



TFG

DECOFICICANDO LAS IMÁGENES.
ANÁLISIS DE LA TRASCRIPTIÓN PICTORIOGRÁFICA
USANDO EL MÉTODO SIMBOLÓGICO DEL CODEX SERAFINIANUS

Presentado por Carles Monzó i Rubió
Tutor: Juan Carlos Domingo Redón

Facultat de Belles Arts de Sant Carles
Grado en Bellas Artes
Curso 2016-2017



**UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA**



UNIVERSITAT POLITÈCNICA DE VALÈNCIA
FACULTAT DE BELLES ARTS DE SANT CARLES

"¿Quiénes fueron los que inventaron Tlön? El plural es inevitable, porque hemos rechazado por unanimidad la idea de un solo creador, algo trascendental Leibnitz trabajando en la oscuridad modesta "

- Jorge Luis Borges, "Tlön, Uqbar, Orbis Tertius"

PREAMBULO

El propósito de este estudio ha sido presentar un nuevo método de investigación de un sistema de escritura no descifrada usando un ordenador, mediante el análisis de las imágenes escritas como imágenes visuales, en lugar de depender de un esquema de la transcripción para asignar caracteres escritos a los números. El ordenador puede extraer y organizar cientos de imágenes escritas más rápida y fiable que un ser humano puede, mientras que la supervisión humana apropiada puede convertir estos datos en puntos de vista sobre el sistema de escritura, evitando los problemas potencialmente introducidos por la capa intermedia de un esquema de la transcripción. El estudio pone en práctica varias aplicaciones que demuestran este principio, utilizando el Codex Seraphinianus como corpus, incluyendo un clasificador tipo y un motor de búsqueda. Las implementaciones son capaces de identificar secuencias recurrentes de fichas en el corpus y proponer una clasificación de fichas en tipos, dos principales sub-problemas en desciframiento. Específicamente, las fichas de cabecera en los tres primeros capítulos se clasifican en 53 tipos. El examen de los resultados arroja datos que apoyan las recientes declaraciones de Serafini que la escritura es artística, no lingüística. El carácter automático de la escritura se examina brevemente en vista de los resultados, anima a las generaciones futuras a descifrarlo.

TABLA DE CONTENIDO

LISTA DE MESAS	v
LISTA DE FIGURAS	vi
DEFINICIÓN DE TÉRMINOS.....	1
NOTA SOBRE EL ENFOQUE.....	1
INTRODUCCIÓN.....	1
Desciframiento y la transcripción de esquemas.....	1
Los problemas con los esquemas de transcripción.....	2
Ordenadores y Transcripción.....	3
El papel de los ordenadores en el desciframiento.....	4
Un problema de desciframiento Asistido por Ordenador	4
Propuesta para resolver el problema.....	4
El Codex Seraphinianus: un sistema de escritura misteriosa.....	6
La escritura asémica	8
La escritura automática.....	9
Declaración de TFG.....	12
TFG primaria	12
Objetivos secundarios.....	12
HISTORIA DE LA ESCRITURA.....	13
El análisis asistido por ordenador de sistemas de escritura no descifrada.....	13
Escritura Indus	13
Rongo Rongo.....	15
Linel A	¡Error! Marcador no definido.
El desarrollo de metodología utilizando un lenguaje conocido.....	19
El reconocimiento por ordenador de imágenes escritas	21
Impreso Reconocimiento de Caracteres	21
Reconocimiento de escritura a mano.....	23
Epigrafía	24
Escritura automática	25
Glosolalia y hablar en lenguas	26
Otros procesos relacionados	27
Resumen	28
MATERIALES.....	28
Hardware y sistema operativo	28
Entorno de desarrollo.....	28
Software de terceros	28
Cuerpo.....	29
Métodos y resultados	30

Texto de cabecera (“letras mayúsculas”)	30
Preparación para la extracción	30
Extracción	31
Identificación de fichas similares	32
Estudio 1: Consolidación en Tipos	34
Estudio 2: motor de búsqueda	43
Cuerpo de texto (“letras minúsculas”)	50
Preparación para la extracción	51
Extracción	51
Estudio 1: Palabras Ordenadas por Longitud	53
Estudio 2: motor de búsqueda	56
Discusión de los resultados	61
Observaciones lingüísticas en la escritura	61
Cabeceras formulaicas	61
Variación libre dentro de un tipo de	62
No repetición de las palabras	64
Resumen	66
Observaciones metodológicas sobre los estudios	67
Resultados	¡Error! Marcador no definido.
Problemas	67
CONCLUSIÓN	68
Desciframiento asistida por ordenador	68
El Codex Seraphinianus	68
Áreas de estudio adicional	69
Referencias	71

LISTA DE MESAS

Tabla 1: Patrones en las cabeceras del Codex.	
..... 37	
Tabla 2: Tipos que se encuentran en los capítulos 1-3, con los nombres y transcripciones sugeridas.	37
Tabla 3: palabras o partes de palabras hechas de tipos comunes en las cabeceras que son únicas.	47
Tabla 4: tabla que muestra los resultados del motor de búsqueda para una palabra objetivo, páginas 10-65.	56
Tabla 5: Tipos de variación en tres de los tipos de cabecera.....	60

LISTA DE FICHAS

Ficha 1: El Codex Seraphinianus.	6
Ficha 2: La secuencia de comandos Indus. 13	13
Ficha 3: La secuencia de comandos en Rongo-rongo. 15	15
Ficha 4: Lineal A. 17	17
Ficha 5: Grafemas marcianos de Smith procedentes de la Flournoy.	25
Ficha 6: Comparación de fichas de cabecera. 33	33
Ficha 7: Aplicación de una matriz de espacio en blanco para el último ejemplo de la figura anterior. ...	34
Ficha 8: App Typing que muestra los resultados para los tres primeros capítulos del códice.	36
Ficha 9: Resultados de la búsqueda de la secuencia común SFA.	43
Ficha 10: Como el anterior pero con Max Dist establece en 1.	44
Ficha 11: Ejemplo de una página de ilustración (13).	46
Ficha 12: Fichas pre-tfg que aparecen en páginas ilustradas en el capítulo uno sobre con plantas similares.	46
Ficha 13: Las muestras de la página 12 del texto del Codex que muestra mezclada con (a) los diagramas, (b) las tablas, (c) los compartimentos de dos líneas, y (d) trazos no de palabras.	49

Ficha 14: una aplicación que muestra todas las instancias de palabras en una página ordenados por longitud, aplicado a la página 13 del códice.	51
Ficha 15: Clasificación de la palabra casos en la página 13 por la longitud	52
Ficha 16: En la parte superior de la palabra resultados son errores de extracción.	53
Ficha 17: Demostración de motor de búsqueda de texto del cuerpo y de primer plano (abajo) de las mejores partidos, páginas 10-48.	54
Ficha 18: Dos fichas cerca de la parte superior de la página 52 que son similares pero no son detectados por el algoritmo de búsqueda, ya que varían en longitud.	57
Figura 19: Fichas en las cabeceras del Codex que parecen difieren sólo por un bucle cerca de un punto final.	60
Figura 20: Ornamentados de fichas cerca uno del otro en la misma página. .	61
Figura 21: Cuatro posibles frases de la página 52.	62

DEFINICIÓN DE TÉRMINOS

En este trabajo, un símbolo cuando se puede identificar como una unidad reutilizable y consistente (como la letra A en el alfabeto latino) se llama un tipo, mientras que una instancia individual que aparece en una muestra por escrito se llama un token. Estos términos son más o menos equivalente a la idea de un grafema y un aloinjerto.

NOTA SOBRE EL ENFOQUE

Se puede suponer con seguridad en este punto, para tomar Serafini en su palabra, que la escritura en el Codex Seraphinianus no codifica una lengua. (Esto se discutirá más adelante.) Sin embargo, en la metodología de este estudio tratará a la escritura como si la pregunta siguiese sin resolverse, con el fin de seguir siendo relevante a la palabra real escrituras sin descifrar que presentan este problema, tales como las secuencias de comandos Indo y Rongo . Por otro lado, cuando se habla del código específicamente, el estudio tendrá en cuenta su escritura asemica.

INTRODUCCIÓN

Desciframiento y la transcripción de esquemas

El mundo está lleno de sistemas de escritura que dejan como misterios para los estudiosos que vienen después, los sistemas que representan las culturas perdidas como el lineal A y B de Creta, el Rongo Rongo de Isla de Pascua, los jeroglíficos egipcios y mayas que se escribieron; y sistemas de escritura que fueron diseñados como misterios desde su concepción. De esa manera, desde el Manuscrito Voynich, estudios de informática en la escritura no descifrada han crecido popularmente, como lo demuestra, por ejemplo, por los recientes debates sobre la escritura del Indo y las piedras pictas (Rao 2009a; 2009b; Lee 2010; Sproat 2010) y por un aparente éxito en el MIT (Snyder 2010; Hardesty 2010). Todos estos casos y otros serán analizados en este documento. Para organizar estos sistemas de escritura con caracteres de manera

diferente a cualquier escritura conocida, los eruditos crean esquemas de transcripción que se asignan a cada personaje aparentemente único un número.

Los estudiosos pueden fácilmente hacer un seguimiento de estos caracteres y compartir sus ideas acerca de la escritura sin describir o dibujar las características reales de la escritura.

Los problemas con los esquemas de transcripción

Un esquema de la transcripción, si bien son útiles porque introducen un nivel de indirecticia entre la escritura y el conocimiento. Cuando los símbolos para ser documentados caen limpiamente en un conjunto manejable de tipos, como las veintiséis letras del alfabeto utilizado para idiomas occidentales, el esquema de la transcripción proporciona una manera rápida y fácil para que se represente. En otros casos, ya que es creado por personas que todavía no entienden lo que están tratando de representar, esta capa intermedia puede realmente obstaculizar la investigación, proporcionando una incorrecta traducción o limitando un marco para el estudio de la escritura original. En algunos casos, el esquema de la transcripción se convierte en un tema constante de debate, teniendo tiempo lejos del estudio de la escritura original. El esquema de la transcripción para el Rongo-Rongo, por ejemplo, está plagado de problemas debido a la naturaleza fluida de los caracteres, que parecen morfinstacticamente fluidos y se combinan de manera que no se entienden, difícilmente compatible con la naturaleza rígida de un mapeo numérico. Por último, a veces las transcripciones se hacen de forma incorrecta, pero se difunde su uso en lugar del texto original, ya que el texto original no está ampliamente disponible en buena calidad o simplemente porque es más fácil de manejar, lo que lleva a malentendidos que afectan a los investigadores que trabajan con el transcrito material (cf. Pozdniakov 2007, Guy 2005, Sproat 2003).

Ordenadores y Transcripción

Unicode se ha convertido en el estándar para la representación de secuencias de comandos del mundo. En el pasado, los investigadores tuvieron que crear sus propias fuentes o encontrar fuentes de casos especiales para los guiones que querían estudiar, lo que resulta en muchos casos, fuentes pequeñas, no estandarizadas. Unicode proporciona un único estándar para la representación de un centenar de pruebas diferentes, incluyendo el alfabeto latino, devanagari, runic, e incluso Lineal B, una antigua escritura de Creta que se entiende sólo en parte (Allen, 2009: 181).

Básicamente, Unicode proporciona un estándar para representar los caracteres, pero el usuario final debe tener una fuente capaz de mostrar los caracteres de acuerdo con estas normas. Es la presencia del carácter lingüístico, no las cualidades visuales de la escritura original, que es importante para las aplicaciones que se benefician de Unicode. Para ilustrar esto, hay que considerar el sánscrito (la escritura devanagari), que aparentemente está compuesto por un conjunto finito de una variedad alfabética de tipos, pero se hace más complejo gráficamente por las conexiones entre ellos. Los vocales del sánscrito están unidas a la consonante que siguen. Cuando dos o más consonantes en sánscrito se colocan en secuencia, se pueden combinar para formar una nueva representación gráfica. Esta duplicación de un carácter también modifica la representación (Coulson 1992: 4-19). Estos son sólo algunos ejemplos. Sin embargo, ninguna de estas consideraciones presenta problemas especiales en la transcripción informatizada. Esto se debe a que es el contenido lo que importa, no las fichas gráficas reales que fueron escritas en algún manuscrito original. El software que interpreta la transcripción decide cómo mostrarlo. De hecho, algunas fuentes capaces de mostrar la secuencia de comandos son capaces de mostrar consonantes combinadas (llamadas ligaduras), mientras que otros no pueden (Chopde (2001).

Lineal B se puede representar por Unicode hoy porque los signos de Lineal B son distintos y los tipos son identificables; ya no es vital para comunicar las fichas originales en toda su variación gráfica en el corpus Lineal B. Esto no es factible para un sistema de escritura desconocida. Sólo la superficie gráfica es observable. La

estructura subyacente no se conoce. Esto se aplica particularmente a un sistema de escritura que no se rompe fácilmente en un pequeño subconjunto de fichas distintas.

El papel de los ordenadores en el desciframiento

Recientemente, los ordenadores están jugando un papel importante en el estudio de los sistemas de escritura misteriosos. Muchos estudiosos han acercado a la escritura del Indo con los ordenadores; más recientemente Rajesh P.N. Rao ha realizado un extenso análisis de la escritura del Indo para mostrar que comparte algunas características con los lenguajes conocidos (Rao 2009a). Sproat (2003) utiliza un ordenador para identificar secuencias paralelas de fichas en el corpus del Rongo-Rongo.

El corpus para las escrituras sin descifrar y semi-descifrado están apareciendo en línea tanto en forma fotográfica y transcrito, lista para el análisis informático (Guy 2005; Younger 2010), por lo que se puede esperar que los equipos continuarán creciendo en este campo. Por otra parte, se puede esperar que no todos los lenguajes descubiertos sean capaces de traducirse, como una escritura olmeca encontrada recientemente demuestra (Rodríguez Martínez, 2006).

Un problema de desciframiento asistido por ordenador

Con el avance del ordenador en este campo, las imperfecciones del esquema de la transcripción están potencialmente explotados muchas veces a través de la miríada de cálculos que permite la tecnología, con ningún supervisor humano; o si hay un supervisor humano, el esfuerzo de verificar cada cálculo es laborioso. Esto puede explicar por qué los ordenadores no han sido adoptados por más investigadores.

Propuesta para resolver el problema

Teniendo en cuenta el respeto que tienen los ordenadores a la imagen digital, no puede ser una solución obvia a este problema: Quitar el esquema de la transcripción. El

reconocimiento óptico de caracteres y el reconocimiento de la escritura están disfrutando de una precisión sin precedentes, con arqueólogos capaces de enseñar al equipo a reconocer antiguas inscripciones romanas ilegibles por los seres humanos. Esta misma tecnología se debe aplicar para avanzar en los ordenadores el siguiente paso en el desciframiento, dando lugar a sistemas que no son más propensos al error, pero mucho menos propensos a errores que los sistemas tradicionales. En resumen, los esquemas de transcripción de los escritos desconocidos son problemáticos, pero los ordenadores son totalmente capaces de organizar las imágenes escritas sin un esquema de la transcripción, mientras que una persona tendría que recurrir a un esquema de transcripción para hacer un seguimiento de todos ellos.

El Codex Seraphinianus: un sistema de escritura misteriosa

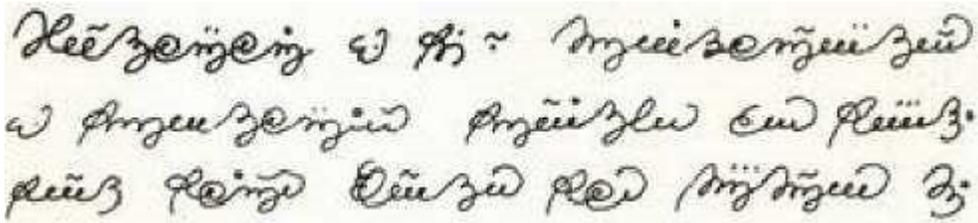


Figura 1: El Codex Seraphinianus.
 Top: páginas 12 y 13. Abajo: detalle página 12.

Creado en la década de 1970 por Luigi Serafini, un arquitecto italiano, el Codex Seraphinianus se ha creído por algunos ser una guía a otro mundo. Lleva más de 300 páginas de fantásticas ilustraciones en color y la escritura no se asemeja a ningún escrito conocido. Las imágenes representan todo tipo de cosas extrañas, de una flor que puede volar en un globo, árboles desarraigados de su raíz que pueden nadar río arriba para desovar, vehículos extraños, incluso reacciones físicas y químicas extrañas. En la portada de una edición popular, un hombre y una mujer copulando se transforman en un cocodrilo.

El *Códice* se organiza en once capítulos. Al comienzo de cada capítulo hay una tabla de contenidos. Hay dos tipos de tablas de contenido, uno que consisten principalmente en el texto a veces con pequeñas imágenes o diagramas y una que consiste principalmente de ilustraciones con leyendas, que en este documento se refieren como páginas de texto y páginas ilustradas respectivamente. En la parte superior de cada página hay un encabezamiento escrito en escritura misteriosa. Este texto de cabecera está hecho de grandes grafías desconectadas que son diferentes entre sí, texto en cursiva conectado que se encuentra en el cuerpo de las páginas. Sin embargo, algunos de los símbolos de cabecera se asemejan a algunas de las fichas del cuerpo, especialmente las fichas iniciales del cuerpo, lo que lleva a la gente a referirse a ellos como letras minúsculas y mayúsculas (Derzhanski, 2004).

Wechsler (1987) y Derzhanski (2004) descifraron de forma independiente el sistema de escritura utilizado para numerar las páginas del Codex, pero los estudios sobre el texto del encabezado y el cuerpo siguen siendo especulativos, con sólo unas pocas observaciones lingüísticas posibles, como la clasificación de las mayúsculas y letras minúsculas.

La escritura en el Codex ha sido elegida para este estudio por varias razones, como un compromiso entre los manuscritos sin descifrar y los formatos escritos, accesibles y conocidos por todos.

1. Los símbolos en el código no se clasifican y parecen estar formados de una manera compuesta con reminiscencias del Rongo-Rongo y algunos de los símbolos en la secuencia de comandos Indo y Linear A. fácilmente
2. El Codex mezcla texto con ilustraciones y el texto se presenta de diferentes tamaños y yuxtaposiciones, por lo que proporciona un reto en el cúmulo de material de origen primario.
3. El texto en el Codex se compila todo junto en un formato organizado de un modo que se puede denominar de una manera familiar, por número de páginas, capítulos,

y la ubicación en la página. Esto permite que se centren en su tesis sin una discusión excesiva en el corpus.

