

## EL TRATAMIENTO DE LA PREHISTORIA EN TWITTER: UNA APROXIMACIÓN A LOS AFECTOS DESDE R

### 1. INTRODUCCIÓN

Las redes sociales se han convertido en uno de los principales canales de comunicación y difusión de contenidos. De acuerdo con los datos proporcionados por Eurostat (2024), en Europa los ciudadanos que usan Internet para participar en las redes sociales han pasado del 40,78 % en 2013 al 59,18% en 2023. En la actualidad, países como Noruega o Dinamarca se aproximan a porcentajes del 90%. Este aumento del uso de las redes sociales no ha pasado desapercibido ni para las Humanidades, en general, ni para la Historia y la Prehistoria en particular, que han visto en estas plataformas digitales un magnífico canal de difusión y divulgación de contenidos (MARTÍNEZ, CHAÍN 2018; PASÍES *et al.* 2021; FERNÁNDEZ, BARRIONUEVO 2022). En este sentido, no faltan trabajos que han investigado sobre la difusión que la prehistoria tiene dentro de las redes sociales, especialmente dentro de Twitter (hoy denominada X), analizándose el contenido de los mensajes, identificándose a los principales “influencers” y su peso dentro de la red, su grado de autoridad y las interrelaciones que se establecen entre ellos y las diferentes comunidades que se forman dentro de la red (VÉLAZ 2023).

Sin embargo, para algunos autores, estos análisis no son suficientes ya que consideran que para que la investigación de un fenómeno como la comunicación de la Prehistoria a través de una red social sea completa, es necesario conocer las pautas de recepción y comprensión en que se basa la relación con los destinatarios de esos mensajes (PASTOR, DÍAZ-ANDREU 2021). Esto es especialmente importante en las redes sociales, en las que se pretende una interacción y una participación social. Lo que ya no queda tan claro, o al menos no existe un criterio establecido, es como hacerlo, es decir, cual es el mejor método para conocer la forma en que un determinado contenido, en el caso de este trabajo sobre Prehistoria, llega al público, o de conocer la forma en que éste se relaciona con los contenidos.

La identificación y análisis de los sentimientos y las emociones es un recurso metodológico de sumo interés para «...estudiar las formas en que los actores – individuales y colectivos – se posicionan frente a los diversos significados que componen el sistema cultural» (RODRÍGUEZ 2008, 158). De este modo, una aproximación desde el plano afectivo a un determinado discurso permitirá un mejor entendimiento e interpretación del mismo, especialmente por parte de los especialistas en la materia. Los sentimientos y las emociones se convierten, de esta forma, en indicadores relevantes para el análisis de estos

contenidos, ya que su significado se comprende mejor si se tienen en cuenta tanto las emociones como los constructos sociales que las desencadenan.

## 2. ANÁLISIS DE SENTIMIENTOS Y EMOCIONES

Las emociones son tantas y tan complejas, que es difícil encontrar una definición o una clasificación omniabarcante. Paul EKMAN (1984, 1992) afirmaba que las emociones son un proceso, un tipo de juicio innato, que se encuentra influido por nuestra vivencia e historia personal, en el que percibimos que algo importante para nosotros está sucediendo. Sea como fuere, y lejos de querer ahondar en cuestiones terminológicas, lo que parece es que dentro de las emociones subyace tanto un componente innato (biológico) como otro adquirido (social-cultural). Así, las emociones parecen encontrarse fuertemente influenciadas por los sistemas de creencias culturales y morales, por lo que no solo proveen una evaluación instantánea de las circunstancias, sino que también influyen en nuestra disposición para responder ante tales circunstancias (RODRÍGUEZ 2008, 152). Del mismo modo, las emociones afectan a la razón y, por lo tanto, a la percepción y la comprensión de los hechos, ya que «... las emociones afectan a la razón porque la razón está vinculada a una serie de procesos cerebrales relacionados con los sentimientos» (DAMASIO 2021, 175). Así separar emoción y razón, o presuponer una posición de poder de la segunda, puede considerarse un error.

El análisis de sentimientos es una rama específica de la minería de textos que pretende inferir la valencia y las emociones expresadas en un determinado texto y que hay que relacionar con técnicas de Procesamiento del Lenguaje natural (NLP) (PANG, LEE 2008). En este sentido, para algunos autores el análisis de sentimientos «...is an umbrella term for the determination of valence, emotions, and other actual states from text or speech automatically using computer algorithms» (MOHAMMAD 2021, 323). De este modo, el primer acercamiento desde el análisis de sentimientos a un determinado texto permite identificar su valencia o polaridad, es decir, el grado en que la opinión expresada en un texto es atractiva (valencia positiva) o, por el contrario, produce rechazo (valencia negativa) o indiferencia (valencia neutra) y en que intensidad lo hace (FRIDIA 1986). Así mismo, un análisis más avanzado, permite ir más allá de la polaridad e inferir los estados emocionales que subyacen en los textos objeto de análisis, pudiendo, de esta forma, identificar si lo que prima es la alegría, el enfado, el miedo u otro tipo de emociones que se estén teniendo en consideración.

El análisis de sentimientos emplea fundamentalmente dos tipos de técnicas: unas relacionadas con el aprendizaje automático (Machine Learning) (ROSENBROCK *et al.* 2021) y otras basadas en diccionarios (Lexicon based) (WANKHADE *et al.* 2022). Las primeras son capaces de analizar y aprender de

ellos para luego tomar decisiones. Las segundas se sirven de un conjunto de algoritmos y diccionarios que vinculan términos con opiniones y sentimientos. Éste es el enfoque desde el que se ha desarrollado este trabajo, como se verá más adelante en el apartado de la metodología. Sea como fuere, y con independencia del enfoque adoptado, lo que parece evidente es que el análisis de sentimientos es una metodología ampliamente puesta en práctica en los ámbitos de las Ciencias Sociales, como sucede, por ejemplo, en el análisis de las redes sociales (ARCE *et al.* 2023; CUERVO *et al.* 2023; SAID *et al.* 2023). Sin embargo, en el campo de las Humanidades y, en concreto, en el campo de la Prehistoria y la Arqueología, el análisis de sentimientos y emociones no se encuentra tan extendido e implementado como en las Ciencias Sociales, abriéndose un abanico de posibilidades que se presenta prometedor a la luz de algunos trabajos que ya vienen tratando estas cuestiones (GARZIA *et al.* 2020, 2022; ALBANESE, GRAZIANO 2021).

