

Un año de edición en Wikipedia: el caso del artículo «COVID-19» en diez lenguas

RICARDO TAVARES LOURENÇO

Universidad Católica Andrés Bello, Universidad Metropolitana

rtavares@ucab.edu.ve, rtavares@unimet.edu.ve

Resumen

Wikipedia, una obra colectiva de gran magnitud por su envergadura idiomática y colaborativa, ha permitido apreciar cómo evoluciona la redacción y edición de los artículos sobre temas emergentes y cambiantes a medida que diversas fuentes revelan más y más datos. Un año después de la declaración oficial por parte de la Organización Mundial de la Salud de la existencia de la pandemia de la COVID-19, y de oficializarse el nombre de esta enfermedad, parece un buen momento para analizar cómo ha sido el proceso de edición del artículo «COVID-19» desde febrero de 2020. Para ello, se revisan las estadísticas disponibles en cada una de las versiones en los diez idiomas de los países con más contagios (inglés, alemán, francés, ruso, italiano, español, chino, portugués, turco e hindi) a fin de apreciar el número de wikipedistas involucrados, el volumen y la frecuencia de las ediciones, y el ritmo de edición mes a mes, entre otros indicadores. Asimismo, se comparan los apartados del artículo en cada idioma estudiado, con el propósito de verificar su orden y jerarquización. Se busca ponderar así no solo la densidad de información disponible, sino también comprender la influencia de unos idiomas sobre otros en los textos de divulgación científica. Se encontró que, a mayor número de editores, más extensos y densos son los artículos, y que se prefiere divulgar información en lenguas de mayor alcance global, independientemente del número de casos de contagio. Además, los editores convienen de común acuerdo el orden de los apartados considerando la audiencia a la que se dirigen.

Palabras clave

Wikipedia; COVID-19; edición en línea; redacción colaborativa; idiomas.

A year of editing at Wikipedia: case «COVID-19» article in ten languages

Abstract

Wikipedia, a big collective and collaborative work by its linguistic magnitude, has permitted to appreciate the writing and editing evolution of articles about emergent and changing themes as more and more data is revealed of different sources. After a year of official declaration of the World Health Organization of the COVID-19 pandemic existence, and the officialization of the disease's name, it is an enough cycle to analyze how the editing of «COVID-19» article has been done since February 2020. Because of this, the available statistics in each version in the ten languages of the countries with more contagions were revised (English, German, French, Russian, Italian, Spanish, Chinese, Portuguese, Turkish and Hindi), and the number of editors involved on the volume and frequency of editions, rate of editing month to month were analyzed. In addition, the subtitles of the article in each studied language were compared to verify their order and hierarchy. The aim of this study is to consider not only the information's density availability, but also to understand the weighing of some languages over others on popular science texts. We found out that the more number of editors, the longer and denser the articles are, and they prefer publishing information in global languages, regardless of the case number of contagions. Also, the editors agree on the order of the subtitles considering the target.

Keywords

Wikipedia; COVID-19; Online Editing; Collaborative Writing; Languages.

Recibido el 27/04/2021

Aceptado el 25/05/2021

1. Introducción

Wikipedia revolucionó el concepto de enciclopedia no solo por la cantidad de lenguas en que se redactan sus versiones —aproximadamente trescientas—, la cantidad de artículos que recoge —más de seis millones solo en la versión en inglés— o por ser de acceso libre, sino por el número de personas que contribuyen en todo el mundo a la redacción de sus textos. Esta descentralización ha hecho posible que mucha gente lea artículos enciclopédicos sobre toda clase de temas y tenga incluso la oportunidad de contribuir con la edición de sus contenidos.

Una temática que exige particular rigor es la salud, pues se debe velar por su precisión y referencialidad a fuentes confiables. Es altamente peligroso difundir información médica que no esté verificada o, aún menos, que carezca del aval de especialistas; por ello Wikipedia siempre incluye en sus artículos un aviso médico¹ para hacer entender al lector que la información que se le presenta es solo referencial y con riesgo de inexactitudes. Además, en la cabecera del artículo encontramos este mensaje en un recuadro:

Este artículo se refiere o está relacionado con un evento de salud pública reciente o actualmente en curso.

La información de este artículo puede cambiar frecuentemente. Por favor, no agregue datos especulativos y recuerda colocar referencias a fuentes fiables para dar más detalles.

Véase el aviso médico.

Uno de los artículos que en este momento tiene enorme actividad en todas las versiones de Wikipedia es «COVID-19», la enfermedad causada por el coronavirus SARS-CoV-2 y que fue declarada oficialmente pandemia por la Organización Mundial de

¹ «Wikipedia contiene artículos sobre cuestiones médicas. Sin embargo, **no es posible garantizar la veracidad** del contenido en los artículos, ya que estos **pueden presentar errores, falsedades o desactualizaciones**. Aunque la información pueda ser correcta o fiable y su contenido estar bien documentado, es posible que lo que se describa no corresponda con una situación de salud específica.

La información médica y terapéutica que ofrece Wikipedia, en especial sobre farmacoterapia, es en el mejor de los casos de naturaleza general y **no puede sustituir el consejo de un médico u otro profesional de la salud**.

Ningún colaborador, administrador, programador, patrocinador ni cualquier otra persona relacionada con Wikipedia es responsable del resultado ni de las consecuencias que pudiera ocasionar cualquier intento de utilizar la información o desinformación que se ofrece en esta enciclopedia. Por lo tanto, **cualquier uso que se haga de la información contenida en cualquier artículo sobre temas de medicina** (especialmente en cuanto al tratamiento de enfermedades se refiere) **será única y exclusivamente responsabilidad del lector**.

