



Wikipedia en español. Comportamiento de la comunidad hispanohablante en el trabajo colaborativo en Internet¹

Florencia Claes²; Luis Deltell³

Recibido: 11 de septiembre de 2018 / Aceptado: 28 de enero de 2019

Resumen. Este artículo aborda una investigación sobre el uso del español en el trabajo colaborativo en Internet. Se trata de una primera fase de un experimento realizado junto a Wikimedia España y apoyado por tres centros de investigación. Se analiza por qué Wikipedia en español, segunda lengua con más hablantes nativos se encuentra en la novena posición en volumen de artículos en dicha enciclopedia. Además, este trabajo aborda el estudio de los sesgos intrínsecos de Wikipedia, tanto los del ámbito internacional como los propios de la versión en el idioma español. Para explorar este comportamiento hemos realizado una investigación con 231 estudiantes universitarios españoles. Dicho experimento pretendía medir la capacidad de creación de los usuarios.

Palabras clave: Wikipedia; internet; trabajo colaborativo; enciclopedia; sesgo; español.

[en] Wikipedia in Spanish. Behaviour of the Spanish-speaking Community in Relation to Collaborative Work on the Internet

Abstract. This article tackles with a research about the use of the Spanish language in relation to collaborative work on the Internet. It describes an early stage of an experiment carried out along with Wikimedia Spain and supported by three research centers. In this research it is analysed why Wikipedia in Spanish, which is the second language with more native speakers, holds the ninth place in relation to the number of articles in that encyclopedia. This work also deals with the study of Wikipedia intrinsic biases, both from the international scene and from the Spanish version. In order to explore the aforementioned behaviour, we have made a research based on 231 Spanish university students aimed at measuring users' creative skills.

Keywords: Wikipedia; internet; collaborative work; encyclopedia; bias; Spanish.

Sumario. 1. Presentación: el español en la era digital. 2. Antecedentes: el español en Wikipedia; 2.1. Sesgos intrínsecos internacionales y domésticos de Wikipedia en español; 2.1.1. Sesgos intrínsecos internacionales. 3. Metodología: corpus y objetivos; 3.1. Relación de objetivos del experimento. 4.

¹ Esta investigación ha sido posible gracias a la financiación de la Universidad Complutense y el Banco Santander en España, Referencia: PR26/16-6B-1, titulado: *PERLAD, Pensamiento y representación literaria y artística digital ante la crisis de Europa y el Mediterráneo*.

² Universidad Rey Juan Carlos (España)
E-mail: florencia.claes@urjc.es

³ Universidad Complutense de Madrid (España)
E-mail: ldeltell@ucm.es

Resultados y discusión; 4.1. Resultado y discusión de los sesgos 5. Conclusiones. 6. Referencias bibliográficas. 7. Agradecimientos y ayudas.

Cómo citar: Claes, Florencia y Deltell, Luis (2019): "Wikipedia en español. Comportamiento de la comunidad hispanohablante en el trabajo colaborativo en Internet". *Estudios sobre el Mensaje Periodístico* 25(3), 1357-1378.

1. Presentación: el español en la era digital

Según las estadísticas más verosímiles, en la actualidad la población mundial se encuentra en torno a unos 7.550 millones de seres humanos (United Nations, Department of Economic and Social Affairs, Population Division 2017a). Lógicamente estas cifras se tratan de aproximaciones logradas por el cómputo de datos masivos y es casi imposible ofrecer un número preciso. De hecho, el resultado varía a cada instante por la resta de las defunciones y la suma de los nacimientos, estas estimaciones se logran utilizando la teoría de datos y no los padrones o las cifras nacionales (United Nations, Department of Economic and Social Affairs, Population Division 2017b). Dentro de la complejidad de estos cálculos, las valoraciones de las lenguas y el número de los hablantes de las mismas se vuelven aun una tarea más ardua. El impacto de los idiomas se puede medir desde diversas perspectivas y variables y cada una de estas formas muestran diferentes realidades.

