

Nuevos paradigmas para realizar análisis de contenido en informaciones sobre salud

New paradigms for conducting content analysis in health-related information

Toda información es importante si está conectada con otra
Umberto Eco

En un mundo donde la información que se genera no cesa, y donde tanto los medios de comunicación como las redes sociales se hacen eco de cuestiones relevantes sobre la salud, resulta muy importante conocer quién está informando, qué se está contando acerca de temas claves como las vacunas, enfermedades, virus, etc. La pandemia lo puso en relieve, los medios de comunicación se convirtieron en nuestros firmes aliados para recibir información acerca de lo que fuera de nuestras casas sucedía, de qué magnitud tenía el virus, qué significaba, qué debíamos hacer. Este puede ser un claro ejemplo de la importancia que pueden tener los mensajes mediáticos en la formación de las representaciones sociales, en cómo la opinión pública genera un imaginario sobre un objeto concreto. De ahí que, como investigadores nos resulte importante conocer cómo se está informando, qué diferencias pueden existir entre uno y otro medio y/o plataforma, hacia dónde vehiculan sus discursos, con qué fuentes, qué temáticas emergen y cómo evolucionan éstas en el eje temporal, estas preguntas forman parte de algunas de las cuestiones que pueden formularse al realizar un análisis de contenido, más específicamente, en un análisis lexicométrico. Lebart y Salem (1988) serán de los primeros autores que empiecen a interesarse por este tipo de análisis metodológico, y lo definirán como «un conjunto de métodos para llevar a cabo reorganizaciones formales de la secuencia textual y análisis estadísticos del vocabulario de un corpus de texto» (p. 183), es decir, es un tipo de análisis que permitirá desgranar el contenido del texto y conocer qué se dice y cómo se dice, basándose en las correlaciones existentes entre el vocabulario del texto, la frecuencia de las palabras y la presencia o ausencia de ellas, entre otras.

Si bien, si nos proponemos aplicar esta metodología para analizar las noticias de prensa que se hayan publicado en diferentes medios sobre por ejemplo la importancia de vacunar para combatir enfermedades como el sarampión, un virus que viene siendo últimamente noticia por los rebrotes que está habiendo y que ocupa un espacio en la parrilla informativa. Por lo tanto, si fuera un tema que quisiéramos explorar ya sea para ver cómo los medios abordan un tema de salud pública de esta magnitud, o cómo se filtran esos mensajes desinformadores en determinadas fuentes, sería muy útil poder hacer una lectura de todo ello aplicando el análisis lexicométrico.

Ahora pues, realizar este trabajo hoy en día precisa de un software que pueda automatizar el proceso, ya que dado el alto volumen de información que se genera, sería muy difícil que un investigador pudiera analizar de manera manual un amplio periodo de tiempo todas esas publicaciones, para ello, el software de libre acceso de análisis textual llamado IRaMuTeQ, creado por Pierre Ratinaud en la Universidad de Toulouse, Francia (Ratinaud, 2024), permitirá explorar un corpus de texto a través de diferentes análisis, que permitirán contestar a distintas preguntas, lo que nos permitirá extraer ciertas conclusiones y trazar la narrativa mediática sobre el tema que estemos analizando. Esto adquiere una relevancia mayor cuando estamos analizando informaciones que versan sobre salud, por ello, resulta interesante comprobar si se está dando una información homogénea, qué enfoques están recibiendo las noticias, sobre qué temáticas principales giran...

Asimismo, tomando como ejemplo un corpus conformado con las informaciones publicadas en la prensa española durante el 17 de mayo de 2025 sobre el aumento de casos de sarampión, si queremos conocer cuáles

Para explorar cómo los medios abordan un tema de salud pública o cómo se filtran mensajes desinformadores, resulta útil aplicar un análisis lexicométrico

son las palabras más frecuentes utilizadas en estas noticias, IRaMuTeQ tiene la opción de realizar una nube de palabras, la cual mostrará la frecuencia proporcional al tamaño, de modo que, si vemos una palabra muy grande significará que está siendo muy utilizada en el corpus, y a la inversa, si su tamaño es muy pequeño denota poco uso. El resultado que arroja la nube de palabras es que hay palabras que destacan mucho en las informaciones, como son caso, sarampión, vacuna, vacunación, enfermedad, población, pese a que no nos muestre la relación entre ellas, sí que nos da ciertas pistas sobre aquello que ocupa una presencia mayor en las noticias acerca de los casos de sarampión notificados recientemente en España.

Ahora bien, si además de conocer de qué se habla queremos saber cómo se hace, en IRaMuTeQ encontramos el análisis de similitud, que está basado en la teoría de grafos y su principal objetivo es analizar las relaciones lexicales y guardar aquellas que tienen un vínculo mayor (Marchand, & Ratinaud, 2012), en el caso de IRaMuTeQ, para poder calcular y representar estas relaciones se basará en el índice de coocurrencia, es decir, aquellas ocurrencias que tienen tendencia a aparecer juntas en el texto. De manera que, en el análisis de similitud el lugar que ocupa cada palabra y el grosor de las artistas que las unen nos aporta cierta información, ya que la cercanía nos indica el campo lexical y cómo se está abarcando una cuestión y además un mayor grosor de las artistas que une a las palabras significa un vínculo más fuerte, es decir, que tienen una mayor tendencia a aparecer juntas en el texto. Todo ello, será muy relevante para poder establecer la macroestructura del discurso que estemos analizando, es decir, los campos lexicales sobre los que se sustenta el mismo.