4. El Codex puede obtenerse como una compilación de escaneos electrónicos de alta resolución de color de las páginas. Por lo tanto, permite el estudio del material de la fuente original de una manera que está organizado, coherente y lista para el análisis del ordenador.

En el momento de iniciar esta investigación, la *Códice* presenta un buen ejemplo de un conjunto de datos debido a su organización, de su longitud, y porque, dado que su autor sigue estando con vida, pero sus labios siguen sellados, cabría la posibilidad de que no sea nula la posibilidad de que un día se verificasen los resultados.

La escritura asémica

Las expectativas para este estudio cambiaron en mayo de 2009, cuando se anunció a Serafini en una charla a los Bibliófilos de Oxford que la escritura en el Codex Seraphinianus no representa ningún idioma, conocido o secreto. Es asémica. En lugar de un sistema de escritura no descifrada, el enfoque del estudio se convirtió en un sistema de escritura asémica. Este desarrollo en realidad reveló nuevas razones para utilizar el Codex como corpus.

1. Al ampliar su gama de intereses, que tiene el potencial para demostrar el valor de la tecnología en áreas distintas de la lingüística: el arte o la psicología, por ejemplo.
2. Porque la escritura codifica ningún lenguaje, el Códice encarna efectivamente el peor de los casos. Si el estudio puede organizar y analizar este pseudo-lenguaje, se puede esperar a trabajar para los escritos que se conocen para codificar idiomas, así como los sistemas de escritura en disputa como el Rongo Rongo-y la escritura del Indo.

Recientemente, la escritura asémica ha ganado un pequeño grupo de seguidores como una forma de arte de vanguardia, pero sigue siendo un fenómeno esotérico. Tim Gaze, editor de *Asemic Magazine*, declaró en una entrevista que la escritura asémica es “algo que se parece a una forma de escritura, pero que no se puede leer” (Alexander 2009 (1)). Técnicamente, nadie puede leer la escritura verdadera asémica, ya que no representa un idioma. Sin embargo, mientras que no se puede leer en un sentido tradicional, que puede proporcionar una experiencia similar: “. . . Estoy tratando de crear cosas que son totalmente abiertas en su significado, sugerente a un espectador, pero sin un sentido privilegiado” (Alexander 2009 (1)). Otro líder en el campo, Michael Jacobson, conocido por la creación de una de 89 páginas de trabajo asémico llamada *The Giant’s Fence*, las llama de manera similar a que “un código que está abierto a la interpretación, sin significado fijo”. Explica que un emergente movimiento poético se centra en romper la manera tradicional de escribir y que la escritura asémica se descompone en las características rudimentarias de la textura y la línea.

Para los bibliófilos de Oxford, Serafini expresó este sentimiento. Como su charla no estaba registrada o publicada, la siguiente es una de las notas escritas por un asistente a ella.

El libro crea una sensación de analfabetismo, que, a su vez, estimula la imaginación, al igual que los niños que ven un libro: que aún no pueden leer, pero se da cuenta de que debe tener sentido (y que de hecho tiene sentido para los adultos) e imaginar lo que debe ser su significado. . . La redacción del Codex es una escritura, no es un lenguaje, aunque da la impresión de ser uno. Parece que quiere decir algo, pero no es así; que está libre de la jaula de un lenguaje y una sintaxis. Se trata de un proceso visual, no un proceso lingüístico (Prodi 2009).

La escritura automática

Además, Serafini afirmó que la experiencia de componer la escritura en el Códice fue similar a la escritura automática (Prodi, 2009). Para crear más de 300 páginas de texto puramente artístico, cuidadosamente deliberado a aparecer para codificar un lenguaje

parece insuperable, así que tiene sentido que la tarea tendría que ser o llegar a ser automatizada; pero ¿qué significa realmente el código de Serafini?

De *The Giant's Fence*, Jacobson comentó: “La principal dificultad con la escritura [del *The Giant's Fence*] sin embargo, era mantener el estilo consistente durante los 2 años que llevó a escribir el libro. Yo no creo que tenga la capacidad o la energía para volver a crear una obra como esa” (Alexander, 2009a). El Codex Seraphinianus es aproximadamente cuatro veces más grande (aunque la comparación es algo inexacta), y se terminó en poco más de dos años (Manetto, 2007). En una entrevista de 2007 de teléfono para El País, Serafini describió el Codex como el agua que brotó de él. Él hizo varias afirmaciones sobre la lengua, lo que implica no sólo que era principalmente estética, sino también que “me di cuenta de que estaba dejando el lápiz solo... Lo hice de repente. Es una visión, un lenguaje de los sueños. El misterio, para mí, es simplemente que radica en el hecho artístico” (Manetto, 2007). Se puede aceptar, entonces, que fue capaz de gestionar una obra tal porque la composición de la escritura, para él, era una especie de tarea automatizada, lo que requiere poco o ningún esfuerzo consciente. La idea de la escritura como resultado de una visión o un sueño no es nueva y se abordará más adelante.

La sección anterior proporciona una discusión de la investigación científica sobre la escritura automática y un fenómeno relacionado, la glosolalia o hablar entre lenguas. En este punto será suficiente para cubrir su faceta no científica. La escritura automática goza de una historia de la mística. A menudo se afirma por un espiritualista como un regalo, como una comunicación o inspiración de alguna otra entidad. Un caso famoso es el de Hélène Smith, conocida como la “Musa de la escritura automática”, que escribiría en lo que afirmaba era un lenguaje de Marte y luego lo traducía al francés. También podría hablar y crear arte de forma automática (Rosenberg, 2000). William Stainton Moses, un líder en la Iglesia de Inglaterra, se fascinó con la escritura automática después de experimentarla por sí mismo. En su libro sobre el tema que da cuentas de primera mano y corroborados como evidencia de una “inteligencia fuera de

un cerebro humano”, incluyendo los casos en los que el lenguaje era desconocido para el psíquico y en el que la escritura se produjo al parecer sin ninguna intervención humana (Moses, 1882). Dos textos religiosos que afirman haber sido escrita de forma automática (canalizado por guías espirituales) son los *Oahspe* por Newbrough y *Evangelio de Acuario* por Dowling. *Un habitante en dos planetas* por Oliver es sólo uno de los muchos libros del siglo 19 supuestamente escritos por antiguos habitantes de la Atlántida que tratan de impartir el conocimiento de su civilización destruida a través de las plumas de los vivos. Dos textos religiosos que afirman haber sido escrita de forma automática (canalizado por guías espirituales) son los *Oahspe* por Newbrough y *Evangelio de Acuario* por Dowling. *Un habitante en dos planetas* por Oliver es sólo uno de los muchos libros del siglo 19 supuestamente escritos por antiguos habitantes de la Atlántida que tratan de impartir el conocimiento de su civilización destruida a través de las plumas de escribir de los vivos.

Sin embargo, también hay manifestaciones menos nobles de la escritura automática. Junto con otra mediación sin procesos, algunos psicólogos freudianos utilizan la escritura automática para revelar el inconsciente. Se propuso por primera vez en 1878 que lo que se atribuye a los seres espirituales es realmente otro yo, disociado. En un caso seminal, Pierre Janet se dio cuenta de que, aunque su paciente Lucie no podía sentir ninguna sensación en su cuerpo, una personalidad que escribiría a través de ella automáticamente cuando se escribió en la firma “Adrienne,” y sintió un todo (Van der Hart, 1989).

El tercer movimiento en el que la escritura automática encuentra su hogar es un recoveco artístico. En su Manifiesto Surrealista, André Breton hace hincapié en todas las formas de automatismo, especialmente la escritura automática, como “. . . el funcionamiento real del pensamiento. . . en ausencia de todo control por la razón. . .” (Breton, 1972). Otros surrealistas bien conocidos, tales como Alexander Brotchie y Benjamin Péret y recomiendan escribir en ausencia de pensamiento controlado como un ejercicio valioso (Brotchie, 2003). Hoy “escritura libre” es un ejercicio de escritura popular no sólo entre los surrealistas.

Declaración de TFG

TFG primario

Al tomar ventaja de la capacidad de un equipo para tratar directamente con imágenes, aplicaciones sencillas y eficaces, éstas se pueden hacer para responder a las preguntas que los lingüistas hacen acerca de los sistemas de escritura sin descifrar sin los problemas potenciales de un sistema de transcripción.

Objetivos secundarios

En este caso, las aplicaciones producidas pueden arrojar luz sobre el sistema de escritura del Codex Seraphinianus, dando lugar a nuevos conocimientos sobre la naturaleza de la escritura y su relación con los sistemas de escritura conocidos. A su vez, esto podría arrojar luz sobre la naturaleza de la escritura automática y asémica en general.

HISTORIA DEL ESTUDIO

El análisis asistido por ordenador de sistemas de escritura no descifrada

Indus script

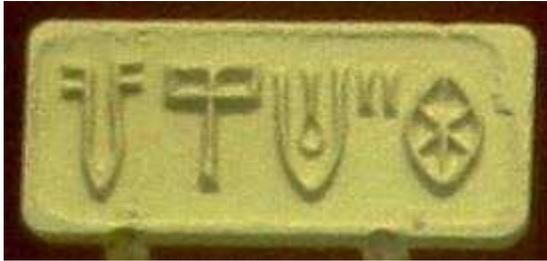


Figura 2: La secuencia de comandos Indus.
Sellar impresión del Museo Británico.

De todas las lenguas sin descifrar, la escritura del Indo (o Indus, también conocido como el alfabeto de Harappa) ha recibido la mayor atención asistido por los ordenadores. Se han hecho intentos de larga duración para usar ordenadores para analizar la secuencia de comandos, incluyendo Knorosov y Parpola independiente (Knorosov 1965; Parpola 1994), y recientemente Yadav y Rao (Por ejemplo, Yadav, 2008; Rao, 2009). Tales estudios utilizan esquemas de transcripción basados en la concordancia Mahadevan (1977), que identifica potencialmente 417 tipos únicos, de los cuales 40 sólo aparecen en contextos dañados o inciertos. El resultado de la mitigación de estos textos dañados y ambiguos es un corpus final de los 1548 textos cortos (líneas individuales de los sellos en su mayoría) y 7000 fichas (Yadav, 2008; Rao, 2009). Muchas señales parecen ser combinaciones de otros signos, pero en el esquema son tratados como señales independientes (Yadav, 2007).

Rao comparó el alfabeto Indus con secuencias de comandos conocidos (más transcritas a través de Unicode), particularmente con respecto a la entropía condicional, que es una medida de la flexibilidad en la ordenación de las fichas. Por ejemplo, un sistema en el que el tipo B debe siempre seguir el tipo A es rígido y tiene una baja entropía, mientras que un sistema en el que cualquier tipo puede seguir de

tipo A con igual probabilidad es flexible y tiene mayor entropía. Ellos calculan valores de entropía condicional para los sistemas lingüísticos y no lingüísticos conocidos para comparar con la escritura del Indo, y encuentran que la entropía condicional para la escritura del Indo es similar a la entropía condicional de los sistemas de escritura utilizados para representar las lenguas humanas naturales (Rao, 2009a). Esto concuerda con los hallazgos de Yadav en el que el orden de los signos no es al azar. Este estudio fue recibido con fuertes críticas en varios aspectos, a partir de datos de origen a malos saltos en la lógica; el debate en torno a la escritura del Indo es una mordaz enconada. Un estudio un poco más tarde usando un modelo de Markov ha tenido más éxito. Evita este estudio que compara la escritura del Indo a otros sistemas conocidos, como las reclamaciones relacionadas con estas comparaciones es lo que atrajo a la mayoría de la crítica (Patel, 2010). A partir del modelo de Markov, Rao y compañía son capaces de confirmar una vez más que la secuencia no es aleatoria y afirman que los símbolos parecen caer en clases de equivalencia funcional. Sugieren signos sin conservante de textos dañados utilizando el modelo (Rao, 2009b).

Recientemente, otro estudio ha utilizado la entropía como el de Rao (2009a) para proponer que los símbolos en piedras pictas pueden codificar una lengua (Lee, 2010). Una crítica detallada de estos estudios basados en la entropía se puede encontrar en Sproat (2010).

Rongo Rongo-

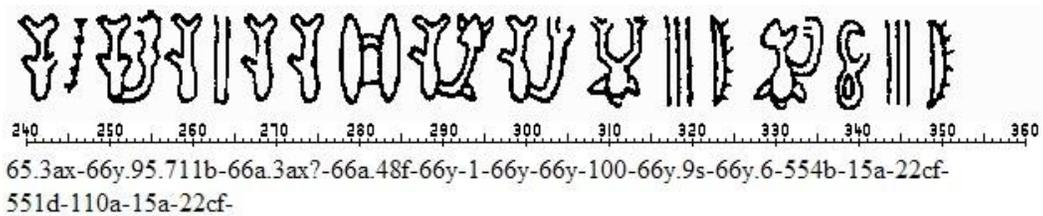


Figura 3: La secuencia de comandos Rongo-rongo.

Arriba: detalle de la tableta P. inferior: parte de los trazados de la tableta de Barthel, con la transcripción.

El Rongo Rongo-se refiere a los glifos misteriosos que se encuentran en artefactos de la isla de Rapa Nui (Isla de Pascua). La lengua hablada Rapa Nui es conocido en una

forma moderna través de algunos documentos de la época colonial, pero la escritura sigue siendo un tema peliagudo. El corpus es escaso y consiste sobre todo en tabletas marcadas en ambos lados de la moda bustrofedon. Los glifos parecen ser animales, perlas, agua, y muchos otros artículos; y son flexibles en el sentido de que, mientras que parece que hay conjuntos de características de composición como cabezas, posición de los brazos, y las colas, los glifos son el resultado de una miríada de configuraciones de estos rasgos. Unos pocos estudiosos observaron lo que llamaron textos paralelos a través de múltiples tabletas o dentro de la misma tableta, donde se repite la misma secuencia o casi la misma secuencia. En 2003, Richard Sproat, entonces en la Universidad de Illinois en UrbanaChampaign, utilizó un ordenador para buscar a estos paralelos, basándose en el trabajo sobre matrices de sufijo para la coincidencia de nexos. Aunque este trabajo nunca se publicó formalmente, está disponible gratuitamente en su página web y bien conocidos por los estudiosos del Rongo Rongo-. Para la entrada, Sproat utiliza transcripciones estándar de los comprimidos, a base de Barthel y modificados por la comisión *Rongorongo*. La transcripción de Barthel asigna 799 asignaciones de señales, con signos similares agrupados juntos, aunque algunas de las asignaciones se dejan vacías. Las diacríticas indican combinaciones o concatenaciones de señales, y las letras minúsculas denotan ciertas variaciones-signo específico (Pozdniakov 2007; Guy 2005). Sproat desliza estas decoraciones para obtener una entrada compuesta estrictamente de números. En los resultados, que se presentan algunas secuencias que son similares, aunque no necesariamente el mismo, debido a la distancia de su entrada desde el texto real. En su búsqueda de coincidencias “aproximadas”. Esta flexibilidad es deseable, ya que conduce a resultados que a continuación pueden ser examinados por los seres humanos. Se señala, no obstante, que en algunos lugares su investigación sufrió sin él saber siquiera que, al principio debido a deficiencias en los datos de origen, donde la transcripción simplemente no se completó.

El estudio más reciente del Rongo Rongo usando un ordenador (Pozdniakov, 2007) supera el trabajo previo de varias maneras. En él se identifican muchas

secuencias más paralelas; se propone alógrafos basadas en ellas, que presentan un inventario de 52 glifos (en lugar de los cientos originales), e investiga la estructura de las palabras y glifos estadísticamente en base a esta propuesta. Por desgracia, la principal conclusión del trabajo es que las posibilidades no se pueden restringir lo suficiente para formar cualquier tipo de desciframiento. Otros resultados son útiles, sin embargo. En particular, el pequeño inventario de glifos nos invita a una mayor investigación. Además, debido al alto número de secuencias paralelas, los autores del estudio sugieren que el contenido del corpus es poco probable que incluya distintos signos de mantenimiento de registros o similares tareas repetitivas nada fiables. La identificación de los glifos y sus alógrafos en este documento ha reunido un poco de crítica académica. Debido a que el método no se describe adecuadamente, los contraejemplos y los argumentos que los críticos están presentando no pueden ser comprobados contra él (véase Miller 2008).

Lineal A



Figura 4: Lineal A. *Detalle en una tableta de Pylos.*

Una de las primeras lenguas de la antigua Creta y la Grecia continental, Lineal A ha sido más fácil de tratar que otros lenguajes desconocidos. La mayor parte de los valores de los símbolos fonéticos se han proyectado hacia atrás desde la tardeoescritura Lineal B, descifrado en 1955, y estudiosos en general, aceptan el resultado como una transcripción adecuada. Además, el formato de las tablas del Lineal A que es similar a las tablas del Lineal B, por lo que los estudiosos son capaces de identificar símbolos pictográficos para el vino, grano, etc., así como identificar qué símbolos representan valores numéricos. Sin embargo, Lineal A tiene símbolos desconocidos (los que no responden a Lineal B), así como símbolos compuestos (ligaduras) que no se entiende completamente todo (aunque muchos producen fácilmente sus componentes, otros no lo hacen). Estos símbolos desconocidos son referidos por enumeraciones estándar al igual que en otras escrituras desconocidas. El problema con Lineal A no es tanto los valores de los símbolos, sino como identificar en qué idioma se supone que representa. Los estudiosos pueden sondear gran parte de la Lineal A, pero no saben lo que significan los sonidos. La mayoría de investigaciones Lineal A se ha preocupado por comparándolo con otros idiomas que surgieron anteriormente.