Nada en Wikipedia.org o en ningún otro proyecto de la Fundación Wikimedia debe ser entendido como un intento de ofrecer o proporcionar opiniones médicas o sobre el ejercicio de la medicina. Debe tenerse en cuenta que Wikipedia es una enciclopedia libre, que cualquier persona puede crear o modificar artículos, y que dicha persona podría no poseer conocimientos exactos de la materia sobre la que versa dicho artículo. De ser posible, **hay necesidad de ser diligente con la información que se encuentre en Wikipedia y verificarla de forma independiente**.

En cualquier caso, esperamos que los datos e información contenidos en Wikipedia resulten de utilidad y, por supuesto, se invita a la colaboración en la edición de artículos sobre medicina o salud, y usar o compartir cualquier material encontrado aquí bajo los términos del esquema de licenciamiento de Wikipedia» (las negritas son del original).

Recuperado de https://es.wikipedia.org/wiki/Wikipedia:Aviso_m%C3%A9dico [27 de marzo de 2021].

la Salud (OMS) el 11 de marzo de 2020. El impacto de esta nueva enfermedad ha sido tal que ha copado la atención en toda clase de ámbitos. Hemos pasado de no saber apenas nada sobre este nuevo virus a conocer ampliamente cómo es, cómo se contagia, las sintomatologías, los tratamientos adoptados, las medidas de prevención, las vacunas, entre otros elementos técnicos, y eso, evidentemente, iba a tener reflejo en Wikipedia. Asimismo, hemos sido testigos en un año de cómo varios países han encabezado la lista de contagiados, de fallecidos y ahora de vacunados, por lo que, intuimos, los ciudadanos afectados estarían interesados en indagar y, sobre todo, en publicar todo lo que se sabe acerca de esta enfermedad.

Ahora bien, debemos considerar que el mismo artículo se ha desarrollado en varias lenguas y, como es bien sabido, las versiones en cada una de ellas varían en extensión, densidad, estructura y actualización, entre otros aspectos. Este fenómeno refleja los niveles socioculturales, ideológicos y geopolíticos de los hablantes de cada una de estas lenguas. Así lo explica Cassany (2012: 158):

El conocimiento no es neutro, imparcial ni universal: está situado y enraíza con el contexto sociocultural y los intereses de las personas. Tras una entrada hay un voluntario anónimo con una cultura, una lengua y una determinada manera de ver el mundo, que escribe para una audiencia particular.

Dada su indiscutible importancia e impacto en estos momentos, y transcurridos doce meses desde su aparición, consideramos de interés analizar cómo ha sido la dinámica de composición y edición del artículo «COVID-19» en diez lenguas, para poder apreciar no solo la variación entre ellas, sino ver qué lenguas son las mejor posicionadas y tratar de indagar el porqué.

1.1 Antecedentes sobre la edición de artículos sobre COVID-19 en Wikipedia

Se han publicado ya algunos estudios sobre aspectos de la edición de Wikipedia, en especial relacionados con el tema COVID-19. Resumimos seguidamente los más relevantes en orden cronológico.

Colavizza (2020) encontró para mayo de 2020 que los editores² de Wikipedia han citado 3 083 artículos científicos de un corpus de 141 783 sobre COVID-19, lo que equivale a un 2 % de ese corpus; aun así, el autor concluye que el número de citas es representativo para sustentar las informaciones sobre el coronavirus. Este hallazgo se corresponde con la valoración hecha por Rumbo-Prieto (2019), quien sostiene lo siguiente:

[...] Wikipedia es una potente enciclopedia online para consultar cuestiones científicas en ciencias de la salud, tan fiable como cualquier otra

² En adelante, utilizaremos el término *editor* —y ocasionalmente *wikipedista*— para aludir a los voluntarios que ingresan a Wikipedia para redactar y editar sus contenidos.

enciclopedia especializada, y cuya utilización como fuente bibliográfica académica, no debería estar «mal vista», sobre todo cuando el contenido tenga un aval científico contrastado (referencias de calidad). Aun así, consideramos que el uso debe ser realizado por profesionales expertos, y siempre de forma racional y limitada; es decir, la información se debe completar con otras fuentes y tener en cuenta, que no se haga un abuso de la literatura gris (referencias de difícil localización o acceso).

Saorín, Pastor-Sánchez y Baños-Moreno (2020) proponen un método para la construcción de vocabularios controlados para los medios de comunicación, utilizando *Wikidata* y *Wikipedia* como fuentes de información terminológica.

Benjakob, Aviram y Sobel (2021) estudiaron las fuentes citadas en la Wikipedia en inglés sobre el coronavirus (enero-marzo 2020) y descubrieron que los artículos relacionados con este tema hacían referencia tanto a medios de comunicación confiables como a investigaciones académicas de alto impacto. Además, los artículos COVID-19 de Wikipedia tenían una clara preferencia por los estudios de acceso abierto publicados en revistas reconocidas.

Ruprecht *et al.* (2021) analizaron 223 millones de ediciones contribuidas de 2018 a 2020 en las versiones de doce lenguas³ de Wikipedia y encontraron que durante las restricciones de la pandemia (primavera de 2020) la comunidad global de voluntarios de Wikipedia aumentó considerablemente tanto en productividad como en número de recién llegados que se unieron ella. De hecho, las contribuciones a la Wikipedia en inglés aumentaron más de un 20 %.

Como puede apreciarse, las investigaciones adelantadas hasta ahora coinciden en que Wikipedia es un referente para los lectores del mundo para informarse sobre esta enfermedad y temas relacionados; de ahí su preocupación por revisar las fuentes citadas y su impacto en los medios de comunicación. Cabe resaltar que la Wikipedia en inglés es la más estudiada por su tamaño y alcance.