Así, uno de los parámetros para medir el valor de las lenguas es contabilizar el número de hablantes de las mismas. Sin embargo, la relevancia de este cálculo cambia de forma significativa si se miden según los hablantes nativos o los hablantes con competencias medias y/o desarrolladas. Lógicamente las autoridades de países con idiomas con más hablantes nativos tienden a citar esta variable – español y portugués, entre otros– frente a los estados cuyas lenguas son aprendidas y usadas por hablantes no nativos que recurren a la variable de personas con capacidades lingüísticas desarrolladas –inglés, hindustaní (o indostánico, nos referiremos a este idioma en nuestra investigación como hindustaní) o francés–.

Un debate no menos importante es la definición y la categoría de qué es una lengua (Alvar 1961) y si esta debe ser homogénea para ser considerada un único idioma o varios dialectos, así como el valor de la cultura y el contexto político (Garayzábal y Zhang 2006). Los casos del árabe, el chino, el indonesio/malayo y el hindustaní son paradigmáticos de esta discusión (Comrie, 2009). Cada uno de ellos se puede considerar como un único idioma o varios según el enfoque que se tenga en el estudio. El caso del hindustaní resulta esclarecedor, aunque su empleo coloquial puede ser comprensible para casi todos los hablantes del hindi y el urdu (entre otras lenguas), su forma escrita, que incluso utiliza diferentes alfabetos: devanagari en la India y árabe en Pakistán y Afganistán, y su manera culta son muy diferentes y, en gran parte, incomprensibles entre los hablantes. Esto, como se verá, ha perjudicado a este idioma en su importancia en la Red y en la comunidad de Wikipedia.

El último gran aspecto a considerar de las lenguas es su reconocimiento internacional, la difusión en el mundo y su oficialidad –legal o de facto– en los distintos estados del mundo. Aunque estos datos son menos complejos de cuantificar que los anteriores, también fluctúan según las organizaciones o los

investigadores que los midan. Pues sus implicaciones políticas y económicas son sustanciales.

Los datos más precisos sobre la extensión de las lenguas en el mundo a 2017 lo ofrece la XXI edición del estudio de Ethnologue. Dicha organización plantea su distinción en las dos variables fundamentales según los hablantes con competencias medias (no necesariamente muy desarrolladas) y hablantes nativos.

Tabla 1. Lenguas ordenadas por hablantes nativos, hablantes con competencias medias y hablantes totales, adquiridas. Fuente: elaboración propia a partir de los datos de los informes de Ethnologue *Languages of the World* (Simons y Fenning 2018).

	Lengua	Hablantes Nativos	Hablantes con competencias medias adquiridas	Total
1	Inglés	378.200.000	743.500.000	1.120.000.000
2	Chino mandarín, mandarín estándar	908.700.000	198.400.000	1.100.000.000
3	Indostánico (hindi/urdu)	329.100.000	368.300.000	697.400.000
4	Español	442.300.000	70.600.000	512.900.000
5	Árabe	290.000.000	132.000.000	422.000.000
6	Francés	76.700.000	208.100.000	284.900.000
7	Indonesio/malayo	77.000.000	204.000.000	281.000.000
8	Ruso	153.900.000	110.400.000	264.300.000
9	Bengalí	242.600.000	19.200.000	261.800.000
10	Portugués	222.700.000	13.800.000	236.500.000
11	Panyabí	148.300.000	sin información	148.300.000
12	Alemán	76.000.000	56.000.000	132.000.000
13	Japonés	128.200.000	131.000.000	128.300.000
14	Persa/farsi	60.000.000	50.000.000	110.000.000
15	Suajili	16.000.000	82.300.000	98.300.000
16	Javanés	84.300.000	sin información	84.300.000

Como se observa en la tabla 1 en esta medición –más correcto decir aproximación– se tienen en cuenta los idiomas en su sentido más amplio. En el caso del chino se han incluido todas las variantes –salvo la cantonesa– y en el caso de indonesio/malayo, hindustaní, árabe y suajili se tratan como idiomas únicos aunque no sean plenamente uniformes. Por último, el término de hablantes con competencias medias se refiere a aquellos individuos que pueden comunicarse en esa lengua, equivalente a un B1 en el modelo del Instituto Cervantes según el Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación (Council for Cultural Cooperation Education Committee 2002), que lo equipara al Preliminary English Level según el sistema de la Universidad de Cambridge.

