le latin (qui comporte en tête de liste E, A, R, I, O) et le breton (eu égard à ses résultats en III.2.a).

Suite à cette réflexion, j'ai cherché empiriquement s'il pouvait exister des méthodes de chiffrement par transposition qui fournissaient des messages où l'algorithme OCPb pouvait bien identifier les voyelles et les consonnes. J'en ai trouvé quelques-unes qui, à partir d'un corpus pseudo-latin (Lorem ipsum), ont abouti à ce résultat :

- La méthode AMSCO⁵⁸ ;
- La méthode de transposition⁵⁹ ;
- La méthode de dérangement alphabétique⁶⁰ ;
- La transformation en anagrammes ;
- Le fait de mélanger aléatoire les lettres au sein de chaque mot ;
- Il en existe sans doute d'autres, mais le but n'est pas ici de dresser une liste exhaustive, seulement de donner des pistes d'analyse.

Parmi ces méthodes, seule la dernière peut respecter la contrainte imposée par la récurrence de AR : ce mot, trop court pour être mélangé, aurait été laissé tel quel, tandis que les autres auraient été davantage mélangés. Cependant, j'estime cette théorie peu probable, car le mot AR aurait lui-même pu être inversé ; les mots de quelques lettres, comme GEM (A9) ou EKGE (A6), quel que soit l'ordre de réarrangement, ne donnent rien dans les dictionnaires des langues candidates restantes. C'est pourtant une piste que l'on pourrait juger prometteuse, puisqu'EROIAL (A14) pourrait être lu comme « roiale » (i.e. royale), congruent avec la symbolique chouanne du Sacré-Cœur.

Hors cette dernière méthode et celle des anagrammes, les méthodes sus-citées ne mélangent pas les mots à chiffrer lettre par lettre, mais plutôt par tronçons de mots. S'il est donc une méthode, elle doit respecter cette contrainte de *modus operandi*. Les espaces ont pu être intégrées au chiffrement ou placés *a posteriori* comme distracteurs.

Plusieurs éléments me font douter de l'utilisation d'un chiffrement par transposition : déjà, comme dit plus haut, la présence de bigrammes fréquents et le placement de la lettre T. À mon sens, aucune méthode transposition pure ne peut justifier de ce placement. Par ailleurs, j'ai observé, dans toutes les méthodes citées plus haut, un grand nombre de lettres doublées, bien plus que dans l'inscription (où cela n'arrive guère que 3 fois).

58 <https://www.dcode.fr/chiffre-amsco>.

59 <https://www.dcode.fr/chiffre-transposition>.

60 <https://www.dcode.fr/derangement-alphabetique>.

L'autre difficulté inhérente au chiffrement par transposition est la nécessité d'une présentation irréprochable : le message est en effet souvent traité en ligne, en colonnes, quelquefois même présenté par groupes de 5-6 lettres chacun. On n'a pas affaire à une présentation aussi nette et tirée au cordeau pour le rocher de l'Anse du Caro : au contraire, l'ordre est chaotique, les lignes tordues et entremêlées, comme erratiques. Cela me semble, non seulement, être un indice que l'inscription n'est pas le fruit d'un chiffrement par transposition, mais aussi un obstacle majeur à tout décryptage : quand bien même on connaîtrait la méthode utilisée et les clés correspondantes, une seule lettre manquante décalerait le traitement de déchiffrement et renverrait un résultat faux, incompréhensible. Or, entre l'érosion du rocher et la difficulté (présente de base) à en déchiffrer correctement les caractères, une telle erreur est vite arrivée...

En bref

- Un chiffrement par transposition est possible, mais doit répondre aux contraintes suivantes :
 - Sauvegarde de la ségrégation voyelle/consonne par OCPb ;
 - Explique la présence de bigrammes fréquents ;
 - Explique le placement particulier de la lettre T ;
- Ces deux derniers éléments, avec la rareté des répétitions et la confusion de l'inscription, sont autant d'arguments *a priori* en défaveur du chiffrement par transposition ;
- La difficulté à retranscrire proprement le message constitue une limite majeure au décryptage du texte s'il est chiffré par transposition.

III.5.d. Surchiffrement

Un surchiffrement consiste à combiner plusieurs méthodes de chiffrement (par substitution, par exemple, et par transposition). Ce cas est plausible, puisqu'aucune des méthodes de chiffrement examinées plus haut ne semble à privilégier.

En fait, l'hypothèse d'un chiffrement par transposition implique lui-même un surchiffrement, attendu qu'aucun alphabet connu ne semble utiliser de tels caractères : une transposition serait de fait combinée à une modification alphabétique (cf. III.4.).

Quelles que soient les méthodes combinées, elles doivent tout de même permettre les caractéristiques énoncées ci-dessus, assez contraignantes : cela réduit considérablement le faisceau de méthodes utilisables.

En bref

- Un surchiffrement est possible (et même nécessaire pour un chiffrement par transposition), mais doit prendre en compte les contraintes énumérées plus haut.

IV. Mise en perspective

IV.1. Synthèse des rubriques « En bref »

- 32 caractères : 25 alphabétiques, 4 chiffres et 3 symboles typographiques.
- L'essentiel de l'alphabet est latin et en majuscule, mais présente certains autres caractères : Ʒ, Ⅱ, d, b, h, Θ et γ.
- Les 10 caractères les plus fréquentes : E, A, R, I, O, 2, V, C, Ⅱ, L (73,03 % de fréquence cumulée).
- Globalement, il semble que les lettres de voyelles désignent bien des voyelles, et les consonnes des consonnes.
- Parmi les caractères particuliers, l'algorithme *OCPb* classe h et γ comme étant des voyelles ; Θ, Ʒ, Ⅱ, b et d comme des consonnes.
- Les caractères 1, 8, 6, K, F, h, + et ' sont très instables selon l'algorithme, la variante et le corpus utilisés.
- Au contraire les lettres stables sont : A, E, I, O, γ (voyelles) ; Θ, S, N (consonnes coronales) ; J et L (consonnes non-coronales).
- 3 répétitions de lettres : 11, EE et Ⅱ.
- Les groupes de voyelles (identifiées en II.3) sont les suivants : EE, EO, EA, EI, EK, OA, OI, AO, (AA), AI, (Aγ), IE, IO, IA, Ⅱ, hE.
- Le bigramme « AR » est étrangement fréquent.
- La lettre T se trouve toujours en fin de mot.
- L'entropie de premier ordre vaut 4,19, ce qui situe l'inscription dans la normale haute des langues naturelles.
- L'entropie de deuxième ordre semble valoir 2,02, soit bien en-dessous de la valeur normale. Ce résultat doit, malgré tout, être considéré avec prudence, dû à de possibles erreurs de procédure (cf. *infra*).
- L'indice de coïncidence se situe entre 0,086 et 0,091.
- Cet indice est légèrement supérieur à l'intervalle naturel.
- L'hypothèse selon laquelle les consonnes désignent bien les sons qui sont les leurs semble partiellement confirmée par les sorties *OCPb*.
- L'entropie de deuxième ordre calculée est en fait seulement un peu plus basse dans l'inscription que dans les corpus étudiés, ce qui indique un défaut de calcul ou de référence. Malgré tout, sa valeur reste inférieure à celle des autres corpus.
- Examen des langues candidates :

Critère/Langue candidate	Breton	Français	Latin	Russe	Grec
Nombre de caractères	3	2	1	4	5
Nombre de caractères (corrigé)	3	1	3	4	3
Fréquence des lettres 1 à 1	1	2	3	5	4
Distribution des fréquences	1	4	3	5	2
Ratio voyelles/caractères	3	2	1	5	4

Couples de voyelles	2	1	3	5	4
Répétabilité des lettres observées	3	3	5	1	3
Propension à la répétition	4	2	1	3	5
Entropie de premier ordre	3	4	2	5	1
Entropie de deuxième ordre	3	1	4	5	2
Indice de coïncidence	2	1	4	5	3
MOYENNE	2,55	2,09	2,72	4,27	3,27

Rang de ressemblance des langues candidates au texte de l'inscription, en fonction des divers paramètres étudiés. Plus le rang est proche de 1, plus la langue désignée ressemble à celle de l'inscription. Les demis rangs correspondent à des cas d'ex æquo.

- Les recherches à partir du texte en clair n'aboutissent à aucun résultat net, allant en faveur de l'hypothèse du chiffrement ou de l'usage d'une langue peu connue.
- Le renversement des lettres est une hypothèse possible mais qui ne peut suffire à décrypter l'inscription.
- L'alphabet armoricain pourrait être utilisé dans l'inscription du rocher, mais certaines de ses équivalences sont contradictoires avec l'algorithme OCPb (O→T, e.g.). Plus globalement, cet alphabet ne suffit pas à décrypter le texte.
- L'équivalence proposée avec l'alphabet grec, pourtant optimisée au niveau paramétrique, ne semble pas convenir, malgré quelques résultats encourageants. Même les mots sans grand difficulté de transcription (EKTE, AOMA, NEIΣ) restent inintelligibles.
- L'équivalence proposée avec l'alphabet russe est lamentable : il est difficile d'établir une correspondance nette, notamment avec les fréquences (souvent très divergentes), et les essais de transcription n'aboutissent pas.
- Un chiffrement mono-alphabétique est possible, mais doit attribuer aux voyelles une valeur de voyelles, et de même pour les consonnes.
- Des tentatives d'analyse fréquentielle prenant en compte ces contraintes n'aboutissent à rien, quelle que soit la langue candidate employée. L'échec apparaît généralement à partir de 4 ou 5 substitutions.
- Un chiffrement poly-alphabétique est hautement incompatible avec l'indice de coïncidence observé.
- Un chiffrement par transposition est possible, mais doit répondre aux contraintes suivantes :
 - Sauvegarde de la ségrégation voyelle/consonne par OCPb ;
 - Explique la présence de bigrammes fréquents ;
 - Explique le placement particulier de la lettre T ;

- Ces deux derniers éléments, avec la rareté des répétitions et la confusion de l'inscription, sont autant d'arguments *a priori* en défaveur du chiffrement par transposition ;
- La difficulté à retranscrire proprement le message constitue une limite majeure au décryptage du texte s'il est chiffré par transposition.
- Un surchiffrement est possible (et même nécessaire pour un chiffrement par transposition), mais doit prendre en compte les contraintes énumérées plus haut.