### 3. MUESTRA

La muestra empleada para la elaboración de este trabajo ha sido recuperada de la red social Twitter, en la actualidad denominada X. Así, a través del Interfaz de Programación de Aplicaciones (API) y desde lenguaje R bajo RStudio, como se explicará más adelante en el apartado de la metodología, se ha tenido acceso a la información existente en la red social procediéndose, de este modo, a la descarga de mensajes que cumplieran estas tres condiciones:

- 1) que contengan el hashtag #Prehistory (o cualquiera de las variantes #prehistory y/o #PREHISTORY). Se entiende que el empleo de la etiqueta en inglés y no en otro idioma favorece la contextualización del trabajo en un panorama internacional;
- 2) que el texto del mensaje esté redactado en inglés. Este requisito obedece fundamentalmente a que las herramientas de análisis empleadas están diseñadas para funcionar bajo este idioma, como se explicará más adelante en el capítulo de la metodología;
- 3) que haya sido publicado en el periodo de un año, que se consideró un marco temporal lo suficientemente amplio y reciente como para poder extrapolar unos resultados válidos.

De este modo, entre las 22:00:36 h del 1 de octubre de 2021 a las 00:38:19 h del 1 de octubre de 2022 se recuperaron un total de 11595 tuits, de los que 10526 (90,78%) presentaban el texto en inglés. De entre estos, el 36,79% eran directos entre usuarios (3873 tuits) y el 63,21% restantes eran retuits, tanto en formato tradicional (6367) como en la forma Quote Tweets o tuits comentados (286). En este sentido, se debe tener presente que la muestra objeto de estudio se basa en aquellos mensajes etiquetados con el hashtag

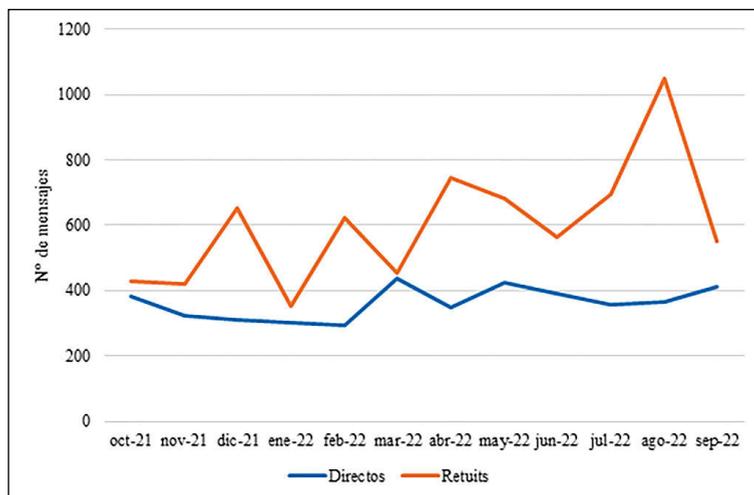


Fig. 1 – Comportamiento del tráfico en la red durante el periodo de estudio para el hashtag #Prehistory.

#Prehistory o cualquiera de sus variantes. Este criterio de selección supone una limitación por cuanto no quiere decir que la muestra contemple todos los mensajes que traten sobre prehistoria en Twitter durante el periodo objeto de análisis. Al contrario, existen discursos que, aun versando sobre la materia, no han sido etiquetados de manera correcta u ortodoxa, de acuerdo con las reglas establecidas para el uso de la red social y hacen uso de otras etiquetas. Aun así, se entiende que el análisis de la muestra ofrece una fotografía fija, real, cierta y clara sobre el tratamiento que de la disciplina se está haciendo en la red social. Además, la consideración de un hashtag para el análisis de un determinado fenómeno en Twitter es un criterio ampliamente aceptado por la literatura científica, tal y como lo ponen de manifiesto los trabajos mencionados en el párrafo anterior y otros más recientes (CARPENTER *et al.* 2023; LA ROCCA, BOCCIA 2023; SAHA 2023) (Fig. 1).

#### 4. METODOLOGÍA

El marco metodológico desde el que se plantea este trabajo puede considerarse como cuantitativo, por cuanto priman técnicas que hay que relacionar con el análisis matemático (HERNÁNDEZ *et al.* 2010). En concreto, el tratamiento y análisis de los datos se ha llevado a cabo desde R (ALAMINOS 2023), lenguaje de programación de código abierto que ha generado una activa comunidad de usuarios y desarrolladores que se articula fundamentalmente a través del proyecto Comprehensive R Archive Network (CRAN). Aunque

en su génesis este lenguaje estuvo orientado y fue utilizado principalmente con fines estadísticos (PARADIS 2005, 3) en la actualidad presenta numerosos desarrollos tanto en las Ciencias Sociales como en el campo de las Humanidades, lo que lo convierte como un recurso metodológico de gran interés y potencial (ELOSUA 2009). En nuestra opinión, la inclusión de R en el ámbito de las Humanidades presenta tres puntos fuertes. El primero, tiene que ver con la posibilidad de automatizar procesos de trabajo, frente a tratamientos de la información realizados desde procesos más manuales. Del mismo modo, permite el análisis de una gran cantidad de información. Finalmente, se minimiza la intervención o interferencia humana durante el proceso de tratamiento de los datos (ARCE, MENÉNDEZ 2018).

El desarrollo del código se ha llevado a cabo haciendo uso del Entorno de Desarrollo Integrado (IDE) RStudio, actualmente en su versión 2023.09.1-494 para sistema operativo Windows. Para proceder con el análisis de los datos recuperados desde Google, se ha manejado la librería Syuzhet (JOCKERS 2023), que permite extraer sentimientos y emociones a partir de textos usando diferentes diccionarios, entre los que se encuentran, además de uno propio generado en la propia librería, otros denominados Affin (NIELSEN 2011), Bing (HU, LIU 2004; LIU *et al.* 2005) y NRC Emotion Lexicon (MOHAMMAD, TURNEY 2010, 2013), que es el que se ha considerado para proceder con los análisis en este trabajo. Se trata de un corpus léxico desarrollado por el National Research Council of Canada (NRC) que comprende 14.182 unigramas en su versión 0.9.2. Cada una de las palabras tiene preasignada una categoría de sentimiento negativo y positivo, así como un conjunto de emociones. Estas hay que relacionarlas con las ocho emociones básicas definidas por Robert PLUTCHIK (1980) y que posteriormente fueron desarrolladas por las teorías de inteligencias múltiples (SAUTER *et al.* 2010). En concreto las emociones que se contemplan en el NRC Emotion Lexicon son: enfado, anticipación (expresión del pensamiento racional), asco o indignación, miedo, alegría, tristeza, sorpresa y confianza. Se trata de un diccionario que se ha diseñado en inglés como lengua principal, aunque permite trabajar en otros idiomas.

