Transcripció del text del Manuscrit de Voynich

Autor:
Data de publicació: 17-05-2019

El principal misteri del MS Voynich és clarament la seva escriptura desconeguda. Aquest tema s'adreça des de tres aspectes diferents, en tres (sèries de) pàgines:

Una mirada al sistema d'escriptura, que descriu les seves principals propietats i les similituds i diferències amb altres sistemes d'escriptura coneguts;
Transcripció del text;
Anàlisi estadística del text de l'EM (subdividit posteriorment en cinc àrees).
Aquesta pàgina tracta la segona part, la transcripció del text. Resumeix els esforços històrics de transcripció i descriu l'estat actual de les coses. Aquesta pàgina és molt rellevant per a aquells que estiguin interessats en la interpretació del text Voynich MS, però es pot ometre per aquells que no ho són.

Anàlisi de text - Transcripció del text
http://www.voynich.net/reeds/

Introducció
El principal misteri del MS Voynich és clarament la seva escriptura desconeguda. Aquest tema s'adreça des de tres aspectes diferents, en tres (sèries de) pàgines:

Una mirada al sistema d'escriptura, que descriu les seves principals propietats i les similituds i diferències amb altres sistemes d'escriptura coneguts;
Transcripció del text;
Anàlisi estadística del text de l'EM (subdividit posteriorment en cinc àrees).

Aquesta pàgina tracta la segona part, la transcripció del text. Resumeix els esforços històrics de transcripció i descriu l'estat actual de les coses. Aquesta pàgina és molt rellevant per a aquells que estiguin interessats en la interpretació del text Voynich MS, però es pot ometre per aquells que no ho són.

Una secció a la part inferior d'aquesta pàgina explica on es poden trobar els fitxers de transcripció principals.

Propòsit
El terme transcripció s'utilitza aquí per descriure la conversió del text manuscrit del MS Voynich en un format (fitxer) llegible per ordinador. El propòsit d'aquesta és permetre que els programes informàtics analitzin el text, per exemple, per obtenir estadístiques o per facilitar la interpretació i la traducció ideal del text. La primera vegada que aquest tipus de transcripció es va exercir va ser per William Friedman després de la Segona Guerra Mundial, per a la seva transformació per les anomenades "màquines tabuladores".

Molts anys abans, als anys trenta, el P. Th. Petersen de la Universitat Catòlica d'Amèrica ja va realitzar una transcripció completa de la mà (és a dir, una còpia escrita a mà) de l'EM utilitzant una fotocòpia completa de l'EM i, en alguns casos, per transcriure les parts difícils, la EM actual. Aquest document, en el qual va afegir moltes anotacions interessants, encara es conserva a la col·lecció William F. Friedman de la Biblioteca Marshall a Lexington (Va) (1) .

Una transcripció s'ha de distingir clarament d'una proposta de traducció del text. L'únic propòsit d'una transcripció és representar cada caràcter escrit a mà en un fitxer llegible per ordinador de manera coherent. Realment no importa quin símbol s'utilitza per a quin caràcter. Des de la dècada dels quaranta, diferents persones han utilitzat convencions diferents o alfabets de transcripció.

Història

FSG

La primera coneguda transcripció de grans parts del MS Voynich va ser realitzada en la dècada de 1940 per l'anomenat First Study Group (FSG) de William Friedman, que es va descriure breument aquí . El grup treballava a partir de còpies fotostàtiques de les pàgines de l'EM, i havien definit un alfabet de transcripció acordat per tots els membres de l'equip. A causa del desig de secret de Friedman, durant molt de temps ningú fora del seu equip era conscient d'aquest exercici de transcripció o de l'alfabet que utilitzava.

La transcripció de Friedman mai no va ser una transcripció completa de tot el text del MS. Van decidir concentrar-se en el text lineal escrit en paràgrafs, i no preocupar-se amb els diagrames complicats amb etiquetes i text escrits en cercles. Això és molt clar a partir d'anotacions escrites a mà realitzades a les còpies d'origen que diuen que "no punxeu això" o "només copegueu això".

L'alfabet de transcripció FSG utilitza lletres majúscules i números. Té un mètode bastant inusual (segons els estàndards posteriors) per transcriure la "forca intrusiva", és a dir, utilitzant un símbol especial (Z) per al pedestal (ch).