Siguiendo con el mismo corpus de ejemplo que en el caso anterior al realizar el análisis de similitud nos permite ver que en torno a las noticias sobre el sarampión se pueden entrever los siguientes campos lexicales, por un lado, estaría todo lo concerniente a los casos de sarampión (qué supone el brote, dónde se está detectando, cuál es su origen), se abre conjuntamente a éste otro campo lexical que aborda la vacunación como medida de precaución importante para la población (prevención, cobertura, reforzar), y otro que destaca una información emitida por el actual secretario de Salud de EEUU, Robert F. Kennedy Jr, quien afirmaba que la vacuna triple vírica está hecha de “células de fetos abortados”, información que ha sido desmentida por varios medios. Por lo tanto, vemos que en este análisis nos permite poder realizar una lectura de la información publicada acerca de esta cuestión, y comprobar qué es lo que está siendo noticia y por ende, permitirnos pensar si hay algún aspecto importante sobre este tema en concreto que no se esté abordando, si así fuera, podríamos tratar de indagar en el porqué, en los efectos a su vez que eso pudiera tener.

Pero además, si queremos realizar nuestro análisis lexical con mayor profundidad, IRaMuTeQ nos ofrece la clasificación descendente jerarquizada que está inspirada en el método Reinert (1983), este análisis trata de arrojarnos las grandes temáticas incluidas en nuestro corpus textual, para ello, a través de una tabla de presencia / ausencia va cruzando los segmentos de texto con las formas activas del corpus, agrupando los segmentos de texto que tienden a compartir el mismo léxico, de manera que las clases obtenidas son agrupaciones de segmentos de texto que tienden a contener el mismo vocabulario (Ratinaud y Marchand, 2012). Estas clases se presentarán en un dendograma, que es un diagrama que muestra la secuencia de divisiones en clasificaciones jerárquicas y nos muestra tanto el tamaño de la clase como el léxico más representativo en cada una de ella. De modo que, siguiendo con el ejemplo que venimos analizando, esta clasificación nos arroja 6 clases lexicales, por un lado estaría la clase 1, que trata sobre el desmentido que se ha realizado al secretario de Salud de EEUU sobre el contenido de la vacuna triple vírica, después en el resto de clases obtenidas veríamos por un lado las clases 5 y 6 que centran en la importancia de potenciar la vacunación en todos los grupos de la población con el fin de tener protegido al mayor número de personas frente a un virus, como puede serlo el sarampión, como apunta la información de El Economista del 17 de mayo de 2025 en el siguiente segmento de texto extraído: “desarrollar estrategias específicas dirigidas a colectivos con baja cobertura vacunal con el objetivo de reducir el número de personas no protegidas frente al virus”. Es otra de las virtudes que nos ofrece IRaMuTeQ, siempre tenemos acceso al texto original, de modo que a la hora de poder interpretar e ilustrar la información, nos es de gran ayuda. Por su parte las clases 2, 3

El software

IRaMuTeQ realiza el análisis de similitud, que está basado en la teoría de grafos, cuyo principal objetivo es analizar relaciones lexicales y conservar aquellas con mayor grado de conexión

y 4 se centrarían más en el registro de casos positivos que se están confirmando en España y también cómo a nivel mundial a la organización de la salud le está preocupando el incremento que se está produciendo.

Todos estos análisis realizados de manera minuciosa y contextualizando los diferentes hallazgos, nos permitirán llegar a concluir cómo se está trazando el discurso mediático en torno a las informaciones que estemos analizando, en el ejemplo que venimos analizando, en torno a los casos positivos que se están registrando de sarampión, pudiendo concluir que se vertebra en tres ejes clave, uno es el puramente informativo dando datos de las localidades donde están apareciendo, explorando cuáles pueden ser los orígenes; otro eje sería el de la vacunación, agregándole el valor para reducir el contagio de enfermedades y por último, estaría el eje de desórdenes informativos, verificando informaciones que se hayan propagado y no sean completamente ciertas, cuestión aún más relevante si cabe cuando se trata de temas de salud.

De manera que, como hemos podido ver a través de este análisis exploratorio, la lexicometría y el uso de software de acceso abierto IRaMuTeQ, supone una herramienta de gran utilidad para los investigadores interesados por hacer un análisis de contenido completo, sin tener limitaciones en cuanto a la cantidad de información que se quiera explorar, de igual modo, hay diferentes idiomas presentes en el software, por lo que además del español, tendremos otra amplia gama de posibilidades para analizar la producción de textos en otros idiomas. Y poder hacer así una lectura sobre una realidad concreta que haya sido publicada ya sea en medios de comunicación, redes sociales u otros lugares que consideremos que el análisis de la información emitida pueda ser relevante, y pueda aportar un valor añadido e informativo a la sociedad.

Referencias bibliográficas

IRaMuTeQ : <http://www.iramuteq.org/>

Lebart, L., & Salem, A. (1988). Analyse Statistique des Données Textuelles. *Questions ouvertes et Lexicométrie*. París: Bordas.

Marchand, P., & Ratinaud, P. (2012). L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française. *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles* (pp. 687-699).

Reinert, M. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les Cahiers de l'analyse des données*, 8(2), 187-198.

Mariola Moreno Calvo

Editora asociada RECS; Departamento de Comunicación, Universidad Carlos III de Madrid, España

 mariomor@hum.uc3m.es