El español se encuentra en la cuarta posición según hablantes con competencias medias y en segundo lugar como lengua de hablantes nativos. El inglés, como resulta obvio, se trata de la lengua franca internacional hablada, o con capacidad para ser hablada con competencias básicas, por una de cada siete personas del mundo. El español, que tiene más hablantes nativos sin embargo, solo lo podría hablar un individuo de cada catorce. Estas cifras son interesantes para entender el siguiente paso que es el impacto de las lenguas en Internet.

En el mismo caso que en el estudio del número total de habitantes del planeta o de hablantes de las lenguas, el impacto de los idiomas en Internet se trata de un dato masivo propio de los estudios *Big Data* y que presenta varias aristas en su cómputo. La cifra básica se logra al monitorizar la cantidad de contenidos escritos en cada lengua en la Red. Es decir, la medición básica del impacto de las lenguas se hace analizando cuál es el idioma principal, o mayoritario, del flujo de contenido en Internet.

Tabla 2. Porcentaje de contenidos en la Red por idioma. Fuente: elaboración propia a partir de los datos de Web Technology Surveys (2018).

	Idioma	% de textos en la Red
1	Inglés	52.4%
2	Alemán	6.3%
3	Ruso	6.2%
4	Español	5.1%
5	Francés	4.1%
6	Japonés	4.0%
7	Portugués	2.8%
8	Italiano	2.5%
9	Persa/farsí	1.9%
10	Chino	1.8%
11	Polaco	1.8%
12	Neerlandés	1.3%
13	Turco	1.2%
14	Checo	1.0%
15	Coreano	0.9%

La tabla 2 nos revela algunos datos importantes. La primera característica indica que el inglés expande su carácter de lengua franca también en la Red. Resulta apabullante el dominio de este idioma ya que más de la mitad del contenido de Internet se encuentra en esa lengua. Su prevalencia supera con creces el porcentaje

de su número de hablantes o de personas con competencias en esa lengua. La segunda idea es que los idiomas menos uniformes (hindustaní, árabe, suajili e indonesio/malayo) no se encuentran representados entre los quince primeros. Incluso el chino no alcanza ni el dos por ciento de los contenidos en Internet. Es cierto que esto no solo se debe a la diversidad idiomática sino también a la posibilidad de acceso a la Red, tanto económico como político y cultural.

La tercera característica es que los estados con medidas para mejorar sus contenidos y su impacto en la Red –Rusia, Alemania, Francia e Italia– logran avances destacados. Por último, se encuentran los idiomas como el portugués y el español que mantienen, o descienden levemente, su impacto en Internet al compararlo con el porcentaje de hablantes –ya sean nativos o con competencias lingüísticas adquiridas–.

Estos datos, o las versiones de otros años, han sido comentados –y celebrados– por las Academias de la Lengua Española (ASALE) y por otras instituciones que hablan del impacto dicha lengua, como el Instituto Cervantes (2017), que se refieren al castellano como una lengua viva. Además, hay otro dato que parece reforzar el impacto de este idioma en la Red, su gran dominio en los medios sociales más importantes (Facebook, Twitter, Whatsapp...) que sitúan al español como la segunda lengua más utilizada en esos espacios.

Sin embargo, la tabla 2 y los cálculos de las redes sociales no miden ni el valor de los contenidos ni la capacidad de creación en Internet. Estos dos aspectos son claves en la importancia real de los contenidos. Por ello, es relevante estudiar cuál es el acceso a Internet de los hablantes de cada lengua, nativos y con competencias. Esto nos permite obtener el rango de penetración, es decir, el porcentaje de los hablantes que tiene accesibilidad real a la Red. Algunas lenguas como el japonés o el alemán muestran que casi todos sus hablantes pueden navegar Internet, mientras solo el 26% de los hablantes del ruso y el 50% de los hablantes del árabe tienen acceso a la Red. Este dato nos permite dar una estadística –no una cifra exacta– del número de usuarios por lenguas y del porcentaje de estos respecto a los usuarios totales de la Red en el planeta.