IV.2. Discussion & suggestions pour les recherches prochaines

Ce dossier s'achève sur une note de déception, bien sûr : le sens de l'inscription (s'il existe) reste encore à déterminer. Cependant, quelques certitudes ont pu être trouvées au long de l'examen de cette inscription : la plupart des hypothèses formulées par le grand public, et glanées sur Internet au long de la constitution de ce dossier, se sont révélées fausses. Ainsi, même les résultats négatifs, en un sens, rapprochent de l'objectif.

Cependant, la recherche est loin d'être terminée : si nous n'avons pas abouti, c'est peut-être que certaines de nos hypothèses de travail sont en fait erronées, et qu'il nous faut envisager d'autres possibilités. Le reste de cette partie sera justement dédiée aux perspectives pour de futures recherches.

Il est, dans un premier temps, nécessaire de conduire de nouvelles recherches à partir du matériel, afin d'obtenir de nouvelles données : fouiller les alentours de l'Anse afin de faire l'inventaire des rochers gravés et, éventuellement, les sauvegarder dans un contexte minimisant leur érosion. Chercher à dater le rocher, par une quelconque méthode (mesure de l'érosion ? analyse iconographique du « voilier » ?), pourrait également apporter de nouvelles informations. D'autre part, les réponses des divers experts et association à la Mairie de Plougastel, sollicités avant l'ouverture du concours, pourraient être recoupées avec les résultats de ce dossier et permettre, qui sait, d'avancer un peu. Bien sûr, la dernière démarche à faire, la plus importante, est de revoir la transcription du texte du rocher, selon les directives suggérées en I.2, afin d'augmenter nos chances de réussite. Durant cette démarche, l'analyse et la reproduction du voilier pourrait s'avérer décisive.

Un autre pan des recherches à venir concerne l'aspect contextuel et historique : se charger, plus en détail, d'analyser les symboles de l'inscription – les usages du cœur surmonté d'une croix (A13), ceux de la simple croix dans un texte (A4).

De même, il faudrait s'informer sur les systèmes de repérages (marins, militaires, etc.) qui pourraient utiliser les chiffres de l'inscription (A4, A13, A15, A21).

Enfin, il serait sans doute très profitable d'étudier les méthodes de chiffrement en vogue à l'époque, et précisément les méthodes de communication utilisées par les armées chouannes.

Un troisième axe d'amélioration concernerait l'optimisation du présent dossier : prioritairement, recalculer l'entropie de deuxième ordre, en comprenant d'où vient l'erreur de calcul observée.

Ce pourrait aussi être profitable de modifier la démarche suivie en III.2. : il ne s'agirait plus de sélectionner soigneusement un faible nombre de langues candidates, mais au contraire d'en constituer un corpus massif dont les caractéristiques seraient calculées (fréquence de lettres, entropie de premier et deuxième ordre, indice de coïncidence) ; le texte de l'inscription serait ensuite situé parmi toutes ces langues afin de déterminer quelle langue lui ressemble le plus (à l'état naturel, non-chiffré).

Autrement, dans l'hypothèse d'une substitution alphabétique (partie III.4), il serait intéressant d'utiliser un algorithme afin d'optimiser les différents critères de substitutions (fréquence, traitement OCPb, ressemblance graphique, caractères manquants par rapport à l'alphabet de la langue candidate, etc.) ; de même, les substitutions mono-alphabétiques pourraient être traitées par algorithme et optimisées avec d'autres paramètres (bigrammes, trigrammes, mots courts, phonotactique).

Il serait sans doute profitable, également, de conduire des analyses détaillées sur le corpus inversé (partie III.4.a).

Enfin, malgré les limites indiquées, on pourrait essayer de suivre les autres pistes d'analyses énoncées en II.6.

Enfin, d'autres pistes d'analyse sont possibles.

Tout d'abord, un certain nombre d'idiomes pourraient faire office de langues candidates, pour la simple raison que ce rocher n'est pas si lointain d'un port, et qu'un port du XVIIIème siècle (dans le cas où notre datation approximative est exacte) peut brasser des gens d'origines très diverses. On relira avec intérêt, donc, les pistes listées en III.3 et III.4.e⁶¹, et y rajouter opportunément le patois vendéen, le gallo. En effet, si l'inscription est chouanne, elle peut être tant bretonne (hypothèse à privilégier du fait de la situation géographique) que vendéenne : d'ailleurs, durant mes recherches, je n'ai jamais trouvé d'utilisation du Sacré-Cœur de la part des Chouans bretons : à ma connaissance, seuls les Vendéens l'ont adopté. Ce pourrait être un indice fort.

D'autre part, s'il est nécessaire d'approfondir les hypothèses de chiffrement, celle de transposition nécessite une attention toute particulière, puisqu'elle est particulièrement complexe (nul moyen, à ma connaissance, de vérifier exhaustivement si l'on a affaire à un chiffrement par transposition ou non ; le décryptage d'un texte chiffré par cette méthode, comme indiqué, est particulièrement difficile). Ayant épuisé les méthodes classiques, il faudra sans doute faire l'hypothèse d'autres méthodes, moins connues, *ad hoc* : comme, par exemple, un renversement du texte entre chaque lettre inversée (Z ou \mathbb{N}), ou un renversement total du texte (lire « RACORG » à la place de « GROCAR », A1), etc.

61 Pour rappel, ladite liste était la suivante : abréviations latines ; alphabet étrusque ; langues siouanes ; algonquin ; copte ; berbère ; zhuang (compliqué) ; cyrillique non-slave ; alphabet latin étendu ; ivbiosakon ; basque ; irlandais ; vieil anglais, anglais.

Il existe, enfin, plusieurs menues pistes à suivre afin d'en comprendre plus sur le texte du rocher : rechercher les noms propres qui pourraient être mentionnés dans le message (roi, dieu, fort, armée, etc.) ; s'étonner de la difficulté à prononcer l'inscription, malgré les allégations de l'algorithme OCPb (par exemple, « CVLES », A11) ; noter les mots courts et variant faiblement, comme IVE (A16), devenant IVEL (A10) ; s'intéresser à la possible existence d'un individu nommé OSCAR CLO(I)VEPRE (A16, transcription de la mairie) ; imaginer certains caractères comme des doubleurs de lettres (« R2 » = « RR », etc.), des distracteurs (lettres inutiles, à supprimer lors du décryptage) ou permettant d'autres modifications (voir quelques curiosités du Grand Chiffre de Louis XIV, comme un caractère annulant le précédent) – au-delà de ces hypothèses un peu sophistiquées, nous n'avons même pas essayé de substituer à une lettre un groupe de lettres (« I » = « NN » ou « 2 » = « CH ») ; pour finir, l'on pourrait voir en la ligne A21 un sorte de colophon, avec une date (I186) et une signature (IIEI2) – qui est ce IIEI2 ?

En somme, beaucoup de possibilités restent à examiner, beaucoup de questions restent sans réponse : mais j'ose espérer que ce dossier sera utile à d'autres pour avancer vers une solution.

– François PARMENTIER

Le 22 août 2019

[L'intégralité de dossier est sous la licence CC-BY-SA 4.0](#)

ADDENDUM N° 1 DU 26 AOÛT 2019
 CONCERNANT L'HYPOTHÈSE D'UN CHIFFREMENT PAR TRANSPOSITION

Cette démarche additionnelle sur la remarque que j'avais faite à propos des répétitions de lettres dans un message chiffré par transposition : je trouvais que les répétitions de lettres y étaient bien plus fréquentes que dans le texte de l'inscription. Ceci dit, cette remarque étant essentiellement qualitative, j'ai cherché à en savoir plus.

Prenons l'hypothèse d'un chiffrement par transposition qui mélange les lettres selon un *pattern* défini. La répartition des lettres va tendre à être aléatoire, non dictée par des exigences phonotactiques : ainsi, la probabilité d'une répétition de lettres vaudra l'inverse de la taille de l'alphabet. Cette probabilité, multipliée par le nombre de bigrammes présents dans le texte, donnera donc un nombre d'apparitions probables. Dans notre cas, si l'on suppose un chiffrement lettre par lettre (c'est-à-dire que chaque lettre du texte est indépendante en position de sa suivante), la probabilité que l'on observe une lettre double vaut $1/32$ (nombre de caractères différents) et, comme le texte comporte 334 bigrammes, l'on devrait observer probablement $334/32$ apparition de lettres doublées, c'est-à-dire un peu plus de 10.

Or, dans notre cas, si l'on supprime les espaces supputatifs du corpus, le nombre de répétitions observé s'élève à 4. Ainsi, si le message est bien chiffré par transposition, il ne semble pas mélanger lettre par lettre, puisque l'écart entre répétition théorique et observée est important (multiplication par 2,5). Cela est d'ailleurs congruent avec nos conclusions provisoires énoncées plus haut : s'il y a chiffrement par transposition, ce n'est pas lettre par lettre, sans quoi l'algorithme OCPb n'identifierait pas les voyelles comme telles. Le mélange se ferait donc au moins par groupe de deux lettres.

Justement, si la transposition se fait par groupe (ou monade) de deux lettres, la fréquence théorique de répétition est moindre : elle vaut toujours l'inverse de la taille de l'alphabet, mais ne s'applique qu'à une lettre sur deux (la première du groupe de deux étant solidarisée avec l'autre, leur succession n'est donc pas aléatoire). Ainsi, la fréquence théorique ne vaut plus $334/32$, mais la moitié, $167/32$, soit 5 et quelques. Si l'on compare cette fréquence à celle observée, on remarque déjà une plus grande proximité. Il est donc plus probable que le chiffrement par transposition, si l'on est bien dans ce cas, s'applique à des monades de deux lettres qu'à une seule.

Répétant ce calcul pour plusieurs tailles de monades, on obtient le tableau suivant :

Monade de ... caractères	1	2	3	4	5	6	etc.
Fréquence théorique	10,44	5,22	3,48	2,61	2,09	1,74	etc.

L'intérêt, c'est que l'on observe que la fréquence observée (4) se situe entre 2 et 3 : l'on peut donc déterminer que, le plus probablement, s'il s'agit d'un chiffrement par transposition, il se base sur des groupes de 2 ou 3 caractères, mais guère plus (et surtout pas moins).

J'ai été tenté de conduire des tests inférentiels, mais les tests Z ne pouvant s'appliquer faute d'un échantillon assez grand (les fréquences théoriques et observées doivent chacune être supérieure à 5), il aurait fallu se rabattre sur un test exact de Fischer, trop sensible à mon goût. Cela n'aurait, de toute façon, fait que révéler ce que l'on a dit : le plus probable sont les monades de 2 ou 3, avec une légère préférence pour la monade de 3.