L'esforç de la transcripció de FSG es descriu en detall a Reeds (1995) (2) . Jim Reeds va descobrir una còpia impresa d'aquesta transcripció a la biblioteca Marshall esmentada anteriorment i, juntament amb Jacques Guy, el va introduir en forma llegible per ordinador. El fitxer resultant està disponible encara per descarregar . Aquesta transcripció està lluny de ser completa per les raons esmentades anteriorment.

Bennett

El 1976, el professor William R. Bennett de la Universitat de Yale va publicar un llibre sobre l'ús de l'ordinador per analitzar certes propietats del llenguatge i del text (3) . En aquest llibre, va dedicar un capítol a una anàlisi del text Voynich MS, i per això va necessitar un alfabet de transcripció. L'alfabet de Bennett no s'ha utilitzat fora d'aquesta publicació. La transcripció es va fer a partir de fotografies i es va emmagatzemar en cinta de paper. Segons va informar Brumbaugh (4) , Bennett va rebre ajuda en l'anàlisi de Voynich MS per Jonina Duker, que aleshores era un estudiant de segon any (5) .

Currier

Aproximadament al mateix temps, el criptoanalista Prescott Currier va iniciar un important esforç de transcripció, usant un alfabet que va dissenyar independentment. Aquest alfabet utilitza les lletres majúscules AZ i els números 0-9 (és a dir, tots els 36). No representa alguns caràcters que fa l'alfabet FSG, i utilitza caràcters individuals per al que sembla que són caràcters compostos a el MS. Quan Currier va presentar els seus descobriments al simposi de 1976 del MS Voynich, Mary D'Imperio va suggerir que seria important que tots els investigadors utilitzessin un alfabet unificat. També havia començat a transcriure parts de l'EM usant el seu propi alfabet, que va abandonar a favor de Currier, i els dos fitxers es van fusionar. La transcripció combinada de Currier i D'Imperio i de l'alfabet de transcripció dissenyat per Currier s'han utilitzat fins als anys 90, també en els primers anys d'Internet. A continuació, es pot fer referència a la transcripció mitjançant l'abreviatura "CD".

La taula següent mostra els tres alfabets esmentats anteriorment.

Char	Bennett	FSG	Currier
Char	Bennett	FSG	Currier

D
4
4

Jo
Jo
Jo

O
O
O

IL
IE
G

S
8
8

IIL
IIE
H

G
G
9

IIIL
IIIE
1

Z
2
2

QI
IR
T

L
E
E

IIQ
IIR
U

Q
R
R

IIIQ
IIIR
0

CT
T
S

U
L
D

ET
S
Z

N
N (*)
N

H
H
P

M
M (*)
M

P
P
B

JO SÓC
IIIL
3

K
D
F

K
J

F
F
V

IK
K

CHT
HZ
Q

IIK
L

CPT
PZ
W

IIIK
5

CKT
DZ

X

(6)
6

CFT
FZ
Y

(7)
7

A
A
A

Y
Y
(n)

C
C
C

V
V
(v)

(*) Nota: Tiltman va utilitzar l'alfabet FSG, però en comptes de N i M va escriure IL i IIL.

Una cosa que apareix immediatament d'aquesta taula és que els diferents transcriptors no van estar d'acord amb allò que constitueix un caràcter únic del text Voynich MS, i això resulta ser un dels punts més difícils per a la transcripció.

Posteriorment, Jim Reeds va mostrar que molts caràcters del MS Voynich no poden representar-se exactament per cap dels alfabets existents. Hi ha alguns caràcters "rars" i, a més, hi ha allò que sembla ser lligats a diversos caràcters. Aquests caràcters tradicionalment s'han anomenat 'weirdoes', encara que aquí utilitzaré el terme "caràcters rars". Jim Reeds va produir un document amb una visió general d'aquests caràcters rars, cridant-los "X1" a "X128". A continuació es mostra un extracte d'aquest document poc conegut.

Transcripció a l'època de la WWW

Amb l'aparició de la "World Wide Web" i l'inici de la llista de correu Voynich, la transcripció de les parts que encara faltaven als esforços de Currier i D'Imperio es va continuar utilitzant l'alfabet de Currier. A aquest alfabet s'han afegit diversos caràcters amb minúscules. Aquests ja estan inclosos (entre parèntesis) a la taula anterior. Com a part d'aquesta nova transcripció, que s'anomenarà "Vnow", s'ha adoptat un format de fitxer per a aquest fitxer de transcripció, sobre el qual es dirà més a continuació.