En este sentido, hay quienes ven un riesgo en el análisis de sentimientos y emociones sobre textos que no estén en inglés, idioma principal sobre el que se ha diseñado el diccionario NRC, por entender que la traducción de una palabra a otro idioma distinto puede llevar implícita un sesgo cultural y la polaridad sentimental supuesta para el idioma original no tiene por qué proyectarse de la misma manera en otro (ISASI 2021). Este factor es el que ha hecho que uno de los condicionantes a la hora de escoger la muestra de estudio sea que los textos de los mensajes estén redactados en inglés. Además de la cuestión idiomática, hay que tener en cuenta otra limitación de este y de otros corpus que pasa por la muestra limitada de términos contemplados. Esto supone que cuando una palabra no se encuentra contemplada en un

determinado corpus, esta adquiere un valor neutro, de ahí la explicación al alto porcentaje de neutralidad que arrojan algunos análisis. Sea como fuere, este mismo corpus ha sido utilizado en otros trabajos, presentando un alto grado de fiabilidad a la luz de los resultados obtenidos (CUERVO *et al.* 2023; SAID *et al.* 2023).

Syuzhet hace uso de un algoritmo que compara el léxico contenido en un determinado texto (sustantivos y verbos), en este caso mensajes recuperados desde Twitter, con un conjunto de valores prefijados, que aumentan o disminuyen en función de otras palabras con significado a su alrededor (adjetivos y adverbios), y consigna un valor numérico global de intensidad para cada mensaje. De este modo es posible obtener el grado de valencia positiva, negativa o neutra existente en cada reseña, su grado de intensidad, así como las diferentes emociones que subyacen en las mismas (MOHAMMAD 2016). Además de Syuzhet, se ha hecho uso de otras librerías, especialmente en la fase inicial de preparación de la muestra de análisis. Así, a través de las librerías Stringr (WICKHAM 2023) y Tm (FEINERER, HORNİK 2023), se ha procedido con una serie de tareas orientadas a la preparación de los textos para su posterior análisis. Entre éstas se encuentran la eliminación de signos de puntuación, símbolos, números, espacios en blanco, etc. Del mismo modo, se han eliminado las stopwords o “palabras vacías” (SARICA, LUO 2021) que se refieren a aquellas palabras, fundamentalmente a artículos, preposiciones o conjunciones, por citar unos ejemplos, que carecen de sentido por sí solas si no van acompañadas de palabra clave.

De acuerdo con lo anterior, podríamos explicar el proceso de análisis y redacción del código del siguiente modo. Primero se cargan las librerías que se van a emplear en el análisis entre las que se encuentran, además de la ya citada Syuzhet, otros paquetes como tm, RColorBrewer, SnowballC, stringr y readxl. Posteriormente se carga la muestra de análisis, que en el caso que nos ocupa se sube en formato .xls a través de readxl. Se procede con la preparación y depurado de los textos de la muestra según se ha explicitado en los párrafos anteriores. Una vez preparada la muestra de análisis se lleva a cabo el análisis de sentimientos a partir de la función `get_sentiment`, y luego se procede con el de emociones a través del comando `get_nrc_sentiment`. Para la elaboración de los gráficos, destacar el empleo de las funciones `plot` y `simple_plot`.

Del mismo modo, se ha procedido con una validación manual de los resultados obtenidos de los análisis efectuados. Para ello, se han escogido aleatoriamente una submuestra de textos y se han analizado manualmente con los datos contenidos en el NRC Emotion Lexicon, tanto a nivel de valencia sentimental como con el grado emocional asignado a cada uno de los términos. En todos los casos los resultados obtenidos del análisis automatizado como del manual son coincidentes.

## 5. RESULTADOS OBTENIDOS

### 5.1 *Análisis de sentimientos*

La Tab. 1 muestra los estadísticos básicos obtenidos del análisis de sentimiento. Puede apreciarse que el valor mínimo de valencia obtenido es -6 y se relaciona con una reseña publicada en el 26 de octubre de 2021 desde la cuenta @sillymickel y que presenta el siguiente texto: «in our desperate struggle to justify the pathetic & meager state in which we find ourselves...& thereby to beat back the pain of realizing how much we have truly lost... READ BOOK... <https://sillymickel.blogspot.com/2020/02/the-consciousness-of-stones-2nd-half-of.html> #devolution #prehistory #primal #pantheism #shamanic #tribal» (<https://twitter.com/sillymickel/status/1452807227375423492>). En el extremo contrario, tenemos un valor máximo de valencia positiva de +7. Se detectan dos mensajes a los que se le ha asignado este nivel de intensidad. El primero de ellos se publica el 2 de octubre de 2021 desde la cuenta @bardicacademic en los siguientes términos: «the new Ravilious exhibition at the lovely Wilts-hire Museum, Devizes, is well worth a visit. The artist's downland work is made especially resonant in the context of the museum's superb Prehistoric collection. A real treasure trove. Exhib on til 30 Jan. #art #museums #prehistory» (<https://twitter.com/bardicacademic/status/1444343777812959234>). El segundo, lo hace el 5 de noviembre del mismo año desde la cuenta @Trevorttg y presenta el siguiente texto: «I ❤️ this! Other versions survive from #prehistory this is the best. Very recognizable #famous #Mesopotamian #sculpture - The #GuennolLioness 5,000 yold, found in mid 1930's in #Iran Powerfully carved by a #Gr8 #craftsman Sold in 2007 for \$52.4 million 🤩 <http://coastconfan.blogspot.com/2011/08/guennol-lioness.html>» (<https://twitter.com/Trevorttg/status/1456552884322119682>).

Si se atiende al número de mensajes asignados a cada una de las categorías de valencia o polaridad (Fig. 2), es posible advertir como de los 10526 mensajes analizados, 1697 presentan una valencia negativa, comprendida entre -6 y -1. Cuatro mil trescientas ochenta y tres mensajes presentan una valencia positiva comprendida entre 1 y 7. El resto de mensajes (4446) se consideran como neutras. En términos porcentuales puede decirse que Syuzhet considera que el 41,64% de los textos analizados tienen una valencia positiva, el 16,12% negativa y el 42,24% restante pueden considerarse como neutras. Este abultado porcentaje de neutralidad puede interpretarse en un doble sentido. En primer lugar, es que pueden reflejar la indiferencia de los usuarios

Valor Mínimo	1er Cuartil	Media	Mediana	Suma	3er Cuartil	Valor Máximo
-6,0000	0,0000	0,4757	0,0000	5007	1,0000	7,0000

Tab. 1 – Estadísticos básicos obtenidos en el análisis de sentimientos.