Reste à savoir comment et où cette séparation s'applique. Cela pose non seulement problème pour les rochers marginaux (B, C, D et E), ornés d'inscriptions laconiques et donc difficiles à remettre en ordre, mais même pour le rocher A, puisque, premièrement, nous ne disposons pas l'intégralité du texte et, deuxièmement, il faudrait tester chaque monade possible de 2 ou 3 lettres et tenter de l'assembler avec d'autres monades, sans connaître réellement la langue cible (les fréquences d'apparition identifiées sont assez peu compatibles avec les langues candidates proposées). L'on pourrait toujours se baser sur les parties imprononçables du message, comme CVLES (A11) ou le EE de DIO2EEVbIO (A2).

Cela, bien sûr, n'aurait même pas la garantie de fonctionner, attendu que certaines méthodes de chiffrement par transposition ne se basent pas sur une taille de monade stable, mais sur d'autres critères (voir les méthodes AMSCO et de dérangement alphabétique).

Ainsi, même si la présente réflexion limite davantage le champ des possibles, et donc le champ d'investigation, bien des hypothèses peuvent être encore formulées sur le thème du chiffrement par transposition. Plus généralement, la perspective d'une transposition, à mon humble avis, reste assez peu probable, malgré sa possibilité : je pense, surtout, ne pas vouloir y croire car, si elle devenait la seule hypothèse possible, elle annoncerait pour ainsi dire notre impuissance à reconstituer le message d'origine – pour toutes les raisons évoquées ci-dessus et dans le présent dossier.

*

**

Il faudrait également ajouter, dans les arguments allant à l'encontre de l'hypothèse de chiffrement par transposition la proximité des chiffres avec eux-mêmes : ne devraient-ils pas être, de fait dispersés ? Dans le cas contraire, on pourrait se baser sur cette propriété pour déduire les paramètres de la méthode utilisée.

V. Annexes

V.1. Glossaire

Chiffrement/déchiffrement/décryptage : on parle de *chiffrement* pour désigner une méthode rendant un message incompréhensible pour toute personne ignorant le procédé utilisé et donc la manière de retourner au message initial. On parle donc de *déchiffrement* quand quelqu'un reconstitue le message initial en connaissant le moyen de le faire. Au contraire, si l'on ne connaît pas initialement la méthode de chiffrement du message mais que l'on tente tout de même d'en percer le sens, on est alors dans un cas de *décryptage*.

Coronal(e) : consonnes dont le lieu d'articulation se trouve entre les dents et le palais dur (inclus). Cependant, la définition utilisée par Huldén dans son algorithme OCP est un peu plus large, puisqu'elle comprend (au moins) les consonnes h et x.

Cyrillique (alphabet) : alphabet créé au Moyen-Âge en Macédoine par le moine Cyrille. Il est de nos jours utilisé pour de nombreuses langues différentes – chacune dispose en fait d'une variante de cet alphabet, contenant des lettres supplémentaires. L'alphabet cyrillique russe et l'orthographe de la langue russe a été réformé de manière importante en 1918, apportant globalement une grande simplification de l'orthographe.

Entropie : mesure de la variabilité d'un texte, reflétant la quantité d'information qu'il contient. L'entropie de premier ordre porte sur la probabilité d'occurrence de chaque lettre ; celle de deuxième ordre de couples de deux lettres.

Hapax : spécimen (lettre, mot, chiffre) n'apparaissant qu'une seule fois dans un corpus donné.

Indice de coïncidence : donnée utile en cryptanalyse, calculant la probabilité de répétition de chaque lettre dans le message chiffré. L'indice de coïncidence est, par propriété, extrêmement réduit dans le cas d'un chiffrement poly-alphabétique (autour de 0,3 – 0,4).

Mono-alphabétique (chiffrement) : Méthode de chiffrement remplaçant chaque caractère du message non-chiffré par un ou plusieurs caractères (e.g. A devient 1, B 2, etc.). La grille de correspondance (affectant à chaque lettre un substitut) ne varie pas au cours du message (dans notre exemple précédent, le A sera toujours remplacé par 1). Ce type de chiffrement est généralement facile à casser, grâce à une analyse fréquentielle.

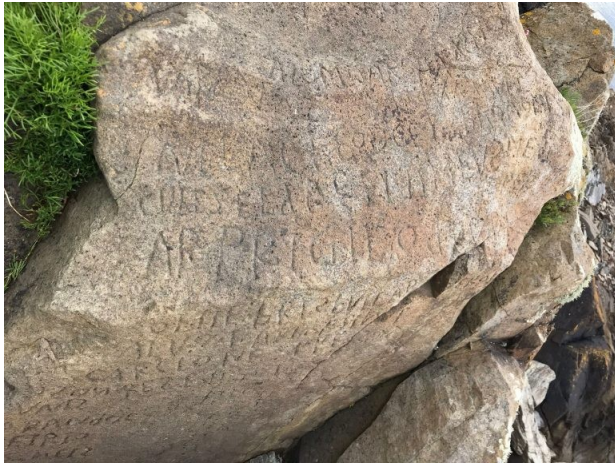
Poly-alphabétique (chiffrement) : Méthode de chiffrement identique au chiffrement mono-alphabétique, à l'exception près que la grille de correspondance change périodiquement (e.g. pour une lettre sur deux, A est remplacé par 1, dans l'autre cas par un 6 ; B par 2 ou 7, etc.). Cette méthode réduit drastiquement l'indice de coïncidence.

Sérif : Empattement visible à l'extrémité de caractères dans certaines polices d'écritures. Comparaison de la lettre I avec et sans sérif : I ; I ; lettre C : C ; C.

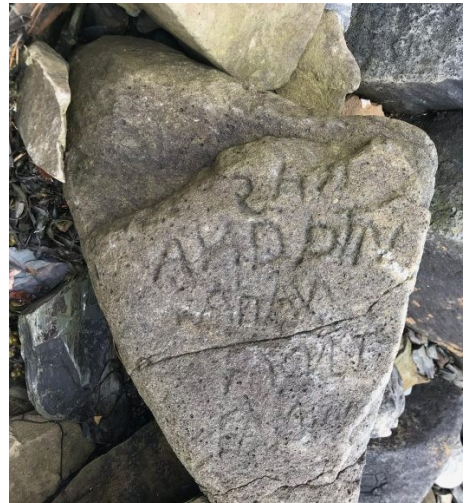
V.2. Système de numérotation

Le décryptage d'une inscription mystérieuse doit toujours passer par une transcription rigoureuse, et la démarche même de transcription appelle un système de numérotation, à des fins de communication et d'interopérabilité.

Seuls les rochers A et B sont exposés dans le dossier photographique constitué par la mairie : le rocher B figure aux pages 17 et 60, le A aux autres. Les autres photographies proviennent d'une internaute du pseudonyme de FrançaiseMQ⁶² :



Rocher A



Rocher B



Rocher D



Rocher E

Le voilier, si je ne m'abuse, se trouve sur un autre rocher.

62 Source : <http://ciphermysteries.com/2019/05/13/plougastel-inscription-photographs>

V.3. Transcriptions précédentes de l'inscription

Les éléments indiqués à gauche sont les correspondances avec la transcription en I.2.

- A1 ROC AR B ...
- A2 ORE AR GRIO SE EVELOH
- A3 AR VIRIONES BAOAVEL ... R 1
- A4-A5 ... GENBICEN DA BEN ESOA ... S E
- A6 DIASBOANT.. .
- A6 ... EKGES ...
- A7 BO ... FET
- A8 DAR
- A7 ASOMGAROPA
- ? VARLAEOERLATQDAO
- A10 NELACIEOD - ET... F ... AOMA
- A11 CVLESEDAREIDIMEVSMES
- A13 ARPRIGILOD1787 (coeur et croix)
- A14 OBIIE: BRISBVILAR ... FROIK ... AL
- A15 ... ALVOA ARBORSINET
- A16-A17 CARCLONEPR ES (SAKI) ASONRES E 1 BEL
- B2-A18 (ANNODIN) VMS
- B3-A19 ... (ABAN) SDARANDOC (SAOU) ADREIRIO
- B4-A21 (FAN) 1786 NEIS
- ? ... CL.. .
- ? ... DOS ...
- ? DARMIS - UT
- ? 1771MESO
- ? EMGAR
- ? DALOTOGREC
- ? TRE : C---N
- A18 1920

Commentaire : Cette transcription a d'abord le mérite d'être la première réalisée sur le rocher. Si, dans les grandes lignes, elle respecte l'inscription, elle se montre malgré tout imprévisible et ne peut servir réellement de base à une analyse rigoureuse. Elle est aussi

remarquable par sa volonté de (trop ?) régulariser les lettres étrangères à l'alphabet latin (lettres inversées, Θ, etc.). Enfin, on s'intéressera particulièrement aux lignes qui ne sont pas retrouvées dans les autres transcriptions (notamment les dernières) et qui semblent quelquefois provenir d'autres rochers.

- A1 GROCAR
- A2 dREAR DIO2EEVbIO
- A3 ARVRFOIE2LAΘENEL
- A4 PEИ AbEИEИE2Θ182E
- A5 GEИbICEИG
- A6 EKGE D
- A7 AZOMOARΘ PA γCDOFET
- A8 dARO
- A8 A
- A10 AIELACHeODCET DA AOMA
- A11 CVLESELDdA RE IdIMEV2ME2
- A13 ARPRIGILOd17(Cœur orné d'une croix)87
- A14 ObИIE bRI2bVILA
- A15 ALVQ4AKbORSIV . T
- A16 OSCAR CLOIVE PRE 2 . T
- A17 ZOI REZE
- A18 VAR
- A19 dARAIIdOL
- A20 AdREIRI9
- A21 1786 ИEИ2

Commentaire : Cette transcription figure dans le dossier de la mairie et, à ce titre, devrait être la plus exacte. Cependant, elle ne porte que sur le rocher A, et pas sur les autres ; par ailleurs, elle se focalise essentiellement sur les lignes les plus lisibles, les plus importantes – ce qui est déjà assez problématique dans le cas d'un message incompris, a fortiori s'il est chiffré par transposition... Enfin, elle peut faire quelquefois appel à des éléments n'apparaissant qu'une fois (Q, 9, etc.), et ce sans nécessité, puisqu'il s'agit d'artéfacts de coloration dus à la pierre et et son érosion. Cela mis à part, la méthode de transcription semble assez rigoureuse et peut permettre de confirmer partiellement notre propre transcription.