En paral·lel a això, es va dissenyar un tipus completament nou d'alfabet de transcripció:

Frogguy

L'alfabet de transcripció 'Frogguy' va ser dissenyat per Jacques Guy el 1991. Utilitza caràcters en minúscula, números i signes diacrítics. En comptes de representar caràcters complets, representa els "traços" del guió del MS Voynich. Com a resultat d'aquest enfocament, aconsegueix produir la semblança més propera del text transcrit amb l'escriptura. Al mateix temps, alguns caràcters del MS que es pot suposar que són un sol caràcter estan representats per diversos usant aquest alfabet. Per la seva naturalesa, és el primer alfabet que permet representar correctament a molts dels caràcters lligats que s'han esmentat anteriorment. Jacques Guy també va introduir l'anomenada "regla de capitalització", el que significa que un caràcter que estigui connectat al següent ha de ser representat per una lletra majúscula. L'alfabet de transcripció de Frogguy i el seu ús s'expliquen amb més detall en una còpia conservada de >> una de les pàgines web de Jim Reeds . Pot ser que us interessi mostrar un exemple d'aquest alfabet:

daiin qotedy dar ochol osaror
8aiiv 40qpc89 8a2 octox osa2o2

Una altra contribució important de Jacques Guy va ser una eina que permetia la traducció dels fitxers de transcripció d'un alfabet de transcrició a un altre. Aquesta eina es deia BITRANS, es va escriure en Pascal i es va executar en DOS (encara ho seria). Podria manejar regles de traducció complexes, com les que es necessiten per a l'alfabet de Currier, on, per exemple, la traducció del caràcter i depèn del context.

L'alfabet Frogguy és l'aspecte més proper al text Voynich MS, i resol la principal inconsistència de l'alfabet de Currier, és a dir, que Currier sintetitza les cadenes de caràcters i consecutius en un sol caràcter, mentre que això no s'ha fet per a cadenes de caràcters consecutius. 6) . No obstant això, té un desavantatge clar en no ser molt adequat per al tipus d'anàlisi numèrica esmentada a la part superior d'aquesta pàgina, principalment pel seu ús freqüent de l'apostrofe, i en menor mesura per la barreja de lletres i números. . No s'han realitzat transcripcions significatives amb aquest alfabet.

Quan Gabriel Landini i jo vam iniciar una nova transcripció del MS, un projecte que vam anomenar "EVMT" en aquell moment, vam decidir que seria necessari un nou alfabet, però un basat en el principi de Frogguy.

Eva

Originalment, EVA era l'alfabet d'European Voynich, però més tard es va convertir en l'alfabet extensible de Voynich. Va ser dissenyat per ser similar a Frogguy, però per utilitzar només caràcters alfabètics. Va ser dissenyat per Gabriel Landini i jo mateix amb importants aportacions i suggeriments de Jacques Guy. El seu disseny formava part d'un esquema més ampli que incloïa:

La necessitat de poder representar el text sencer del MS Voynich, incloses les parts no transcrites anteriorment, per exemple, alguns text fora dels paràgrafs estàndard.

Estandardització del format del fitxer de transcripció, incloent-hi una manera estàndard per identificar la ubicació de tots els elements de text a l'MS (en els llocs anomenats a continuació).

Mantenir la compatibilitat amb l'ús de BITRANS

L'alfabet havia de ser un superconjunt de FSG, Currier i Frogguy, de manera que seria possible i reversible una traducció de BITRANS de tota la transcripció existent a Eva (7) .

Una manera clara d'identificar caràcters rars, com ara la llista feta per Reeds.

La capacitat de dissenyar una font de tipus True True que representa amb exactitud el text Voynich MS (8) .

"Eva bàsica" és el conjunt de caràcters en minúscula identificats i "Eva estesa" inclou la representació de tot tipus de caràcters rars. L'alfabet bàsic d'Eva s'ha triat de manera que el text transcrit sigui gairebé pronunciable. Aquesta excel·lent idea de Gabriel no era poder "parlar Voynichese", però fa molt fàcil al cervell humà reconèixer i recordar paraules transcrites.

