

#noviembreHD. Cuarto congreso de la Asociación Argentina de Humanidades Digitales (AAHD). Asociación Argentina de Humanidades Digitales, Buenos Aires, 2021.

Análisis cuantitativo de un corpus textual de historia oral utilizando Voyant Tools.

Ignacio MORENO NAVA.

Cita:

Ignacio MORENO NAVA (2021). *Análisis cuantitativo de un corpus textual de historia oral utilizando Voyant Tools. #noviembreHD. Cuarto congreso de la Asociación Argentina de Humanidades Digitales (AAHD). Asociación Argentina de Humanidades Digitales, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/noviembrehd/18>

ARK: <https://n2t.net/ark:/13683/ehed/qyy>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-sa/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Análisis cuantitativo de un corpus textual de historia oral utilizando Voyant Tools

Quantitative Analysis of an Oral History Textual Corpus Using Voyant Tools

Coordinación

Gimena del
Río Riande

Romina De
León

Ignacio MORENO NAVA

imoreno@ucemich.edu.mx

Universidad de La Ciénega del Estado de Michoacán de Ocampo

<https://orcid.org/0000-0003-2537-6840>

1. DESCRIPCIÓN DEL CORPUS

El antecedente de este estudio de caso es una investigación, cuyo objetivo fue digitalizar registros de testimonios orales que contuvieran datos de la leyenda de Martín Toscano provenientes del Archivo de Historia Oral (AHO) de la Unidad Académica de Estudios Regionales de la Coordinación de Humanidades de la Universidad Nacional Autónoma de México (UAER-COHU-UNAM) en Jiquilpan, Michoacán, México, y aplicar tecnología de reconocimiento óptico de caracteres (OCR) para generar un corpus textual digital susceptible de ser explorado mediante búsquedas con cadenas de caracteres y herramientas de las Humanidades Digitales (Moreno-Nava, 2020). Los contenidos del AHO fueron generados en la década de 1980, contando con adultos mayores como informantes y cuyas fechas de nacimiento oscilaban en un rango que iba de 1888 a 1920.

El legendario bandolero Martín Toscano, referido generalmente en la oralidad como una aparición sobrenatural que resguarda tesoros y riquezas, fue un hombre de carne y hueso, que acompañado de sus compañeros de gavilla asaltaba conductas y haciendas españolas para impedir el saqueo de estas tierras, proveer de provisiones a sus soldados e iniciar con ello el movimiento preinsurgente en la región Ciénega de Chapala.

Martín Toscano González nació el 13 de noviembre de 1754 en Atoyac, Jalisco. Morisco, pero autonombrado Mestizo. Hijo de Blas Toscano, clasificado como español y Juana Catharina González, mulata (Moreno-Nava, 2017). Capitán de gavilla, compartió correrías y aventuras junto a Francisco Gil y Marcos Coronel, también capitanes. Encarnaron el sentir de un pueblo hartado de las vejaciones perpetradas por los españoles. Sus acciones mostraban un

fuerte sentimiento de amor a su tierra, sentimiento contrario al que demostraban a los “gachupines” como ellos mismos los llamaban. Haciendas y conductas españolas eran blanco de sus acciones, por lo cual se ganaron el desprecio y el odio de estos. Dignos representantes de los que fueran los albores de un *bandolerismo social* como lo llamaría Hobsbawm (1968). Fueron aprendidos en varias ocasiones y muchas más escaparon de las garras españolas hasta que, a finales de 1795, Martín Toscano es aprendido de manera definitiva durante la resistencia de Sayula. Francisco Gil y Marcos Coronel corren la misma suerte meses después y se abre una causa penal contra ellos. La corona española tuvo que intervenir preguntando quiénes eran esos hombres que hacían temblar los territorios de la Nueva Galicia y Valladolid. Sus vidas terminaron el día 12 de enero de 1803 al ejecutarse sobre ellos una sórdida sentencia. Sus cabezas fueron exhibidas a manera de escarmiento en los últimos puntos de sus capturas (Moreno-Nava, 2019).