Packard (1974) fue el primero en publicar un análisis exhaustivo en ordenador de Lineal A, mediante el cálculo de las frecuencias para cada símbolo, con uno de sus objetivos es para verificar que las proyecciones del Lineal B son razonables. Al igual que en otros estudios desciframiento, se utilizó la transcripción numérica como entrada; pero debido a la situación única de Lineal A también se es capaz de analizar y presentar sus resultados organizados por el sonido fonético. Mientras hubo un corpus frustrado sin desarrollar en el tiempo, su procedimiento fue elogiado, y sus resultados parecen favorecer las asignaciones aceptadas (Packard, 1974; Chadwick, 1977). Packard también identificó secuencias paralelas y realizó algunos otros análisis. Recientemente, Brent Davis ha aplicado el procedimiento de Packard para un nuevo objetivo: mediante el cálculo de las frecuencias de símbolos en el contexto de la posición de la palabra, para identificar sufijos morfológicos. Estos elementos

morfológicos podrían compararse con otras lenguas para tratar de identificar el idioma de la familia Lineal A (Davis, 2010).

El desarrollo de metodología utilizando un lenguaje conocido

Para desarrollar metodologías que se pueden mostrar para ser eficaz, una opción es utilizar un lenguaje conocido, tratándolo como si fuera desconocido. En un intento exitoso reciente, Ugaritic fue tratado como una lengua desconocida. Utilizando estadísticas recogidas del hebreo, un sistema de lenguaje y escritura estrechamente relacionados, el equipo identificó 29 de 30 cartas y 60% de los significados de las palabras correctamente (60% de esas palabras que tienen cognados en hebreo). Para este estudio, en lugar de utilizar un sistema de transcripción alfanumérico, un sistema especial fue desarrollado para introducir símbolos Ugaritic utilizando un teclado moderno. Este sistema se da sólo una mención en un artículo de noticias; no hay detalles aparecen en el papel (Snyder, 2010; Hardesty, 2010).

Resumen

Aunque todavía no existen muchas lenguas sin descifrar, los que sí están recibiendo atención ordenadora. Idiomas existentes sin descifrar tienen grandes inventarios de símbolos y corpus dispersos, y un ordenador puede ayudar a organizar y minar estas amplias series de datos. Además, se puede esperar que no todos los idiomas, incluso los que se han descubierto, como una escritura olmeca encontrado recientemente demuestra puedan ser transcritos (Rodríguez Martínez, 2006).

Así como las demandas de una lengua picta (Lee, 2010).

Diferentes idiomas sin descifrar tienen diferentes problemas. Eruditos en el Rongo Rongo saben algunas cosas sobre el lenguaje Rapa Nui, pero no puede leer la escritura, mientras que en Lineal A los estudiosos pueden leer el escrito sin conocer su idioma. Sin embargo, hay problemas que comparten ambas lenguas. Los temas comunes en el desciframiento de ordenador incluyen la dependencia de un esquema de

la transcripción para proporcionar entradas manejables y la búsqueda de secuencias paralelas de símbolos para tratar de identificar las estructuras lingüísticas significativas. Otras preguntas se refieren a la clasificación de la lengua: ¿Es definitivamente una lengua? ¿A qué otro idioma se parece? Los ordenadores permiten el análisis estadístico que puede ayudar a responder a estas preguntas. Enumerado en beneficio de la claridad, aquí están algunas de las preguntas que se le preguntó acerca de los sistemas de escritura sin descifrar. Algunas son coincidencias. Para mostrar la importancia de secuencias paralelas, las preguntas de tal análisis que pueden ayudar a responder estas cuestiones están marcados con un asterisco (*). Cuando se citan las obras, no son más que ejemplos para ayudar a iluminar la naturaleza de la cuestión. Muchas otras referencias están disponibles en estos temas.

1. ¿Es un idioma? (Sproat, 2010; Rao, 2009a; 2009b Rao; Lee, 2010)
 - a. ¿Sus símbolos siguen las reglas que se encuentran en las lenguas conocidas?
 - Número de tipo y frecuencia
 - La ley de Zipf
 - entropía
 - Los modelos estadísticos
2. ¿Qué otros idiomas están más estrechamente relacionados con? (Packard, 1974)
 - a. ¿Qué sistemas de escritura no se asemejan gráficamente?
 - b. ¿Pueden ser asignados a los valores de los símbolos de una manera que parezcan crear palabras o estructuras que se encuentran en algún otro idioma?
 - c. * ¿Se pueden identificar las reglas o estructuras aparte de las reglas o estructuras paralelas en algún otro idioma?
3. ¿Cuáles son los tipos de sistema de escritura? (Coe, 1999)
 - a. * ¿Los tipos varían según el contexto o son compatibles?
 - b. * ¿Son los tipos distintos o pueden ser combinados para crear materiales compuestos o ligaduras? Si existen materiales compuestos, ¿cómo se descomponen?

4. ¿Es principalmente alfabético, silábico, o ideográfico? ¿Qué son los símbolos fonéticos y cuáles son ideográficos? (Chadwick, 2004)
 - a. ¿Cuántos tipos se encuentran en el sistema de escritura?
 - b. ¿Qué reglas de colocación no siguen los tipos?
5. ¿Cuál es la estructura morfológica de la lengua?
 - a. ¿Hay tipos más prominentes en una posición particular en palabras (por lo general principio o al final)? (Davis, 2010; Rao, 2009 b)
 - b. ¿Hay algún tipo que se encuentre comúnmente en conjunción con otros tipos?
 - c. * ¿Hay secuencias comunes que difieren en sólo uno o dos símbolos? (Chadwick, 2004)
6. ¿Es útil el esquema de la transcripción?
 - a. ¿Cubre todos los tipos en el sistema de escritura?
 - b. ¿Cubre la variación en los tipos?
 - c. ¿Representa las relaciones entre los tipos que podrían ser significativos?
 - d. ¿Cómo organizar y comunicar estas características de manera efectiva?

El reconocimiento por ordenador de imágenes escritas

Reconocimiento impreso de Caracteres

Los primeros estudios serios en el reconocimiento óptico de caracteres tuvieron lugar en la década de 1960, y hoy en día la tecnología ha llegado a un nivel tal que un documento mecanografiada o impreso puede ser escaneado y su contenido lingüístico digitalizado con una precisión reclamada de 97-99% (caracteres reconocidos correctamente) (Mori, 1999: xiii, 6, 7). El reconocimiento de caracteres de hoy sigue generalmente tres pasos (Mori, 1999; Cheriet 2007):

1. Preprocesamiento: La imagen se prepara para el análisis.
2. Extracción de características: Identificación de las características de la imagen y del carácter que recoge. Una de las decisiones más importantes en el reconocimiento de caracteres es lo que se tiene en cuenta para decodificar; los

investigadores han explorado la optimización de la selección a través de procesos automatizados.

3. La coincidencia de patrones: Las características que se utilizan para clasificar la imagen. Esto puede ser por un árbol de decisiones, una red neuronal, o cualquier otro algoritmo de toma de decisiones. Las características utilizadas para el reconocimiento de caracteres pueden ser cualquier cosa que ayude a describir la ficha (o token). En los métodos de imagen de dominio público se utiliza cada píxel; una imagen contadora se compara con otra mediante la medición de las diferencias entre sus píxeles (Manmatha, 2003). Una transformación afín puede ser aplicada para permitir una cierta flexibilidad, aunque este proceso es lento (Manmatha, 1995; Barmpoutis, 2009). Estos enfoques ahorran trabajo en la extracción ya que no se necesita ningún cálculo adicional, pero las comparaciones pueden ser largas ya que las características son numerosas. El método de cruce, el trazado de líneas a intervalos a través de la imagen y el registro de cuántas veces se cruzan el primer plano, es una técnica simple de extracción de características. Se trata de un menor número de coincidencias que son rápidas de realizar, y que supera ciertos tipos de variaciones, pero codifica menos información. Por lo tanto, la decisión sobre qué técnica a utilizar debe ser hecha con cuidado (Mori, 1999: 21). El objetivo no es sólo la eficiencia sino clasificación correcta, y para ello todo tipo de características se han tratado para permitir el reconocimiento pesar de la variación del modo, para capturar la forma esencial de un tipo, sobre todo el contorno y la corriente siguiente, las matrices de orientación, y los descriptores matemáticos. Para más información ver Mori (1999) y Cheriet (2007). La impresión de caracteres de reconocimiento confluye con los problemas básicos que se encuentran en todos los casos de reconocimiento de la escritura, sobre todo la extracción de palabras y caracteres de texto y la separación de los gráficos (cf. Fletcher 1988).

Los reconocimientos de caracteres dependen de la fuente, la imagen del carácter puede ser que aparezca exactamente o casi exactamente en una configuración conocida, por lo que el análisis de imagen sólo es simple. Una comparación píxel a píxel sería lo más acertado. El reconocimiento de caracteres independientes requiere sistemas más complejos, como la imagen real que está siendo analizada podría estar en cualquiera de un gran número de configuraciones y todavía necesita ser reconocido como el tipo correcto. La dirección del contorno, dividida en una matriz espacial, es una opción popular en estos casos (Cheriet, 2007: 59). Sin embargo, los investigadores de reconocimiento de caracteres de forma rutinaria mejoran la precisión al considerar la palabra completa o incluso van más allá de la palabra. Para los pequeños dominios, la palabra puede ser igualada contra un léxico. De lo contrario, el léxico puede ser reorganizado para permitir búsquedas más rápidas, o un modelo de Markov puede ayudar a anticipar el tipo del carácter dado el contexto (Cheriet, 2007: 234-237).

La mayoría del trabajo hasta la fecha de reconocimiento de caracteres se ha llevado a cabo en el alfabeto inglés y chino mandarín, aunque se han hecho progresos significativos en las escrituras de la India (Bhardwaj, 2009; cf. Bruel, 2009, Kompalli, 2007).

Reconocimiento de escritura a mano

El problema del reconocimiento de escritura a mano comparte más características con el problema de escrituras no cifradas que el simple reconocimiento de caracteres porque está más relacionado con las imágenes originales del manuscrito y, por supuesto, la mayoría de los scripts no cifrados son escritos a mano en lugar de mecanografiados. De escritura a mano no es necesariamente constante, lo que significa representaciones del mismo tipo podría variar sustancialmente dentro del mismo documento. La segmentación es por supuesto más complicada que en el reconocimiento de caracteres impresos, como fichas que pueden conectar o se superponen, también de manera inconsistente (cf. Nicchiotti, 2000, Yanikoglu, 1998). Un modelo de Markov se puede utilizar para segmentar la escritura, al mismo tiempo

que el reconocimiento de ésta (Cheriet, 2007: 237). También hay enfoques que evitan la segmentación, como la comparación con un léxico, factible para un dominio limitado (Cheriet, 2007: 251).

Manmatha muestra cómo agrupar palabras escritas a mano en un índice sin preocuparse por segmentación o reconocimiento, un proceso que ellos llaman *palabra manchada* o localización de palabras. Una vez que las palabras se han clasificado, el usuario final puede navegar por todas las imágenes de una determinada palabra en un documento o serie de documentos. El reconocimiento automatizado no se realiza; más bien las palabras pueden ser identificadas por un ser humano después de la indexación. Manmatha y los académicos relacionados han probado muchos algoritmos de comparación de imagen para detectar palabras. El mapeo distancia euclídeana, su primer intento y la línea de base, toma como entrada el resultado XOR de las dos imágenes de comparación y mide la distancia de cada píxel de primer plano a un píxel de fondo (Manmatha, 1995; 1996). Un éxito reciente utiliza la deformación dinámica del tiempo, que modela las imágenes de dos palabras a lo largo del eje horizontal como distorsiones entre sí (Rath, 2003). Sin embargo, un intento aún más reciente de detectar las esquinas es más rápido y casi tan preciso (Rothfeder, 2003).

Epigrafía

Los ordenadores se han utilizado durante mucho tiempo en el almacenamiento y la propagación de las representaciones de las inscripciones antiguas, ya sea a partir de fotografías o de opresiones, que son herramientas populares para la toma de impresiones de inscripciones en el papel. Estas técnicas ni siquiera existen para la captura de información tridimensional, a partir de fotografías tomadas en múltiples ángulos (Meyer, 2004), de opresiones escaneadas con varios equipos de luces (Barmpoutis, 2009), y de los procedimientos de escaneo láser más caros (Bodard, 2009). El ordenador puede entonces ser utilizado para identificar los elementos separados. En un estudio de las inscripciones griegas de Epidauro, el equipo segmentó

con éxito y agrupó las fichas de acuerdo a su similitud gráfica, creando un atlas de todas las fichas por tipo. Una alta precisión se consigue haciendo uso de los datos tridimensionales a partir de las opresiones de éstas. A continuación, se usó la agrupación de proponer un esquema de citas para las inscripciones (Barmoutis, 2009). Otro estudio ha demostrado cómo los jeroglíficos egipcios pueden ser indexados por sus constituyentes curvas de Bezier (Meyer, 2006), y parece que sólo sea un pequeño paso para el reconocimiento automático de los jeroglíficos a través de patrones.

Escritura automática

Como no ha habido un tratamiento científico directo de la escritura automática, el autor utiliza la investigación de los fenómenos relacionados para formar una teoría. Por lo tanto esta sección es necesariamente larga.

Si bien no ha habido ningún tratamiento científico de la escritura automática, se han realizado investigaciones intelectuales en las afirmaciones místicas de la escritura automática. Estos escritos místicos pueden ser en un idioma conocido por el sujeto o pueden estar en un lenguaje misterioso. En un caso famoso, el profesor Théodore Flournoy siguió a la notable Hélène Smith, que dijo estar recibiendo correspondencias en un lenguaje marciano. A medida que se traduciría al francés, fue capaz de llevar a cabo un estudio. Encontró ciertas semejanzas al francés. A pesar de esto, o incluso debido a esto, seguía convencido de que no era una estratagema deliberada sino un verdadero fenómeno inconsciente. Los fonemas marcianos están contenidos completamente dentro del inventario de fonemas franceses, por ejemplo, y los grafemas de Marte también se asignan a los mismos sonidos de grafemas franceses. La sintaxis es exactamente el mismo que en francés (Flournoy, 1900: 246-251). Flournoy describe otros ejemplos de la escritura automática de Smith de personajes conocidos (no marcianos) (cf. Flournoy, 1900: 288), pero el idioma marciano es relevante porque aparece en un sistema de escritura extraña, similar a la del Codex Seraphinianus. Cabe destacar que Serafini no hace ninguna afirmación mística sobre el código. A partir de

Pierre Janet, los psicólogos han confirmado que la capacidad de escribir inconscientemente existe (Van der Hart, 1989).



Figura 5.: Grafemas marcianos de Smith procedentes de las investigaciones de Flournoy. (Flournoy, 1900: 208)

Glosolalia y hablar en lenguas

Hay un proceso estrechamente relacionado con la escritura automática que ha recibido atención en algunos círculos: Es el fenómeno conocido como “hablar en lenguas” exhibidos por algunos espiritistas cristianos. Se cae en una clase de comportamientos llamados glosolalia. Al igual que la escritura automática, se le atribuye a la mística. (Smith no sólo escribía en su lenguaje marciano, sino que también lo hablaba.) El sujeto, a menudo en medio de la oración o de culto, balbucea de forma espontánea. En algunos casos, el murmullo se afirma que es en un idioma desconocido para el hablante, pero es reconocido y entendido por alguien más. En otros casos se acepta como un lenguaje “espiritual”, desconocido para nadie en la tierra (Cutten, 1927: 116, 164, 165). La glosolalia ocurre en otras culturas y se ha producido a lo largo de la historia. Específicamente con respecto a la práctica cristiana e fenómenos tan paranormales (Xenoglosia), que más tarde adaptado estos para aplicar a casos no cristianos (mayo de 1956). Al igual que en la escritura automática es algo novedoso se produce con un mínimo de esfuerzo consciente. Especialmente en el siglo XX, cuando se está generalizando, las personas estudiosas podían hablar en lenguas en un contexto psicológico. Las conclusiones son breves y necesariamente no no han sido probadas:

resulta que cuando el ser consciente se desintegra, permitiendo que el inconsciente pase a controlar las funciones motoras. En respuesta a las lenguas no conocidas por el sujeto que habla, estos estudios de casos demuestran que podría mejorarse el hecho de que la memoria permite que salgan en la ausencia del yo consciente: Por ejemplo, una camarera en una ocasión habló latín y el griego, y recordó que en un trabajo pasado que había trabajado de barrendera mientras su maestro recitaba pasajes largos en diversos idiomas (Cutten, 1927: 176). Mientras que a principios del siglo XX se utilizaron términos como el éxtasis y la histeria, en la década de 1970 este tipo de desintegración de la conciencia se había dado a conocer como un episodio disociativo (Arkin, 1981: 63, 64).

Como nota al margen, podría valer la pena considerar el arrullo y el balbuceo de los bebés humanos como un proceso automático espontáneo y creativo, basado en las observaciones que todos los bebés participan en al menos la etapa del arrullo, incluso los bebés de padres sordos, y que el progreso de una etapa a la siguiente no requiere la práctica (Lenneberg, 1967). Si es así, tal vez los mecanismos que producen la glosolalia son restos de este desarrollo temprano.

Otros procesos relacionados

Procesos similares se han estudiado científicamente más. Al igual que la escritura automática y la glosolalia, la improvisación musical crea algo nuevo, a veces sin control consciente (Nettl y Russell, 1998: 82, 171, 339). Soñando, también, puede ser visto como el cerebro crea algo nuevo con la ausencia de un control consciente. Los escáneres cerebrales (TAC, resonancia) y estudios de lesiones de estos procesos sugieren que una zona prefrontal es inhibida mientras que otra área prefrontal, el área responsable de interés y la curiosidad, se excita (Limb, 2008; Dietrich 2003; Solms 1997). Esta teoría se ha extendido a hacer garabatos (Sheridan, 2002). La idea principal es que el cerebro entra en un estado en el que está interesado, pero que no está regulado. A continuación, utiliza todos los recursos disponibles a la misma (la memoria y reglas simbólicas) para manifestar algo concreto (Solms, 1997).

Resumen

Sobre la base de investigaciones previas, se puede esperar que si la escritura del Codex Seraphinianus fue generado automáticamente entonces tiene sus raíces en la experiencia adquirida por Serafini con el lenguaje. Esta idea se puede tener en cuenta durante la visualización de los resultados del análisis que sigue.

MATERIALES

Hardware y sistema operativo

Todos los procedimientos descritos en este documento se realizaron con un Pavilion HP (Modelo dv6119us) con una pantalla de 15 pulgadas, que ejecuta Microsoft Windows XP Media Center Edition, versión 2002, Service Pack 3. El procesador es AMD Turion (TM) 64 X2 Mobile, 803 MHz (960 MB de RAM).