Els caràcters rars es poden classificar en quatre categories:

caràcters singulars inusuals o rars
Ligaments d'un sol caràcter d'Eva bàsica
Ligaments que inclouen caràcters rars únics
Alguna cosa més

Tots ells van ser identificats en el transcurs de l'exercici de transcripció, que es basava en dos documents: la transcripció de la mà de Petersen de l'EM ja mencionada (vegeu la nota 1) i el "copyflo" de Yale (9) . Com va resultar, hi havia tres caràcters únics que es podrien anomenar inusuals perquè no apareixien en cap dels alfabets de transcripció prèviament definits, però que es van produir més de 10 vegades a la EM. Aquests tres caràcters: b , u i z van ser assignats a la seva pròpia lletra "Eva bàsica" (b, u i z respectivament). Aquí s'inclou la taula de Basic Eva:

Es va assignar un caràcter "alt ASCII" a caràcters rars individuals o rares parts de lligadures i es va definir la manera d'incloure-les en fitxers de transcripció de la manera següent: & 185; per al codi ascii 185. La imatge següent mostra tots aquests caràcters rars (10) .

Les ligatures dels caràcters es poden representar de dues maneres. La primera és la regla de capitalització anteriorment esmentada introduïda per Jacques Guy. Com mostra la taula de Eva bàsica anterior, aquest mètode permet una representació exacta del text utilitzant el tipus de lletra Eva True Type. Un caràcter en majúscula sempre es connecta a la següent. A més, hi ha els casos especials següents:

c / c i q / q ja es connecten a la dreta, de manera que no hi ha necessitat de fer-los amb majúscules
h / h i H / H també es connecten a l'esquerra
E / E es connecta a la part inferior.

Sovint, els caràcters Sh, cTh, cKh, cPh i cFh s'escriuen simplement com sh, cth, ckh, sph i cfh, respectivament, la qual cosa resulta que la seva interpretació a la font de True Type no és òptima.

La segona manera d'indicar les lligatures en els fitxers de transcripció, que és més intuïtiva per al lector humà, és incloure els caràcters connectats entre parèntesis. La taula següent mostra exemples d'això. La representació d'escriptura d'aquesta taula utilitza la regla de majúscules.

EVA	EVA amb majúscules	Ús de TTF
(ao)		
Ao		
Ao		
(ee)		
Ee		
Ee		
(cthh) ey		
cthhey		
cthhey		

(Ith)
ITh
ITh

(oy)
Oy
Oy

sh 1
Sh
Sh

(yk)
Yk
Yk

1 Exemple d'ús estàndard encara que no òptim.

Takeshi Takahashi
L'alfabet Eva, principalment en la seva forma bàsica, va trobar una gran recepció a la comunitat i va ser utilitzat pel japonès Takeshi Takahashi per produir una nova transcripció, també basada en el "copyflo" de Yale (vegeu la nota 9) , que era el primer que es podria considerar essencialment complet. Va utilitzar "Eva majúscula" per representar tots els caràcters de referència o pedestallats, per exemple, "Sh" per Sh . També ha actualitzat la seva transcripció al llarg del temps, basant-se en informes dels usuaris.

El fitxer de transcripció interlineal
Basat en un esforç inicial de Jim Reeds, Gabriel Landini va recollir les transcripcions més antigues importants esmentades anteriorment en un únic fitxer interlinear , el que significa que es van presentar juntes per línia. Tots es van convertir en Eva, ja que aquesta conversió és coherent i revisable, com es va esmentar anteriorment.

El brasiler Jorge Stolfi va agafar aquest fitxer interlinear i va fer un esforç molt important per millorar-lo, afegint transcripcions addicionals. Això va incloure la transcripció completa de Takeshi Takahashi (estat de 1999), on va convertir la capitalitzada Eva a minúscules. També va afegir les seves pròpies transcripcions i transcripcions parcials de diverses altres, per exemple, John Grove. Com a resultat, aquest fitxer interlinear s'ha convertit en la font de fet per a les dades de transcripció. A continuació, es coneix com a fitxer Interlinear Landini-Stolfi, amb l'abreviatura "LSI".

v101
Uns anys més tard, Glen Claston va iniciar una transcripció completa del MS Voynich i va idear un alfabet de transcripció que ell va anomenar Voynich 101 (aquí v101). L'alfabet v101 s'ha dissenyat per mantenir les combinacions de traços que li semblen signes individuals com a caràcters individuals. A més, distingeix entre diverses variants de caràcters que es van considerar un i el mateix en tots els alfabets anteriors. Va transcriure el text complet del MS Voynich usant aquest alfabet. A continuació es mostra la seva definició de l'alfabet i l'assignació a valors ASCII.