V.4. Sorties des analyses pour la détermination de voyelles sur l'inscription

Analyse 1a :

- 0 (voyelles)
 - 00
 - 000 : I K
 - 001 : A h ' ʎ
 - 01
 - 010 : 1 6 O
 - 011 : + E
- 1 (consonnes)
 - 10
 - 100 : C M S b d Θ
 - 101 : - 8 N P V H
 - 11
 - 110 : R T 2
 - 111 : 4 D F G J L

Analyse 1b :

- 0 (voyelles)
 - 00
 - 000 : I h ʎ
 - 001 : 6 A F
 - 01
 - 010 : 1 O
 - 011 : E
- 1
 - 10 (consonnes non-coronales)
 - 100
 - 1000 : M H
 - 1001 : L
 - 101
 - 1010 : V
 - 1011 : + 8 D J K R
 - 11 (consonnes coronales)
 - 110
 - 1100 : b
 - 1101 : - G S T d Θ
 - 111
 - 1110 :
 - 11100 : C P
 - 11101 : N
 - 1111 : 4 2'

Analyse 2a : les données brutes ont été perdues, elles sont ici reconstruites.

- 0 : E A I O 1 h K + γ 6
- 1 : R ʒ V C И L d T M b G P D N 8 Θ J S – F ' 4

Analyse 2b :

- 0 (consonnes)
 - 00 (consonnes non-coronales)
 - 000 : - F G V d Θ
 - 001 : 4 8 C J M N P S b
 - 01 (consonnes coronales)
 - 010 : 2
 - 011 : D L R T И '
- 1 (voyelles)
 - 10
 - 100 : O ʏ
 - 101 : 1 E
 - 11
 - 110 : 6 A K h
 - 111 : + I

Analyse 3a :

- 0 (voyelles)
 - 00
 - 000 : E
 - 001 : O ʏ
 - 01
 - 010 : I h
 - 011 : A
- 1 (consonnes)
 - 10
 - 100 : M V d Θ И
 - 101 : C K N P S b
 - 11
 - 110 : D F J R T ʒ
 - 111 : G L

Analyse 3b :

- 0 (voyelles)
 - 00 :
 - 000 : O
 - 001 : E
 - 01 :
 - 010 : I h √
 - 011 : A
- 1
 - 10 (consonnes non-coronales)
 - 100
 - 1000 : L
 - 1001 : D F J K M И
 - 101
 - 1010 : V
 - 1011 : R
 - 11 (consonnes coronales)
 - 110
 - 1100 : G S T d Θ
 - 1101 : b
 - 111
 - 1110 : C P
 - 1111
 - 11110 : ʒ
 - 11111 : N

Algorithme de Sukhotine (classé par qualité et par score) :

- Voyelles : E (32), A (26), I (14), O (10), 8 (3)
- Consonnes : R (-26), V (-10), b (-8), G (-7), d (-7), И (-7), D (-5), L (-5), M (-5), 1 (-4), P (-3), 4 (-2), - (-2), F (-2), J (-2), ʒ (-2), S (-2), h (-1), Θ (-1), + (0), C (0), N (0), T (0)
- Indéterminés : ' √ K 6 (1 ex æquo)

V.5. Résultats de l'analyse phonotactique

	1	4	6	8	A	b	C	d	D	E	F	G	h	I	J	K	L	M	N	И	O	Θ	P	R	S	2	T	V	γ	-	+	'	.		
1	1			2																														1	
4																																		1	
6																																		1	
8	1		1																							1									
A					1	1	1						1			2	1	2	3	1	1		9	1						1			8		
b									1				3							1			1				1								
C				2				1	3			1				1											1							3	
d	1			3									1									1		2											
D				1		1							1									1												1	
E				1		1			1				2	1	1	2	2		4	2			2	1	4	5	2							12	
F									1																										
G				1					2				1										1												1
h									1																		1								1
I	1			1	1		2	1	2			1	1			2	1	1		3			1		2		3							5	
J				1																															1
K												1																							
L				1			1		1									1		1							1								3
M				1					2		1																								3
N							1	1	1																										2

	1	4	6	8	A	b	C	d	D	E	F	G	h	I	J	K	L	M	N	И	O	Θ	P	R	S	2	T	V	γ	-	+	'	.		
И						1				4		1		1										1										4	
O		1			2	1	2		1					1				2	1	3			1	2		1						1	6		
Θ													1	1																					
P					1					1				1										3											
R					1	1	1			7				4							3				1			1						9	
S														1																				1	
2						1		1		1			1				1	1			1	1												5	
T																																		8	
V					1	1				5				1			1				1			1		2									
γ							1																												
-					1																														
+																																			1
'											1																								
.					15	1	3	3	2	9		2		8	1					2	9		5	5		4	3	4	1			1			

Les espaces sont figurées par un point médian : ·

Les lignes donnent la première lettre du bigramme, les colonnes la seconde. Ainsi, par exemple, le bigramme 18 (ligne du 1, colonne du 8) apparaît 2 fois dans l'inscription.

V.6. Résultats de l'analyse fréquentielle sur les langues candidates

V.6.a. Breton

Caractère	Fréquence	Occurrences
Espace	18,68 %	3004
E	11,29 %	1815
A	9,64 %	1550
N	6,85 %	1102
O	6,83 %	1098
R	6,23 %	1002
T	4,72 %	759
I	4,58 %	737
D	3,65 %	587
U	3,51 %	564
S	3,16 %	508
H	3,13 %	503
L	3,06 %	492
C	3,05 %	491

Caractère	Fréquence	Occurrences
M	2,44 %	392
P	1,92 %	309
V	1,70 %	274
G	1,53 %	246
B	1,08 %	174
F	0,73 %	117
Z	0,63 %	102
K	0,53 %	86
'	0,53 %	85
-	0,20 %	32
J	0,17 %	28
Y	0,14 %	22
X	0,01 %	1

V.6.b. Français

Caractère	Fréquence	Occurrences
Espace	16,98 %	4076
e	14,84 %	3562
r	6,64 %	1593
a	6,26 %	1502

Caractère	Fréquence	Occurrences
v	0,98 %	236
q	0,88 %	212
f	0,85 %	205
'	0,89 %	214

s	2,90 %	695
f	3,27 %	785
n	5,83 %	1400
i	5,54 %	1331
t	5,21 %	1251
u	5,07 %	1218
o	4,31 %	1035
d	4,19 %	1005
l	3,98 %	955
c	2,69 %	645
p	2,41 %	578
m	2,40 %	577

g	0,79 %	190
b	0,61 %	147
&	0,57 %	136
h	0,47 %	113
y	0,45 %	108
j	0,44 %	106
x	0,34 %	82
z	0,12 %	30
k	0,07 %	16
4	0,00 %	1
2	0,00 %	1

V.6.c. Latin

Caractère	Fréquence	Occurrences
Espace	13,42 %	4059
i	10,05 %	3039
e	9,60 %	2904
v	7,64 %	2311
a	7,16 %	2166
t	6,77 %	2049
s	6,29 %	1903
r	6,00 %	1815
n	5,97 %	1806

Caractère	Fréquence	Occurrences
p	2,26 %	684
b	1,41 %	427
q	1,32 %	400
g	1,19 %	360
f	0,79 %	240
h	0,56 %	168
j	0,36 %	110
x	0,36 %	110
-	0,02 %	5

o	5,57 %	1685
m	4,09 %	1236
c	3,89 %	1176
d	2,79 %	844
l	2,45 %	740

1	0,01 %	3
z	0,01 %	3
4	0,01 %	2
y	0,01 %	2
k	0,00 %	1

V.6.d. Russe

Caractère	Fréquence	Occurrences
Espace	15,12 %	3983
O	9,84 %	2592
И	5,93 %	1563
A	5,79 %	1525
E	5,35 %	1411
H	5,02 %	1323
C	4,74 %	1250
P	4,19 %	1105
Ъ	4,11 %	1084
B	4,09 %	1079
T	3,84 %	1011
Л	3,09 %	814
M	2,87 %	756
K	2,78 %	732
Д	2,58 %	681
П	2,33 %	615

Caractère	Fréquence	Occurrences
I	1,69 %	445
У	1,55 %	408
З	1,37 %	360
Ы	1,34 %	354
Ч	1,05 %	276
Б	1,04 %	275
X	0,82 %	216
Ж	0,80 %	210
Ь	0,75 %	197
Й	0,61 %	161
Ю	0,58 %	153
Ш	0,28 %	75
Ц	0,27 %	72
Щ	0,27 %	71
Ф	0,17 %	46
-	0,02 %	6

Я	2,05 %	539
Г	1,89 %	497
Ђ	1,75 %	462

Ә	0,01 %	2
V	0,00 %	1

V.6.e. Grec

Caractère	Fréquence	Occurrences
Espace	16,00 %	4087
E	9,58 %	2448
A	8,57 %	2188
I	8,11 %	2072
O	7,67 %	1960
N	7,19 %	1837
Σ	6,28 %	1605
T	6,25 %	1596
H	3,93 %	1004
Υ	3,86 %	985
P	2,92 %	747
Λ	2,65 %	676
Π	2,61 %	667

Caractère	Fréquence	Occurrences
K	2,51 %	640
Ω	2,39 %	610
Δ	2,29 %	585
M	2,26 %	576
Γ	1,48 %	377
Θ	1,22 %	311
X	0,65 %	165
B	0,50 %	128
Φ	0,45 %	114
E	0,29 %	74
,	0,15 %	39
Z	0,12 %	30
Ψ	0,08 %	21

V.7. Résultats de l'analyse de voyelles sur corpus de langues candidates

V.7.a. Breton

- 0
 - 00 (consonnes non-coronales)
 - 000
 - 0000 : N
 - 0001
 - 00010 : S
 - 00011 : Z

- 001
 - 0010 : R
 - 0011
 - 00110 : H
 - 00111 : L X
- 01 (consonnes coronales)
 - 010
 - 0100 : T
 - 0101
 - 01010
 - 010100 : P
 - 010101 : F
 - 01011 : D
 - 011
 - 0110
 - 01100
 - 011000 : B Y
 - 011001 : G
 - 01101 : C
 - 0111
 - 01110 : M
 - 01111
 - 011110 : V
 - 011111 : J K
 - 1 (voyelles)
 - 10
 - 100
 - 1000 : A
 - 1001 : '
 - 101
 - 1010 : U
 - 1011 : I
 - 11
 - 110 : E
 - 111
 - 1110 : O
 - 1111 : -