Tabla 3. Usuarios en internet por lenguas con competencias medias del idioma y rango de penetración de la lengua en la Red. Fuente: elaboración propia a partir de Internet World Stats (2108).

	Hablantes nativos y con competencias medias	Usuarios de internet	Rango penetración	Porcentaje de internautas mundiales
Inglés	1.120.000.000	806.400.000	72%	25.3 %
Chino	1.100.000.000	609.400.000	55.4%	19.4 %
Español	512.900.000	335.949.500	65.5%	8.1 %
Árabe	422.000.000	212.266.000	50.3%	5.3 %
Portugués	236.500.000	139.771.500	59.1%	4.1 %
Indonesio / malayo	281.000.000	158.484.000	56.4%	4.1 %
Japonés	128.300.000	119.703.900	93.3%	2.9 %

Francés	284.900.000	21.6808900	76.1%	2.7 %
Ruso	264.300.000	70.303.800	26.6%	2.8 %
Alemán	132.000.000	117.744.000	89.2%	2.2 %

La tabla 3 no indica la importancia de las lenguas en la Red, sino el idioma que emplean los usuarios en Internet. Esto nos permite entender que lenguas muy habladas como el hindustaní no son las más usadas en la Red⁴ –ya que posiblemente el rango de penetración es muy bajo. Otras, como el árabe y el indonesio/malayo con una gran población que usa Internet no logran generar gran cantidad de contenido en la Red.

Así, estos datos previos nos sirven de antecedentes para la investigación. La lengua que prevalece de forma rotunda en Internet es el inglés que gestiona más de la mitad del flujo de información. Además, solo los siete primeros idiomas controlan el ochenta por cien de los contenidos. El español se encuentra en una quinta posición –que si bien se aproxima a la cuarta según el número de hablantes con competencias medias– se queda alejada de la segunda posición que le corresponde por hablantes nativos.

2. Antecedentes: el español en Wikipedia

Sin duda uno de los proyectos más importantes del trabajo colaborativo online ha sido la enciclopedia en red inaugurada por Jimmy Wales y Larry Sanger en enero de 2001. Wikipedia es la enciclopedia global más importante en Internet, se escribe en 300 idiomas y contiene más de 48.409.360 artículos. (Colaboradores de Wikipedia 2018). A pesar de ser una web específica de conocimiento y de carácter divulgativo y no lucrativo, se encuentra entre las cinco más visitadas del mundo (Alexa 2018). Su impacto social y cultural es incuestionable. Por ejemplo, en protesta a la votación del 5 de julio de 2018 en Bruselas sobre el uso de los derechos de autor y nuevas políticas de la Unión Europea en el tráfico de información de Internet, Wikipedia en español decidió cerrar por 36 horas. Esta clausura temporal provocó una alarma social y tuvo una enorme repercusión mediática (Koch 2018).

Es cierto que el proyecto creado por Wales y Sanger ha sido criticado, primero por la ideología anarco-capitalista de sus fundadores (Ippolita 2012), segundo porque se basa en el desarrollo de un trabajo gratuito y voluntario y supone en último caso un modo de explotación, o auto explotación como define el teórico germano-coreano Buyng Chul Han (2013). También el trabajo colaborativo y no remunerado ha sido criticado por teóricos como Remedios Zafra, que no menciona

⁴ Las estadísticas sobre el acceso real a Internet son meras aproximaciones, Internet World Stats y el Berkman Center for Internet & Society en la Universidad de Harvard ofrecen un anuario sobre las mismas. Esta capacidad para navegar en la Red depende de múltiples factores: económicos, desarrollo social y cultural, políticos, ideológico... Según estas aproximaciones, en Pakistán y la India, países centrales del hindustaní, el acceso a Internet, el rango de penetración, es del 15,51% y del 29,55% respectivamente. La posibilidad de uso de la Red es aún muchísima más baja en los países africanos (Clark, Faris y Jones 2017).

directamente a esta organización, quien señala el hábito del trabajo gratuito como uno de los modelos más claros de capitalismo digital (Zafra, 2017).