També per a l'alfabet v101 s'ha dissenyat un tipus de lletra True Type i, igual que el tipus de lletra Eva, permet la representació d'alta qualitat del text Voynich MS en documents electrònics (Word, Excel). També es pot descarregar aquest tipus de lletra al mapa del lloc . A continuació, aquesta transcripció s'indicarà amb l'abreviatura "GC".

Format de fitxer de transcripció

Per a l'anàlisi de les diverses transcripcions de text de Voynich MS, és molt desitjable que aquestes transcripcions estiguin contingudes en fitxers amb un format (estandarditzat) ben definit. Aquests fitxers haurien de ser "anotats", el que significa que inclouen les anomenades metadades, que no és el text en si, sinó informació sobre el text. Aquesta metadada ha d'indicar per a cada peça de text en què es troba el fulletó o la pàgina i on es troba a la pàgina. Es pot incloure informació rellevant addicional sobre les pàgines o el text. El programari d'anàlisi hauria de poder interpretar aquesta informació o ignorar-la.

En els primers dies d'Internet, quan s'estava ampliant la transcripció per part de Currier, es va acordar un format per a aquest nou fitxer. Aquest format es va ampliar en el transcurs de la transcripció per Gabriel Landini i jo. Va ser modificat per Jorge Stolfi i utilitzat en aquesta versió final en el recurs principal: el fitxer interlineal Stolfi-Landini.

Malauradament, les diferències són bastant significatives i el fitxer de transcripció v101 utilitza una altra representació. Només es poden donar algunes convencions generals aquí.

Símbol (s) Significat Comentari

(hash) al començament de la línia
Tota la línia és un comentari.
No s'utilitza a "GC".

{...}
Tot el text en claudàtors és igualment un comentari. Això pot aparèixer en qualsevol lloc d'una línia.
No s'utilitza a "GC".

<...>
Aquest és un indicador de lloc i apareix a l'inici de cada segment de text. Explica on es pot trobar aquest text.
Si no inclou un període (.), És l'inici d'una nova pàgina

[...]
S'utilitza per a la lectura alternativa o incerta. [ao] significa que pot ser un o un o, però el transcriptor no és cert. Les opcions poden estar separades per un | que és opcional en cas que només hi hagi dos, però obligatoris quan hi hagi més, per exemple, [r | s | d]. Si és possible, l'opció més probable hauria de ser la primera.
S'utilitza principalment a "Vnow" i "LZ".

.
Un espai de paraules (separador de paraules).
Utilitzat per tots

,
Un espai de paraules incert.
Utilitzat per la majoria

-
Intrusió del dibuix en el text.
En algunes transcripcions també s'utilitza per marcar el final d'una línia

= al final de la línia.
Final d'un paràgraf.

Utilitzat per tots

*
Un sol caràcter il·legible.
"GC" utilitza un?

L'estructura dels tipus de locus varia molt entre fitxers i fins i tot dins del fitxer "LSI", que no es pot proporcionar resum. El fitxer "GC" identifica loci per un número de línia o una identificació més descriptiva en altres casos.

Jorge Stolfi també va introduir la característica útil en el fitxer "LSI", on un únic lloc pot ser representat diverses vegades, tal com es transcriu per diferents persones. En aquest cas, l'indicador de locus és post-fixat per un "identificador d'autor", que és un caràcter en majúscules. Es troba dins dels <> claudàtors i precedit per un punt i coma.

Eina de transcripció
Per (pre-) processar un fitxer de transcripció per a l'anàlisi estadística, he desenvolupat una eina de línia de comandes anomenada VTT (Voynich Transcription Tool). Això es basa en el format del fitxer de transcripció del fitxer "LZ" i permet diverses operacions útils sobre el text, principalment l'eliminació dels diferents tipus d'anotacions i la selecció de pàgines en funció de les seves propietats.