V.7.b. Français

- 0 (voyelles)
 - 00
 - 000

- 0000 : o
 - 0001 : a
 - 001 : e
- 01
 - 010
 - 0100 : 4 u
 - 0101 : k'
 - 011 : i
- 1
 - 10 (consonnes non-coronales)
 - 100
 - 1000
 - 10000 : q
 - 10001 : d
 - 1001 : r
 - 101
 - 1010
 - 10100 : l
 - 10101 : 2 j z
 - 1011
 - 10110 : n
 - 10111 : h
 - 11 (consonnes coronales)
 - 110
 - 1100
 - 11000
 - 110000 : f x
 - 110001 : '
 - 11001
 - 110010 : y
 - 110011 : s
 - 1101
 - 11010 : f
 - 11011 :
 - 110110 : m
 - 110111 : v
 - 111
 - 1110
 - 11100
 - 111000 : c
 - 111001 : g
 - 11101 : p
 - 1111

- 11110 : t
- 11111 : b

V.7.c. Latin

- 0 (voyelles)
 - 00
 - 000 : 3 v
 - 001
 - 0010 : a
 - 0011 : o
 - 01
 - 010 : 5 i
 - 011 : 2 e y
- 1
 - 10 (consonnes non-coronales)
 - 100
 - 1000 : r
 - 1001
 - 10010 : l s
 - 10011 : h
 - 101
 - 1010
 - 10100 : 1
 - 10101
 - 101010 : x
 - 101011 : q
 - 1011 : n
 - 11 (consonnes coronales)
 - 110
 - 1100
 - 11000
 - 110000 : g
 - 110001 : f k
 - 11001 : m
 - 1101
 - 11010 : 4 b
 - 11011 : 9 c k
 - 111
 - 1110
 - 11100 : d z
 - 11101 : p
 - 1111 : t

V.7.d. Russe

- 0
 - 00 (consonnes non-coronales)
 - 000
 - 0000
 - 00000 : K
 - 00001 : - 9 T
 - 0001
 - 00010 : 6 Ц
 - 00011 : H
 - 001
 - 0010
 - 0011
 - 00110 : 3 Л
 - 00111 : Ш
 - 01 (consonnes coronales)
 - 010
 - 0100
 - 01000 : C
 - 01001 : Ж
 - 0101
 - 01010 : П
 - 01011
 - 010110 : 3
 - 010111
 - 0101110 : 2 V Г
 - 0101111 : 4 Ц
 - 011
 - 0110
 - 01100
 - 011000
 - 0110000 : Ф
 - 010001 : Ю
 - 011001
 - 0110010 : X
 - 0110011 : Ё
 - 01101 : B
 - 0111
 - 01110 : M
 - 01111
 - 011110
 - 0111100 : Ч
 - 0111101 : Б

- 011111 : Δ
 - 1 (voyelles)
 - 10
 - 100
 - 1000 : 1 8 A
 - 1001 : E
 - 101
 - 1010
 - 10100 : Э Я
 - 10101 : Ъ
 - 1011
 - 10110 : Ь
 - 10111 : 5 У
 - 11
 - 110
 - 1100
 - 11000 : Ъ
 - 11001 : Ы
 - 1101 : 7 И
 - 111
 - 1110 : 0 I
 - 1111 : O

V.7.e. Grec

- 0
 - 00 (consonnes non-coronales)
 - 000
 - 0000 : Σ
 - 0001 : N
 - 001
 - 0010 : P
 - 0011
 - 00110 : Z Ψ
 - 00111 : Λ
 - 01 (consonnes coronales)
 - 010
 - 0100 : T
 - 0101
 - 01010 : Θ
 - 01011
 - 010110 : Γ
 - 010111 : Δ

- 011
 - 0110
 - 01100
 - 011000 : X
 - 011001
 - 0110010 : Φ
 - 0110011 : B
 - 01101 : Π
 - 0111
 - 01110
 - 011100 : Ξ
 - 011101 : K
 - 01111 : M
- 1
 - 10
 - 100
 - 1000
 - 10000 : '
 - 10001 : Υ
 - 1001 : I
 - 101 : E
 - 11
 - 110
 - 1100 : O
 - 1101 : H
 - 111
 - 1110 : Ω
 - 1111 : A

V.8. Analyses phonotactiques sur les corpus de langues candidates

Breton :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	X	Y	Z	'	-	.
A	1	14	12	47	23	3	2	5	22	5	1	77	43	357	2	13	295	41	45	6	18			6			368
B	18				65				7			39			16		19	1		7						1	1
C	69		5		25			95	16		2	11			9		19		9	7					34		19
D	138				149				19		2	2			44		33			21	1				1	2	76
E	21	13	42	32	12	6	15	12	17	2	3	94	67	291	41	31	175	15	263	124	91	1	6	43	7	2	298
F	15				17	19			29			2			17		1	1		5							2
G	47				26			1	16			8		13	57		19			53							6
H	175				48	1			26			5			148		7			14					14	7	58
I	9	2	4	13	54	9	39			2	13	32	59	13	76	6	21	92	78	1	13			1			65
J	3				1										8					7							
K	3				54				1																		19
L	74		4		76			1	74			91			32			1	3	2	3						131
M	79	17		2	138				18				1		23	56		3	2	11				6	3		24
N	49		6	27	114	11	6	1	43	1	9			2	63			21	18	5	1				12	2	475
O	67	1	31	1	29	17	1	5	9			5	76	18		2	4	17	9	26	2		1	1	2		333
P	46				96			6	12			12			39	4	24		3	13					1	3	5
R	81	7	13	17	29	1	13	3	92		2	1	35	14	19	1	27	19	3	32	14			2		3	277
S	27		2	1	55	1			23		7				57	18		44	5	17						6	182

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	X	Y	Z	'	-	.
T	49				84			9	6						3		59	1	12	14						6	435
U	23	12	6	43	98	2	12		33	12		19	6	35	3	1	9	8	19	1	2		1		1		56
V	4				78				7			1			56		2			9							
X					1																						
Y	16				4														2								
Z	3				31				12						16		2			3							35
'	8		1		7			49	2				1		8	1	1		1	1							4
-								13	2				5	3	2			4		2	1						
.	489	99	148	44	311	47	139	33	35	6	47	48	9	85	143	167	141	78	53	3	128		5	34			

Français :

	e	r	a	s	f	n	i	t	u	o	d	l	c	p	m	v	q	f	'	g	b	&	h	y	j	x	z	k	4	2	.
e	5	275	2	357	171	429	18	125	136	1	38	97	81	34	165	31	21	61		32	1		1		3	24	24				1346
r	537	66	82	65	14	8	133	66	5	117	34	15	1	1	38	14	3			1	2		7	9	1		1				356
a		248		14	12	233	141	59	169	4	16	26	28	6	46	48	2	1		36	19			35	2						32
s	2		6	6			14	3	2	1	1							3													673
f	165		75	1	71		98	139	72	85	17	1	8	6	3		3		5												
n	168	1	47	19	4	89	31	341	1	45	11	2	8		1	18	4	16	12	5				5	1	1					268
i	222	9	4	58	11	124	2	246	3	19	14	82	52	7	27	32	9	13	2	2	8					21					48

	e	r	a	s	f	n	i	t	u	o	d	l	c	p	m	v	q	f	'	g	b	&	h	y	j	x	z	k	4	2	.
t	26	121	58	38	1		162	23	42	47				1	1								2								494
u	161	244	14	2	47	84	92	116		6	22	22	23	2	15	13	5	1	38	45	15			13	1	35	5				179
o	5	98		2	34	334	157	19	168		8	9	34	11	64	9	2	9		4	17		3	37						1	
d	524	16	73	5			14		133	7	1				1				57						2					18	
l	334		225	15			63	2	23	35		83					3		76				1	2						83	
c	161	31	46			1	43	3	27	195		19	14				9		3				59							7	
p	67	157	142	1			1	18	15	87		23		51									1							6	
m	243	26	78	1		18	48	1	22	38				26	49						26									1	
v	15	19	25				29		2	55																				1	
q									212																						
f	42	5	41	2	1		54		14	17	1	1		1				17												9	
'	66		58				32		25	9			3										7								
g	94	19	12			1	13	1	17	6		16																		2	
b	2	33	13	1	4		15		6	18		33			3															2	
&																															136
h	42		37				1		6	23														3						1	
y	34		11	1										3																58	
j	8		8						72	14									4												
x	7		1				9	4				1	2	2	2															52	

	e	r	a	s	f	n	i	t	u	o	d	l	c	p	m	v	q	f	'	g	b	&	h	y	j	x	z	k	4	2	.
z	7	1			1																										22
k	1	1	2											1	5				1	3		1			1						
4																															1
2																													1		
.	217	142	38		288	7	58	57	36	52	75	521	31	372	163	63	151	85	93	36	45	136	29	4	93			16	1		

Latin :

	e	a	r	i	o	v	c	l	d	t	m	b	g	p	n	h	j	s	-	k	f	4	q	x	z	y	.
e		73	452	26	31	32	15	15	6	228	226	19	11	32	345	5	26	22			34		26	89			619
a	227		14	1		143	97	114	95	392	189	76	48	36	226	1	3	75			7		6	2			282
r	39	235	3	386	199	189	11		33	64	11	7	2	6	12	2		41			7						179
i	51	161	36	81	253	13	198	94	8	266	143	228	51	72	382	9		414			26		28	7			35
o	6		299		1	13	61	66	54	29	88	51	22	6	46	5		15			23		2	6			271
v	298	134	175	198	95	14	44	12	84	167	429	3	7	2	126		15	38			6			4			27
c	132	126	77	143	257	114	26	27		159					2	22							1				9
l	9	98		331	48	25		59		41		1	2	1				2									42
d	257	51		246	74	85			5		4			2	1	7	7	2					5				97
t	252	264	9	489	157	271				16							25						23				456
m	76	94		97	6	13			4	2	21	4		52	33			2					33				654

	e	a	r	i	o	v	c	l	d	t	m	b	g	p	n	h	j	s	-	k	f	4	q	x	z	y	·
b	64	17	76	7	7	118		9	1	6					1		3	17									38
g	59	91	84	49	7	31		4			6				29												
p	117	42	21	42	16	38		24		47				2		4		39									4
d	257	51		246	74	85			5		4			2	1	7	7	2					5				
n	226	14		215	152	83	81		158	347			93		1	2	5	144			8		18				147
h	12	51	7	32	46	2																					
j	5	8		3	2	92																					
s	173	88		21	57	183	75		12	168	11			27				97					29			2	
-																											3
k		1																									
f	58	31	9	76	7	26		1													32						
4																											1
q						4																					
x	8	16		21	1		7			5				18		5		1						2			26
z	3																										
y															2							1					
·	395	479	124	32	122	228	426	135	257	86	14	11	26	338	177	8	51	374	5	1	97		29	2	3		