Por último, el proyecto de conocimiento en abierto presenta una sobreinformación, debate y en último caso discusión y ruido sobre algunos temas determinados que cambian según la versión idiomática consultada. Estos temas de controversia terminan siendo bloqueados por los usuarios y gestionados directamente por los administradores y la comunidad.

Sin embargo, a pesar de las críticas a la organización, lo cierto es que Wikipedia es la gran enciclopedia de la actualidad y, sin duda, uno de los proyectos claves del trabajo colaborativo y la difusión del conocimiento como se reconoció en los premios Princesa de Asturias 2015 en la modalidad de Cooperación Internacional (Fundación Princesa de Asturias 2015).

Por ello consideramos a Wikipedia como uno de los lugares de la Red donde mejor se puede medir el impacto de una lengua y, sobre todo, dos aspectos claves para esta investigación: el uso colaborativo y el valor de los contenidos de los artículos que hacen los hablantes de ese idioma.

Al igual que hemos observado en las mediciones anteriores las cifras que utilizamos son monitorizaciones de estudios de datos masivos y contienen resultados que son importante explicar y definir. En la tabla 4 se ofrece el tamaño de las distintas Wikipedias según su idioma. Se clasifican en orden a la magnitud de artículos publicados (y aceptados), pero también del total de textos (no necesariamente artículos, también páginas de usuarios, talleres y otras); del número de ediciones; del sumatorio de administradores por idiomas; del cómputo de usuarios (perfiles creados con el fin de generar contenido en la enciclopedia)⁵; del número de usuarios activos (aquellos que han realizado alguna edición en el último mes); y, por último, de la profundidad de cada versión (este concepto, *depth* en inglés, no es de carácter científico o académico, sino que mide el volumen de trabajo y de usuarios que participan de promedio en la redacción de cada artículo).

Tabla 4. Ordenación de las quince Wikipedias por idioma según el volumen de artículos.
Fuente: elaboración propia a partir de datos publicados por Wikipedia (Colaboradores de Meta 2018) y Wikistasts (2018).

Idioma	Wiki ⁶	Nº artículos	Total	Nº ediciones	Adm.	Usuarios registrados	Usuarios activos	Depth
Inglés	en	5.688.386	45.442.770	847.471.568	1.214	34.116.007	120.106	909,23
Cebuano	ceb	5.382.623	8.964.564	23.134.840	5	49.648	159	1,14

⁵ En esta investigación, siguiendo el criterio de Wikipedia, emplearemos el concepto usuario solo para lo perfiles creados con intención de generar contenidos. No se considera por tanto usuario al internauta que es mero lector de la enciclopedia online.

⁶ Wiki: sigla del idioma de la enciclopedia. Nº artículos: se refiere a la cantidad de voces en la enciclopedia. Total: todas las páginas del wiki, incluidas las de discusión, redirecciones, páginas de usuario, etc. Nº ediciones: cantidad de ediciones que se han hecho desde que se abrió Wikipedia en ese idioma. Adm.: administradores por wiki, se trata de usuarios con facultades especiales, denominados actualmente bibliotecarios. Usuarios registrados: usuarios registrados en la wiki. Usuarios activos: usuarios registrados que han tenido actividad en los últimos treinta días. *Depth*: profundidad de las ediciones y trabajo por artículo.