1.2 Problema de investigación

Es bien sabido que Wikipedia tiene versiones de un mismo artículo en diversas lenguas, lo que incide en la organización, el tamaño y la densidad de sus contenidos. Considerando que el COVID-19 es en este momento un tema de elevado interés para la ciudadanía, nos interesa saber cómo se presentan los artículos sobre esta enfermedad y, al mismo tiempo, verificar si a mayor cantidad de contagios en los primeros diez países sus respectivas lenguas tendrían un posicionamiento similar, o si, por el contrario, este factor no tiene nada que ver con ello. Además, quisimos indagar cómo ha sido la dinámica de la edición de los artículos durante el primer año de la pandemia, dado que es un tema en constante actualización.

³ Las lenguas estudiadas fueron las siguientes: inglés, francés, alemán e italiano (grandes Wikipedias); japonés, sueco, neerlandés y coreano (medias Wikipedias), y serbio, noruego, finés y danés (pequeñas Wikipedias).

Los hallazgos que arroje esta investigación nos permitirán comprender no solo el peso de las lenguas en la difusión de temas de corte científico, sino el modo como las respectivas comunidades lingüísticas organizan y presentan una misma información.

2. La edición en Wikipedia

La redacción y edición de artículos en Wikipedia se caracteriza principalmente por su perfil colaborativo. Dependiendo del tema tratado, la dinámica de la edición será más o menos elevada. Muchas veces alguien toma la iniciativa y comienza a redactar un nuevo artículo; luego otros editores continúan agregando contenidos o simplemente corrigiendo el texto producido. Los tipos de edición más comunes son los siguientes: corrección, remoción de un texto duplicado en una sección, división de una sección en dos con subtítulos, adición de información y fuentes, adición de material y cambio del título de la sección (Broughton, 2008).

Uno de los pilares que caracteriza a esta megaenciclopedia es el punto de vista neutral, el cual es definido así en la Wikipedia en español (2021a):

[Se] intenta conseguir que los artículos no aboguen por un punto de vista en concreto. Esto requiere ofrecer la información desde todos los ángulos posibles, presentar cada punto de vista de forma precisa, dotar de contexto los artículos para que los lectores comprendan todas las visiones, y no presentar ningún punto de vista como «el verdadero» o «el mejor». Esto implica citar fuentes autorizadas que puedan verificarse siempre que sea posible, especialmente en temas polémicos. Cada vez que aparezca un conflicto para determinar qué versión es la más neutral, debe declararse un periodo de reflexión mediante un cartel de discutido en el artículo. Se aclararán los detalles en la página de discusión y se intentará resolver la disputa con calma.

En el espacio denominado «Discusión», que aparece en una pestaña en la parte superior izquierda del artículo, los editores justifican sus cambios en él y, en caso de controversia, se procura llegar a acuerdos en aras de conservar el punto de vista neutral. Ahora bien, ¿por qué puede haber controversia en un artículo? Broughton (2008) explica que los contextos de los wikipedistas son diversos. Para empezar, los editores proceden de muy distintos países del mundo, sobre todo en la edición en inglés de Wikipedia; sus redactores pueden ser más adolescentes que adultos; en Wikipedia no hay prerequisites para ser editor; y, finalmente, las creencias de los editores difieren unas de otras, por lo que Wikipedia exige que se reflejen todos los puntos de vista y se llegue a acuerdos. El autor citado añade que muchos editores tienen diferentes motivaciones para editar: desde el vandalismo, pasando por la curiosidad o el estar personalmente involucrado —lo que conlleva un conflicto de intereses—, hasta la experticia o pasión por un tema. La otra razón de peso que Broughton (2008) señala como causante de controversias es el desconocimiento por parte de algunos editores sobre las políticas editoriales de Wikipedia: punto de vista neutral, las investigaciones no son originales y los contenidos o datos deben ser verificables. Esto significa que un editor que haya tenido la iniciativa de empezar un nuevo artículo y de editarlo

constantemente no significa que deba rechazar o revertir sistemáticamente las ediciones de otros por gusto.

2.1 Los bots en Wikipedia

Un pequeño porcentaje de las ediciones de los artículos de Wikipedia es ejecutado por bots, es decir, programas informáticos que llevan a cabo un conjunto de tareas repetitivas de forma automática. Zheng *et al.* (2019) estudiaron los bots utilizados en Wikipedia y encontraron que hasta febrero de 2019 estaban registradas 1 601 cuentas bot, de las cuales 25,36 % ejecutaron más de diez mil ediciones acumuladas y 24 bots hicieron más de un millón de ediciones desde sus inicios. Zheng *et al.* (2019) establecieron una clasificación de los bots según el tipo de rol que cumplen en la edición:

- 1) *Generador (Generator)*: genera páginas con base en plantillas predefinidas. Entre sus funciones están redireccionar páginas y crear páginas basadas en otras fuentes.
- 2) *Arreglador (Fixer)*: corrige errores con enlaces, contenidos, archivos y parámetros en las plantillas y categorías.
- 3) *Conector (Connector)*: conecta Wikipedia con otras wikis y sitios web.
- 4) *Etiquetador (Tagger)*: etiqueta o categoriza estatus del artículo, evaluación del artículo, wikiproyectos y estatus multimedia.
- 5) *Secretario (Clerk)*: actualiza estadísticas, documenta el estatus de los usuarios, actualiza el mantenimiento de las páginas y emite una alerta del artículo.
- 6) *Archivador (Archiver)*: archiva contenidos y limpia contenidos de la zona de pruebas (*sandbox*).
- 7) *Protector (Protector)*: identifica violaciones de las políticas editoriales, *spam* y vandalismo.
- 8) *Asesor (Advisor)*: provee sugerencias para wikiproyectos, usuarios y saluda a los recién llegados.
- 9) *Notificador (Notifier)*: envía notificaciones a los usuarios.

Esta herramienta permite comprender por qué en ciertos aspectos los artículos entre algunas lenguas se asemejan en parte o en su totalidad. También permite entender cómo es el proceso de edición de la Wikipedia, en el que se unen elementos automatizados y humanos en pos de una composición más ágil de los artículos.