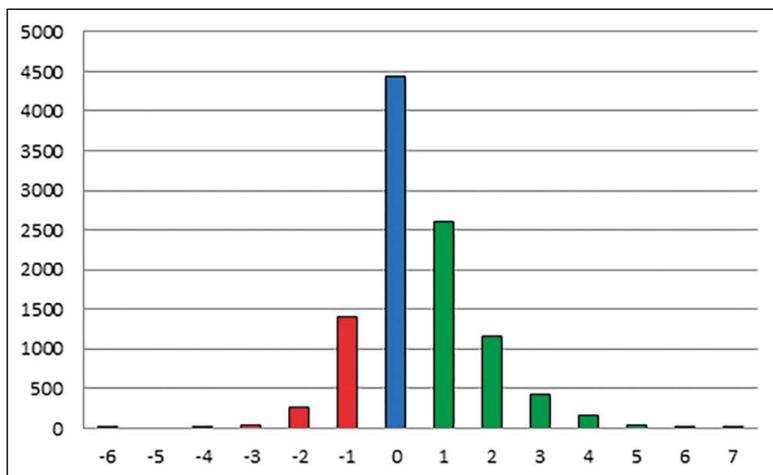


Fig. 2 – Número de mensajes asignados a cada nivel de valencia.

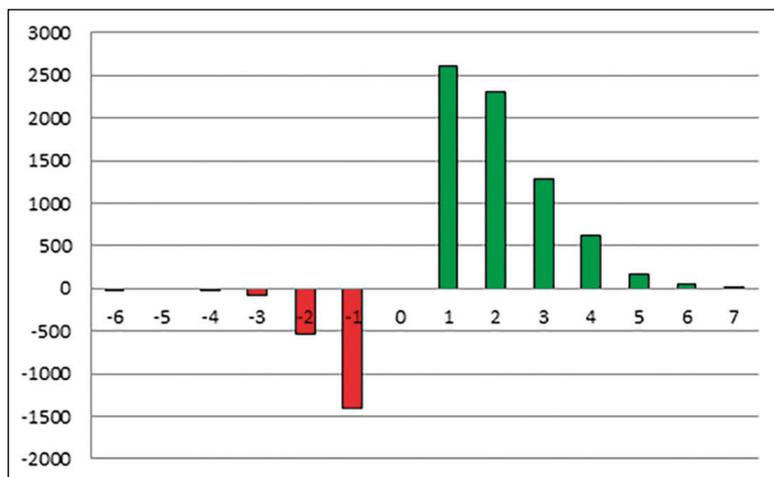


Fig. 3 – Valencia acumulada en cada una de las categorías.

para la cuestión objeto de este trabajo, es decir, que desde un punto de vista de la polaridad sentimental para la mayor parte de la muestra la cuestión de la prehistoria despierta indiferencia. Pero este alto porcentaje de neutralidades puede entenderse desde la propia limitación inherente al corpus empleado, tal y como se ha señalado líneas atrás, por cuanto contempla una muestra terminológica limitada y puede suceder que muchos términos empleados en

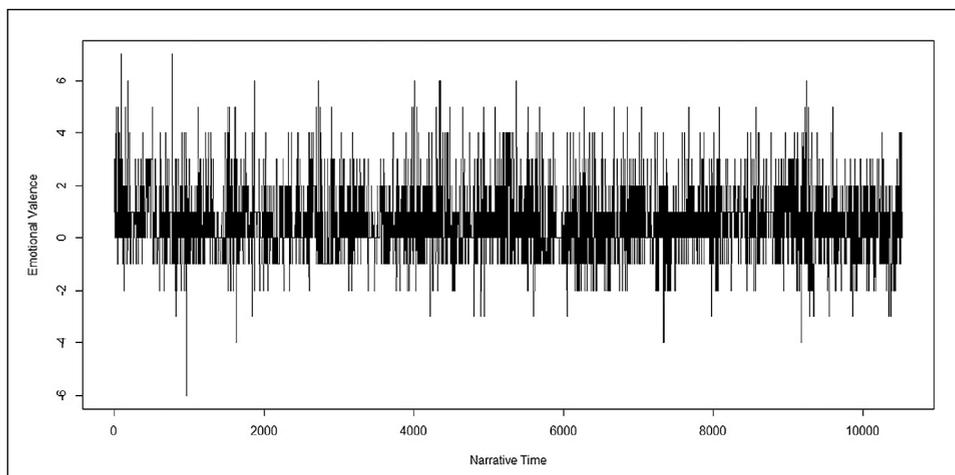


Fig. 4 – Trayectoria de las valencias a lo largo del tiempo.

los mensajes no estén contemplados en el diccionario, lo que explica la alta presencia de los valores neutros.

La valencia acumulada en cada una de las categorías (Fig. 3) junto con los estadísticos básicos obtenidos (Tab. 1), permite afirmar como la polaridad general que subyace en los mensajes analizados es manifiestamente positiva. Esta afirmación está fundada tanto por la media obtenida, que se encuentra dentro de valores positivos (0,4757), como por la suma de las polaridades en todas las categorías, que aporte igualmente un valor positivo de 5007.

Hasta el momento, los datos estadísticos que se han manejado aportan una serie de información de interés, pero dicen poco en términos de cómo está estructurada la narrativa de las diferentes reseñas y cómo los sentimientos positivos y negativos se suceden a lo largo de la misma. En este sentido, resulta más efectivo trazar los valores en un gráfico donde el eje X representa el paso del tiempo mientras que el eje Y mide los grados de sentimiento positivo y negativo (Fig. 4). Puede apreciarse el comportamiento de los diferentes grados de valencia a lo largo de los 10.526 mensajes analizados que, al haberse ordenado de más antiguo a más reciente, permite ver la evolución a lo largo del tiempo.

No obstante, y aunque este tipo de representaciones gráficas resultan efectivas para identificar la presencia y disposición de las polaridades o valencias más extremas a lo largo de la muestra, no lo son tanto a efectos de análisis global de las trayectorias de las reseñas a lo largo del tiempo. De este modo, a efectos de una representación que facilite el análisis gráfico, procede eliminar el ruido y proceder con una forma más simplificada de la trayectoria