Sueco	sv	3.781.861	7.689.862	43.241.891	63	622.761	2.178	5,97
Alemán	de	2.202.492	6.184.200	178.324.804	192	2.960.974	17.852	94,33
Francés	fr	2.004.397	9.585.607	150.247.889	162	3.166.224	16.163	223,85
Neerlandés	nl	1.937.216	3.985.182	51.846.705	44	931.667	3.448	14,52
Ruso	ru	1.485.829	5.745.679	93.909.450	86	2.338.805	10.097	134,06
Italiano	it	1.450.321	5.873.879	98.450.037	113	1.692.417	7.617	155,42
Español	es	1.439.925	6.405.030	109.104.513	71	5.080.407	14.722	203,82
Polaco	pl	1.291.743	2.883.221	53.827.880	102	899.948	3.871	28,29
Samarreño	war	1.263.166	2.876.343	6.191.458	2	37.511	72	3,51
Vietnamita	vi	1.183.514	13.499.971	41.272.242	23	611.728	1.638	329,40
Japonés	ja	1.113.862	3.300.469	69.159.679	42	1.356.733	13.100	80,65
Chino	zh	1.015.341	5.468.942	50.298.355	78	2.551.276	8.195	176,33
Portugués	pt	1.002.216	4.674.559	52.577.968	77	2.123.376	6.305	150,79

La tabla 4 nos muestra el tamaño de las versiones por idiomas y clasificadas según el número de artículos publicados en cada edición. Estas cifras revelan, como hemos observado previamente en todos los datos, el enorme impacto de la lengua inglesa que presenta el mayor número de artículos, de usuarios, de administradores y de profundidad en sus contenidos.

El segundo dato interesante es la aparición en la tabla de idiomas minoritarios como el cebuano –segundo puesto–, el samareño y, en menor medida, el vietnamita. Sin embargo, como se observa al contemplar el número de usuarios de las mismas; el total de administradores y de ediciones; y, sobre todo, la profundidad de los artículos podemos discernir que son el resultado de estrategias políticas y regionales que utilizan la traducción automatizada de textos directamente del inglés y no de una verdadera construcción colaborativa. No se trata, por tanto, de páginas realmente generadas sino de estrategias de traducción idiomática de poco peso, a pesar de su elevado número de artículos.

El tercer aspecto importante es que los grandes idiomas asiáticos y africanos han desaparecido de la lista. El árabe, el hindustaní, el indonesio/malayo y suajili no aparecen entre las versiones con más artículos. Así estos cuatro idiomas a pesar de ser hablados por cientos de millones de personas presentan un corpus mínimo de artículos en las versiones de Wikipedia, característica que coincide con su escaso impacto en el total de contenidos en Internet (tabla 2).

El cuarto dato importante, las lenguas europeas con más contenidos en la Red (ruso, francés, alemán, italiano e, incluso, portugués) sobresalen, también, como los idiomas con más artículos en la enciclopedia, aunque en ambas estadísticas se posicionan en distinto orden.

El último aspecto relevante es precisamente el que motiva esta investigación y la colaboración con Wikipedia en español y Wikimedia España⁷: la baja situación de la edición en este idioma. Como se observa, se sitúa en la novena posición algo que no se corresponde con su número de hablantes nativos (segunda) ni con el volumen de contenidos en internet (cuarta). Pero antes de abordar este punto clave, resulta interesante ofrecer una perspectiva más.

En la tabla 5 se muestra la clasificación de las versiones de la enciclopedia online con más volumen de contenidos y se compara con el número de administradores de cada una de ellas. En este caso se entiende por hablantes aquellos que pueden redactar de forma coherente y lúcida un artículo, se trata de personas con competencias lingüísticas muy elevadas (C2 en el sistema del Instituto Cervantes y en el modelo de la Universidad de Cambridge, Proficiency in English).

Tabla 5. Relación de artículos en cada Wikipedia según hablantes de la lengua y administradores de cada enciclopedia. Fuente: elaboración propia en base a datos obtenidos de Wikipedia (Colaboradores de Meta 2018) y Ethnologue Languages of the World (Simons y Fenning 2018).