Esto puede considerarse una máquina de gama baja en comparación con las máquinas que se ofertan hoy en día.

Entorno de desarrollo

Las aplicaciones fueron desarrolladas en Microsoft Visual Studio 2005 Professional Edition, Service Pack 2, utilizando el lenguaje C #.

Software de terceros

Para ahorrar el tiempo y el esfuerzo implicados en el desarrollo de herramientas complejas de análisis gráfico, se utilizó AForge.NET (versión 1.7.0) como un conjunto de herramientas de manipulación de imágenes, añadido como una referencia al

proyecto. AForge.NET es un kit de desarrollo de software con un conjunto completo de utilidades de inteligencia artificial, incluyendo imágenes, aprendizaje automático, robótica y algoritmos. Sólo para el espacio de nombres AForge.Imaging se utilizó en este estudio. El marco entero está disponible gratuitamente en <http://code.google.com/p/aforge/>. Se afirma claramente cuando los procedimientos hacen uso de AForge.NET.

Cuerpo

El Codex Seraphinianus fue adquirido en formato Adobe Acrobat Reader (PDF) y se convierte en 369 imágenes individuales, una por página, usando PDF to Image Converter 1.0 por PDF-TIFF-Tools.com. Las imágenes de salida miden 1573 por 2169 píxeles, con una resolución de 150 dpi y una profundidad de bit de 24.

Para evitar las primeras páginas del códice que tienen el formato no estándar, la primera página se utiliza para el estudio es la página 10.

Métodos y resultados

Texto de cabecera (“letras mayúsculas”)

Por estas implementaciones, las fichas de cabecera se extrajeron utilizando el software de terceros. Una función fue ideada para medir la similitud gráfica entre fichas. En una implementación, las fichas en los tres primeros capítulos del código fueron clasificadas en tipos utilizando un algoritmo automatizado sencillo combinado con una supervisión humana. En otra implementación, un motor de búsqueda demuestra la posibilidad de localizar secuencias paralelas, una tarea común para el análisis de lenguas no descifradas, con base en la similitud gráfica solo.

Preparación para la extracción

Cuando una imagen de una de las páginas del código se procesa primero por el conjunto de aplicaciones desarrollado en este estudio (descrito en el resto del papel), tres filtros AForge.NET se aplican a la misma. Un filtro GrayscaleRMY se utiliza para convertir a 8bpp, el formato que es requerido por los algoritmos AForge.NET. Un filtro Threshold aumenta la definición de las líneas de fichas a un contraste en blanco y negro. En tercer lugar, un filtro invertido se utiliza para obtener las fichas que deben ser blancas sobre un fondo negro para que puedan ser analizadas correctamente. Los valores por defecto AForge.NET se utilizaron para cada uno de estos filtros.

Para ahorrar una cantidad significativa de tiempo durante el proceso de análisis, se recorta la imagen antes de la extracción de un área que puede es razonable predecir que abarque el texto de cabecera y no mucho contenido superfluo, utilizando un filtro de cultivos AForge.NET.

El rectángulo utilizado para el cultivo es la ubicación (0, 0) y el tamaño (1572, 270).

Extracción

Un algoritmo analizado está destinado a separar imágenes de primer plano de un fondo mediante la identificación de formas contiguas. El resultado del análisis es un conjunto de manchas que representan estas formas. En el caso de las cabeceras del Codex, una gota se produce para cada ficha.

La clase es AForge.NET BlobCounter utilizado para determinar la localización y extensión de las fichas sobre la página. Para extraer las letras mayúsculas, la BlobCounter está configurado con los siguientes valores:

```
FilterBlobs = true  
MinHeight = 20  
MinWidth = 20  
MaxHeight = 300  
MaxWidth = 300
```

Estos ajustes sólo producen piezas que caen en el rango de tamaño correcto para las letras mayúsculas. Además, la aplicación ordena las manchas de más alta primero y considera sólo las piezas que caen dentro de un arco de veinte píxeles de la burbuja superior. Esto garantiza que sólo las piezas en la cabecera son consideradas (no piezas que pueden aparecer más abajo en la página en las ilustraciones, por ejemplo). Cualquier burbuja que pasa todos estos filtros se considera un símbolo de letra mayúscula válida. Su número de página, la ubicación, extensión, y los datos visuales se almacenan en una matriz que se puede acceder por la suite de aplicaciones.

Esta técnica no funciona en las páginas en las que la cabecera no está cerca de la parte superior, incluyendo la portada y las páginas con orientación horizontal. Estas páginas se descartaron del estudio.

En raras ocasiones, debido a la integridad de la página procesada, la extracción falló en fichas particulares. Por ejemplo, de las 264 fichas en el primer capítulo (páginas 10-43), cinco fueron no extraídas correctamente, y de las extraídas, solo se extrajo una parte de esas 5 fichas. Estas extracciones fallidas se tuvieron en cuenta.

Identificación de fichas similares

El conjunto de aplicaciones compara dos fichas utilizando un recuento de los espacios en blanco sencillo con fluctuación de cómo se describe a continuación. En primer lugar, se escala a las dos imágenes del mismo tamaño (70 x 70) usando un filtro AForge.NET ResizeBilinear (una práctica llamada normalización). Más información sobre el tamaño normalizado es a continuación. A continuación, toma la proporción de píxeles blancos en la imagen XORed a los píxeles blancos en las dos imágenes originales combinadas (las fichas son de color blanco sobre un fondo negro debido a los filtros). Debido a que las imágenes no están necesariamente alineadas, para empezar, se utiliza la fluctuación como se describe en el texto de Manmatha (1995) para encontrar la mejor alineación entre ellos. Específicamente, una de las imágenes se prueba con x offsets de -4, -2, 0, 2, 4 e y offsets de -4, -2, 0, 2, 4, lo que resulta en 25 ensayos diferentes en lugar de uno. (Se utilizó el filtro de AForge.NET CanvasMove para crear las imágenes offset.) El valor flotante resultante se utiliza como una medida de la similitud entre las dos imágenes. Este valor puede variar desde cero (si las dos imágenes coinciden exactamente) a uno (si las imágenes no se superponen en absoluto). Un algoritmo de comparación alternativa también se trató, la cartografía de distancia euclidiana como se describe en el texto de Manmatha (1995). Si bien se ha demostrado dar buenos resultados para las palabras escritas a mano, que parecía no producir resultados significativamente mejores en esta aplicación, probablemente debido a la dificultad relativa de estas imágenes (letras en vez de palabras).

La reducción del tamaño normalizado, por supuesto, reduce el número de píxeles y por lo tanto reduce el tiempo de procesamiento. (Cuando se hace esto los valores de fluctuación también ser necesario ajustarlos.) Sin embargo, hay una compensación debido a que la resolución de la imagen también se reduce. Probablemente, un algoritmo de comparación más intrincada puede ser aplicado en investigaciones futuras para permitir que las imágenes mucho más pequeñas, ahorrando tiempo de

procesamiento. Las imágenes y el título siguiente a continuación muestran el algoritmo de comparación de XOR en el trabajo para dos fichas diferentes en grandes y pequeños tamaños normalizados.

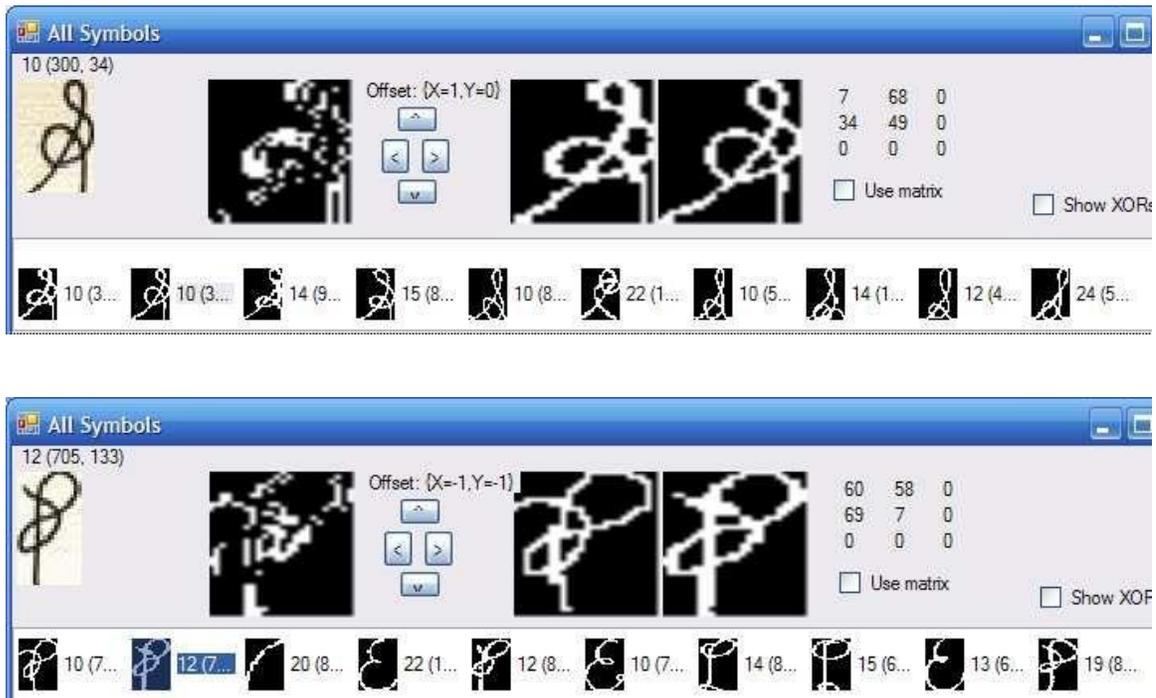


Figura 6: Comparación de tokens de cabecera.

La primera ficha en la fila inferior de cada imagen es el objetivo. En cada caso, el segundo contador ha sido seleccionado para mostrar algunas de las características de comparación de la herramienta. El cuadro de la izquierda muestra el símbolo seleccionado tal como aparece en el código. Las cajas a la derecha muestran el objetivo (a la izquierda, se extendía a una gran plaza) y el seleccionado (a la derecha, se extendía a una gran plaza) fichas de lado a lado. La caja restante muestra el resultado de la operación XOR. La herramienta también informa el hecho de compensar el mayor éxito y permite al usuario obtener la imagen XORed utilizando los botones de flecha para verificar esto. En orden: página de destino 10 de contadores 3, tamaño 70x70; página objetivo 10 de contadores 3, tamaño 20x30; página objetivo 10 contador 7, tamaño 70x70; página objetivo 10 contador 7, 20x30 tamaño. (La curva aislada que aparece en los resultados es un error de extracción.)

Un intento para parchear el algoritmo es la incorporación de una matriz de espacio en blanco, que se divide la imagen en nueve secciones y considera la diferencia blanco-recuento en cada algoritmo al adoptar su decisión. Si bien esto funcionó como se esperaba, no ayudó a los resultados e incluso los degrada cuando dos fichas del mismo tipo varían demasiado.

La técnica en última instancia usada para proceder con el estudio utiliza el tamaño de 70x70 sin consideraciones adicionales.

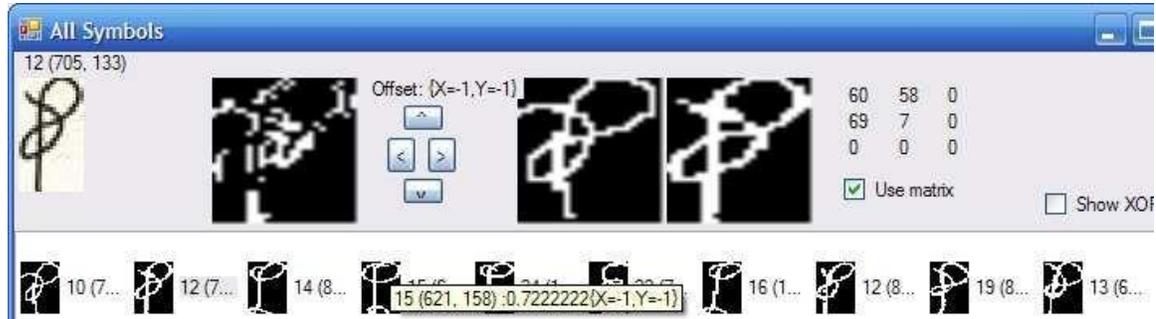


Figura 7: Aplicación de una matriz de espacio en blanco para el último ejemplo de la figura anterior. Los números más a la derecha representan los recuentos de espacio en blanco en cada sección de la matriz para el contador seleccionado. Hay que tener en cuenta que, mientras que los resultados coinciden con el objetivo de cerca con los espacios en blanco en cuenta el equilibrio, la mayor parte de los resultados en el medio de la fila parecen ser de un tipo diferente.

Estudio 1: Consolidación en Tipos

Justificación: Si las fichas del sistema de escritura se pueden separar en un número finito de tipos, este número podría indicar si se trata principalmente alfabético, silábico, o ideográfico. Además, la identificación de los tipos es necesario para muchas formas de análisis lingüístico incluyendo análisis de grafías y modelos de Markov. Sin embargo, es importante señalar que para identificar los tipos hay que entrar en las limitaciones de un esquema de la transcripción que se describe en la introducción.

Fichas similares se agrupan automáticamente juntas por el valor se ha explicado anteriormente en un esquema simple, rápido y de plana agrupación: Para cada contador, si cumple un umbral de similitud ($<0,65$) con el primer miembro añadido de un tipo existente, éste es añadido a ese tipo; de lo contrario se crea un nuevo tipo. No se hace ningún intento de mantener un medoide o centroide. El enlace único de agrupamiento también fue estudiado; como se esperaba, produjo resultados algo mejores, pero tardó un tiempo mucho más largo que el esperado. Esta es un área de más investigación.

Debido a la naturaleza de la escritura en el códice, no se hizo ningún intento de medir la precisión del esquema de agrupamiento. Aunque algunas de las fichas sin duda parecen ser del mismo tipo, la clasificación de muchas de las fichas es ambigua. Desde la escritura es asémica, esto no es una sorpresa. Está claro que estas decisiones acerca de las fichas no se pueden hacer sin supervisión posterior.

Una vez que los tipos se identificaron tentativamente por el ordenador, el proceso de supervisión humana comienza. Un interfaz fue desarrollado para permitir que el usuario humano para revisar los tipos, combinar tipos, y fácilmente mover fichas entre los tipos existentes o utilizarlos para crear nuevos tipos principalmente de arrastrar y soltar. Además, cada ficha se puede ver en su contexto original en la página. Esta es una etapa importante durante el cual el usuario humano tiene la libertad para examinar las fichas en estos dos contextos: similitud con otras fichas y el uso de la página.

Las fichas de cabecera de los tres primeros capítulos del códice (páginas 10-123) fueron clasificadas en tipos utilizando esta aplicación. Debido a que la cantidad de interacción humana fue sustancial, como algunas de las decisiones de clasificación se consume mucho tiempo, el ahorro y la funcionalidad de carga fue ideado de manera que cada capítulo se pudo procesar y se guarda por separado y luego se combina con los otros más tarde. Esto permite al usuario elegir trabajar con un pequeño trozo de los datos a la vez. Además, lo que limita la cantidad de datos de trabajo mantiene la comparación automáticamente manejable, ya que cada nuevo ficha debe en el peor de los casos puede comparar con cualquier otro símbolo. Aunque no es grande, la colección resultante podría utilizarse como datos de entrenamiento para ayudar en la clasificación del resto del códice. Se describe a continuación.

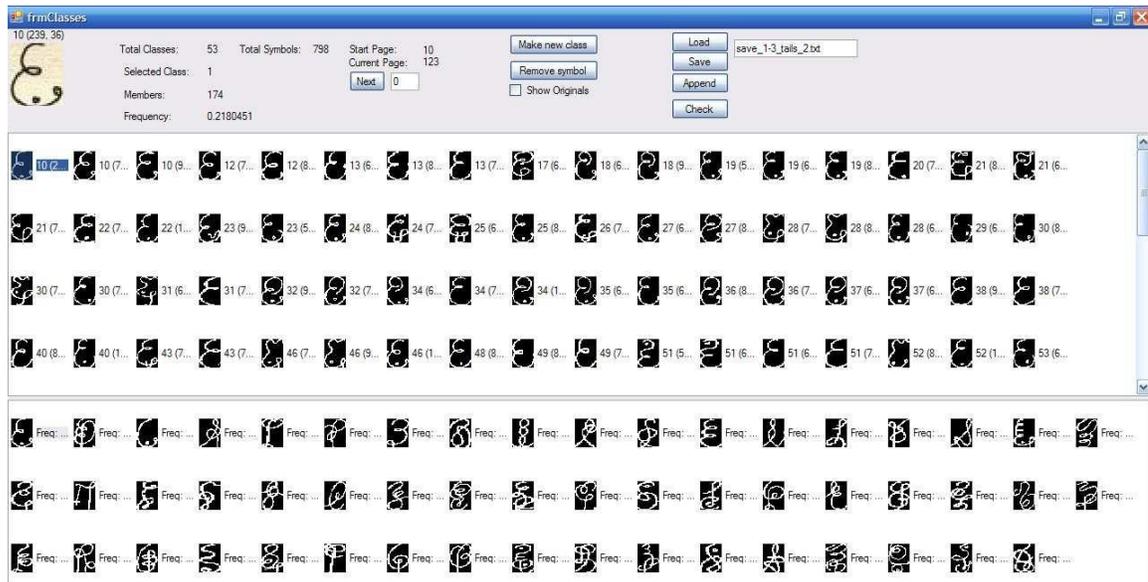


Figura 8: App Typing que muestra los resultados para los tres primeros capítulos del código. *El panel inferior muestra los finales 53 tipos. El panel superior muestra todas las fichas del primer tipo.*

Varias observaciones de superficie sólo de los primeros capítulos. Capítulo 1 (páginas 1043) muestra el dominio de unos pocos tipos, y hay muchos tipos que sólo aparecen una vez. Capítulo 2 (páginas 46-103) ofrece con muchas más señales los tres capítulos e introduce mucha más diversidad de fichas. Las 438 fichas no podrían clasificarse cómodamente por el autor a primera vista en ningún menos de 70 tipos, 34 de los cuales contenía sólo una única ficha. Capítulo 3, con las fichas de menor cantidad, encaja en su mayoría en las clases ya creados a partir de los capítulos anteriores. Cuando se combinaron los tres capítulos, había 798 fichas clasificadas en alrededor de 70 tipos. Este número limitado de tipos es el resultado del autor ignorando las pequeñas variaciones entre las fichas, tales como un punto que aparece en el interior de la convexidad, un bucle que aparece en un punto final, y ciertas diferencias en curva de contornos. De lo contrario no habría sustancialmente más tipos.