2.2 Manual de estilo

Otro instrumento que guía el trabajo de los editores tanto en la estructura como en la redacción y política editorial es el manual de estilo. Cada versión lingüística de Wikipedia posee el suyo y también se elabora de manera colaborativa. Por supuesto, las grandes Wikipedias son las que han desarrollado estas guías, que varían dependiendo

de los rasgos característicos de sus respectivas lenguas. En caso de que alguna de ellas no cuente con uno, el artículo sobre las políticas editoriales de Wikipedia hace las veces de manual de estilo.

De cualquier modo, en esta fuente de consulta se pueden encontrar directrices clave para la edición de los artículos: estilo enciclopédico, lengua estándar, lenguaje técnico, inserción de recursos multimedia, hipertexto, referenciación y categorización. Además, existen subsecciones con indicaciones específicas según la temática, entre ellas medicina. Sobre este último caso encontramos reglas que explican no solo cómo titular apropiadamente estos artículos, sino también las secciones sugeridas según el tema tratado en ellos. Por ejemplo, en textos sobre enfermedades, desórdenes o síndromes, las secciones sugeridas son las siguientes: clasificación, signos y síntomas, causas, mecanismo, diagnóstico, prevención o rastreo, tratamiento o manejo, epidemiología, historia, sociedad y cultura, investigaciones, grupos especiales de población, otros animales (Wikipedia, 2021b).

3. Metodología

A continuación se describe el método adoptado para llevar a cabo el presente estudio, el cual es de carácter descriptivo y documental.

3.1 Selección de los idiomas

Para la selección de los idiomas, se partió del criterio de revisar los 10 países que el 15 de marzo de 2021 encabezaban la lista de mayor número de contagiados por COVID-19, con base en los datos recopilados por la Universidad Johns Hopkins en su portal web <https://coronavirus.jhu.edu/map.html>. Esta elección tiene como propósito indagar si una comunidad altamente afectada por COVID-19 influiría en la publicación y difusión de mayor cantidad de información sobre esta enfermedad en Wikipedia (ver tabla 1).

Tabla 1. Lista de los países infectados por COVID-19 (a 15 de marzo de 2021)

PAÍS	NÚMERO DE INFECTADOS	LENGUA OFICIAL O PRINCIPAL
Estados Unidos	24 428 075	Inglés
Brasil	11 439 558	Portugués
India	11 359 048	Inglés, hindi
Rusia	4 341 381	Ruso
Reino Unido	4 271 710	Inglés
Francia	4 131 872	Francés

Italia	3 223 142	Italiano
España	3 183 704	Español
Turquía	2 879 390	Turco
Alemania	2 578 537	Alemán

Fuente: Universidad Johns Hopkins. Elaboración propia.

Como puede apreciarse, hay tres países cuya lengua principal u oficial es el inglés: Estados Unidos, Reino Unido e India. Para tener el número completo de diez idiomas para este estudio, se decidió incluir el chino mandarín, considerando que fue en China donde comenzó la propagación de la enfermedad. En el caso de India, incluiremos además el hindi, por ser lengua cooficial junto con el inglés en ese país.

En tal sentido, se mencionan las lenguas con las que se trabajaron en esta investigación:⁴ inglés (EN), portugués (PT), hindi (HI), ruso (RU), francés (FR), italiano (IT), español (ES), turco (TR), alemán (DE) y chino (ZH).

3.2 Elicitación de datos

Mediante la página web XTools (<https://xtools.wmflabs.org>) se extrajeron los datos estadísticos de edición de cada una de las diez versiones del artículo «COVID-19», los cuales arrojan, entre otros indicadores, los siguientes: tamaño del artículo, número de editores, cantidad y promedio de las ediciones, cantidad de ediciones mayores y menores, bots utilizados, enlaces empleados, editores con más aportes, etc. De todas estas estadísticas, solo extrajimos aquellas informaciones relacionadas con la edición propiamente dicha.

Por otra parte, para comprender cómo se organizan los artículos, se revisaron los índices de cada versión, con el propósito de analizar qué apartados se han adoptado y en qué orden aparecen en cada caso. Observar este indicador nos permite ver la jerarquización de la información en cada caso.

4. Caso de estudio: artículo «COVID-19»

4.1 Estadísticas de edición

A continuación, se presenta un conjunto de tablas que recogen datos interesantes que revelan no solo el tamaño y la densidad de contenidos, sino el peso de cada lengua a la hora de congregarse editores y componer el artículo.

⁴ En adelante, en las tablas, cuadros y gráficos utilizaremos los códigos ISO 639-1 para identificar los idiomas estudiados en esta investigación.

Tabla 2. Número de editores (a 15 de marzo de 2021)

EN	DE	FR	RU	IT	ZH	ES	TR	PT	HI
1 221	713	709	443	582	289	256	145	72	35

Fuente: XTools. Elaboración propia.

Tabla 3. Número absoluto de ediciones totales (a 15 de marzo de 2021)

EN	DE	FR	RU	IT	ZH	ES	TR	PT	HI
6 493	3 239	3 154	2 092	1 792	1 376	1 017	410	350	73

Fuente: XTools. Elaboración propia.

Tabla 4. Promedio de ediciones diarias (a 15 de marzo de 2021)

EN	FR	DE	RU	IT	ZH	ES	TR	PT	HI
20,4	9,9	9,7	6,6	5,7	4,4	3,2	1,3	1,1	0,3

Fuente: XTools. Elaboración propia.