Lengua	Artículos	% de Admin.	Artículos / 1000 hablantes nativos o C2
Inglés	5.684.414	1,01	11,30
Cebuano	5.382.624	3,29	340,50
Sueco	3.781.359	2,94	411,10
Alemán	2.200.902	1,07	28,10
Francés	2.002.859	0,98	26,70
Neerlandés	1.936.778	1,24	88,30
Ruso	1.484.533	0,84	8,90
Italiano	1.448.935	1,46	22,80
Español	1.435.907	0,48	3,10
Polaco	1.290.971	2,61	33,40
Samarreño	1.263.131	2,70	493,40

⁷ Wikipedia en español, como la mayoría de las otras versiones idiomáticas, tuvo que competir con otros modelos de enciclopedia virtual. El rival más destacado de esta Wikipedia fue la Enciclopedia Libre Universal en Español (Enciclopedia contributors. 2018), que surgió como una escisión de Wikipedia en español, y fue auspiciada por la Universidad de Sevilla (Tkacz 2011). Desde su creación en febrero de 2002 y durante su primer año de existencia, esta pudo competir en número de artículos con Wikipedia. Pero en la actualidad la diferencia en la cantidad de contenidos es enorme entre ambas, la sevillana tiene escasamente 51.100 artículos frente a los más de 1.400.000 de la Wikipedia en español. Así, como ocurre en otras lenguas, la creada por Wales y Sanger se considera, sin lugar a dudas, como la enciclopedia online más importante del castellano. Existe otra enciclopedia en español de cierta importancia que es EcuRed. Esta fue auspiciada por el Ministerio de Informática y comunicaciones de Cuba y, a pesar de contener más de 160.000 artículos, ha sido criticada en diversas ocasiones por su sesgo ideológico (Jones 2010).

Vietnamita	1.182.683	1,45	16,10
Japonés	1.113.240	0,32	8,90
Chino	1.014.675	0,96	0,80
Portugués	1.001.852	1,23	4,60

Como se observa el español y el chino son los idiomas que muestran el porcentaje más bajo de creación de artículos (entre las versiones más destacadas de Wikipedia) en relación a su número de hablantes. La cifra del castellano es bajísima: solo se escriben 3,1 artículos por cada mil personas con competencias lingüísticas elevadas o nativos. De hecho, idiomas como el inglés, le cuadriplican en producción, pero lenguas como el francés, el italiano y el alemán le superan de forma aplastante (26,7; 22,8; 28,7 y 33,4). Esta ratio tan baja de actividad nos indica el porqué, entre los idiomas con más contenidos en la Red, el español sufre el mayor retroceso en Wikipedia.

No menos importante es el porcentaje de administradores que tiene el español en Wikipedia, tan solo el 0,48 de cada mil hablantes (únicamente es inferior el caso de la versión japonesa 0,32). De nuevo la ratio es muy baja comparada con los demás idiomas: inglés (1,01), alemán (1,07), italiano (1,46). Los administradores de la enciclopedia, como se sabe, son usuarios voluntarios esenciales para el buen funcionamiento del sistema. Su trabajo es central para controlar los ataques, los vandalismos y, sobre todo, para subir la calidad de la versión.

Los datos de las tablas 3 y 4 resultan significativos y muy interesantes. Son en realidad el tema central de nuestra investigación, ¿por qué el español se encuentra en una posición tan baja en Wikipedia? ¿Por qué se genera un volumen tan escaso de artículos dentro de esta comunidad de hablantes?

Para acercarnos a esta problemática se deben revisar algunas de las cifras ofrecidas. La Wikipedia en español tiene una profundidad alta (tabla 4), lo cual demuestra que es un espacio activo y que sus artículos se construyen con cuidado, aunque no comparables a la versión en inglés. Pero, sobre todo, la versión en castellano presenta un elevado número de usuarios que al menos han colaborado alguna vez con la enciclopedia. Sin embargo, estos no han realizado nunca un artículo completo, sino solo pequeñas ediciones parciales. Tomando en cuenta al número de personas que han contribuido de algún modo en Wikipedia en español, esta se situaría como la segunda con más actividad. No obstante, este dato tan relevante se pierde al comprobar que los usuarios activos son muy escasos y el total de administradores mínimo en relación al número de hablantes.