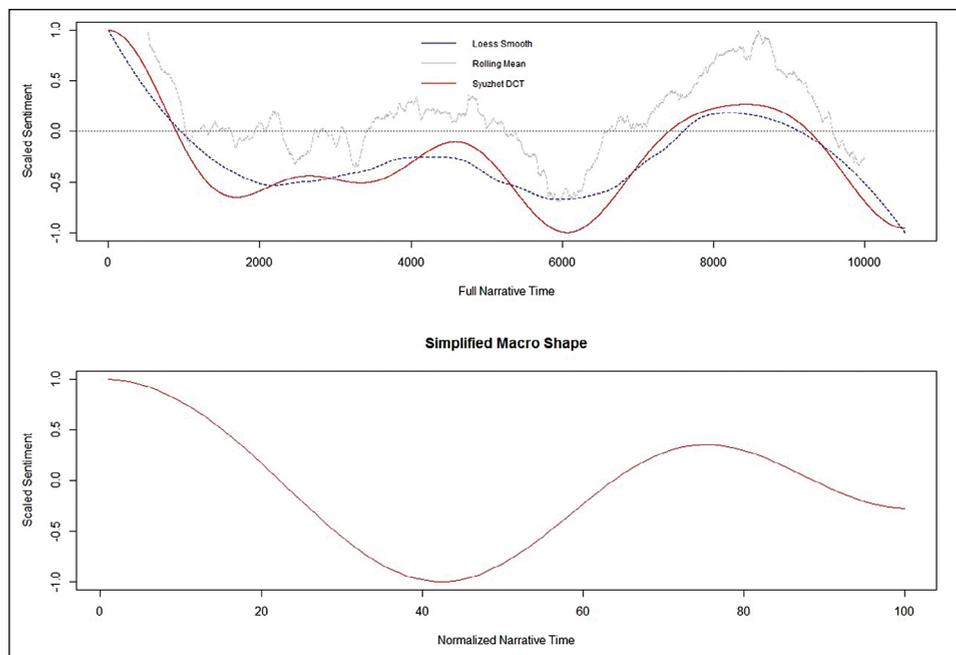


Fig. 5 – Evolución de los sentimientos a partir de la aplicación de técnicas de suavizado de datos.

(WADE 1994; JOCKERS 2023). Para ello, Syuzhet permite aplicar una serie de técnicas de suavizado de los datos mediante las que se consigue eliminar el ruido y comportamientos no deseados o anómalos de los datos. La Fig. 5 muestra en su eje horizontal X el total de los 10.526 mensajes considerados en la muestra de análisis ordenados de más antiguo a más reciente. El eje vertical Y muestra la fluctuación de la valencia de sentimientos sólo que en un rango normalizado, es decir, con unos valores comprendidos entre 0 y 1 para los sentimientos positivos y entre 0 y -1 para los negativos. Las representaciones gráficas se corresponden con tres covariables. La primera, es el resultado de aplicar una regresión local (“loess smooth”) que es un tipo de ajuste de curvas a datos mediante suavizados en los que el ajuste en X se realiza utilizando únicamente observaciones en el entorno de X (GARCÍA *et al.* 2023, 24).

De este modo es posible advertir tendencias y ciclos en los datos que pueden ser difíciles de advertir en curvas de tipo paramétrico (línea de picos en gris). La segunda, es la consecuencia de proceder con el cálculo de la media móvil (“rolling mean”) que es una lista de números en la cual cada uno es el promedio de un subconjunto de los datos originales y evidencia el promedio de las fluctuaciones de los sentimientos (línea azul) (GARCÍA *et*

al. 2023, 24). Finalmente, la línea continua en rojo se corresponde con la transformada discreta de coseno (Discrete Cosine Transform - DCT) que se basa en la Transformada de Fourier discreta, pero utilizando únicamente números reales. Se trata de una de las representaciones gráficas más utilizadas para el análisis de datos como los que se manejan en este trabajo, ya que permite una representación de los datos mejor que la de Fourier para la observación de la variación de la frecuencia de los sentimientos (JOCKERS 2023). Esta misma gráfica se reproduce en la parte inferior de la Fig. 5, solo que sobre un eje horizontal normalizado dividido en cien fragmentos (ISASI 2021; JOCKERS 2023).

Puede apreciarse como el conjunto de mensajes analizados comienza con unas valencias marcadamente positivas que va descendiendo de manera paulatina hasta la reseña 4500, aproximadamente, donde se alcanza el nivel más bajo de polaridad. Desde este punto, el tono sentimental de los mensajes vuelve a tender hacia valores positivos, alcanzando su máximo en torno al mensaje 8500. Desde este, hasta el final de la muestra, la polaridad vuelve a decaer terminando en valores ligeramente negativos. Esta tendencia se suaviza desde este punto hasta el final de la muestra, donde nos encontramos una tendencia ligeramente negativa en lo que a la valencia general de los textos se refiere.

## 5.2 *Análisis de emociones*

El análisis de las emociones devuelve una serie de estadísticos básicos para cada una de las emociones (Tab. 2). Si nos fijamos en las medias, es posible advertir como las emociones que aparecen más representadas dentro de los textos analizados son la anticipación (0,4649) y la confianza (0,3999). En todos los casos, la mediana aparece como cero. Esto puede significar dos cosas: que en el diccionario que se ha utilizado para el análisis, el NRC, aparecen pocas de las palabras usadas en las reseñas, o que, por el contrario, pocas de las palabras cuentan con una asignación de sentimiento o emoción en el diccionario (ISASI 2021).

Emoción	Valor Mínimo	1er Cuartil	Media	Mediana	3er Cuartil	Valor Máximo
Enfado	0,0000	0,0000	0,1763	0,0000	0,0000	4,0000
Anticipación	0,0000	0,0000	0,4649	0,0000	1,0000	5,0000
Disgusto	0,0000	0,0000	0,1047	0,0000	0,0000	3,0000
Miedo	0,0000	0,0000	0,2588	0,0000	0,0000	5,0000
Alegría	0,0000	0,0000	0,2980	0,0000	0,0000	4,0000
Tristeza	0,0000	0,0000	0,1869	0,0000	0,0000	4,0000
Sorpresa	0,0000	0,0000	0,1501	0,0000	0,0000	3,0000
Confianza	0,0000	0,0000	0,3999	0,0000	1,0000	5,0000

Tab. 2 – Estadísticos básicos obtenidos en el análisis de emociones.

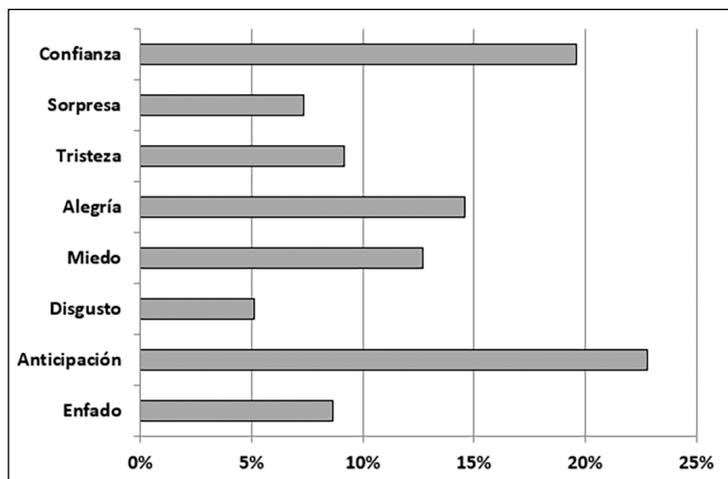


Fig. 6 – Presencia de cada una de las emociones en la muestra de análisis.