Aquesta "opció de selecció de pàgines" es basa en la disponibilitat de comentaris específics a les capçaleres de pàgines que defineixen "variables de pàgina". Aquests s'utilitzen consistentment als fitxers "LSI" i "LZ". Sense entrar en tots els detalls, la següent línia en un fitxer de transcripció:

{$ L = A $ H = 1 $ I = H}
estableix les tres variables de pàgina L, H i I als valors A, 1 i H respectivament. (Això significa, respectivament: llenguatge currier = A, mà de currier = 1, tipus d'il·lustració = herbal). Es pot instruir l'eina VTT per incloure o excloure pàgines amb qualsevol combinació de variables establertes a determinats valors. Aquesta característica útil també s'ha inclòs al "Navegador d'informació Voynich" d'Elias Schwerdtfeger, que es mostra a continuació .

L'estat de coses el 2017
Per resumir l'estat dels recursos de transcripció disponibles per a l'any 2017, tenim èxits importants, però també àrees problemàtiques molt importants.

Assoliments
Hi ha disponibles diverses transcripcions independents del MS Voynich. Hi ha tres que estan gairebé complets:

La transcripció de Landini-Zandbergen ("LZ"), que no va ser publicada mai.
El fitxer Landini-Stolfi Interlinear ("LSI"). La transcripció de Takeshi Takahashi, inclosa en ella, és gairebé completa.
La transcripció v101 de Glen Claston ("GC").

A continuació es mostra una llista més completa dels esforços de transcripció.

Àrees de problemes
La transcripció del MS Voynich és particularment difícil, ja que no es pot llegir el text. Un no té ajuda del context. El major problema és decidir si dos caràcters gairebé semblants són dos, o dues representacions del mateix, on la variació només és causada per variacions d'escriptura. Mentre es transcriu, un és obligat contínuament a prendre aquestes decisions i és inevitable que siguin subjectius. Hi ha un problema similar per decidir sobre espais de paraules. De vegades, els caràcters es compensen una mica i és difícil decidir si hi ha un espai o no. Fins i tot la introducció de la coma per representar un espai incert no resol completament aquest problema. L'única manera d'avistar d'aquests dos problemes és que aquestes decisions s'adoptin de manera més objectiva, és a dir, mitjançant un programari (OCR), però estem lluny de ser capaç d'aconseguir-ho.

Un segon problema general és la manca de normes. Com es va esmentar anteriorment, els fitxers de transcripció tenen diferències força significatives en el format. No hi ha cap lloc on es recullen tots ells. Del fitxer interlinear també hi ha diverses versions.

Actualment és impossible presentar la transcripció "GC" en cap dels formats estàndard, ja que molts dels símbols que tenen un significat especial al fitxer "LSI" (i "LZ") són representacions dels caràcters de text Voynich MS a la Transcripció "GC". Cal trobar una solució a aquest problema.

Gairebé sempre és impossible repetir les anàlisis fetes per altres, ja que no sabem quines dades van utilitzar, i realment no hi ha cap manera clara de com podrien haver-lo especificat. El format (s) de fitxer utilitzat actualment està lluny de les normes de representació modernes. Ja fa més d'una dècada, Rafa? Prinke va suggerir l'ús de TEI (Text Encoding Initiative), però mai no va sortir. Hi ha alguns comentaris més sobre aquest tema general a la pàgina final d'aquest lloc i es proporciona una descripció més detallada en una pàgina independent.

Transcripció de text de pròxima generació

Nova iniciativa
Inventari del text transcrit
Com es va esmentar anteriorment, existeixen tres transcripcions gairebé completes del text. Per establir la cobertura del text Voynich MS per cadascun d'ells, he comparat els tres amb l'altre i amb les imatges digitals de l'EM. Això va mostrar que hi ha algunes inconsistències menors en el recompte dels diferents llocs , que es podrien resoldre.

Nou format de fitxer de transcripció

L'anàlisi va mostrar que la transcripció "LZ", que es va beneficiar fortament de l'obra de Theodore Petersen, és més completa. Només falta:

Els pocs caràcters de l'escriptura al marge dret de f1r (3 són visibles)
Les paraules curtes sota la pintura d'una fulla a f2r (en tots els fitxers existents).
Una entrada al marge esquerre de f11v (també falta a tots els fitxers existents).
Les paraules de l'escriptura Voynichese que s'inclouen al marge superior de f17r , només són visibles sota llum ultraviolada (igualment a tots els fitxers de transcripció existents).
Les dues paraules Voynichese de f116v , que només es van incloure a "LSI".