Sin embargo, esta clasificación inicial condujo a un descubrimiento. Después de examinar el gran número de tipos que sólo se producen una vez, el autor se dio cuenta de que la mayoría de estos precede el tablero en una página de ilustración y, por tanto,

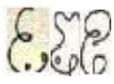
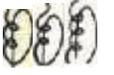
no tienen ningún valor lingüístico. Ellos efectivamente significarían algo así como: “Este es un identificador único para esta página ilustración y no tiene otro propósito.” Esta relación será visitada aún más en la sección de discusión. Descartando estas fichas únicas proporcionan una clasificación final de 53 tipos.

Es útil señalar aquí las plantillas o patrones que se repiten con frecuencia en los encabezados de página. Esto hará que sea más fácil hablar acerca de dónde los tipos se han encontrado. No se necesita un ordenador para notar estos, pero la aplicación de maquetación sin duda ayuda destacarlos ya que ciertos tipos aparecen principalmente en estos patrones. Los usos para estos patrones se solidificaron por la aplicación del motor de búsqueda se describe en la siguiente sección. Las transcripciones se basan en las sugerencias de la tabla de los tipos, que sigue inmediatamente. Se dirá más acerca de los patrones posteriores.

Tabla 1: Patrones en las cabeceras del Codex.

primera aparición	Trans.	Uso
página 10	DSD	Termina encabezado en la primera página de un capítulo
página 10	ESSH3	Inicia encabezado en la primera página de un capítulo
página 11	DE2C	Tabla de cabecera contenidos
página 12	DS	inicia la cabecera
página 12	SFA	termina cabecera
página 88	L2F	inicia la cabecera
página 88	TE	termina cabecera

Tabla 2: Tipos que se encuentran en los capítulos 1-3, con los nombres y transcripciones sugeridas.

	Frecuencia	Posible Nombre	Posible Trans.	Ejemplos	Encontró
1	174 (0.218)	tipo E	mi		En todo En el patrón de TE
2	103 (0.129)	tipo D	re		En todo En el patrón de DS

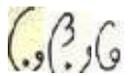
3	84 (0.105)	tipo C	do		En todo
4	68 (0.085)	tipo S	S		En todo En la DS patrones y SFA

Tabla 2 (continuación)

5	49 (0.061)	tipo de Phi	F		En todo En el L2F patrones y SFA
6	48 (0.060)	tipo P	PAG		En todo
7	29 (0.036)	3 tipo	UN		Principalmente en el patrón de SFA
8	28 (0.035)	8 tipo	segundo		Principalmente capítulos 2-3
9	26 (0.033)	tipo 3-nudo	H3		En todo
10	21 (0.026)	3 hojas (trébol) Tipo	T		Principalmente en el patrón de TE
11	18 (0.023)	tipo de gancho	L2		Sólo en el patrón L2F
12	18 (0.023)	Doble-E Tipo	E2		En todo
13	16 (0.020)	tipo 2-nudo	H2		A lo largo del capítulo 2
14	12 (0.015)	tipo L	L		A lo largo de los capítulos 1-2
15	10 (0.013)	tipo PS (O tipo Harp)	P2		A lo largo de los capítulos 2-3
dieciséis	9 (0.011)	tipo SC (O la mitad Harp)	S2		A lo largo de los capítulos 2-3
17	6 (0.008)	Tipo de Triple-E	E3		capítulos 2-3
18	5 (0,006)	tipo de dragón	J		Capitulo 2
19	5 (0,006)	Doble-3 Tipo	A2		Capítulo 2 y distorsionado en la página 25

Tabla 2 (continuación)

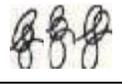
20	5 (0,006)	tipo de Pi	norte		todos los capítulos
21	4 (0,005)	tipo de caracol	GRAMO		Capítulos 2-3, sólo tres palabras
22	4 (0,005)	(Invertida) tipo de agudos	Bo		Capitulo 2
23	4 (0,005)	tipo E8	eb		Páginas 108 y 109
24	4 (0,005)	Double-Y Tipo	Y2		Capitulo 2
25	3 (0,004)	Caracol-8 tipo	gb		Páginas 62 y 67
26	3 (0,004)	tipo Y	Y		Capitulo 2
27	3 (0,004)	tipo de molinete	x		Página 18 y distorsionado en la página 71
28	3 (0,004)	Tipo L-Hook	LL2		Páginas 50 y 55
29	3 (0,004)	E-stick tipo	Ei		Páginas 16 y 46
30	2 (0,003)	Y-8 tipo	Yb		página 96
31	2 (0,003)	-Rot E Tipo	Nosotros		Páginas 122 y 123 (una palabra)
32	2 (0,003)	Tipo de resorte o de tipo arbusto	yo		Páginas 26 y 27
33	2 (0,003)	tipo 3-bar	Ai		Páginas 70 y 71

Tabla 2 (continuación)

34	2 (0.003)	tipo R	R		Páginas 64 y 66
35	2 (0.003)	Tipo Dragón-L	jl		página 87
36	2 (0.003)	L-8 tipo	lb		Páginas 19 y 93
37	2 (0.003)	tipo Triple-C	C3		Páginas 12 y 13
38	2 (0.003)	tipo de EL	El		Páginas 122 y 123 (una palabra)
39	1 (0.001)	Tipo de trébol-palo	Ti		página 120
40	1 (0.001)	tipo de CD	Discos compactos		página 96
41	1 (0.001)	tipo de humo	W		página 89
42	1 (0.001)	2-nudo E Tipo	H2e		página 101
43	1 (0.001)	Izquierda-8 Tipo	B2		página 23
44	1 (0.001)	tipo S-bar	Si		página 67
45	1 (0.001)	tipo de 6	O		página 57
46	1 (0.001)	3-8 tipo	ab		página 19
47	1 (0.001)	Tipo izquierda 3-8	Ab2		página 69

Tabla 2 (continuación)

48	1 (0.001)	tipo de engranajes	Ir		página 65
49	1 (0.001)	-3 Phi tipo (Árbol)	Fa		página 40
50	1 (0.001)	<Ninguno>	Sy?		página 53
51	1 (0.001)	tipo espiral-stick	Oi		página 89
52	1 (0.001)	tipo Phi-H	FH		página 27
53	1 (0.001)	Phi-E Tipo	ef		página 51

Estudio 2: motor de búsqueda

Justificación: Una gran parte de los esfuerzos sobre los idiomas no descifrados se gasta en la identificación de secuencias paralelas de fichas en todo el corpus, en la expectativa de que cuantos más contextos en los que una secuencia se puede encontrar mejor entendida. Esta aplicación se entiende como una prueba de concepto de que estas secuencias se pueden razonablemente encontrar buscando en las imágenes.

Independiente de la interfaz para la identificación de tipos, una interfaz de motor de búsqueda se produce, basado en el mismo valor de similitud que se explicó anteriormente. Se trata de un motor de búsqueda completamente visual, no se basa en ningún tipo de transcripción o escribiendo régimen. El usuario coloca los datos de imagen en el portapapeles de Microsoft Windows (copiando una ficha o secuencia de símbolos de una de las páginas del Codex en cualquier programa de imagen) y luego inicia la búsqueda. La aplicación intenta encontrar las secuencias de fichas que coinciden con las secuencias de fichas en el portapapeles, tomando el promedio de similitud entre la secuencia en el portapapeles y cada secuencia en la cabecera de la página. Cualquier secuencia que satisface el umbral de similitud ($<0,65$) se devuelve en los resultados de búsqueda. Por ejemplo, si la secuencia en el portapapeles es dos fichas de largo, la aplicación prueba cada bigrama en la cabecera. No tiene en cuenta el espacio vacío para que no consideren dos fichas separadas por cualquier distancia del espacio vacío como un bigrama válido para probar en contra.

La búsqueda se ejecuta en un hilo separado de manera que el usuario puede ver los resultados de búsqueda, mientras que el programa todavía se está realizando el análisis en las páginas restantes.

Al especificar los criterios de búsqueda, el usuario no quiera escalar imágenes simbólicas todos al mismo tamaño como se describe anteriormente. Esto se debe a que el tamaño de la imagen es una pista importante para la coincidencia. Cuando un pequeño punto se escala para 70 píxeles cuadrados, por ejemplo, puede parecer que el algoritmo para ser similar a una barra vertical larga también reducido a 70 píxeles cuadrados. Por otra parte, sería deseable perder coincidencias válidas simplemente debido a pequeñas variaciones en el tamaño. Por esta razón, la interfaz de búsqueda ofrece la opción al usuario en forma de una casilla de verificación “no escalar” (*do not scale*) a escala o no a escala las imágenes utilizadas para la comparación. En lugar de la escala, la aplicación crea una nueva imagen del tamaño deseado y escribe la ficha en él, el relleno de forma efectiva. (Un error a destacar es que fichas superiores a 70 píxeles se recortan de acuerdo con este esquema de relleno).

El motor de búsqueda permite el ajuste de dos valores de umbral para el control de los resultados: umbral individual y la tolerancia general. Cada ficha debe coincidir con el umbral individual, y la media de todas las fichas deben cumplir con el mínimo total. El ajuste del umbral individual es ligeramente más bajo que el mínimo total que se obtienen los mejores resultados, ya que permite a cualquier señal única que falle la coincidencia ligeramente mientras las otras muestras superan la coincidencia suficiente para recuperarse.

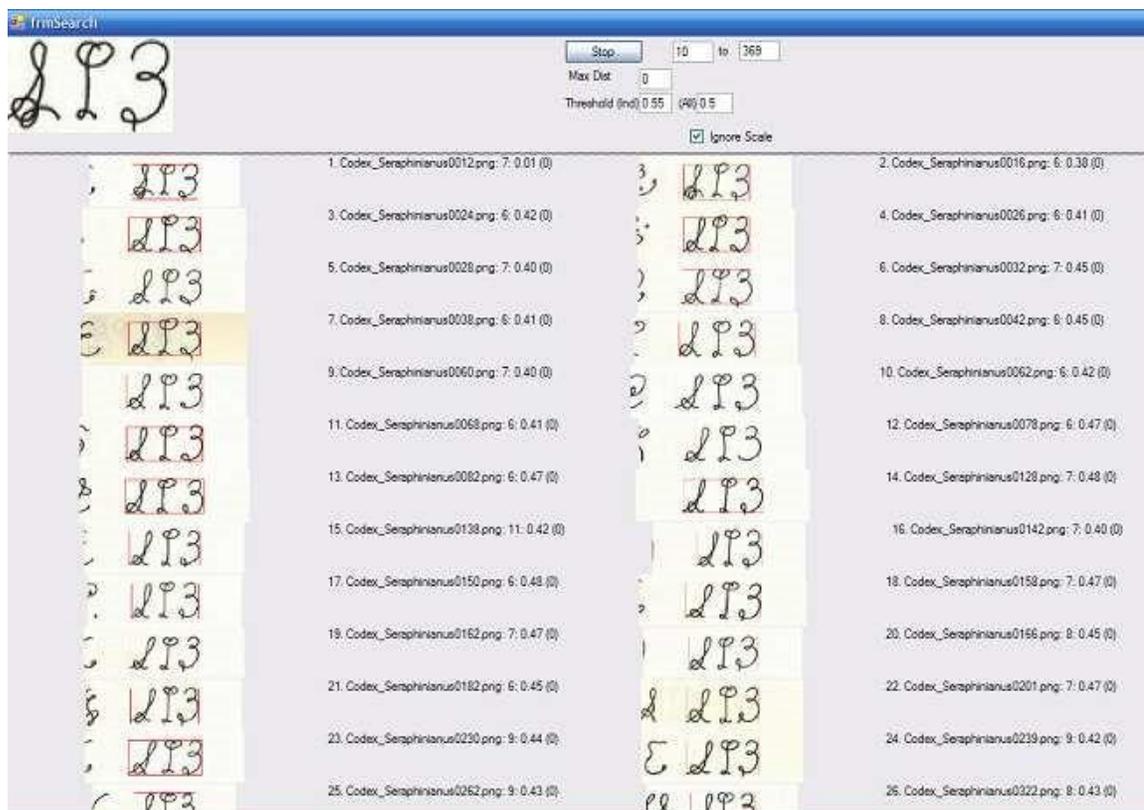


Figura 9: Resultados de la búsqueda de la secuencia común SFA.

Umbrales: Individual = 0,55, All = 0,5. Max Dist (véase más adelante) se establece en 0. Los tres primeros resultados son de las páginas 12, 16 y 24. El último resultado totalmente visible es de la página 239.

Por último, para una secuencia multi-modo, el usuario puede especificar el número máximo de fichas que pueden no coincidir (“dist max”). Al establecer este valor a 1 permite que los resultados para los que el algoritmo puede haber dado una falso computo negativo en una ficha, pero no se identificó correctamente los otros.

del libro, y porque todas estas coincidencias se producen en al parecer exactamente el mismo contexto, parecería que estos símbolos son libremente intercambiables.

También podemos ver la variación en la ficha E. En casi todos los casos termina en un pequeño lazo, con una excepción, en la página 78. Debido a la omisión de un bucle accidentalmente no es improbable, y debido a que sólo se produce en un caso aislado, esta podría ser descartado como un error.

También debe tenerse en cuenta que el motor de búsqueda hace evidente de inmediato el uso repetido de la palabra a través de las cabeceras del Codex, siempre en la misma posición. Si bien esto podría ser fácilmente observado por un investigador, que llevaría más tiempo para documentar cada caso.

Además de explorar las similitudes entre las cabeceras, el motor de búsqueda puede explorar singularidad en las cabeceras. Cuando el tipo de clasificador produjo varios tipos de una sola manera, se encontró que estas fichas de un solo disparo suelen aparecer antes de la escritura en las páginas de imágenes. Exploraremos más a fondo este fenómeno, utilizando el motor de búsqueda para encontrar todas las variaciones escritas en el códice es una manera fácil de navegar por todas las fichas pre-escritura. Los resultados muestran que las fichas de pre-escritura son casi todas únicas. Incluso entre las páginas que parecen mostrar ilustraciones estrechamente relacionados, las fichas no se parecen entre sí.

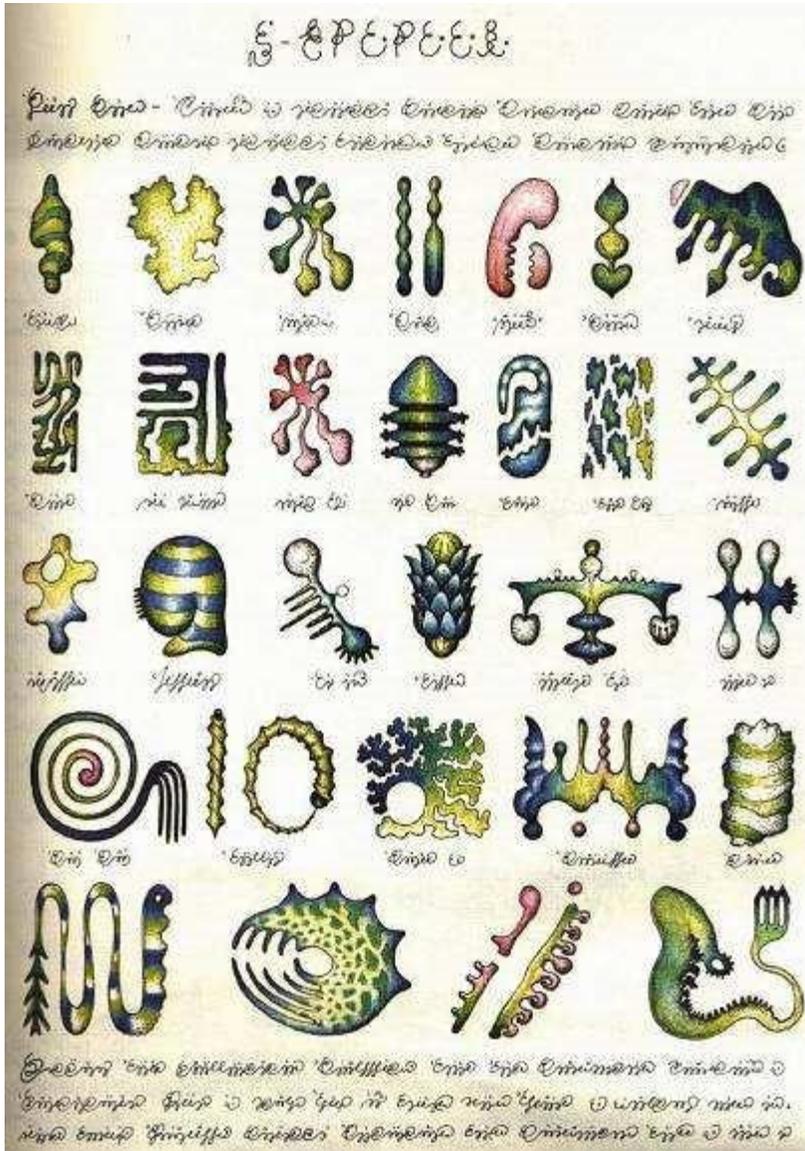


Figura 11: Ejemplo de una página de ilustración (13).
 La cabecera se inicia con una ficha única y un guion.