La presencia de cada una de las emociones en el conjunto de mensajes analizados (Fig. 6) pone de manifiesto como, frente a la mayor presencia de la anticipación y la confianza, se encuentran las emociones relativas a la sorpresa y al disgusto, que son las que se encuentran menos representadas. Del mismo modo, es posible identificar las reseñas concretas a las que se les han asignado los valores máximos de emoción. Esto es especialmente interesante en aquellas que tienen una mayor presencia. Así, entre las reseñas a las que se les ha asignado un nivel de emoción anticipación más alto, igual a 5, se encuentra una subida a la red social el 29 de agosto de 2002 desde la cuenta @marie\_orange24 que presenta el siguiente texto: «First quick attempt at creating a 3D model with Polycam and pretty happy with the result! #obsidian #core on a secondary deposit cobble from #syunik found in Kültepe I #Azerbaijan #caucasus #prehistory #Archaeology #3dmodeling» ([https://twitter.com/marie\\_orange24/status/1564309341745958918](https://twitter.com/marie_orange24/status/1564309341745958918)). Del mismo modo existe otra reseña con unos niveles semejantes. En este caso, se trata de un texto subido a Twitter el 13 de septiembre de 2022 desde @Trevorttg que se redacta en los siguientes términos: «#tumulus #mound #prehistory #burial #megalithic #ArchaeoAstronomy #archeology #season #mythology #Folklore #saga #FairyTail #Fairy #MythAndFact #Time moves 🙌 pushing on our #summer #sun ☀️ thru to it's 3rd quarter 🌑 #Fairies glad watch as #barrow light shades to #Autumn glow. 🌻» (<https://twitter.com/Trevorttg/status/1569717513579073538>). Por ilustrar otro ejemplo de entrada que presenta una puntuación alta para la emoción anticipación es posible mencionar el mensaje subido el 9 de abril de 2022 que presenta el siguiente texto:

«Excited to start preparations for excavations this summer at the Illyrian hill fort in Vičja luka on the island of Brač, #Dalmatia. The welcome committee from the last campaign is waiting for us with excitement 📍 #Archaeology #prehistory» (<https://twitter.com/MajaMise/status/1512727844350476291>).

En relación con la segunda emoción mejor representada, la confianza, es posible identificar varios tuits que presentan un valor alto asignado para esta emoción, igual a 5. El primero de los mensajes se sube el 11 de diciembre de 2021 desde la cuenta @DrJenniferBates y presenta el siguiente texto: «Our paper looks at how food lump analysis can help us think about food and foodways in South Asian #prehistory. We present a pilot study from Southern Indian Neolithic and Iron Age Kadebakele where we found numerous food lumps in the #archaeobotany samples» (<https://twitter.com/DrJenniferBates/status/1469487466080272386>). El Segundo de los mensajes, subido a Twitter el 23 de marzo de 2022 desde @NExSpain, dice: «Mighty Moat. Dug by hand in the Iron Age, this moat protected a hill fort on a coastal peninsula, effectively cutting it off from the mainland. #Spain #IronAge #hillfort #archaeology #prehistory #historytravel» (<https://twitter.com/NExSpain/status/1506719944645488648>). Además de los citados, es posible mencionar más ejemplos de textos con un alto nivel para la emoción confianza. Es el caso del texto subido el 5 de octubre de 2021 desde la cuenta @John22786385 en los siguientes términos: «t's 22 years since #WalkingwithDinosaurs landed on our screens 🦖🦕. All these years on, the iconic #dinosaur documentary still inspires our passion for #prehistory and #paleontology. Thank you to the crew behind the show, that made the impossible real!» (<https://twitter.com/John22786385/status/1445086319638683662>).

Además de la anticipación y la confianza, la tercera de las emociones mejor representada es la alegría. Presenta un nivel máximo de 4 que es posible encontrar en varias reseñas. Entre ellas, se identifica un texto subido a Twitter el 29 de agosto de 2022 desde la cuenta @antje\_wilton en los siguientes términos: «very happy and a litte proud to have finished this successfully 🥰 all online, but so much fun and with lots of support from the wonderful people at @ArchAncHistLeic #archaeology #prehistory #postgraduate #heritage #distancelearning #lifelonglearning» ([https://twitter.com/antje\\_wilton/status/1564168864216252416](https://twitter.com/antje_wilton/status/1564168864216252416)).

De acuerdo con el modelo de emociones propuesto por Robert Plutchick, es posible identificar, a tenor de los resultados obtenidos, lo que el autor denomina como diadas primarias, concepto con el que se hace referencia a las emociones que... «...result when two or more fundamental emotions are combined, in the same way that red and blue make purple» (PLUTCHIK 2001, 350). De este modo, la consecuencia de combinar anticipación y alegría supone la aparición de una nueva emoción y que Plutchick relaciona con el optimismo. Del mismo modo, de la combinación de a alegría con la confianza, surge el amor.

Aunque es manifiesto un predominio de emociones de carácter positivo, el miedo es la emoción mejor representada entre las de carácter negativo con una media de 0,2588. Entre los mensajes que presentan niveles más elevados para esta emoción se encuentra uno subido a Twitter el 19 de junio de 2022 y que dice: «A mother being able to run with an infant, as opposed to having the infant in the womb at that time, would no doubt also have contributed to survival READ BOOK <https://angelsinnature.wordpress.com/2022/02/13/prodigal-human-the-descents-of-man-by-michael-adzema-2016-complete-book-free-downloadable/> #birth #evolution #anthro #devolution #bipedalism #psyche #pelvicbones #prehistory» (<https://twitter.com/sillymickel/status/1538578979254259713>). No sorprende que el mensaje fuese subido desde @sillymickel ya que se trata de una de las cuentas mejor posicionadas de acuerdo con los resultados obtenidos del análisis de centralidad llevado a cabo para la muestra objeto de este análisis (VÉLAZ 2023). Otra de las cuentas mencionadas en el trabajo de este autor como de alto impacto es @Trevorttg, a la que se ha hecho referencia líneas atrás.