He creat un fitxer nou que inclou només la meva pròpia part de l'esforç de transcripció LZ i l'he actualitzat per incloure tots els 5389 llocs . A continuació es coneixerà com a fitxer "ZL" (Zandbergen-Landini) i reconec la important influència de Gabriel Landini en la creació de la meva pròpia transcripció.

Tots els loci han estat distribuïts en diferents tipus, classificats per dos caràcters. A continuació es mostra la distribució dels llocs sobre aquests tipus per a tot el MS (11) .

Escriviu Subtipus Significat Comptar

P
-
Tot el text en execució en paràgrafs
4130

-
P0
Text alineado a l'esquerra estàndard
3885

-
P1
Línies sangrades per mitja pàgina a causa de dibuixos o un altre text
72

-
Pb
Text dislocat en paràgrafs flotants lliures
134

-
Pc
Línies centrades en paràgrafs normals
17

-
Pr
Línies justificades a la dreta en paràgrafs normals
12

-
Pt
Títols del text del paràgraf normal
10

L
-
Totes les etiquetes i paraules o caràcters dislocats
1033

-
L0
Paraules o caràcters individuals no prop d'un element de dibuix
297

-
La
Etiquetes d'elements astronòmics o cosmològics que no són estrelles o etiquetes del zodíac
7

-
Lc

Etiquetes de contenidors a la secció farmacèutica
40

-
Lf
Etiquetes de fragments d'herbes a la secció farmacèutica
195

-
Ln
Etiquetes de les nimfes a la secció biològica / balneològica
63

-
Lp
Etiquetes de plantes completes a la secció d'herbes
3

-
Ls
Etiquetes d'estrelles
76

-
Lt
Etiquetes de "tubs i banyeres" a la secció biològica / balneològica
47

-
Lx
Peces individuals d’escriptura "externa (?)"
6

-
Lz
Etiquetes d’elements del zodíac
299

C
-
Tot escrivint al llarg dels cercles
84

-
Ca
Escriure en sentit antihorari al llarg dels cercles
1

-
Cc
Escriure en sentit horari al llarg dels cercles
83

R
-
Tot escrivint al llarg dels radis dels cercles
142

-
Ri
Escriptura cap a l'interior al llarg dels radis dels cercles
75

-
Ro
Escriptura externa al llarg dels radis dels cercles
67

Total

5389

Nou format de fitxer de transcripció

Per poder representar totes les transcripcions existents en un únic format de fitxer, utilitzant el seu propi alfabet de transcripció original, cal resoldre una sèrie de problemes, principalment relacionats amb el joc de caràcters de l'alfabet v101, que s'ha mostrat anteriorment .

La taula següent explica les convencions utilitzades per al nou format, que he anomenat IVTFF (format intermedi de fitxers de transcrició Voynich).

La descripció del format completa es pot trobar aquí .

(hash) al començament de la línia
Tota la línia és un comentari.

Amb el caràcter Aquest és un indicador de lloc i apareix a l'inici de cada segment de text. Explica on es pot trobar

aquest text. A continuació s'explica el format dels indicadors de locus .

...>
Amb el <* NOT * a la primera posició d'una línia. Tot el text entre els delimètres és un comentari. Això pot aparèixer en qualsevol lloc d'una línia.

.
Un espai de paraules (separador de paraules).

,
Un espai de paraules incert.

[...]
S'utilitza per a la lectura alternativa o incerta. [ao] significa que pot ser un o un o, però el transcriptor no és cert. Les opcions poden ser separades per un: el qual és opcional si només hi ha dos, però obligatòria quan hi hagi més, per exemple, [r: s: d]. Si és possible, l'opció més probable hauria de ser la primera.

{...}
Una lligadura de caràcters estàndard. Ús únicament amb els alfabets Eva i Frogguy

@nnn;
Un codi d'alta ascii. nnn ha d'estar entre 128 i 255

<->
Intrusió del dibuix en el text.

<$> al final de la línia.
Final d'un paràgraf.

/ al final de la línia
Aquest lloc es continua a la següent línia del fitxer. Aquesta línia també ha de tenir un / en la primera posició.

?
Un sol caràcter il·legible.

???
Un nombre incert de caràcters il·legibles.

Hi ha dos tipus d'indicadors de locus :

Un indicador de lloc sense un període significa l'inici d'una nova pàgina, en aquest cas f17r. No hi haurà text Voynich MS després d'aquest, però pot haver-hi un comentari, amb metadades referides a tota la pàgina.