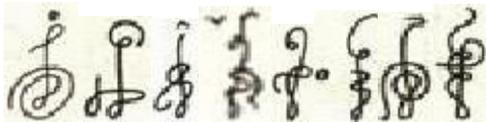
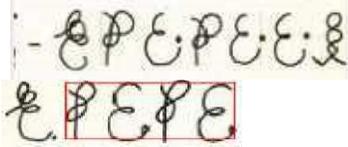
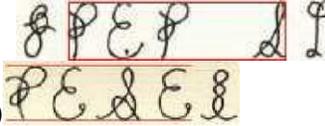
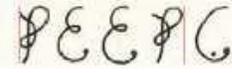
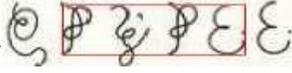
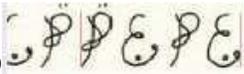


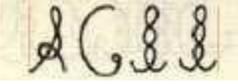
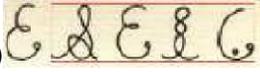
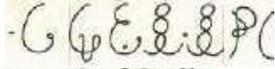
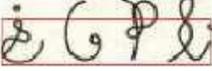
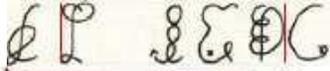
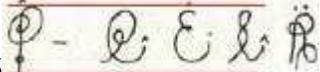
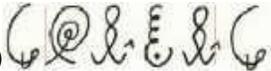
Figura 12: Fichas únicas (no repetidas) pre-escritura que aparecen en páginas ilustradas en el capítulo uno trata de plantas similares.
 Páginas: 17, 18, 19, 20, 21, 23, 25, 27. La última ficha es similar al tipo I (a la derecha), pero no el mismo.

Los resultados iniciales del motor de búsqueda sugirieron que, aparte de las obvias palabras repetidas, la recurrencia en las secuencias de cabecera es rara. Esto fue explorado brevemente mediante la búsqueda de secuencias compuestas de tipos comunes en los encabezados que se podría esperar que se repitan. Los resultados se pueden ver a continuación. Aparte del objetivo y su página de ilustración correspondiente, la apariencia de la palabra o secuencia es única.

Tabla 3: palabras o partes de palabras hechas de tipos comunes en los encabezados que sin embargo son únicas.

La columna de la izquierda muestra la página de ilustración asociada al objetivo y en el que se repite con frecuencia. El lado derecho muestra algunos otros resultados de búsqueda. Ninguno de los resultados de la búsqueda son exactamente la misma secuencia que el objetivo. A la izquierda están las páginas números en los que se producen las cabeceras.

secuencia diana y página ilustración	otros hallazgos
<p style="text-align: right;">12</p> 	<p>10</p> 
<p>13</p>	<p>48</p> 
	<p>60</p> <p>147</p> 
	<p>269</p> 

<p>112 </p> <p><No ilustración en la página></p>	<p>10 </p> <p>177  </p> <p>204</p> <p>295 </p>
<p>52 </p> <p><No ilustración en la página></p>	<p>66  </p> <p>80</p> <p>210 </p> <p>320 </p>

Cuerpo de texto (“letras minúsculas”)

Para el cuerpo del texto, palabras enteras fueron extraídas. A primeras órdenes de ejecución todas las palabras en una sola página de longitud para permitir algunas impresiones iniciales. A continuación, una función fue ideada para medir la similitud gráfica entre las palabras. Un motor de búsqueda demuestra la validez y deficiencias de esta función y permite nuevas observaciones relativas a la distribución de las palabras en el códice.

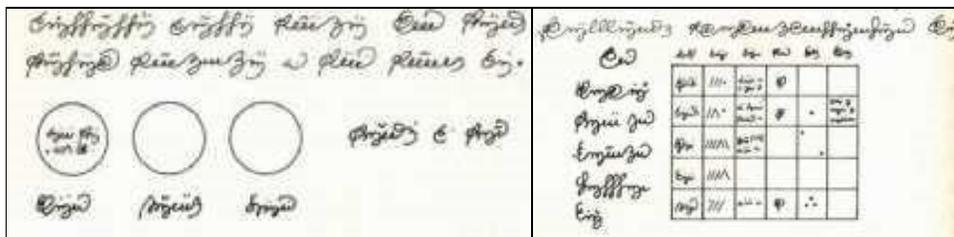
Preparación para la extracción

El cuerpo del texto fue preparado para la descryptación en la misma forma que el texto del encabezado, con pequeñas diferencias entre ensayo y error. En primer lugar, se aplicó un filtro de AForge.NET GrayscaleBT709, a continuación, un filtro de umbral con un valor umbral de 170, y por último el filtro Invertir.

Extracción

Las palabras de texto de cuerpo fueron extraídas por AForge.NET de la misma manera que los caracteres de texto de encabezado. Tabla de contenido y páginas de orientación horizontal fueron descartadas. Debían implementarse varias técnicas para mitigar el diseño impredecible del texto del cuerpo. Los factores de complicación son:

1. palabras demasiado cerca verticalmente
2. palabras apretadas entre líneas
3. palabras que desembocan en el margen de la página o de la sombra arrojada
4. palabras rotas accidentalmente durante el procesamiento previo
5. puntos y otros signos diacríticos que aparecen por encima de las palabras
6. números, trazos y símbolos aislados que aparecen a servir a un propósito matemático
7. tablas, diagramas y otras configuraciones no convencionales de fichas
8. ilustraciones



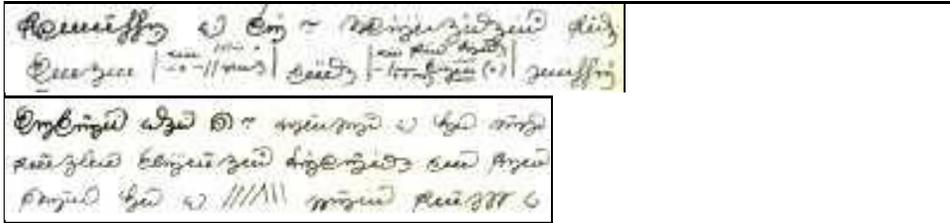


Figura 13: Las muestras de la página 12 del texto del Codex que muestra mezclada con (a) los diagramas, (b) las tablas, (c) los compartimentos de dos líneas, y (d) trazos no de palabras.

Para hacer frente a estos factores se utilizó el siguiente método.

1. Cualquier manchorrón o borrón que es más alto de lo esperado (> 40) y se superpone una mancha abajo se reduce de modo que no se solapa con la mancha a continuación.
2. Cualquier borrón que todavía es más alto de lo esperado es candidato para ser dividido por la mitad horizontalmente. La fila más negra se encuentra, y si el espacio en blanco por encima y por debajo de esta fila es aproximadamente igual (a menos de 300) la extracción se divide en la primera fila. Esta es fijar las palabras que son demasiado cerca verticalmente.
3. Las manchas que están muy cerca horizontal se fusionan. Esto es para fijar las palabras rotas durante el procesamiento previo y tratar de unir las manchas adicionales formadas en la etapa 2 a sus verdaderos dueños. Para ello, una caja de la longitud de la gota y 6 píxeles de alto se extrae a través del medio de la gota. A continuación, se infla 6 píxeles en cada lado, y cualquier intersección con cualquier otro resultado en una combinación con la extracción de la mancha. Cualquier mancha más alta que 40 píxeles no puede mezclarse.
4. Manchas que se encuentran totalmente dentro de otras manchas se fusionan. Esto es para asegurarse de que los puntos diacríticos y se incluyen. Para ahorrar tiempo, esto se combina con el siguiente paso.
5. Manchas que están muy cerca verticalmente se fusionan. Esta fusión es para cualesquiera mancha adicional restante de la división en el paso 2 y también ayuda a combinar las manchas compartimentados para que puedan ser filtrados. La

mancha se compara con todas las demás manchas; cualquier manchas que se cruzan su parte inferior se fusionan si cualquiera de (a) las manchas tienen exactamente la misma posición x en la página o (b) si ambas manchas son finas ($H < 15$). Al igual que con la fusión horizontal, ninguna burbuja más alto que 40 píxeles no puede mezclarse.

6. Paso 1. Se repite este proceso potencialmente desde que se han formado nuevas manchas.
7. Cualquier mancha no puede ser una palabra, debido a la altura (< 20 o > 60), el tamaño total de ($H + W < 30$), o relación de aspecto ($H / W > 1,2$) se descartan. Es de esperar que elimina la mayoría de las ilustraciones y las marcas de páginas perdidas de consideración.

El resultado final es que todas las palabras de las páginas con orientación vertical que no se cruzan al margen de la página o de la sombra del canal, incluyendo fichas individuales, pies de fotos y palabras que se encuentran dentro de gráficos, y sin incluir los compartimientos en línea ya que estos generalmente se fusionan y filtran. El autor no está preparado para verificar su integridad.

Ciertamente, las palabras indicadas faltan, pero podrían ser fácilmente otras pérdidas también.

Para esta investigación, no es esencial que cada palabra fuera capturada.

Estudio 1: Palabras Ordenado por Longitud

Justificación: La meticulosa observación revela que dentro de una página del Códice, el mismo tipo de formas de palabras se repiten con frecuencia, como si las palabras fuesen afines. Comparándolos podría conducir a la identificación de las características morfológicas.

Las palabras son demasiado pequeñas e intrincadas a someterse a la misma métrica de similitud utilizadas en la cabecera. Una aplicación para ordenar todas las palabras del cuerpo en una página de longitud es un intento de identificar rápida y fácilmente estas palabras similares para determinar los próximos pasos.

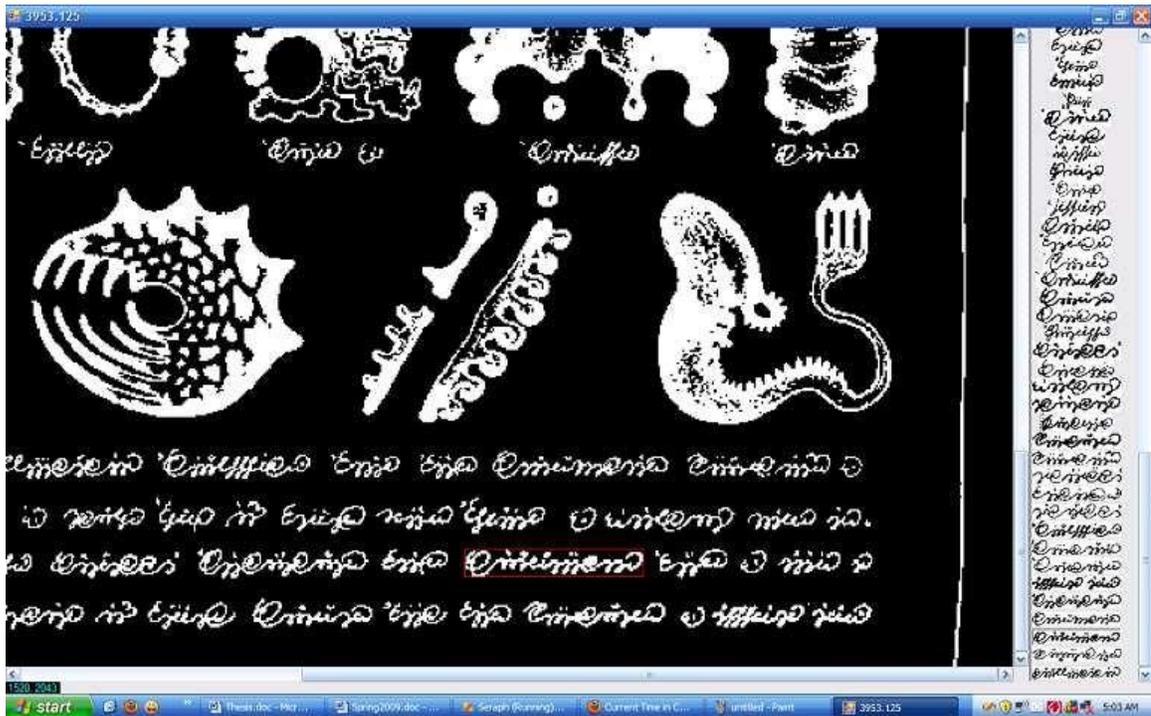


Figura 14: una aplicación que muestra todas las instancias de palabras en una página ordenados por longitud, aplicada a la página 13 del códice.

Selección de una palabra desde su aparición subrayada en rojo. (ensayos programados 3828, 3828, 3844ms.)

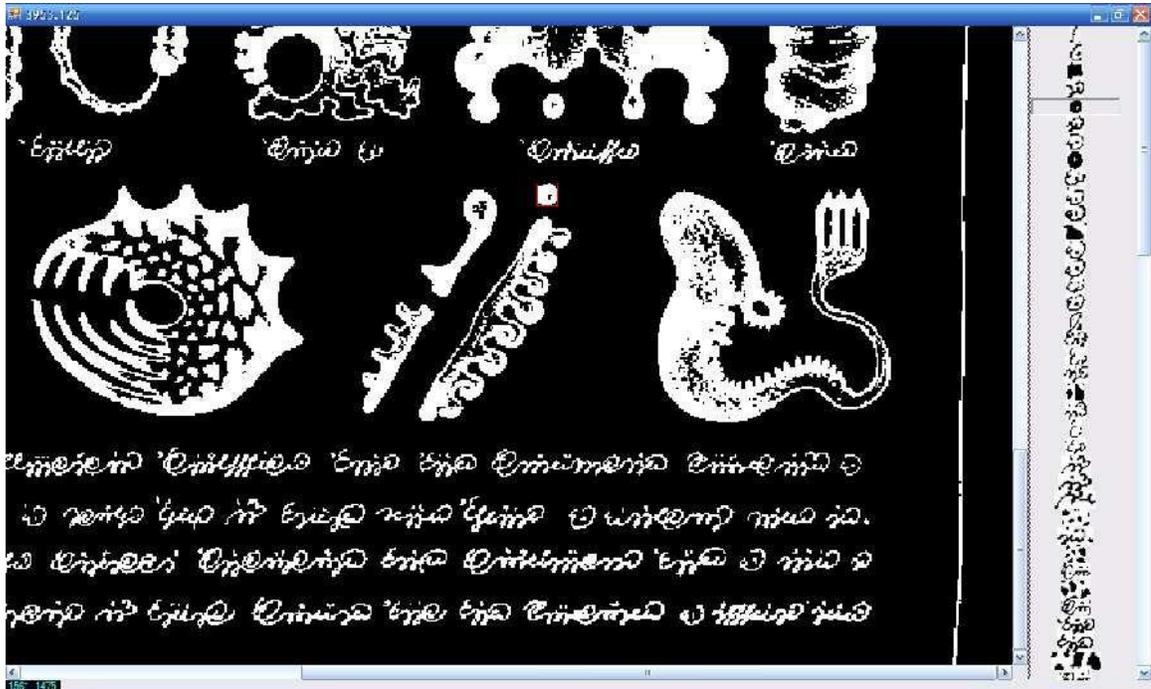


Figura 16: En la parte superior de la palabra los resultados son errores de extracción. Estos no son en realidad las palabras, sino más bien piezas de las ilustraciones. La investigación adicional podría encontrar una manera de filtrar éstos hacia fuera, posiblemente por el negro a la proporción de blanco.

Estudio 2: motor de búsqueda

Ordenar palabras de longitud proporcional a una buena mirada inicial a las similitudes de palabras en una página. La indexación de las palabras a través de múltiples páginas debe producir resultados más exactos en un área más grande del corpus. Como un primer paso hacia esta palabra manchada, se utilizó una técnica de esquina a juego, similar a, pero más simple y más rápido que la de Rothfeder (2003), para construir un motor de búsqueda. No se utilizaron intensidades de color gris, aunque sin duda hay margen para tal esfuerzo en la investigación futura. Más bien, la evaluación considera sólo las posiciones de los puntos de esquina. Para este estudio se utilizó el algoritmo detector de esquina 'Susan', ya que está incluido en AForge.NET. Cada punto de las esquinas de la imagen de destino se hace coincidir con el punto de esquina más

cercano en el examinado, si es que existe dentro de los 6 píxeles. El valor total partido se calcula como sigue: La diferencia total de la distancia para todos los pares partidos se divide por el número total de combinaciones para conseguir una distancia media. Al igual que en el texto de Rothfeder, esto se multiplica por el número total de puntos de esquina en el objetivo sobre el número de coincidencias, para favorecer a las exámenes con más coincidencias. Un problema potencial se encuentra en que, si sólo se encontró un par de coincidencias, pero su distancia era cero, esto daría lugar a un valor cero. Esto se evita forzando la distancia de pares de coincidencias para siempre por lo menos 0,01.

En una aplicación de motor de búsqueda, este algoritmo de validez queda demostrado. Al igual que en el motor de búsqueda de símbolos de cabecera, se lee en el portapapeles de Windows para su objetivo. Una palabra objetivo en la página 13 fue elegido por el autor porque sabía que era una palabra muy similar que estaba cerca, que difiere principalmente en el grosor de línea y signos diacríticos. El motor de búsqueda ha encontrado la palabra-objetivo y la palabra similar a la primera y segunda coinciden respectivamente. Las coincidencias comparten la misma forma general que la palabra-objetivo. Esta palabra incluye signos diacríticos. La investigación futura podría llevarse a cabo para eliminar los signos diacríticos de la comparación o para compararse por separado.



Figura 17: Demostración de motor de búsqueda de texto del cuerpo y de primer plano (abajo) de los resultados más importantes, páginas 10-48.

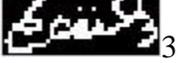
La palabra-objetivo está en la segunda fila, a la última fila de la página 13, elegido porque se conoce por el autor a ser una palabra muy similar que estaba cerca. Los puntos rojos en la imagen de destino muestran los puntos de esquina. La palabra de destino y palabras similares aparecen como se esperaba ya en las dos primeras coincidencias. Tenga en cuenta que coincida con 3 y 4, aunque no son tan similares, comparten formas de palabras similares con el objetivo, especialmente cuando se considera el posicionamiento de marcas diacríticas.

Es significativo que sólo hay dos casos de esta palabra en el códice. La singularidad de las palabras no se limita a este caso. Mientras que las secuencias particulares claramente se producen una y otra vez, incluso en la misma página, y hay señales muy cortas que aparecen a lo largo del códice, recurrencias de palabras que parecen raras a primera vista.

Para explorar más a fondo singularidad de la palabra en el Codex y la eficacia del algoritmo de búsqueda, se realizó una búsqueda de una palabra corta que es gráficamente simple en comparación con el corpus de palabras en el códice. La hipótesis radica en que incluso el uso de esta palabra sería raro o único. Los resultados se pueden ver en la página siguiente. A modo de referencia, hay más de 200 palabras en la página 12 del códice. En las páginas 10-65, en el motor de búsqueda se reunieron 131 grafías dierentes. De estos, 18 se asemejan a la palabra objetivo. (Varios de estos

resultados fueron clasificados a pesar de los que no se parecen a la palabra objetivo. Esto indica que el algoritmo es demasiado simple, en el futuro, debe ser juzgado con intensidades en escala de grises como en el texto de Rothfeder.) El motor de búsqueda muestra que aparece esta palabra a lo largo del código, desacreditando la hipótesis.