La historia oral genera perspectivas y aspectos distintos a los considerados por la historia tradicional. Es una ventana a las vivencias y recuerdos expresados por el habla de aquellos que comparten su vertiente mediante la lengua hablada. Abre posibilidades de reconstrucción histórica entre sectores que no transmiten su experiencia por escrito, coadyuva a una construcción de la memoria personal proporcionando un paisaje más completo y complejo de los procesos históricos, rescata la voz de los marginados (campesinos, arrieros, pescadores, etc.).

Existen al menos dos maneras de generar historia oral. La primera son los archivos de la palabra (archivos de oralidad) para construir fuentes de consulta mediante el depósito de las entrevistas procesadas (Collado, 1994). Es el caso del Archivo de Historia Oral de la Unidad Académica de Estudios Regionales de la Coordinación de Humanidades de la Universidad Nacional Autónoma de México, sede Jiquilpan, en el estado de Michoacán, en México. La segunda manera involucra la recopilación de fuentes orales, además de consultarlas. Es importante remarcar que, a la par de la obtención de materiales, se generan fuentes para los investigadores del futuro.

2. TIPO DE TRABAJO SOBRE LOS DATOS (CUALITATIVO Y CUANTITATIVO)

Se utilizó la herramienta Voyant Tools¹ para efectuar un trabajo de tipo cuantitativo sobre los datos y se realizó un análisis cualitativo somero en términos de la frecuencia de palabras asociadas con aspectos representativos de la leyenda. Previo a ello, el corpus textual fue curado; se eliminaron de los contenidos los datos de identificador clave, páginas, descripción de la entrevista, fecha y datos del entrevistado. En lo referente a los contenidos de las entrevistas se quitaron las iniciales del entrevistador y el informante, de manera tal que se dejaron solamente los contenidos de la conversación.

A continuación, se muestra un ejemplo de un fragmento de registro de testimonio oral del AHO (UAER-COHU-UNAM). Resaltados en gris se encuentran los datos que fueron eliminados durante la fase de curación del texto, previa al análisis cuantitativo.

¹ Accesible desde: <https://voyant-tools.org/>.

Aguilera Flores, Sabas

(AHOCLC-Z1-E: 93/65 pp.)

Entrevista con el señor Sabás Aguilera Flores, realizada por Griselda Villegas M. los días 13 y 22 de septiembre de 1983 en Jiquilpan, Mich. (2 sesiones).

Nació en 1900; originario de Jiquilpan, Mich.

Martín Toscano (pp. 50 - 51).

G.V. Don Sabás, ¿Y usted oyó hablar de ese personaje que se aparece y que anda vestido de negro?

S.A.F. No, no.

G.V. ¿Martín Toscano?

S.A.F. ¡Ah, Martín Toscano! Sí, ese sale de la nada, al menos allí por la virgencita, pero se aparece en las cuevas donde robaba cerca de "Las Higueras". Anda montado en un caballo negro y usa un sombrero del mismo color.

G.V. ¿Y a usted nunca se le llegó a aparecer?

S.A.F. A mí no me salió para que le echo mentiras, pero a mi hermano sí, él iba pasando por la virgencita para llegar a la siembra.

G.V. ¿Y cómo lo cuenta al que le salió?

S.A.F. Cuando lo vio, lo miró, pero no le dijo nada. A otros sí les hablaba para que sacaran el dinero y gozaran. Traía gente para que robaran todo el dinero ni las Higueras y el troncón dejaron.

G.V. ¿Y cómo les decía?

S.A.F. – Saca ese dinero que esta allá - y te llevaba de la mano.

G.V. ¡Ah!, te llevaba de la mano.

S.A.F. Sí, te llevaba de la mano y si tú tenías valor te ibas con él.

G.V. ¿Muerto?

S.A.F. Era su fantasma. Tenías que tener valor; uno no pudo, se volvió loco y se murió, pero allí estaba enterrado el dinero. Tenías que llevártelo todo. El hijo de Manuel Orozco, Trino Lúa dijo: – Vamos a desenterrar el dinero que me encontré.

Sacaron una camioneta de pura plata de la de antes, aquí en Las Higueras. Toscano robaba en Cojumatlán, Tizapán, El Carpintero, en Contla, por todos los caminos reales. Antes no había tren, ni coche, solo animales de carga. Cuando la guerra de los Ornelas, los franceses solo tenían mulas para cargar, por eso aquí en Jiquilpan todavía hay mucho dinero enterrado.