Tabla 5. Promedio de ediciones mensuales (a 15 de marzo de 2021)

EN	FR	DE	RU	IT	ZH	ES	TR	PT	HI
490,1	237,5	231,8	158,7	137,3	105,7	77,9	32,3	27	6,2

Tabla 6. Número absoluto de ediciones por bots y su porcentaje con respecto al total de ediciones (a 15 de marzo de 2021)

EN	FR	IT	RU	ES	TR	ZH	DE	PT	HI
165	63	57	52	43	43	26	14	5	3
TR	ES	HI	IT	EN	RU	FR	ZH	PT	DE
10,5%	4,2%	4,1%	3,2%	2,5%	2,5%	2%	1,6%	1,5%	0,4%

Fuente: XTools. Elaboración propia.

Tabla 7. Porcentaje de ediciones menores con respecto al total de ediciones (a 15 de marzo de 2021)

RU	PT	FR	ZH	ES	DE	HI	TR	EN	IT
41,5%	37,4%	35,8%	33%	31,1%	30,6%	30,1%	28,4%	18,6%	17,1%

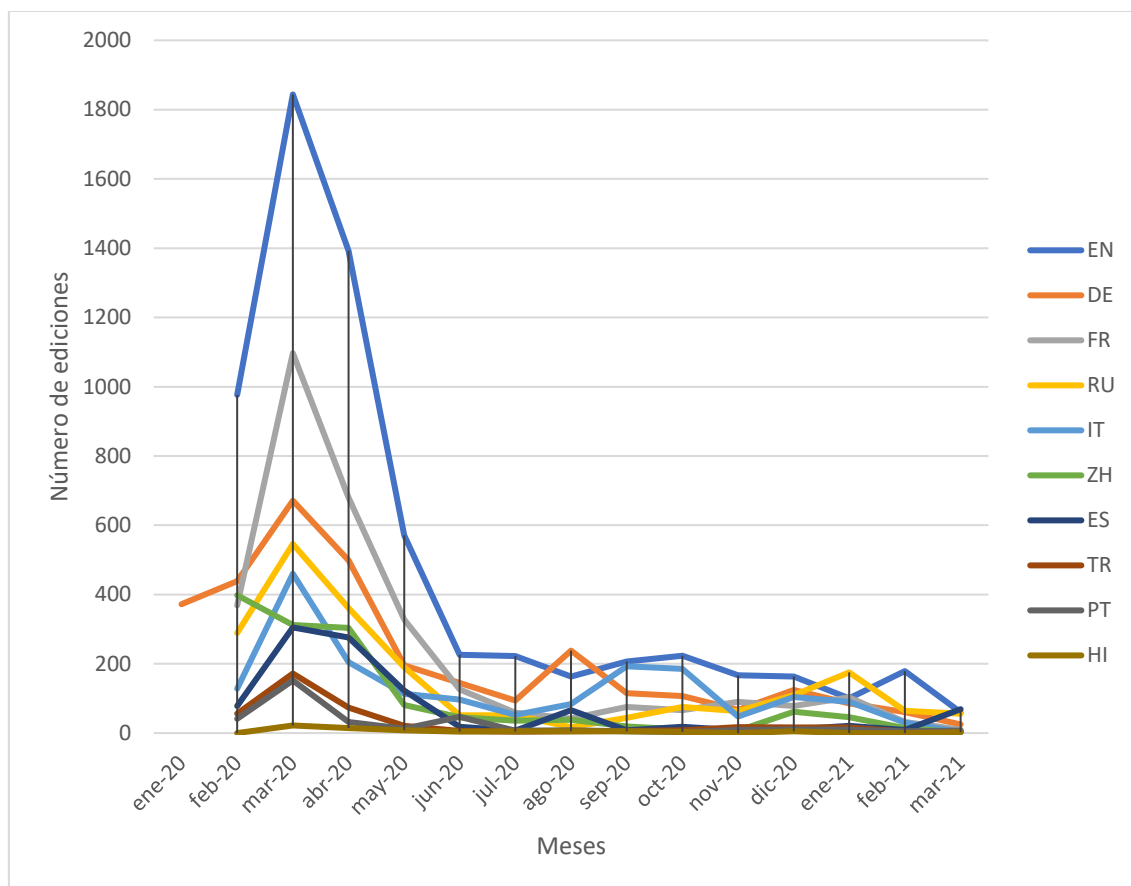
Fuente: XTools. Elaboración propia.

Tabla 8. Número de enlaces al artículo «COVID-19» (a 15 de marzo de 2021)

EN	ZH	ES	FR	PT	DE	IT	RU	TR	HI
14 002	3 049	2 881	2 081	1 825	1 633	1 616	1 320	1 380	30

Fuente: XTools. Elaboración propia.

Gráfico 1. Número de ediciones mensuales por lengua



Fuente: XTools. Elaboración propia.

4.2. Interpretación de los datos estadísticos

4.2.1 Editores y cantidad de ediciones

Las tablas mostradas en el apartado anterior permiten apreciar varias cosas. La primera de ellas es que el artículo «COVID-19» en inglés es, de lejos, el que cuenta con mayor cantidad de editores; de hecho, la cantidad de editores en todas las versiones es directamente proporcional al volumen de ediciones, bien sean totales, mensuales o diarias. En cierta medida, estas cifras también explican el tamaño y la densidad de las versiones del mismo artículo. La versión en hindi se ubica en último lugar en todos estos indicadores.

Siguiendo con los datos presentados en las tablas 2, 3, 4 y 5, todo apunta a que los países con más contagios hasta el momento en que hacemos este estudio no influyen en el tamaño y la densidad de los artículos. Si bien Estados Unidos encabeza el número acumulado de contagios en el mundo, los idiomas de los países que lo siguen no cuentan con versiones del artículo «COVID-19» en la misma proporción. Esto es notable en el caso del portugués: aunque Brasil se encuentra en segundo lugar en la lista de países con más contagios, la versión en portugués se sitúa en noveno lugar. El artículo «COVID-19» en hindi es muy básico y escueto; esto incluso es llamativo si consideramos que es la cuarta lengua con más hablantes nativos en el mundo, según datos de Ethnologue. En el caso del francés y del alemán, las versiones del artículo en estudio son bastante extensas y completas, pese a que sus países de referencia (Francia y Alemania) no figuran entre los cinco con más contagios acumulados.