Aquest és l'inici d'una peça de text del lloc "Ab". El valor de N ha d'augmentar monòtonament, començant per 1 per a cada pàgina. El significat de @ s'explica a la descripció detallada del format. El locus tipus Ab és com a la Taula que es mostra sota el títol "Inventari del text transcrit", anterior.

Igual que anterior, però identificant el transcriptor pel caràcter T.

Per preprocessar tots els fitxers en format IVTFF, l'eina VTT ha estat modificada i es diu IVTT.

Eina de transcripció Voynich

Ubicació dels principals recursos de transcripció
A continuació trobareu enllaços als recursos de transcripció originals que tinc coneguts i a les noves còpies en format IVTFF. Es presenten aproximadament en ordre cronològic.

Codi Descripció Alfabet Original Còpia en IVTFF Notes

FSG
Còpia de la transcripció de la FSG, creada per Friedman's First Study Group, en el format preparat per Jim Reeds i Jacques Guy.
FSG
>> Enllaç
versió beta
(12)

CD
La transcripció original de D'Imperio i Currier
Currier
>> Enllaç
versió beta
(13)

Vnow
La versió actualitzada d'aquest fitxer, realitzada durant els primers anys de la llista de correu Voynich MS.
Currier
>> Enllaç
Temporalment no disponible !
(14)

TT
La transcripció original de Takeshi Takahashi.
Eva
A un conjunt de pàgines separades del seu lloc web.

Versió 1999
(15)

LSI
El fitxer Landini-Stolfi Interlinear
Eva
A una pàgina >> a la pàgina de Stolfi amb enllaços a aquest fitxer en diversos formats comprimits.
versió beta (23/08/2017)

GC
El fitxer de transcripció v101.
v101
Còpia local .
- Ordre de locus original
- ordenats
(16)

ZL
El "Zandbergen" forma part de l'esforç de transcripció LZ.
Eva
-
versió 1b
(17)

El navegador d'informació >> Voynich d' Elia Schwerdtfeger permet extreure transcripcions individuals del fitxer "LSI", amb moltes opcions útils.

Notes
1 Està inclòs en els ítems 1615.1 i 1615.2 de la col·lecció William F. Friedman. Per obtenir més informació sobre la Biblioteca Marshall, vegeu aquí . 2 Vegeu Reeds (1995) . 3 Vegeu Bennett (1976) . 4 Vegeu Brumbaugh (1978) . 5 Estic agraït a Jonina Duker pels seus records d'aquests esdeveniments de fa 40 anys. 6 Per descomptat, no tenim cap manera de saber si una suposició és correcta o no. 7 D'altra banda, hauria de quedar clar que no totes les característiques d'una transcripció feta directament a Eva poden convertir-se en els alfabets de transcripció més antics de manera reversible. 8 Aquest tipus de lletra ha estat dissenyat per Gabriel Landini i s'utilitza àmpliament en aquest lloc web. Vegeu el mapa del lloc . 9 Una impressió en blanc i negre d'un microfilm fet a la dècada dels setanta, vegeu també aquí . 10 Gabriel Landini va afegir tots ells a la font True Type. 11 Vegeu també la pàgina sobre el sistema d'escriptura . 12 Extret del fitxer LSI i convertit de nou en l'alfabet FSG. Encara no s'ha comprovat completament l'original. 13 Convertit de l'original. 14 Extret del fitxer LSI i convertit a l'alfabet de Currier. Resulta que el fitxer LSI estava incomplet, per la qual cosa s'ha de tornar a fer. Important, ja que aquest fitxer va ser utilitzat per Reddy i Knight, i part d'ella per Hauer i Kondrak. 15 S'ha extret la versió de 1999 del fitxer LSI. Les versions actuals encara s'han de processar. 16 S'ha verificat el fitxer. L'únic canvi d'ordre de lloc està relacionat amb l'ordre de les pàgines, és a dir, que la pàgina de rosetes s'ha mogut entre f85r2 i f86v4. 17 La versió 1b del 24/09/2017 té tots els errors coneguts. Les versions abans de les 0h tenien la majoria de textos ciculars llargs corromputs.

Copyright René Zandbergen, 2018
Comentaris, preguntes, suggeriments? Els vostres comentaris són benvinguts.
Última actualització: 02/02/2018