Tabla 4: tabla que muestra los resultados del motor de búsqueda para una palabra objetivo, páginas 10-65. El número que precede a la imagen es el rango asignado por el motor de búsqueda. Siguiendo la imagen es el número de página y entre paréntesis las coordenadas en la página. Sólo se muestran resultados semejantes a la palabra objetivo. De 131 resultados encontrados por el motor de búsqueda, estos 18 se asemejan a la meta. En la pagina 60 tiene 6 ocurrencias, mientras que 36 y 52 tienen cada uno 3; 62 tiene 2.

1.  52 (310, 472) (objetivo)	85.  60 (480, 1076)
2.  52 (1474, 428)	86.  36 (715, 1367)
10.  60 (273, 461)	93.  24 (856, 455)
11.  62 (560, 407)	109.  60 (1019, 660)
12.  36 (122, 507)	
36.  60 (117, 415)	
39.  32 (1118, 144)	
41.  40 (333, 105)	
50.  62 (1023, 711)	

53.  52 (468, 1132)	
59.  22 (1445, 402)	
sesenta y cinco.  60 (416, 1426)	
70.  36 (682, 1566)	
80.  60 (505, 1019)	

Sobre la base de los resultados anteriores, el algoritmo claramente podría beneficiarse de técnicas investigadas por el grupo de Rothfeder. Es sobre todo hay que señalar que este algoritmo no parece resolver coincidencias así en su estado actual de palabras cortas. Además, no hace falta trabajo para las palabras que son similares que hayan sido elaboradas con diferentes longitudes, como el siguiente par acerca de la parte superior de la página 52. Además, la aplicación de los algoritmos analizados en la palabras llamadas ‘mancha’ o ‘borrón’ en los papeles sin duda podría ayudar a estos problemas.



Figura 18: Dos fichas cerca de la parte superior de la página 52 que son similares pero no son detectados por el algoritmo de búsqueda, ya que varían en longitud.

Discusión de los resultados

Observaciones lingüísticas en la escritura

Cabeceras formularias

El motor de búsqueda hace que sea fácil darse cuenta de que las cabeceras siguen fórmulas. Dado que esto se señaló previamente por Derzhanski y se puede observar, simplemente accionando a través de los encabezados, no es un notable éxito del estudio. Sin embargo, el conjunto de aplicaciones fue capaz de añadir algunas ideas. En concreto, cuando el tipo de clasificador produjo varios tipos de grafías de una sola manera. Se encontró que estas fichas de un solo disparo siempre aparecen antes de la escritura en las páginas de imágenes. Esto podría ser la hipótesis de un académico, pero se necesitaría mucho tiempo para confirmar que las fichas de pre-escritura son únicas. El clasificador hizo esto evidente. Exploraremos más a fondo este fenómeno, utilizando el motor de búsqueda para encontrar todos los guiones en el código es una manera fácil de navegar por todas las fichas pre-escritura. Los resultados confirman que las fichas de pre-escritura son únicas. Incluso entre las páginas que parecen mostrar ilustraciones estrechamente relacionadas, las fichas no se parecen entre sí. Una posibilidad es que estos símbolos particulares son ideogramas. sistemas de escritura ideográficos que representan una idea con una sola ficha y por lo tanto requieren un gran número de fichas. Hay lenguas que se mezclan símbolos ideográficos con fichas fonéticas (Packard, 1974). Sin embargo, un sistema de escritura que no se correlaciona con la representación ortográfica de significado semántico sería inmanejable. Los algoritmos chinos reconocen miles de diferentes ideogramas, pero aprenden a reconocer y recordar a través de sus elementos constitutivos (Shu, 1997). Desde que Serafini es el único usuario responsable de la lengua en el código, esto no es un problema para él. Se puede hacer un nuevo

ideograma para cada página de ilustración y asignarlo, es decir, sin preocuparse de que está siendo utilizado por una comunidad más amplia. Es imposible que él podría haber pensado todos estos ideogramas antes de empezar el libro. Mediante la improvisación ellos, incluso si él no asigna conscientemente el significado que se las escribe, Serafini está creando una enorme colección de fichas únicas de las que extrae su significado, si lo tiene, de su uso particular. Tenían como bien podría no estar allí en absoluto. Se está estableciendo una práctica para su sistema de escritura que se traduce en más trabajo sin nada ganado, la de hacer una nueva ficha para cualquier nuevo conjunto de elementos, incluso si ese conjunto de elementos similar a uno que ya ha sido asignado. Esto tendría sentido si se utilizaran las fichas de otro lugar para hacer referencia a los artículos, pero no lo es. También hay que señalar que la práctica de que los hace únicos no es sostenible, al menos en lo que se limita al código. Eventualmente, la comunidad de la escritura comenzaría a la reutilización de fichas, ya sea consciente o inconscientemente. Por esta razón, debe estar al alcance de la Ley lingüística actual (libro, artículo, discusión, etc.).

Esta observación demuestra dos cosas. En primer lugar, si se pretende descifrar una lengua, Serafini creó un sistema de lenguaje que solo en parte él comprende. En segundo lugar, él estaba más preocupado por la apariencia o la novedad de la lengua de su utilidad; de lo contrario no habría creado más trabajo por sí mismo sin ningún beneficio funcional.

Variación libre dentro de un tipo

Se ha demostrado mediante el seguimiento de la secuencia de SFE a través del código que el tipo S se puede escribir con bucles ornamentados o sin estos bucles. Esta distinción no parece ser contextual, o que se encontraría en algo más que los primeros tres apariciones de esta palabra en el código. Más bien parece accidental (libremente

intercambiables). Esta observación es importante porque sugiere que otras fichas que se diferencian sólo por un adorno de bucle cerca de un punto final es probable que sean variaciones accidentales del mismo tipo. Esto sucede a menudo en el código.



Figura 19: Fichas en las cabeceras del Codex que parecen diferir solamente por un bucle alrededor de un punto final.

Especialmente desde las versiones en bucle son raros, esto podrían ser variaciones libremente intercambiables. Si bien la clasificación, el autor decidió hacerlos del mismo tipo, lo que conduce en última instancia a 53 tipos.

Esta observación es interesante porque el conjunto de variaciones permitidas parece estar bien definida, por lo que no es completamente descifrado sin restricciones. Si se añade un adorno, parece que sólo puede ser de unos pocos tipos, incluso a través de fichas. Si las variaciones eran completamente intercambiables, no habría necesidad de restringirlos de tal manera.

Tabla 5: Tipos de variación en tres de los tipos de cabecera.

Ninguna			
Lazo			
Sobresalir			
Extensión			
punto en su interior			
punto por encima			
Gancho			

otras combinaciones			
---------------------	---	--	---

Esta observación también es interesante porque en varios casos, el símbolo ornamentado y la ficha sin adornos se encuentran en la misma cabecera cerca unos de otros. Esta inconsistencia parecería implicar que la variación es a propósito y por lo tanto no es accidental.

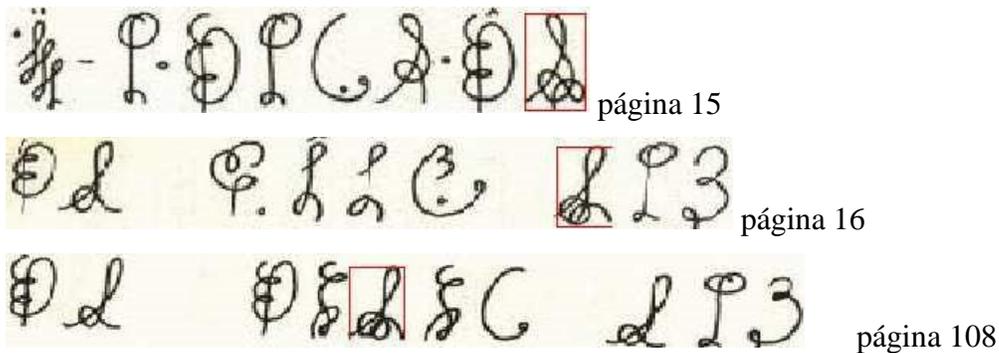


Figura 20: Fichas ornamentadas y sin adornos cerca uno del otro en la misma página. Ejemplos similares se pueden encontrar para los otros tipos.

Por supuesto, es posible que las variaciones son tanto libremente intercambiables y estrictamente definidas y que el autor es libre de la elección de utilizar lo que él considera una mezcla saludable de todas las variaciones. Si es así, simplemente se hace hincapié en la preocupación de Serafini con la aparición de la lengua: Al permitir la variación libre que da vida a la apariencia visual, pero restringiendo los tipos de variación que mantiene la apariencia de reglas lingüísticas.

No repetición de las palabras

Mientras que las cabeceras observan fórmulas y hay palabras cortas ciertamente repetidas (a primera vista supuestamente funciona palabras), hay una notable singularidad de palabras en el Codex, tanto en las cabeceras y el cuerpo del texto. Por

lo que el cuerpo del texto, se ha sugerido y se muestra en una muestra muy limitada que puede haber varias palabras que tienen características similares en una página, pero que no son exactamente lo mismo. Esto podría ser fácilmente verificado adicionalmente por una exploración más larga. Una situación similar se ha sugerido para las cabeceras: Las palabras que no son fijas por las fórmulas aparecen no más de dos veces, una vez en una página de texto y una vez en la página de ilustración. No se han encontrado contraejemplos. De hecho, la mayoría de las palabras en las cabeceras, además de contener tipos comunes, también contienen tipos muy poco comunes, un fenómeno que sugiere que su repetición en otros lugares es poco probable. Incluso aquellos compuestos en su totalidad de tipos comunes no parecen repetirse.

Que las palabras en las cabeceras no se repitan es concebible. Sería el equivalente de un libro sobre animales en los que cada página tiene un título completamente diferente, como “osos”, “gatos”, “perros”, “calamar”, etc. Son palabras que llevan tanto tiempo en el cuerpo del texto que no se repiten en la misma página mientras que las piezas de ellas son es más difícil de explicar su procedencia y significado. La coincidencia en la figura 17 podría haber sido dos instancias de la misma palabra a pesar de que la diacrítica está apagada, pero esta situación para las palabras de más de cuatro caracteres parece rara. Las secuencias se repiten, pero, por lo que este estudio ha encontrado, no las palabras largas. Podrían ser diferentes formas de la palabra. Esto indica o bien una lengua flexiva con tantas inflexiones que no se vuelven a utilizar, lo que parece poco probable, o una lengua aglutinante en la que una palabra contiene una cantidad sustancial de información transmitida a través de la combinación de varias raíces y morfemas. Es este rasgo combinatorio que guardan las palabras únicas. Si los puntos siguientes algunas palabras en el código se toman como períodos, las frases que delinean a veces contienen muchas palabras largas, e incluso dentro de una oración estas palabras largas a veces se parecen entre sí. Serafini parece estar reduplicando la

palabra en una forma ligeramente diferente para enfatizar o para algún otro propósito gramatical o semántica.

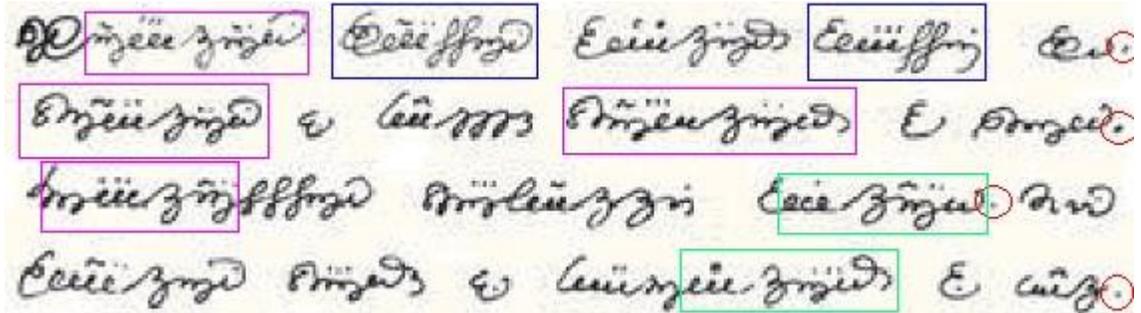


Figura 21: Cuatro posibles frases partir de la página 52.

Marcado de hecho en MS Paint. Las confluencias tienen un círculo rojo. Las palabras que aparecen en correspondencia entre sí están encuadradas en los colores correspondientes. Hay que tener en cuenta que incluso dentro de una sentencia de la misma palabra puede corresponder con una palabra muy similar.

Resumen

Ninguno de los resultados discutidos demuestra que el idioma del códice no es un lenguaje, pero los resultados demuestran que la preocupación de Serafini para imágenes era más fuerte que su preocupación por la construcción de un lenguaje funcional. En primer lugar, se crea una ficha única para cada página de ilustración. En segundo lugar, él define conjuntos restringidos de variaciones de tipo que parecen llenar ningún propósito lingüístico, pero se intercalan libremente. Estas técnicas no añaden ningún valor práctico, pero no se añade valor visual a la escritura, que no se requiere más trabajo por parte del autor. Serafini debe haber sentido que la disminución de la utilidad valía un aumento en el interés visual. En tercer lugar, Serafini repite secuencias de caracteres sin repetir las palabras exactas. De manera similar a la variación dentro de los tipos de caracteres, esto es una variación dentro de los tipos de palabras.

Observaciones metodológicas sobre los estudios

Resultados

Estas implementaciones muestran que es posible estudiar una lengua usando un ordenador visualmente, sin tener que recurrir a un esquema de la transcripción. Este hallazgo es muy significativo, ya que logra el objetivo principal de la investigación. No sólo es posible extraer y clasificar fichas cuando están claramente escritas, consistente y sin interrupciones, ya que la mayoría de las fichas de cabecera se encuentran en el código, sino también para extraer fichas plagados de los problemas tipográficos del texto del cuerpo en tan sólo unos segundos. En este caso, los algoritmos utilizados son elementales, implementado por alguien sin experiencia previa en el reconocimiento de imágenes o extracción. Sin duda, mucho más podría lograrse mediante un académico más avanzado en estos campos.

Problemas

La insuficiencia de los algoritmos ya se ha señalado, en particular la búsqueda y clasificación de las palabras en el texto del cuerpo, así como ciertos errores de extracción con el cuerpo del texto. Mientras que las implementaciones se reunieron con éxito con el texto de cabecera, otras escrituras sin descifrar plantearán problemas adicionales. En las tabletas de Rongo, por ejemplo, las líneas no son rectas, y las fotos de las tablillas en Lineal A no siempre son claras. Afortunadamente, rectas, planos claros de algunos de estos cuerpos están disponibles en línea, pero no todos. Otras cuestiones se abordarán en la conclusión.

CONCLUSIÓN

Desciframiento asistido por ordenador

Este estudio demuestra que es posible estudiar un sistema de escritura desconocida por un ordenador de forma gráfica, sin necesidad de utilizar un esquema de la transcripción. La extracción de caracteres y técnicas de reconocimiento han alcanzado un nivel de éxito que permite la clasificación asistida por ordenador de símbolos desconocidos y la identificación de secuencias recurrentes, dos sub-problemas primarios en desciframiento. (Es importante recordar que los algoritmos utilizados en este estudio son sólo el más simple de implementar. Existen alternativas mucho más refinadas que pueden investigarse. Dado que ninguna de transcripciones tiene que ser producida por el usuario humano, este tipo de sistema promete menos esfuerzo y menos posibilidades de error de los sistemas existentes. Porque no solo permite examinar la escritura en su estado original, sino que también abre la puerta a nuevas ideas sobre la escritura.

El Codex Seraphinianus

Como se indica en la discusión de los resultados, el único hallazgo claro acerca de la escritura en el código es que se pretende que sea visualmente interesante. Serafini puso tiempo y esfuerzo en las características visuales que al parecer no sirven para nada visto desde un modo lingüístico. Estas características visuales están equilibradas de una manera que crea patrones sin repetición continua de palabras y símbolos. Serafini era un artista y no un lingüista, Serafini puede haber pensado que este equilibrio visual sería la mejor manera de crear un pseudo-lenguaje. Es posible que haya querido pasar su tiempo de trabajo en el aspecto artístico, no perder el tiempo en el aspecto lingüístico que no se sentía capacitado para capturar. Esta conjetura es apoyada por la

declaración de Serafini que la lectura del código está destinado a ser un proceso visual, no uno lingüística (Prodi, 2009). Si bien afirma que visualmente él piensa que mezcla elementos del árabe, cuneiformes y algunas lenguas muertas (Manetto, 2007), que nunca da ningunas opiniones sobre el aspecto lingüístico de la escritura, a pesar de que claramente algunos podrían pensar que se parece a un alfabeto, o que las palabras parecen repetirse en una forma flexionada. Él no parece preocupado por el aspecto lingüístico (o pseudolingüístico) de la escritura.

Los conocimientos obtenidos de este análisis informático podría ser una valiosa adición a un estudio interdisciplinario del código. Por ejemplo, las características de la escritura que no sirven para nada lingüístico están aún en contra de lo que cabría esperar lingüísticamente podría ser un tema intrigante para una investigación artística o psicológico. Además, ahora que se han identificado los tipos de fichas de cabecera y tipos de variación en las fichas de cabecera, una investigación sobre las influencias inconscientes de la escritura puede tomar ventaja de esta información.

Áreas de estudio adicional

Hay varias partes de esta investigación que podrían beneficiarse de estudios adicionales, especialmente bajo la guía de un experto en análisis de imágenes, en particular de documentos y análisis de la escritura. Algunos de ellos se han mencionado ya y se reiterarán aquí junto con los demás.

Este estudio se ha basado en métodos de bajo costo de extracción en cuanto a tiempo y dinero. La calidad de las imágenes extraídas, en especial del texto del cuerpo, sufrió a causa de esto. Incluso con una amplia gama pre-procesamiento, los diseños de algunas de las palabras y las ilustraciones impidieron la extracción. Estos problemas podrían resolverse por alguien experto en el análisis de documentos que podrían hacer

que las herramientas de segmentación y extracción del corpus en lugar de depender de software de terceros.