Podría parecer que una excepción a esta tendencia es Italia. En marzo de 2020, Italia era uno de los países más afectados por el COVID-19 y, en el gráfico 1, se puede observar que se ubica en el quinto lugar en número de ediciones de su versión del artículo. Luego, entre septiembre y octubre de 2020, se constata un incremento importante de ediciones —casi igualando al volumen de ediciones en inglés—, etapa que coincide con la segunda ola de la pandemia en ese país.

El caso del artículo en chino merece un comentario especial. Según los datos recogidos, esta versión tiene un estatus intermedio, muy similar a la española. El texto estudiado en chino, además, se sitúa en segundo lugar en número de enlaces hacia el artículo, aunque muy lejos del inglés. Otro indicador interesante es el de las ediciones mensuales: en febrero de 2020 estaba en tercer lugar, tras el artículo en inglés y en alemán, pero después se da una disminución acentuada de ediciones; de hecho, en marzo de 2020 es la única versión que no crece más, contrariamente a la tendencia de las demás lenguas (ver gráfico 1). ¿Por qué ocurre esto? El acceso a Wikipedia desde China es limitado y muchas veces ha sido bloqueado por el gobierno de ese país; por ello, los editores activos de lengua china proceden en su mayoría de Taiwán, según las estadísticas de Wikimedia (<https://stats.wikimedia.org>), lo que explicaría la baja participación, a pesar de ser la lengua materna con más hablantes del mundo.

Estas cifras nos sugieren que son otras las razones por las cuales una lengua está más favorecida que otra. Para empezar, el inglés ha sido desde hace varias décadas la lengua preferida para publicar estudios e informaciones sobre ciencia, como se demuestra en este caso. Hasta el 15 de marzo de 2021 han intervenido 1 221 editores de múltiples países, no exclusivamente estadounidenses o británicos. Esto explica por qué el hindi se ha visto menos favorecido: hay muchos editores indios que prefieren intervenir en la Wikipedia en inglés, considerando que este idioma es también oficial en su país.⁵ Otra razón que ratifica el peso del artículo «COVID-19» en inglés la encontramos en la tabla 8, la cual revela claramente que dicha versión es el principal referente, pues más de catorce mil enlaces dirigen a ella, lo que representa el 46,95 %

⁵ Este dato se puede corroborar en la página de estadísticas de Wikimedia: en la Wikipedia en hindi apenas participan cuatrocientos editores indios, mientras que en la Wikipedia en inglés participan más de seis mil desde India (<https://stats.wikimedia.org/>).

de las diez versiones estudiadas. Por lo tanto, aquí el elemento clave es que aquellas comunidades lingüísticas con arraigada cultura investigativa de alto nivel son las que desarrollan los contenidos más extensos y densos, y dichas comunidades se expresan principalmente en inglés, francés y alemán.

4.2.2 Dinámica de las ediciones

Con respecto a la dinámica de la edición, las estadísticas obtenidas ofrecen datos interesantes. El gráfico 1 muestra que el mayor volumen de ediciones del artículo «COVID-19» en todas las versiones estudiadas —excepto en la china, como ya se acotó en el apartado anterior— se da en el mes de marzo del año 2020, encabezadas por las versiones en inglés (1 844), francés (1 097) y alemán (671). Esto es de esperar, pues justamente el 11 de marzo de ese año la OMS declaró oficialmente la pandemia del coronavirus, como ya se mencionó. Desde el 11 de febrero de 2020 a la enfermedad se le llama oficialmente COVID-19.

A partir de junio se estabiliza en todos los casos el volumen de ediciones. Solo en dos momentos el número de ediciones en una lengua sobrepasa al inglés: en agosto de 2020 hubo 228 ediciones en alemán —en inglés hubo 163— y en enero de 2021 hubo 175 ediciones en ruso —en inglés hubo 100—. En el caso del alemán, durante el mes de agosto de ese año se dio un significativo incremento en los contagios y, al mismo tiempo, se extendieron las protestas de los ciudadanos por las medidas de restricción y confinamiento que el Gobierno alemán estaba estableciendo. Al revisar el historial de las ediciones de ese mes, se observa que hubo varias inserciones revertidas, entre otras razones porque se tomaban de medios de comunicación poco rigurosos, porque la terminología no era correctamente manejada o porque se acotaban informaciones poco relevantes para el artículo. En el caso del ruso, en enero de 2021, se verificó el inicio de la vacunación masiva en Rusia. Al revisar el historial de ediciones, se reprodujo una situación similar al caso alemán: ediciones revertidas y debates sobre la confiabilidad de las fuentes citadas o uso correcto de la terminología.

Con respecto a la proporción de ediciones menores —entendidas estas como pequeñas modificaciones de forma que no afectan al fondo o al contenido—, se observan algunas tendencias. La tabla 7 muestra que, de promedio, el 30 % de las ediciones en Wikipedia son menores, lo que indica que los editores han incorporado o eliminado gran cantidad de contenido. La versión en ruso ha alcanzado el mayor número de este tipo de edición, con 41,5 %, mientras que en italiano apenas fue del 17,1 %. Esto se puede deber no solo a la naturaleza de Wikipedia —hay que considerar que los artículos no están definitivamente concluidos—, sino a que el tema sobre el COVID-19 es tan novedoso y cambiante que demanda constante actualización.

4.2.3 La edición con bots

Las estadísticas recopiladas con respecto a la edición con bots señalan datos que conviene comentar. En general, las ediciones efectuadas con este mecanismo son muy bajas, pues en promedio apenas sobrepasan el 3 %. Llama la atención que en el artículo

en turco, si bien se han hecho apenas 43 cambios, estos representen el 10,5 %. Los bots más utilizados —Khutuck Bot e InternetArchiveBot— se centran en correcciones de texto, en corregir redirecciones dobles y en rescatar enlaces muertos. La versión que menos bots utiliza es la alemana: apenas 0,4 % de todas las ediciones del artículo.