## 6. DISCUSIÓN Y CONCLUSIONES

De los 11595 mensajes recuperados de Twitter con el hashtag #Prehistory entre los meses de octubre de 2021 y 2022, 10526 fueron objeto de análisis por encontrarse escritos en inglés. Desde un punto de vista del análisis de sentimientos, puede afirmarse que la valencia que subyace en la muestra de datos analizada puede considerarse como marcadamente positiva, presentando extremos, a lo largo de su trayectoria, que van del -6 al +7. El análisis de las emociones evidencia un predominio de las de carácter positivo, especialmente la anticipación, la confianza y la alegría. De acuerdo con el modelo de las emociones propuesto por Plutchick, la alta presencia de estas tres emociones lleva a pensar en la generación de diadas primarias que tienen como resultado la presencia de otras emociones que hay que relacionar con el optimismo y el amor. Por todo ello, así como por los diferentes resultados obtenidos en este trabajo, puede afirmarse que la valencia sentimental y las emociones que subyacen en los discursos que sobre prehistoria se están generando en Twitter son manifiestamente positivos.

Es cierto que el contenido existente en una determinada red social sobre una determinada materia, como es el caso de la Prehistoria en Twitter, no es inocente. Pueden subirse, y de hecho se suben, contenidos por parte de personas que carecen del conocimiento científico suficiente o que tienen intereses que van más allá de la divulgación de contenidos sobre la materia. Del mismo modo, la existencia de bots o cuentas automatizadas pueden alterar y modificar los discursos sobre una determinada materia, apartándolos del rigor y la ortodoxia científica y adentrándolos en veredas en las que subyacen

intereses comerciales o de otro tipo que van más allá de la transmisión del conocimiento científico sobre la materia (VÉLAZ 2023). Sin embargo, el objeto de este trabajo es identificar los sentimientos y las emociones que subyacen en los diferentes textos.

Sea como fuere, el lenguaje R, en general, y el análisis de sentimientos y emociones, en particular, vislumbran un escenario prometedor tanto para el campo de la Historia como para la Prehistoria y la Arqueología. Es posible analizar una gran cantidad de información, en diferentes idiomas y de una forma ágil y automatizada. Los datos son tratados de una manera matemática por lo que se les devuelve parte de la objetividad robada por la inferencia humana y eso la investigación lo agradece. Del mismo modo, en un contexto más amplio, puede afirmarse que este tipo de análisis son un “traje a medida” para las Humanidades Digitales por cuanto permiten pasar de los números a los significados a través de un proceso fundamentado.

DAVID VÉLAZ CIAURRIZ

Escuela Superior de Ingeniería y Tecnología (ESIT)  
Universidad Internacional de La Rioja (UNIR)  
david.velaz@unir.net

## BIBLIOGRAFÍA

- ALAMINOS A. 2023, *Introducción a la minería de texto y análisis de sentimiento con R*, Alicante, Universidad de Alicante.
- ALBANESE V.E., GRAZIANO T. 2021, *The role of cultural heritage in wellbeing perceptions: A web-based software analysis in two Italian provinces*, «Il Capitale Culturale. Studies on the Value of Cultural Heritage», 24, 293-324 (<https://doi.org/10.13138/2039-2362/2724>).
- ARCE S., DÍAZ J., CAMBRONERO B. 2023, *Online hate speech and emotions on Twitter: A case study of Greta Thunberg at the UN Climate Change Conference COP25 in 2019*, «Social Network Analysis and Mining», 13, 48, 1-13 (<https://doi.org/10.1007/s13278-023-01052-5>).
- ARCE S., MENÉNDEZ M.I. 2018, *Aplicaciones de la estadística al framing y la minería de texto*, «Estudios de comunicación, Información, Cultura y Sociedad», 39, 61-70.
- CARPENTER J.P., MORRISON S.A., ROSENBERG J.M., HAWTHORNE K.A. 2023, *Using social media in pre-service teacher education: The case of a program-wide Twitter hashtag*, «Teaching and Teacher Education», 124, 104036 (<https://doi.org/10.1016/j.tate.2023.104036>).
- CUERVO T., ARCE S., ORVIZ N. 2023, *Corporate social responsibility and its communication on Twitter: Analysis of the discourse and feelings generated in society*, «Management Letters», 23, 1, 63-73 (<https://doi.org/10.5295/cdg.211639tc>).
- DAMASIO A.R. 1999, *El error de Descartes: la razón de las emociones*, Barcelona, Andres Bello.
- EKMAN P. 1984, *Expression and the nature of emotion*, in K.R. SCHERER, P. EKMAN (eds.), *Approaches To Emotion*, New York, Psychology Press, 319-343.
- EKMAN P. 1992, *An argument for basic emotions*, «Cognition and Emotion», 6, 3, 169-200.
- ELOSUA P. 2009, *Existe vida más allá del SPSS? Descubre R*, «Psicothema», 21, 652-655.
- EUROSTAT 2024, *Individuals using the Internet for participating in social networks* (<https://ec.europa.eu/eurostat/databrowser/view/tin00127/default/table?lang=en>). <https://doi.org/10.2908/TIN00127>).