Russe :

	О	И	А	Е	Ъ	І	У	З	Ы	Ь	Й	Э	Н	С	Р	В	Т	Л	М	К	Д	П	Я	Г	Ъ	Ч	Б	Х	Ж	Ю	Ш	Ц	Щ	Ф	-	V	.	
О	11	39	1	57			2	41			68		74	21 2	21 4	21	18 1	15 5	15 5	51	26	52	18	13		37	11	8	5	29	8	3	2	3	3		479	
И		3		1			1	82					14	12 2	25	28	11	8	14 3	61	13	5	5	11		35	9	68	9	1	14	12	9	14			67	
А	1	2	1	34		1	2	59			2		15 7	55	11 1	93	71	11	67	6	55	11	42	85		69	6	33	9	1	1	17	1	14	1		345	
Е	13	2	9	7				17			41		25 5	6	13 3	82	56	11 3	13 7	15	35	11	4	47		11	7	6	47	18	1	8	11	1			264	
Ъ		2												1		1								2													178	
І	6	93	14	12 1		2	8				33				3										12 6	1					36						3	
У		2		6				1			1		14	32	9	1	14	1	11	5	31	28	1	9	12	7	7	6	13	12	2		5	3	1		174	
З	13	2	35	36	49	8	13	4	15	9			64	6	1	2		2	3	1	33	4	17	3	2	1	3		3	4	3						6	
Ы				18							14		23	11	11	32	6	32	22	6		1	22			1		32	1		6	1					115	
Ь		1		5									23	18		2	1	1	8	8			5		2			3		3	3						114	
Й													2	17							2	1							1		2						136	
Э																				1								1										
Н	21 5	91	31	99	38	18	21	1	15	9			73	27	2		6			1	25		61	4	51	1	2		1	2		6		1				
С	73	45	56	92	49	45	18		2	37			9	36	3	5	38	76	3	16 9		9	97	1	37	3		8		3		3						

	О	И	А	Е	Ъ	І	У	З	Ы	Ь	Й	Э	Н	С	Р	В	Т	Л	М	К	Д	П	Я	Г	Ъ	Ч	Б	Х	Ж	Ю	Ш	Ц	Щ	Ф	-	V	.	
Р	38	99	17 3	17 7	15	36	36		24	12			21	24		8	6		41	7	13	1	14	13	48	1	4	8	3	5	2	1		3			2	
В	22 1	71	12 4	17	24	16	21	3	31	3			38	45	33	2		44	1			5	27	1	56		1				2						4	
Т	27 1	64	12 7	48	11 3	24	1		28	47			14	74	65	68	6		1	2	5		8		2	1	2			5	1	1	2	4				
Л	14 6	18 2	13	8	51	7	14		7	6			1	9			3	2	4		2	4	31	8	54			1	2	14		1					1	
М	92	93	54	82	21 1	9	68		13	5			18	1		1		13	3	2		1	12		65							2	1				1	
К	22 5	66	14 2	11	11	35	28			1			49	4	27	7	2	3	3	1					13				4		1							
Д	82	51	14	6	51	4	41		38	9			22	39	48	17	1	13	1	13	1	6	2		64	1					3	9					1	
П	28	42	2	41	2	3	12		1					3	18 8		2	14		2		1	1		3													
Я				3				37					43	6	13	1	26	7	13	13	3		6	4				11	8	2			4				339	
Г	26 3	21	28	15	7	5	15						2	1	4	2		15			73				4				2		1						3	
Ъ		1		15				13			2		19	46	9	24	3	43	9	42	16	6	4	2		3	4	5	3	2	2		11				151	
Ч	2	62	5	62		5	15			5			5		3		63	2							1												1	
Б	34	23	8	2	11	5	7		66				1	9	19			26					1		25			2						8			1	

	О	И	А	Е	Ъ	І	У	З	Ы	Ь	Й	Э	Н	С	Р	В	Т	Л	М	К	Д	П	Я	Г	Ъ	Ч	Б	Х	Ж	Ю	Ш	Ц	Щ	Ф	-	V	.
Х	37	2	8	14	4	1							1	13																							5
Ж		23	23	83	18	1	1						15	6					2	33				1		1	2										1
Ю							3							3		7			8			7		3	2	6					7					17	
Ш		19	4	31		7	1					5			1	1	4		1					1													
Ц	1		6	11	6	24	1		5									8						1													
Щ		9	12	43	1	6																															
Ф	4	6	7	1	2	8	1		1				3				3		1																		
-	2																						3													1	
V						1																															
.	28	44	88	36		9	71	99				2	26	37	13	43	12	64	13	26	12	46	33	16	1	12	12	24	46	7	5	8	1	3	1	1	
	3	7										3	3	2	7			1	1	6	9		9			6											

Grec :

	Ε	Α	Ι	Ο	Ν	Σ	Τ	Η	Υ	Ρ	Λ	Π	Κ	Ω	Δ	Μ	Γ	Θ	Χ	Β	Φ	Ε	'	Ζ	Ψ	.
Ε	48	53	297	79	327	231	14	14	15	116	15	117	79	79	28	53	78	25	35	15	5	4		2	5	46
Α	3		434	1	233	227	126		114	195	133	94	41		33	49	41	27	12	8	12	13		11	4	37
Ι	32	59	1	127	194	213	42	19		28	72	21	1	49	53	28	9	21	16	4	4	1	3	11	1	859
Ο	8		36	2	396	232	34	4	441	42	82	18	24	2	14	94	19	6	1	6	15	4			1	27
Ν	113	146	31	167	7		175	66	18					36	42			41								979

	E	A	I	O	N	Σ	T	H	Υ	P	Λ	Π	K	Ω	Δ	M	Γ	Θ	X	B	Φ	Ε	,	Z	Ψ	.
Σ	69	154	28	53		18	11	2	36			15	13	15		8	1	76	19	6	25					746
T	257	335	9	486			2	216	22	38				13									2			2
H	2		113	7	246	27	35			36	18	2	25		13	44	18	13	3			2				154
Υ	13	2		17	86	159	21	5		42	37	42	25	6	45	2	39	11	14	8		4			3	176
P	57	95	11	149	4	11	8	77	18	3		19	7	65	17	6	25	4	11	5		5	5			6
Λ	15	72	35	9				91	4		16	4	4	32		2		13	2	12	14		9			9
Π	136	127	86	145	1		12	21	7	81	23	3		14									11			11
K	8	285	16	86	7		1	19	8	41	18	4	1	23	8	1	1			1	1					3
Ω	1		73		252	95	31		12	27	3	24			5	21	2	1	2		1					59
Δ	278	41	68	52				67	3	18				19		3							9			9
M	248	76	61	55	15			69	1			8		11		6				8	4				5	
Γ	86	85	16	47				37	59	4	2		6	2		4	5		6							
Θ	85	81	9	17	4			47	18	35	1			1		4										
X	4	9	15	25				15	5	32				16		1		7								
B	6	5	28	3				4	1	1	1			3	4											
Φ	16	12	27	29				5	6	4	2			4				9								
Ε	23	23	12					2	2					1												11
,																										39
Z	1	1		12				1						6												

	E	A	I	O	N	Σ	T	H	Υ	P	Λ	Π	K	Ω	Δ	M	Γ	Θ	X	B	Φ	Ξ	'	Z	Ψ	.
Ψ	5	14							1																	
.	672	443	45	273	65	148	666	113	32	4	73	287	31	66	323	23	139	56	43	55	33	5		6	1	

V.9. Synthèse des résultats de l'analyse de voyelles sur corpus

L.	Inscription	Breton	Français	Latin	Russe	Grec
E	V.	V.	V.	V.	V.	V.
A	V.	V.	V.	V.	V.	V.
R	C. 2	C. non-coronale	C. non-coronale	C. non-coronale	-	-
I	V.	V.	V.	V.	V.	V.
O	V.	V.	V.	V.	V.	V.
ʒ	C. 1	-	-	-	C. coronale (ʒ)	-
V	C. 2	C. coronale	C. coronale	V.	C. coronale	-
C	C. 1	C. coronale	C. coronale	C. coronale	C. coronale (C)	C. coronale (Σ)
И	C. 2	-	-	-	V.	-
L	C. non-coronale	C. non-coronale	C. non-coronale	C. non-coronale	C. non-coronale (ɹ)	C. non-coronale (Λ)
d	C. 1	C. coronale (D)	C. non-coronale (D)	C. coronale (D)	C. coronale (ɹ)	C. coronale (Δ)
T	C. 1	C. coronale	C. coronale	C. coronale	C. non-coronale	C. coronale
M	C. 2	C. coronale	C. coronale	C. coronale	C. coronale	C. coronale
b	C. 1	C. coronale (B)	C. coronale (B)	C. coronale (B)	V. (β) ; C. coronale (β) ; C. coronale (β)	C. coronale (B)
G	C. 1	C. coronale	C. coronale	C. non-coronale	C. coronale (Γ)	C. coronale (Γ)
P	C. 1	C. coronale	C. coronale	C. coronale	C. non-coronale	C. non-coronale
D	C. 2	C. coronale	C. non-coronale	C. coronale	C. coronale (ɹ)	C. coronale (Δ)
N	C. coronale	C. non-coronale	C. non-coronale	C. non-coronale	C. non-coronale (H)	C. non-coronale
1	Variable	-	-	C. non-coronale	V.	-

h	V.	C. non-coronale (H)	C. non-coronale	C. non-coronale	C. non-coronale (H)	V. (H)
8	C.	-	-	-	V.	
Θ	C. coronale	-	-	-	C. coronale (Φ) ⁶³	C. coronale (Θ)
J	C. non-coronale	C. coronale	C. non-coronale	C. coronale	C. coronale (Ж)	-
S	C. coronale	C. non-coronale	C. coronale	C. non-coronale	C. coronale (C)	C. coronale (Σ)
-	C. 1	V.	-	-	-	-
K	Variable	C. coronale	V.	C. coronale	C. non-coronale	C. coronale (K)
F	Variable	C. coronale	C. coronale	C. coronale	-	-
+	Variable	-	-	-	-	-
'	Variable	V.	Variable	-	-	V.
ʏ	V.	-	-	-	-	-
6	Variable	-	-	-	C. non-coronale	-
4	C. 1	-	V.	C. coronale	C. coronale	-
Q	-	-	C. non-coronale	C. non-coronale	-	-
X	-	C. non-coronale	C. coronale	C. non-coronale	Variable	C. coronale
Z	-	C. non-coronale	C. non-coronale	C. coronale	C. coronale (3)	C. non-coronale (Z)
Y	-	C. coronale	C. coronale	V.	V. (ʏ)	V. (ʏ)
U	-	V.	V.	-	-	-
CH	-	C. non-coronale	-	-	-	-
C'H	-	C. non-coronale	-	-	-	-
Ъ	-	-	-	-	V.	-