3. RESULTADOS, ALCANCES, CONSTATAción O NUEVAS CONCLUSIONES, DESAFÍO Y LIMITACIONES

De manera inicial, se digitalizaron y se sometieron a procesamiento de OCR así como a limpieza y revisión de transcripción 25 registros de testimonios orales que contenían datos de la leyenda de Martín Toscano. El corpus textual digital generado constó de 50 cuartillas, 11,144 palabras, 49,158 caracteres sin espacios, 580 párrafos y 1,235 líneas.

Posteriormente, en Voyant Tools se realizó un primer análisis para obtener un panorama general cuantificatorio del corpus. Se consiguieron los siguientes resultados desde el marco teórico de la lectura distante (Moretti, 2015), efectuando algunas operaciones para generar automáticamente diversas visualizaciones:

Sumario: 9,564 total de palabras y 1,934 formularios de palabra única

Densidad del vocabulario: 0.202

Promedio de palabras por oración: 14.9

Palabra más frecuente en el corpus: *a* (271); *no* (172); *dinero* (111); *le* (108); *me* (76) (figura 1).

ción de resultados, de manera tal que luego se podrá abundar en ellos utilizando algunas otras funciones de Voyant Tools.

Partiendo del análisis desde el enfoque de la lectura distante (Moretti, 2015) deberá procederse posteriormente a profundizar aspectos del corpus que podrán ser observados mediante la lectura cercana. De igual forma, el corpus se presta muy bien para comenzar su marcado utilizando TEI-XML, para integrarlo en visualizaciones desde la perspectiva de las geohumanidades, así como su exploración mediante análisis de redes. Estos desafíos sin duda alguna permitirán generar nuevas conclusiones sobre el análisis semiótico convencional de este tipo de textos y ampliarán el alcance investigativo desde la óptica de las Humanidades Digitales.

REFERENCIAS BIBLIOGRÁFICAS

- Collado, C. (1994). ¿Qué es la historia oral? En G. de Garay (Comp.), *La historia con micrófono. Textos introductorios a la historia oral* (pp. 13-32). Instituto de Investigaciones Dr. José Ma. Luis Mora.
- Gutiérrez De la Torre, S. (2019). Análisis de corpus con Voyant Tools. *The Programming Historian en español*, (3). <https://programminghistorian.org/es/lecciones/analisis-voyant-tools>
- Hobsbawm, E. J. (1968). *Rebeldes primitivos: estudio sobre las formas arcaicas de los movimientos sociales en los siglos XIX y XX*. Ariel.
- Moreno-Nava, I. (2017). *La leyenda de Martín Toscano: una investigación integrativa desde el Pensamiento Complejo y la Transdisciplina*. [Tesis de Doctorado, Multiversidad Mundo Real Edgar Morin]. Repositorio institucional - Multiversidad Mundo Real Edgar Morin.
- Moreno-Nava, I. (2019). La fiesta de “La Plata del Rey” en Pajacuarán, Michoacán. En V. M. Mendoza Sánchez (Recop.), *Fiestas, tradiciones y devociones populares en los estados de Colima, Jalisco y Michoacán. Memoria del séptimo coloquio regional de Crónica, Historia y Narrativa* (pp. 361-379). Centro de Preservación del Patrimonio Histórico de Tuxpan, Jalisco, A. C. - Asociación de Cronistas de Pueblos y Ciudades del Estado de Colima, A. C.
- Moreno-Nava, I. (2020). Tecnología OCR para la transcripción de registros de testimonios orales de la leyenda de Martín Toscano. *Interpretextos*, 13(24), 209-228. <http://ww.ucol.mx/interpretextos/resultados.php?idarti=425>
- Moretti, F. (2015). *Lectura distante*. Fondo de Cultura Económica.
- Sinclair, S. y Rockwell, G. (2016). *Voyant Tools*. <http://voyant-tools.org>
- Unidad Académica de Estudios Regionales de la Coordinación de Humanidades de la Universidad Nacional Autónoma de México, sede Jiquilpan (UAER-COHU-UNAM). Archivo de Historia Oral (AHO). En Archivo Histórico. <http://uaer.humanidades.unam.mx/archivo-historico/introduccion/>