4.3 Comparación de los índices

Los índices de los artículos nos indican cómo está organizada y jerarquizada la información. En esta investigación se revisaron dichos índices para averiguar qué apartados utilizan, en qué orden y hasta qué punto coinciden en las distintas versiones.

De los diez documentos en estudio, el texto en francés es el más extenso y detallado —16 apartados de grado uno, con abundantes subtítulos de segundo y tercer grado de jerarquía—, mientras que la versión en hindi es la más corta y simple —7 apartados de grado uno, sin títulos de segunda jerarquía—.

4.3.1 Orden de los subtítulos

El primer subtítulo grado uno en que coinciden cuatro versiones es «Síntomas»: inglés (*Signs and symptoms*), ruso (Симптомы), turco (*Belirti ve Bulgular*) y portugués (*Sinais e sintomas*). Este mismo apartado aparece en segundo lugar en alemán, chino e hindi; y en cuarto lugar aparece en francés, italiano y español (ver cuadro 1). Cuando nos detenemos a ver qué títulos privilegian las demás lenguas, encontramos que los editores de las otras seis lenguas han preferido brindar al lector un poco de contexto antes de hablar de síntomas; caso contrario a las lenguas inglesa, rusa, turca y portuguesa, que han optado por comenzar por los síntomas e introducir después el contexto.

Esta organización no es azarosa. Efectivamente, determinar este orden forma parte de las discusiones que los editores sostienen en la plataforma. La decisión de hablar primero de los síntomas obedece a que la comunidad de wikipedistas asume que esa será la información que buscará primero el lector, por tratarse de una enfermedad. En consecuencia, la organización de los contenidos del artículo puede deberse a los hábitos de lectura y expectativas de lectura que tendrá el lector.

Cuadro 1. Primeros cinco subtítulos de grado 1 del artículo «COVID-19» en diez lenguas

DE	FR	IT	ZH	ES
1. Causa y desarrollo de la enfermedad (<i>Ursache und Krankheitsentstehung</i>)	1. Nombre de la enfermedad (<i>Nom de la maladie</i>)	1. Epidemiología y notas históricas (<i>Epidemiologia e cenni storici</i>)	1. Nombre (名稱)	1. Nombre
2. Síntomas clínicos y signos químicos de laboratorio (<i>Klinische Symptome und laborchemische Krankheitszeichen</i>)	2. Historia de la enfermedad (<i>Historique de la maladie</i>)	2. Etiología (<i>Eziologia</i>)	2. Síntomas (症狀)	2. Historia
3. Diagnóstico (<i>Diagnostik</i>)	3. Causa: coronavirus SARS-CoV-2 (<i>Cause : coronavirus SARS-CoV-2</i>)	3. Patogénesis (<i>Patogenesi</i>)	3. Causas (病因)	3. Fisiopatología
		4. Anatomía patológica clínica (<i>Anatomia patologica clinica</i>)	4. Detección (检测)	4. Signos y síntomas
			5. Prevención (预防)	5. Complicaciones

4. Opciones de tratamiento (<i>Behandlungsmöglichkeiten</i>)	4. Síntomas y manifestaciones clínicas (<i>Symptômes et manifestations cliniques</i>)	5. Manejo y tratamiento de Covid-19 (<i>Gestione e trattamento del Covid-19</i>)		
5. Perspectiva de curación (<i>Heilungsaussicht</i>)	5. Diagnóstico (<i>Diagnostic</i>)			
EN	RU	TR	PT	HI
1. Signos y síntomas (<i>Signs and symptoms</i>)	1. Síntomas (Симптомы)	1. Signos y síntomas (<i>Belirti ve Bulgular</i>)	1. Señales y síntomas (<i>Sinais e sintomas</i>)	1. Expansión (फैलाव)
2. Causa (<i>Cause</i>)	2. Epidemiología (Эпидемиология)	2. Causa (<i>Sebebi</i>)	2. Causas (<i>Causas</i>)	2. Signos y síntomas (संकेत और लक्षण)
3. Fisiopatología (<i>Pathophysiology</i>)	3. Proceso infeccioso (Инфекционный процесс)	3. Epidemiología (<i>Epidemiyoloji</i>)	3. Mecanismo (<i>Mecanismo</i>)	3. Causa (कारण)
4. Diagnóstico (<i>Diagnosis</i>)	4. Profilaxis (Клиническая картина)	4. Diagnóstico (<i>Tani</i>)	4. Diagnóstico (<i>Diagnóstico</i>)	4. Prevención (रोकथाम)
5. Prevención (<i>Prevention</i>)	5. Diagnósticos (Профилактика)	5. Prevención (<i>Önlenmesi</i>)	5. Prevención (<i>Prevenção</i>)	5. Ver también (इन्हें भी देखें)

Fuente: Elaboración propia.

5. Conclusiones

Como se ha podido apreciar y demostrar, Wikipedia es una megaenciclopedia que deja al descubierto muchas brechas lingüísticas, lo que afecta la calidad de los artículos. Estudiamos el artículo «COVID-19» por ser un tema de altísimo interés para el mundo y quisimos saber cómo ha sido la dinámica de la edición en las lenguas de los países con más contagios de esta enfermedad. Encontramos que influye más seleccionar escribir en la lengua de mayor alcance global que ser uno de los países con más contagios, tal como ocurre con el portugués respecto de Brasil.