Las técnicas de formación de manchas de palabras utilizadas en este estudio son simplistas. Mientras que los algoritmos se desempeñaron bien en las fichas de cabecera, probablemente mejores resultados pudieron obtenerse tanto para la cabecera y el cuerpo del texto, considerando los avances recientes.

Debido a su naturaleza interconectada, el cuerpo del texto se estudió como palabras completas. Sin conocer los componentes de las palabras no se pueden segmentar correctamente. Por lo tanto, este estudio parece mostrar tanto cuando un texto desconocido está conectado de esta manera las fichas deben ser examinados, al menos inicialmente, a nivel de palabra tanto como es conocido. Una vez que el motor de búsqueda se mejora al estado de que las palabras se pueden clasificar con éxito (si se desea), la investigación adicional podría investigar algún tipo de búsqueda heurística que pudiese mirar a todos los tipos de palabras y proponer posibles segmentaciones.

Sólo los tres primeros capítulos se han utilizado para determinar los tipos de cabecera. Los 53 tipos resultantes deben ser utilizados como datos de entrenamiento para que los restos de las cabeceras del Codex se puede clasificar. Si los estudios futuros están de acuerdo con la clasificación, los análisis deben realizarse en estas características. Debido a que es una lengua falsa, los estudios podrían comenzar inmediatamente con la clasificación propuesta aquí durante los tres primeros capítulos, y más tarde los capítulos siguientes, y así sucesivamente, para medir la consistencia de Serafini. La modelización estadística puede ayudar a aclarar cómo el lenguaje falso en el código se relaciona con otros idiomas. En particular, un modelo de Markov se podría utilizar para descubrir las reglas Serafini lo utiliza para producir las cabeceras. Se cree que durante los procesos automáticos del lóbulo parietal todo se convierte en

un producto de reglas simbólicas y en un manifiesto, y el modelo de Markov podría mostrar como de similares son estas reglas.

Referencias

Alexander, Lynn (2009a). Michael Jacobson. Entrevista. Agosto de 2009. Entrevista serie. Nueva York, EE.UU.: <http://fullofcrow.com/prate/2009/08/michael-jacobson/>.

Alexander, Lynn (2009b). Tim Gaze, Revista Asemica. Entrevista serie. Nueva York, EE.UU.: <http://fullofcrow.com/prate/2009/11/tim-gaze/>.

Allen, Julie D., ed. (2009) El estándar Unicode / El consorcio Unicode, versión 5.2. Mountain View, CA: Unicode, Inc.

Arkin, Arthur M. (1981). Somniloquias: La psicología y psicofisiología. Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates, Inc.

Barmpoutis, Angelos, Eleni Bozia y Robert Wagman (2009). Un nuevo marco para la reconstrucción y el análisis de las antiguas inscripciones romanas 3D. *Visión Artificial y Aplicaciones*. Publicado en línea.

Bekedam, DJ, GHA Visser, JJ de Vries y HFR Precht (1985). El comportamiento del motor en el crecimiento fetal retardado. *Desarrollo Humano Temprano* 12 (2): 155-165.

Bhardwaj, Anurag, Sriranguraj Setlur y Venu Govindaraju (2009). Técnicas de manchado de palabras clave de documentos sánscrito. En G. Huet, A. Kulkarni, y P. Scharf (eds.), *Sánscrito CL 2007/2008*, LNAI 5402 (403-416). Berlín y Heidelberg: Springer-Verlag.

Bodard, Gabriel, y Ryan Baumann (2009). Las oportunidades para la epigrafía en el contexto de la digitalización 3D. Presentación en AIA 110ª Reunión Anual.

Breton, André (1972). *Manifiestos del surrealismo* (Richard Seaver y Helen R. Lane, trad.). Ann Arbor: University of Michigan Press.

Brotchie, Alastair, y Mel Gooding (2003). *Juegos surrealistas*. Boston y Londres: Shambhala Press.

- Bruel, Thomas M (2009). Aplicando el sistema OCR para OCRopus académica literatura sánscrita. En G. Huet, A. Kulkarni, y P. Scharf (eds.), *Sánscrito CL 2007/2008*, LNAI 5402 (391-402). Berlín y Heidelberg: Springer-Verlag.
- Chadwick, John (1977). Revisión de minoica lineal A por David W. Packard. *Los ordenadores y las Humanidades* 11 (1): 51.
- Chadwick, John (2004). Lineal B. En JT Hooker (eds.), *La lectura del pasado*. London: The British Museum Press.
- Cheriet, Mohamed, Nawwaf Kharma, Cheng-Lin Liu, y Ching Y. Suen (2007). *Los sistemas de reconocimiento de caracteres: Una guía para los estudiantes y profesionales*. Hoboken, Nueva Jersey: John Wiley & Sons, Inc.
- Chopde, Avinash (2001). interfaz de línea de ITRANS. <http://www.aczoom.com/itrans/online/>. Último actualizado 12 de de junio de 2001.
- Coe, Michael (1999). *Rompiendo el código Maya*. Nueva York, Nueva York: Thames & Hudson Inc.
- Cohen, Jonathan D., James L. McClelland y Kevin Dunbar (1990). En el control de procesos automáticos: Una cuenta de procesamiento distribuido en paralelo del efecto Stroop. *Psychological Review* 97 (3): 332-361.
- Coulson, Michael (1992). *Sánscrito: Un curso completo para principiantes*. Chicago, Illinois: Contemporary Books.
- Cutten, George Barton (1927). *Hablar en lenguas: Históricamente considerada como psicológicamente*. EE.UU.: Yale University Press.
- Danielsson, PE (1980). mapeo de distancia euclidiana. *Computación Gráfica y Procesamiento de Imágenes*, 14: 227-248, 1980.
- Davis, Brent (2010). Linear A: toques de morfología Minoan flexional. Presentación en la reunión anual de AIA de 2010.
- Derzhanski, IA (2004). *Codex Seraphinianus: algunas observaciones*. <http://www.math.bas.bg/~iad/serafin.html>. Última actualización 2004.
- Dietrich, Arne (2003). *neuroanatomía funcional de los estados alterados de conciencia:*

La hipótesis de la hipofrontalidad transitoria. *Consciousness and Cognition* 12: 231-256.

Fletcher, Lloyd A., y Rangachar Katsuri (1988). Un algoritmo robusto para la separación cadena de texto a partir de imágenes Mezcla de texto / gráficos. *IEEE Transactions on Análisis de patrones y la máquina de Inteligencia* 10 (6): 910-918.

Flournoy, Théodore (1900). Desde la India hasta el planeta Marte: (. Daniel B. Vermilye, trans) Un estudio de un caso de sonambulismo con la glosolalia. Nueva York, Nueva York: Harper & Brothers.

Chico, Jacques (2005). El Rongorongongo de Isla de Pascua. <http://www.rongorongongo.org/>. Último actualizado hace 8 febrero, 2005.

Hardesty, Larry (2010, 30 de junio). Computadora descifra automáticamente el idioma antiguo. MITnews. <http://web.mit.edu/newsoffice/2010/ugaritic-barzilay-0630.html>.

Hobson, J. Allan (1995). Dormir. Nueva York, Nueva York: Scientific American Library.

Knorozov, Yuri (ed.) (1965). *Predvaritel'noe soobshchenie ob issledovanii protoindijskih tekstov*. Moscú: Institut Etnografii, Akademiya Nauk SSSR.

Kompalli, Suryaprakash (2007). Un marco estocástico para la fuente independiente de devanágari OCR. Disertación, Universidad Estatal de Nueva York en Buffalo. UMI número: 3.244.288.

Laberge, S. (2000). El sueño lúcido: la evidencia y la metodología. *Behavioral and Brain Sciences* 23 (6), 962-3.

Laberge, S. y H. Rheingold (1990). *Explorando el mundo del sueño lúcido*. Nueva York: Ballantine.

Lee, Rob, Philip Jonathan, y Pauline Ziman (2010). símbolos pictos revela como un lenguaje escrito mediante la aplicación de la entropía de Shannon. *Proceedings of the Royal Society A: Matemáticas, Física y Ciencias de la Ingeniería*, 31 de Marzo: 1-16.

Lenneberg, Eric H (1967). *Fundamentos biológicos del lenguaje*. Nueva York: J. Wiley and Sons.

Leymarie, F. y MD Levine (1992). Una nota sobre la “trama rápida propagación de la distancia de exploración en la red rectangular discreta.” CVGIP: Interpretación de Imágenes, 55: 85-94.

Miembro, Charles J. y Allen R. Baum (2008). sustratos neurales de la actuación musical espontánea: Un estudio de resonancia magnética funcional de la improvisación de jazz. PLoS ONE 3 (2): e1679.

Manetto, Francesco (2007). Historia de un libro raro. El País.com.
http://www.elpais.com/psp/index.php?module=elp_pdapsp&page=elp_pda_noticia&idNoticia=20071111elpepspag_13.Tes+y+seccion+=cul.

Manmatha, R., Chengfeng Han y EM Riseman (1995). Palabra manchado: un nuevo enfoque para la escritura de indexación. Informe Técnico CS-UM-95-105, Ciencias de la Computación Departamento de la Universidad de Massachusetts en Amherst, MA.

Manmatha, R., Chengfeng Han, EM Riseman y el WB Croft (1996). La indexación de escritura mediante la comparación de textos. Bibliotecas digitales '96: primero ACM Conferencia Internacional de Bibliotecas Digitales.

Marslen-Wilson, William (1973). La estructura lingüística y el habla sombreado con latencias muy cortos. Naturaleza 244 (5417): 522-523.

Madura, Patrice Dawn (1996). Las relaciones entre el rendimiento vocal de improvisación del jazz, el conocimiento teoría de jazz, la capacidad imitativa, la experiencia musical, la creatividad y el género. Revista de Investigación en Educación Musical 44 (3): 252-267.

Mayo, L. Carlyle (1956). Una encuesta de la glosolalia y fenómenos relacionados con las religiones no cristianas. Antropólogo americano 58 (1): 75-96.

Meyer, Elise, Pierre Grussenmeyer, Temy Tidafi, Claude y Jean PARISEL Revez (2004). La fotogrametría para la encuesta epigráfica en la gran sala hipóstila del templo de Karnak: Un nuevo enfoque. Archivos Internacionales de fotogrametría teleobservación y espacial Information Sciences 35 (5): 377-382.

Meyer, Elise, Pierre Grussenmeyer, Claude Parisel, Jean Revez y Temy Tidafi (2006). Una solución computarizada para el estudio epigráfico en Egipto Templos. Journal of Archaeological Science. 33 (11): 1605-1616.

- Miller, Kirk y Jacques Guy (2008). Algo mal con el inventario básico de Pozdniakov de rongorongo. foro de discusión en KRR_Study Yahoo Group, inició el 25 de junio, 2008.
- Mori, Shunji, Hirobumi Nishida y Hiromitsu Yamada (1999). Reconocimiento óptico de caracteres. Nueva York, Nueva York: John Wiley & Sons, Inc.
- Moisés, William Stainton (1882). Psicografía: Un tratado sobre una de las formas objetivas de los fenómenos psíquicos o espirituales (2ª ed.). Londres: La Asociación de la Prensa de Psicología.
- Nettl, Bruno y Russell, Melinda (eds.) (1998). En el curso de ejecución: estudios en el mundo de la improvisación musical. Chicago: University of Chicago Press.
- Nicchiotti, G. y C. Scagliola (2000). Un método de segmentación de palabras cursiva simple y eficaz. Proc. 7 de IWFHR, Amsterdam.
- O'Neill, Daniel (2005). Etiología de la plaga baile. *Intercultura: un Interdisciplinary Diario* 2 (3): 1, 7-12.
<http://interculture.fsu.edu/pdfs/oneill%20dancing%20plague.pdf>.
- Packard, David W. (1974). *Minoica lineal A*. Berkeley, California: University of California Press.
- Parkes, JD (1985). *El sueño y sus trastornos*. London: WB Saunders.
- Parpola, A. (1994). *Descifrar la escritura del Indo*. Cambridge University Press, Cambridge.
- Patel, Samir S. (2010). El enigma del Indo. *Arqueología* 63 (2): 18, 58, 60, 65, 66.
- Pozdniakov, Konstantin e Igor Pozdniakov (2007). la escritura y el idioma rapanui rapanui: Los resultados preliminares de un análisis estadístico. *Foro de Antropología y Cultura* 3: 3-36.
- Prechtl, Heinz FR y Brian Hopkins (1986). transformaciones de desarrollo de los movimientos espontáneos en la primera infancia. *Desarrollo Humano Temprano* 14 (3-4): 233.238.
- Prodi, Enrico (2009). El correo electrónico personal al autor. Sep. de 10 del 2009.

- Rao, RPN, N. Yadav, MN Vahia, H. Joglekar, R. Adhikari y I. Mahadevan (2009a). entrópico pruebas para la estructura lingüística en la escritura del Indo. *Ciencia* 324 (5931): 1165. Material de apoyo en línea en www.sciencemag.org/cgi/content/full/1170391/DC1.
- Rao, RPN, N. Yadav, MN Vahia, H. Joglekar, R. Adhikari y I. Mahadevan (2009b). Un modelo de Markov de la escritura del Indo. *PNAS* 106 (33): 13685-13.690.
- Rodríguez Martínez, Ma. del Carmen, Ponciano Ortíz Ceballos, Michael D. Coe, Richard A. Diehl, Stephen D. Houston, Karl A. Taube, y Alfredo Delgado Calderón (2006). la escritura más antigua del Nuevo Mundo. *Ciencia* 313 (5793): 1610-1614
- Rosenberg, Daniel (2000). Hablando de Marte. *Gabinete Revista 1: inventó Idiomas*. Invierno 2000/01.
- Rath, TM y R. Manmatha (2003). imagen Palabra coincidentes con el alineamiento de tiempo dinámico. *Proc. de la Conf. en Visión por Computador y Reconocimiento de Patrones (CVPR) 2: 521-527*. Madison, WI, 18-20 de de junio de 2003.
- Rosenzweig, Mark R., Arnold L. Leiman y S. Marc Breedlove (1999). *psicología biológica: Una introducción a la neurociencia conductual, cognitiva y clínica*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Rothfeder, JL, S. Feng y TM Rath (2003). El uso de correspondencias característica esquina para clasificar imágenes de la palabra por similitud. *Proc. del Taller de Análisis de documentos y recuperación de imágenes (DIAR)*, Madison, WI, 21 de Junio del 2003.
- Saver, Jeffrey L., y John Rabin (1997). Los sustratos neurales de la experiencia religiosa. *El Diario de Neuropsiquiatría y Neurociencias Clínicas* 9 (3): 498.510.
- Serafini, Luigi (1983). *Codex Seraphinianus (primera edición americana)*. Nueva York: Abbeville Press.
- Sheridan, Susan (2002). La importancia neurológica del dibujo infantil: El garabato hipótesis. <http://www.marksandmind.org/scribble.html>.

- Sheridan, Susan (2010). Garabatos: El eslabón perdido en una teoría del lenguaje humano en el que las madres y los niños juegan un papel importante. Las hipótesis presentadas al médico Diario. <http://www.marksandmind.org/scribbs.html>.
- Shu, Hua y Richard C. Anderson (1997). Papel de la conciencia radical en el carácter y la palabra adquisición de los niños chinos. *Reading Research Quarterly* 32 (1): 78-89.
- Silvestri, Rosalía (2007). terrores Somnambulation, Somniloquy, y el sueño. En Culebras, Antonio (ed.), *Trastornos del sueño y enfermedades neurológicas*. Nueva York, Nueva York: Informa Healthcare EE.UU., Inc.
- Snyder, Benjamin, Regina Barzilay y Kevin Knight (2010). Un modelo estadístico para perdido desciframiento del lenguaje. Presentado en la Reunión Anual de la Asociación de Lingüística Computacional.
- Solms, Mark (1997). *La Neuropsicología de los sueños*. Mahwah, Nueva Jersey: Lawrence Erlbaum Associates.
- Sproat, Richard (2003). cadena aproximada coincide en el corpus Rongorongo. <http://compling.ai.uiuc.edu/rws/ror/>. Último actualizado 11 de enero, 2003.
- Sproat, Richard (2010). símbolos antiguos, lingüística computacional, y las prácticas de revisión de las revistas científicas generales. *Computational Linguistics*, 36 (3).
- Van der Hart, Onno, y Rutger Horst (1989). La teoría de la disociación de Pierre Janet. *Journal of Traumatic Stress* 2 (4): 397-412.
- Van Lancker, Diana (1979). Idiomatica frente a la interpretación literal de frases ambiguas ditropically. Manuscrito.
- Visser, GHA, RN Laurini, JIP de Vries, DJ Bekedam, y HFR Prechtl (1985). conducta motora anormal en los fetos anencefálicos. *Desarrollo Humano Temprano* 12 (2): 173-182.
- Wechsler, Allan (1987, de 23 de julio). SF-AMANTES digest V12 # 334. Difusión en la lista de correo SF-amantes. <http://groups.google.com/group/rec.arts.books/msg/25e55b7771903c1d>.
- Williams, J. Mark G., Andrew Mathews y Colin MacLeod (1996). La tarea Stroop emocional y psicopatología. *Psychological Bulletin* 120 (1): 1, 3-24.

Yadav, N. (2007). escritura del Indo: la búsqueda de la gramática. Presentación realizada por primera vez en el Indo Guión: problemas y perspectivas de 2006. Actualizado el 23 de diciembre de 2007. Disponible en <http://www.harappa.com/script/tata-writing/tata-indus-script-research.html>.

Yadav, N., MN Vahia, I. Mahadevan y H. Jogelkar (2008). Un enfoque estadístico para el patrón de búsqueda en la escritura del Indo. *Revista Internacional de Dravidian Linguistics*, 37 (1): 39-52.

Yanikoglu, B. y P.A. Sandon (1998). Segmentación de off-line letra cursiva usando programación lineal. *Patt. Reconoce.* 31 (12): 1825-1833.

Younger, John (2000). textos en lineal A en la transcripción fonética. <http://people.ku.edu/~jyounger/LinearA/>. Último actualizado 14 de marzo de 2009.