63 La lettre Θ est en effet prononcée /f/, et absente du texte de Lomonosoff.

Ы	-	-	-	-	V.	-
Э	-	-	-	-	V.	-
Я	-	-	-	-	V.	-
Ђ	-	-	-	-	V.	-
Ц	-	-	-	-	C. non-coronale	-
Ш	-	-	-	-	C. non-coronale	-
П		-	-	-	C. coronale	C. coronale
Й	-	-	-	-	C. coronale	-
Ч	-	-	-	-	C. coronale	-
Щ	-	-	-	-	C. coronale	-
Ю	-	-	-	-	C. coronale	-
Ψ	-	-	-	-	-	C. non-coronale
Е	-	-	-	-	-	C. coronale
Ф	-	-	-	-	C. coronale	C. coronale
Ω	-	-	-	-	-	V.

Synthèse des analyses de détermination de voyelles avec l'algorithme OCPb. C : consonne ; V : voyelle. C. 1 : consonne coronale dans les corpus 1 et 3 mais non-coronale dans le corpus 2 (voir. II. 2). C. 2 : Opposé de C. 1.

L.	Inscription	Breton	Français	Latin	Russe	Grec
E	V	V	V	V	V	V
A	V	V	V	V	V	V
R	C	C	C	C	-	-
I	V	V	V	V	C	V
O	V	V	V	V	V	V
2	V	-	-	-	V (3)	-
V	C	C	C	V	?	-
C	C	C	C	C	V	C. (Σ)
И	C	-	-	-	V	-
L	C	C	C	C	C. (Λ)	C. (Λ)
d	C	C (D)	C (D)	C (D)	C. (Δ)	C. (Δ)
T	C	C	C	C	C.	C.
M	C	C	C	C	C.	C.
b	C	C (B)	V (B)	C (B)	V (Б); C (Б); C. (Б)	C. (Б)
G	C	C	C	C	C. (Γ)	C. (Γ)
P	C	C	C	C	C.	C.
D	C	C	C	C	C. (Δ)	C. (Δ)
N	?	V	C	V	C. (H)	C.
1	C	C	-	C	-	-
h	C	C (H)	C (H)	C (H)	C. (H)	V (H)
8	V	-	-	C	-	-
Θ	C	-	-	-	C. (Φ)	C. (Θ)
J	C	C	C	C	C. (Ж)	-
S	C	C	C	C	C. (C)	C. (Σ)
-	C	V	-	-	?	-
K	?	C	V	C	C.	C. (K)
F	C	C	C	C	-	-
+	C	-	-	-	-	-
'	C	V	C	-	-	C.

У	?	-	-	-	-	-
6	?	-	-	-	-	-
4	С	-	-	С	-	-
Q	-	-	С	С	-	-
Х	-	-	V	С	V	С.
Z	-	-	С	С	V (3)	С. (Z)
У	-	-	С	V	V (У)	V (У)
U	-	V	V	-	-	-
Ъ	-	-	-	-	С	-
Ы	-	-	-	-	V	-
Э	-	-	-	-	С	-
Я	-	-	-	-	V	-
Ъ	-	-	-	-	V	-
Ц	-	-	-	-	V	-
Ш	-	-	-	-	С	-
П	-	-	-	-	С	С.
Й	-	-	-	-	С	-
Ч	-	-	-	-	С	-
Ш	-	-	-	-	С	-
Ю	-	-	-	-	V	-
Ψ	-	-	-	-	-	С.
Е	-	-	-	-	-	С.
Ф	-	-	-	-	С	С.
Ω	-	-	-	-	-	V.

Synthèse des analyses de détermination de voyelles avec l'algorithme de Sukhotine.

V.10. Analyses sur le corpus inversé

V.10.a. Analyse fréquentielle

Caractère	Fréquence
Espace	20,70 %
E	11,56 %
A	8,87 %
R	7,53 %

Caractère	Fréquence
P	1,61 %
D	1,34 %
1	1,08 %
H	0,81 %

I	7,53 %
O	6,72 %
S	4,57 %
N	4,57 %
B	4,03 %
V	3,76 %
C	3,23 %
L	2,42 %
T	2,15 %
M	1,88 %
G	1,61 %

8	0,81 %
Θ	0,54 %
J	0,54 %
-	0,27 %
F	0,27 %
'	0,27 %
γ	0,27 %
+	0,27 %
K	0,27 %
6	0,27 %
4	0,27 %

V.10.b. Analyse par l'algorithme OCPb

- 0
 - 00 (consonnes non-coronales)
 - 000
 - 0000 : D R
 - 0001 : Θ
 - 001
 - 0010 : M N
 - 0011 : + 8 F J K V
 - 01 (consonnes coronales)
 - 010
 - 0100
 - 01000 : B T
 - 01001 : G
 - 0101 : - L
 - 011
 - 0110 : 4 S
 - 0111 : C P
- 1
 - 10
 - 100 : 1 E
 - 101
 - 1010 : H
 - 1011 : O
 - 11
 - 110 : I'
 - 111 : 6 A γ

V.11. Résultats de substitution mono-alphabétique

V.11.a. Breton

V.11.a.1. Corpus brut

vnodan
lnean bioreetmio
antneoserhax'ted
ges ameseserxikre +
smidesv
_oasi eKve
aropvanoga ydbo'Fec
lan oa
o i e ep __ vep e jaiej
iteh ad'eobdec ba-aopa
dthey ehla ne ilipetrper
i __n en
an gnivih o lzzkz
omiie mnirmtihf enoiah
ahto4 anmonyit_c
dandho ite gne r c
tros ner e _____ i
ta_r
rlanaflod
alneinio
izk6 seir
r'_
afb gif
_a_as
_a_tec
_ap__
____a
_ gnec
on
ofe_as_ __
__e_
o
tec
i__c
__osn
__nid
__n_

__de_

V.11.a.2. Corpus renversé

pnohan
dnean gioneesdio
ansneorenla'useh
ver adereren'i8ne +
rdiherp
_oari eKpe
anompanova yhgoj-ec
dan oa
o i e em __ pem e zaiez
isel ahueoghec gakaoma
hslen elda ne idimesnmen
i __n en
an vnipil o d1181
odiie dnindsilr enoial
also4 andonnis_c
hanhlo ise vne n c
snor nen e_____ i
sa_n
ndanardoh
adneinio
i186 rein
nu_
arg vir
_a_ar
_a_sec
am
____a
_vnec
on
ore_ar_ __
__e_
o
sec
i__c
__orn
__nih
__n_

__he_

V.11.b. Français

V.11.b.1. Corpus brut

qrutar
crear giuseenviu
arnreudeslajhnet
fed avededesjiyse +
dvitedq
_uadi ezqe
asumqarufa ytgu'kep
car ua
u i e em __ qem e xaieX
inel atheugtep ga-auma
tnleS elca re icimensmes
i __r er
ar friqil u cggyg
uviie vrisvnilb eruial
alnu4 arvurSin_p
tartlu ine fre s p
nsud res e _____ i
na_s
scarabcut
acreiriu
igy6 deis
sh_
abg fib
_a_ad
_a_nep
am
____a
_ frep
ur
ube_ad_ __
__e_
u
nep
i__p
__udr
__rit
__r_

__te_

V.11.b.2. Corpus renversé

vrular
trear fiuseedtiu
ardreunescabydel
qen atenesbigse +
ntilenv
_uani exve
asumvaruqa ylfu'jep
tar ua
u i e em __ vem e haieh
idec alyeuflep fa-auma
ldces ecta re itimedsmes
i __r er
ar qrivic u t11g1
utiie trisdicn eruiac
acdu4 artursid_p
larlcu ide qre s p
dsun res e_____ i
da_s
starantul
atreiriu
i1g6 neis
sy_
anf qin
_a_an
_a_dep
am
____a
_ qrep
ur
une_an_ __
__e_
u
dep
i__p
__unr
__ril
__r_

__le_

V.11.c. Latin

V.11.c.1. Corpus brut

qtanet
dtiet fvasiirbva
etrtiamisce-jrin
gim ebimimis-vxsi +
mbvnmq
_aemv ikqi
esapqetage onfa'_il
det ae
a v i ip __ qip i zeviz
vric enjiafnil fe-eape
nrciS icde ti vdvpirspis
v __t it
et gtvqvc a d11x1
abvvi btvsbrvch itavec
ecra4 etbatSvr_l
netnca vri gti s l
rsam tis i _____ v
re_s
sdetehdan
edtvtva
v1x6 mivs
sj_
ehf gvh
_e_em
_e_ril
_ep__
____e
_ gtil
at
ahi_em_ __
__i_
a
ril
v_l
__amt
__tvn
__t_

__ni_

V.11.c.2. Corpus renversé

btacet
ntiet gvasiimnva
etmtiarisdehomic
qir eniririshvfsi +
mncirb
_aerv iKbi
esapbetaqe ycga'-il
net ae
a v i ip __ bip i jevij
vmid ecoiagcil gexeape
cmdis idne ti vnpimspis
v __t it
et qtvbvd a n11f1
anvvi ntvsnmvdri itaved
edma4 etnatsvm_l
cetcdavmi qti s l
msar tis i_____ v
me_s
sneternac
entivtva
v1f6 rivs
so_
erg qvr
_e_er
_e_mil
_ep__
____e
_ qtil
at
ari_er_ __
__i_
a
mil
v__l
__art
__tvc
__t_