Si bien Estados Unidos está a la cabeza de los contagios y fallecimientos en el mundo, vemos que el artículo forma parte de una Wikipedia en la que intervienen personas de muchos países; ya de entrada, el artículo «COVID-19» en inglés estuvo a la cabeza desde el inicio de la pandemia. Además, las comunidades lingüísticas con cultura investigativa contribuyen a que sus lenguas destaquen, no solo en inglés, como ya hemos visto, sino en francés y alemán, cuyas versiones se componen esencialmente en Francia y Alemania, respectivamente, países de elevada tradición científica. Que la lengua esté entre las más habladas del mundo tampoco es suficiente: lo vemos en el caso del chino, el hindi e incluso el español y el portugués, cuyos hablantes nativos superan en número al inglés. En síntesis: a menor número de editores para una determinada versión de un artículo, menor será la extensión y densidad de sus artículos. De hecho, este orden de proporciones coincide en buena medida con el tamaño de las Wikipedias según el número de artículos (ver tabla 9) y ello se debe a que el artículo «COVID-19» es de interés mundial y no regional o local.

Tabla 9. Tamaño de las Wikipedias por número de artículos (a 15 de marzo de 2021)

EN	DE	FR	RU	IT	ES	ZH	PT	TR	HI
6 308 335	2 582 333	2 333 936	1 728 289	1 696 714	1 688 663	1 200 202	1 065 988	405 752	147 908

Fuente: List of Wikipedias (https://meta.wikimedia.org/wiki/List_of_Wikipedias). Elaboración propia.

También investigamos el ritmo de ediciones de cada uno de estos artículos. Encontramos que el mes de marzo de 2020, por ser el mes de la declaración oficial de pandemia por la OMS, estimuló el máximo volumen de ediciones mayores en las diez lenguas estudiadas. Desde junio de ese año se observa una estabilización en la frecuencia de las ediciones en general, con tendencia a la baja.

Predominan en todas las versiones ediciones mayores que menores, lo que indica que el tema «COVID-19», por ser novísimo en el mundo de la ciencia, exige actualización constante en diversos elementos: síntomas, tratamiento, prevención, formas de contagio, entre otros.

Las ediciones con bots se usan en muy bajo nivel y apenas para corregir de forma automática errores menores, enlaces muertos, duplicaciones, etc., lo que indica el elevado control humano en la edición.

La organización de los apartados en los artículos obedece a consensos entre los editores, atendiendo, entre otros criterios, al interés de los lectores y a sus expectativas de lectura. Esto también deja entrever que unas versiones son más “enciclopedistas”, en tanto que prefieren empezar por el contexto de la enfermedad, y otras son más “informativas”, al escoger como primer apartado la sintomatología del COVID-19.

Finalmente, este estudio nos ha permitido comprender que las lenguas por sí solas no favorecen o desfavorecen la calidad de los textos; lo determinante es el criterio de la sociedad del conocimiento para escoger la lengua con mayor impacto global, terreno donde el inglés gana ampliamente, o la tradición académica arraigada en comunidades lingüísticas como la francesa o la alemana. Si queremos que otras lenguas, incluido el español, figuren como lenguas referentes de ciencia, debemos empezar por cultivar en la ciudadanía el gusto y el interés por la investigación y el rigor académico. Al entender eso, estimularemos expresar nuestros hallazgos en muchas lenguas y lograremos una mayor equidad para beneficio de muchos lectores.

Referencias bibliográficas

- BENJAKOB, O., R. AVIRAM & J. SOBEL (2021). «Meta-Research: Citation needed? Wikipedia and the COVID-19 pandemic». *bioRxiv*, March 1st, 2021, 1–22. <https://doi.org/10.1101/2021.03.01.433379>
- BROUGHTON, J. (2008). *Wikipedia: The Missing Manual*. Sebastopol: O'Reilly.
- CASSANY, D. (2012). *En línea. Leer y escribir en la red*. Barcelona: Anagrama.
- COLAVIZZA, G. (2020). «COVID-19 research in Wikipedia». *Quantitative Science Studies*, 1(4), 1349–1380. https://doi.org/10.1162/qss_a_00080

- ESTADÍSTICAS DE WIKIMEDIA (2021) [sitio web en línea]. Recuperado de <https://stats.wikimedia.org>
- ETHNOLOGUE (2021). «What is the most spoken language?». Recuperado de <https://www.ethnologue.com/guides/most-spoken-languages>
- RUMBO-PRIETO, J. M. (2019). ¿Es la Wikipedia® una fuente bibliográfica fiable para la investigación en ciencias de la salud? *Enferm Dermatol*, 13(37): 66-68. DOI: 10.5281/zenodo.3408550
- RUPRECHTER, T., M. H. RIBEIRO, T. SANTOS, F. LEMMERICH, M. STROHMAIER, R. WEST & D. HELIC (2021). «Volunteer contributions to Wikipedia increased during COVID-19 mobility restrictions». *arXiv:2102.10090*.
- SAORÍN, T.; J. A. PASTOR-SÁNCHEZ; M. J. BAÑOS-MORENO (2020). «Uso de Wikidata y Wikipedia para la generación asistida de un vocabulario estructurado multilingüe sobre la pandemia de Covid-19». *Profesional de la Información*, 29(5), e290509. <https://doi.org/10.3145/epi.2020.sep.09>
- WIKIPEDIA (2021a). «Los cinco pilares» [sitio web en línea]. Recuperado de https://es.wikipedia.org/wiki/Wikipedia:Los_cinco_pilares [Fecha de consulta: 30 de marzo de 2021].
- WIKIPEDIA (2021b). «Manual of Style. Medicine Related Articles» [sitio web en línea]. Recuperado de https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Medicine-related_articles [Fecha de consulta: 30 de marzo de 2021].
- ZHENG, L., C. M. ALBANO, N. M. VORA, F. MAI, & J. V. NICKERSON (2019). «The Roles Bots Play in Wikipedia». *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, artículo 215 (November 2019), 20 páginas. <https://doi.org/10.1145/3359317>