- FERNÁNDEZ A., BARRIONUEVO L. 2022, *La difusión de la producción científica en el ámbito de las Humanidades: el caso de la Universidad de León*, «Investigación Bibliotecológica: archivonomía, bibliotecología e información», 36, 90, 47-65 (<http://dx.doi.org/10.22201/iibi.24488321xe.2022.90.58486>).
- FREINER I., HORNICK H. 2023, *tm: Text Mining Package*, R package version 0.7-11 (<https://cran.r-project.org/web/packages/tm/tm.pdf>).
- FRIDJA N.H. 1986, *The Emotions*, Cambridge, Cambridge University Press.
- GARCÍA L.A., ITURRALDE E., RAMOS J.A. 2023, *Polarización del movimiento feminista en México a partir de los métodos digitales, el análisis de sentimientos y los hashtags #UNAMFeminista y #UNAMSinTransfobia*, «Paakat: Revista de Tecnología y Sociedad», 13, 25, 1-25 (<http://dx.doi.org/10.32870/Pk.a13n25.800>).
- GARZIA F., BORGHINI F., BRUNI A., LOMBARDI M., MIGHETTO P., RAMALINGAM S., RUSSO S.B. 2020, *Emotional reactions to the perception of risk in the Pompeii archaeological park*, «International Journal of Safety and Security Engineering», 10, 1, 11-16 (<https://doi.org/10.18280/ijssse.100102>).
- GARZIA F., BORGHINI F., BRUNI A., LOMBARDI M., MINO L., RAMALINGAM S., TRICARICO G. 2022, *Sentiment and emotional analysis of risk perception in the Herculaneum archaeological park during COVID-19 pandemic*, «Sensors», 22, 8138 (<https://doi.org/10.3390/s22218138>).
- HERNÁNDEZ R., FERNÁNDEZ C., BAPTISTA P. 2010, *Metodología de la investigación*, México, McGrawHill.
- HU M., LIU B. 2004, *Mining and summarizing customer reviews*, in ASSOCIATION FOR COMPUTING MACHINERY, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, ACM, 168-177 (<https://doi.org/10.1145/1014052>).
- ISASI J. 2021, *Análisis de sentimientos en R con 'syuzhet'*, «Programming Historian en español», 5 (<https://doi.org/10.46430/phes0051>).
- JOCKERS M.L. 2023, *Introduction to the Syuzhet Package*, 2023 (<https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>).
- LA ROCCA G., BOCCIA A.G. 2023, *Interpreting the changeable meaning of hashtags: Toward the theorization of a model*, «Frontiers in Sociology», 7, 1104686 (<https://doi.org/10.3389/fsoc.2022.1104686>).
- LIU B., HU M., CHENG J. 2005, *Opinion observer: Analyzing and comparing opinions on the web*, in A. ELLIS, T. HAGINO (eds.), *Proceedings of the 14<sup>th</sup> International World Wide Web Conference*, New York, Association for Computing Machinery, 342-351 (<https://doi.org/10.1145/1060745.1060797>).
- MARTÍNEZ L., CHAÍN C. 2018, *Humanidades digitales para el aprendizaje y difusión del Patrimonio Naval*, «Revista de Educación a Distancia», 18, 57, 1-19.
- MOHAMMAD S. 2016, *Sentiment analysis: Detecting valence, emotions, and other affectual states from text*, in H.L. MEISELMAN (ed.), *Emotion Measurement*, Sawston, Woodhead Publishing, 201-237 (<https://doi.org/10.1016/B978-0-08-100508-8.00009-6>).
- MOHAMMAD S. 2021, *Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text*, in H.L. MEISELMAN (ed.), *Emotion Measurement*, Swaston, Woodhead Publishing, 323-379 (<https://doi.org/10.1016/B978-0-12-821124-3.00011-9>).
- MOHAMMAD S., TURNEY P. 2010, *Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon*, in D. INKPEN, C. STRAPPARAVA (eds.), *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, Association for Computational Linguistics, 26-34.
- MOHAMMAD S., TURNEY P. 2013, *Crowdsourcing a word-emotion association lexicon*, «Computational Intelligence», 29, 3, 436-465 (<https://doi.org/10.1111/j.1467-8640.2012.00460.x>).

- NIELSEN F.Å. 2011, *A new ANEW: Evaluation of a word list for sentiment analysis in micro-blogs*, «arXiv», 1103.2903 (<https://doi.org/10.48550/arXiv.1103.2903>).
- PANG B., LEE L. 2008, *Opinion mining and sentiment analysis*, «Foundations and Trends in Information Retrieval», 2, 1-2, 1-135.
- PARADIS E. 2005, *R for Beginners*, Montpellier, Institut des Sciences de l'Evolution, Université Montpellier II.
- PASÍES T., MARTÍNEZ A., PAVÓN F. 2021, *Difusión de la conservación y restauración de bienes culturales en el Museu de Prehistòria de Valencia a través de las redes sociales*, «Revista del Instituto de Prehistoria y Arqueología Sautuola», 26, 309-326.
- PASTOR A., DIAZ-ANDREU M. 2021, *Conservación (crítica) social en Arqueología*, «Chungara. Revista de Antropología Chilena», 54, 1, 165-179 (<http://dx.doi.org/10.4067/s0717-73562021005002602>).
- PLUTCHIK R. 1980, *A general psychoevolutionary theory of emotion*, in R. PLUTCHIK, H. KELLERMAN (eds.), *Emotion: Theory, Research, and Experience* Cambridge, Academic Press, 3-33.
- PLUTCHIK R. 2001, *The nature of emotions*, «American Scientist», 89, 4, 344-350 (<http://www.jstor.org/stable/27857503>).
- RODRÍGUEZ T. 2008, *El valor de las emociones para el análisis cultural*, «Papers», 87, 145-159 (<https://doi.org/10.5565/rev/papers/v87n0.793>).
- ROSENBROCK G., TROSSERO S., PASCAL A. 2021, *Técnicas de análisis de sentimientos aplicadas a la valoración de opiniones en el lenguaje español*, in M.I. MAC GAUL (ed.), *Memorias del Congreso Argentino en Ciencias de la Computación-CACIC*, Salta, Universidad Nacional de Salta, 291-300.
- SAHA A. 2023, *Twitter imparting and reinforcing gender-based identities of the Aboriginal Australia women*, in C. CHAKRABORTY, D. PAL (eds.), *Gender Inequality and its Implications on Education and Health*, Bingley, Emerald Publishing Limited, 223-234 (<https://doi.org/10.1108/978-1-83753-180-620231020>).
- SAID E., ARCE S., MOTTAREALE D. 2023, *Polarización sentimental en Twitter durante el Paro Nacional de 2021 en Colombia*, «Cuadernos.Info», 55, 281-309 (<http://dx.doi.org/10.7764/cdi.55.50483>).
- SARICA S., LUO J. 2021, *Stopwords in technical language processing*, «PlosONE», 16, 8, e0254937 (<https://doi.org/10.1371/journal.pone.0254937>).
- SAUTER D., EISNER F., EKMAN P., SCOTT S.K. 2010, *Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations*, «Proceedings of the National Academy of Sciences», 107, 6, 2408-2412 (<https://doi.org/10.1073/pnas.0908239106>).
- VÉLAZ D. 2023, *La comunicación de la Prehistoria a través de las redes sociales: El caso de Twitter*, «Complutum», 34, 2, 561-581 (<https://doi.org/10.5209/cmpl.92268>).
- WADE G. 1994, *Signal Coding and Processing*, Cambridge, Cambridge University Press.
- WANKHADE M., RAO A.C.S., KULKARNI C. 2022, *A survey on sentiment analysis methods, applications, and challenges*, «Artificial Intelligence Review», 55, 5731-5780 (<https://doi.org/10.1007/s10462-022-10144-1>).
- WICKHAM H. 2023, *Package 'stringr': Simple, Consistent Wrappers for Common String Operations* (<https://cran.r-project.org/web/packages/stringr/stringr.pdf>).

## ABSTRACT

This work aims to investigate the feelings and emotions that Prehistory generates in the social network Twitter (today called X). To do this, the messages that, during a period of one year, contain the hashtag #prehistory, together with its possible variants, have been considered. The analysis has been carried out from an R programming environment, mainly through the *Syuzhet* library, that allows extrapolating feelings and emotions that underlie texts. However,

the methodological proposal made in this paper is susceptible to being applied to different data samples related to other cultural manifestations or assets. From the point of view of the results obtained, it can be stated that the valence underlying the data sample analyzed can be considered as markedly positive. Regarding emotions, there is a predominance of anticipation and trust, followed by joy according to the model proposed by Plutchick, which is the one used in this work.