__ci_

V.11.d. Russe

V.11.d.1. Corpus brut

ГНЕВИН
МНОИН ДАЕСООРПАЕ
ИНРНОЕКОСЛИБЯРОВ
ЗОК ИПОКОКОСБАЧСО +
КПАВОКГ
_ЕИКА ОЙГО
ИСЕДГИНЕЗИ ЪВДЕ'ШОК
МИН ЕИ
Е А О ОД __ ГОД О ХИАОХ
АРОЛ ИВЯОЕДВОК ДИ-ИЕДИ
ВРЛОЖ ОЛМИ НО АМАДОРСДОС
А __Н ОН
ИН ЗНАГАЛ Е МЪЪЧЪ
ЕПААО ПНАСПРАЛН ОНЕАИЛ
ИЛРЕ4 ИНПЕНЖАР_К
ВИНВЛЕ АРО ЗНО С К
РСЕК НОС О ____ А
РИ_С
СМИНИНМЕВ
ИМНОАНАЕ
АЪЧ6 КОАС
СЯ_
ИНД ЗАН
_И_ИК
_И_РОК
ИД
__И
_ ЗНОК
ЕН
ЕНО_ИК_ __
__О_
Е
РОК
А_К
__ЕКН
__НАВ
__Н_

__ВО_

V.11.d.2. Corpus renversé

ПНЕЛИН
ВНОИН ЗАЕСОТВАЕ
ИНТНОЕРОСМИБЯТОЛ
ГОР ИВОРОРОСБАЧСО +
РВАЛОРП
_ЕИРА ОЙПО
ИСЕДПИНЕГИ њЛЗЕ'ЖОК
ВИН ЕИ
Е А О ОД __ ПОД О ХИАОХ
АТОМ ИЛЯОЕЗЛОК ЗИ-ИЕДИ
ЛТМОС ОМВИ НО АВАДОТСДОС
А __Н ОН
ИН ГНАПАМ Е ВЪЪЧЪ ОНЕАИМ
ЕВААО ВНАСВТАМР
ИМТЕ4 ИНВЕНСАТ_К
ЛИНЛМЕ АТО ГНО С К
ТСЕР НОС О_____ А
ТИ_С
СВИНИРВЕЛ
ИВНОАНАЕ
АЪЧ6 РОАС
СЯ_
ИР3 ГАР
_И_ИР
_И_ТОК
ИД
__И
_ ГНОК
ЕН
ЕРО_ИР_ __
__О_
Е
ТОК
А_К
__ЕРН
__НАЛ
__Н_

__ЛО_

V.11.e. Grec

V.11.e.1. Corpus brut

ΘΝΟΡΑΝ
ΚΝΕΑΝ ΒΙΟΣΕΕΤΓΙΟ
ΑΝΤΝΕΟΛΕΣΠΑΞΗΤΕΡ
ΧΕΛ ΑΓΕΛΕΛΕΣΞΙ8ΣΕ +
ΛΓΙΡΕΛΘ
_ΟΑΛΙ ΕΚΘΕ
ΑΣΟΜΘΑΝΟΧΑ γΡΒΟ'ΦΕΔ
ΚΑΝ ΟΑ
Ο Ι Ε ΕΜ __ ΘΕΜ Ε ΖΑΙΕΖ
ΙΤΕΠ ΑΡΗΕΟΒΡΕΔ ΒΑ-ΑΟΜΑ
ΡΤΠΕΨ ΕΠΚΑ ΝΕ ΙΚΙΜΕΤΣΜΕΣ
Ι __Ν ΕΝ
ΑΝ ΧΝΙΘΠΙ Ο Κ1181
ΟΓΠΕ ΓΝΙΣΓΤΠΦ ΕΝΟΙΑΠ
ΑΠΤΟ4 ΑΝΓΟΝΨΙΤ_Δ
ΡΑΝΡΠΙΟ ΙΤΕ ΧΝΕ Σ Δ
ΤΣΟΛ ΝΕΣ Ε _____ Ι
ΤΑ_Σ
ΣΚΑΝΑΦΚΟΡ
ΑΚΝΕΙΝΙΟ
Ι186 ΛΕΙΣ
ΣΗ_
ΑΦΒ ΧΙΦ
_Α_ΑΛ
_Α_ΤΕΔ
ΑΜ
___Α
_ ΧΝΕΔ
ΟΝ
ΟΦΕ_ΑΛ_ __
__Ε_
Ο
ΤΕΔ
Ι_Δ
__ΟΛΝ
__ΝΙΡ
__Ν_

__ΡΕ_

V.11.e.2. Corpus inversé

ΓΝΟΠΑΝ
ΡΝΕΑΝ ΧΙΟΣΕΕΛΡΙΟ
ΑΝΛΝΕΟΤΕΣΚΑΦΗΛΕΠ
ΘΕΤ ΑΡΕΤΕΤΕΣΦΙΒΣΕ +
ΤΡΙΠΕΤΓ
_ΟΑΤΙ ΕΨΤΕ
ΑΣΟΜΓΑΝΟΘΑ γΠΧΟ'ΖΕΔ
ΡΑΝ ΟΑ
Ο Ι Ε ΕΜ __ ΓΕΜ Ε ΞΑΙΕΞ
ΙΛΕΚ ΑΠΗΕΟΧΠΕΔ ΧΑ'ΑΟΜΑ
ΠΑΚΕΣ ΕΚΡΑ ΝΕ ΙΡΙΜΕΛΣΜΕΣ
Ι __Ν ΕΝ
ΑΝ ΘΝΙΓΚ Ο Ρ11Β1
ΟΡΙΕ ΡΝΙΣΡΛΙΚΤ ΕΝΟΙΑΚ
ΑΚΛΟ4 ΑΝΡΟΝΣΙΑ_Δ
ΠΑΝΠΚΟ ΙΛΕ ΘΝΕ Σ Δ
ΛΣΟΤ ΝΕΣ Ε_____ Ι
ΛΑ_Σ
ΣΡΑΝΑΤΡΟΠ
ΑΡΝΕΙΝΙΟ
Ι1Β6 ΤΕΙΣ
ΣΗ_
ΑΤΧ ΘΙΤ
_Α_ΑΤ
_Α_ΛΕΔ
ΑΜ
___Α
_ΘΝΕΔ
ΟΝ
ΟΤΕ_ΑΤ_ __
__Ε_
Ο
ΛΕΔ
Ι_Δ
__ΟΤΝ
__ΝΙΠ
__Ν_

__ΠΕ_

TABLE DES MATIÈRES

I. Données du problème.....	1
I.1. Contexte historico-géographique.....	1
I.2. Transcription du texte.....	2
I.3. Éléments symboliques de l'inscription.....	4
II. Analyses de l'inscription.....	5
II.1. Analyse fréquentielle.....	5
II.2. Algorithme de détermination de voyelles.....	6
II.3. Analyse phonotactique.....	9
II.4. Analyse de l'entropie du texte.....	9
II.5. Analyse de l'indice de coïncidence.....	10
II.6. Autres pistes d'analyse.....	10
III. Tentatives de décodage.....	10
III.1. Détermination <i>a priori</i> de langues candidates.....	10
III.2. Analyse des langues candidates.....	12
III.2.a. Analyse fréquentielle.....	12
III.2.b. Détermination de voyelles.....	13
III.2.c. Analyse phonotactique.....	14
III.2.d. Entropie.....	17
III.2.e. Indice de coïncidence.....	18
III.3. Diverses tentatives sans modification.....	19
III.4. Chiffrement par simple modification alphabétique.....	19
III.4.a. Simple renversement de lettres.....	19
III.4.b. Alphabet armoricain.....	21
III.4.c. Alphabet grec.....	23
III.4.d. Alphabet cyrillique ancien.....	26
III.4.e. Autres pistes alphabétiques.....	28
III.5. Hypothèses de chiffrement.....	29
III.5.a. Chiffrement mono-alphabétique.....	29
III.5.a.1. Breton.....	30
III.5.a.2. Français.....	30
III.5.a.3. Latin.....	31
III.5.a.4. Russe.....	31
III.5.a.5. Grec.....	31
III.5.b. Chiffrement poly-alphabétique.....	32
III.5.c. Chiffrement par transposition.....	32
III.5.d. Surchiffrement.....	34
IV. Mise en perspective.....	35
IV.1. Synthèse des rubriques « En bref ».....	35
IV.2. Discussion & suggestions pour les recherches prochaines.....	37
V. Annexes.....	42

V.1. Glossaire.....	42
V.2. Système de numérotation.....	42
V.3. Transcriptions précédentes de l'inscription.....	44
V.4. Sorties des analyses pour la détermination de voyelles sur l'inscription.....	46
V.5. Résultats de l'analyse phonotactique.....	48
V.6. Résultats de l'analyse fréquentielle sur les langues candidates.....	51
V.6.a. Breton.....	51
V.6.b. Français.....	51
V.6.c. Latin.....	52
V.6.d. Russe.....	53
V.6.e. Grec.....	54
V.7. Résultats de l'analyse de voyelles sur corpus de langues candidates.....	54
V.7.a. Breton.....	54
V.7.b. Français.....	55
V.7.c. Latin.....	57
V.7.d. Russe.....	58
V.7.e. Grec.....	59
V.8. Analyses phonotactiques sur les corpus de langues candidates.....	60
V.9. Synthèse des résultats de l'analyse de voyelles sur corpus.....	71
V.10. Analyses sur le corpus inversé.....	75
V.10.a. Analyse fréquentielle.....	75
V.10.b. Analyse par l'algorithme OCPb.....	76
V.11. Résultats de substitution mono-alphabétique.....	76
V.11.a. Breton.....	76
V.11.a.1. Corpus brut.....	76
V.11.a.2. Corpus renversé.....	77
V.11.b. Français.....	78
V.11.b.1. Corpus brut.....	78
V.11.b.2. Corpus renversé.....	79
V.11.c. Latin.....	80
V.11.c.1. Corpus brut.....	80
V.11.c.2. Corpus renversé.....	81
V.11.d. Russe.....	82
V.11.d.1. Corpus brut.....	82
V.11.d.2. Corpus renversé.....	83
V.11.e. Grec.....	84
V.11.e.1. Corpus brut.....	84
V.11.e.2. Corpus inversé.....	85