Nos encontramos, por tanto, con las dos preguntas claves en esta investigación, aunque ahora podemos reformularlas con más precisión: ¿por qué Wikipedia en español contiene un número de artículos tan bajo a pesar del impacto de este idioma en la Red, al ser la segunda lengua con más hablantes nativos? Y, ¿por qué los usuarios de esta versión idiomática de Wikipedia dejan de ser activos con tanta frecuencia?

2.1. Sesgos intrínsecos internacionales y domésticos de Wikipedia en español

Wikipedia, como enciclopedia y obra de trabajo colectivo, se ha encontrado con varios problemas ideológicos y de sesgos desde su fundación en 2001. Ya en los encuentros Wikimanía de 2005 en Fráncfort, Alemania (Colaboradores de Wikimanía 2005) hasta la actualidad se ha debatido por superar los sesgos intrínsecos del proyecto. El análisis de estos es pertinente en nuestra investigación ya que explican el comportamiento y la actividad de los usuarios, así como parte de la profundidad y calidad de los artículos. Existen tres sesgos intrínsecos básicos de carácter internacional que se dan en todas las versiones lingüísticas con más usuarios y dos propios de la edición en español.

2.1.1. Sesgos intrínsecos internacionales

S.I.1. Urbano versus rural. Refleja la diferencia en el volumen total de artículos que versan sobre aspectos urbanos frente a los que tratan temas rurales. Este sesgo se percibe en todas las versiones idiomáticas, evidenciándose en las lenguas de uso mayoritario en Internet.

S.I.2. Sesgo occidental y oriental. Si bien las categorías occidente y oriente ya son en sí un estereotipo, la Wikipedia mantiene y define este sesgo como artículos que tratan contenidos relaciones con Europa, América y Oceanía, frente a los que tratan sobre los de los otros continentes. Así, las entradas sobre los tres primeros continentes son ampliamente mayoritarias en la enciclopedia. Este sesgo fue especialmente tratado en el encuentro Wikimanía 2018 en Suráfrica donde se planteó abiertamente que no sólo la enciclopedia online sino la propia Red sufría de un colonialismo, o neocolonialismo. (Colaboradores de Wikimanía 2018a, 2018b).

S.I.3. Carácter masculino. Wikipedia refleja con más frecuencia el trabajo y las biografías de los hombres que los de las mujeres. Este sesgo no solo se percibe en el número de entradas dedicadas a varones, sino que también se extiende a los artículos centrados en las películas, los libros, las obras artísticas o científicas, que mayoritariamente se centran en creaciones que han sido dirigidas, pintadas o compuestas por hombres⁸.

Además, Wikipedia en español plantea dos sesgos intrínsecos propios o domésticos:

S.P.1. Sesgo hispanoesfera. Esta versión detecta cómo los contenidos sobre temas o personas cuya lengua principal es el castellano son mayoritarios frente a los otros idiomas. Dicho comportamiento provoca que la enciclopedia esté sobredimensionada en biografías, temas y lugares del ámbito de América y España.

⁸ Así en nuestro estudio el sesgo masculino sigue las recomendaciones de las organizaciones feministas (AEIHM, ATGENDER y CIMA) y se han catalogan no solo biografías sino también las obras realizadas por hombres o mujeres. Es decir, una película como *El reino* (2018) se considera masculina pues su director y guionista es un hombre. Fuera de este sesgo se encuentran los conceptos y muchos temas de carácter neutro como topónimos y definiciones.

S.P.2. Sesgo deportivo y, especialmente, futbolístico. Desde 2010 se detecta que uno de los temas principalmente tratados en la enciclopedia en español es el deporte. De hecho, una investigación publicada en 2013 sobre los artículos más debatidos en las distintas versiones de Wikipedia, mostró cómo la edición en castellano era la única que incluía temas de deportes entre las diez entradas más polémicas. No solo presentaba artículos sobre este aspecto, sino que el debate en torno al fútbol representaba el cincuenta por cien de los más discutidos (Yasseri et al. 2013).

3. Metodología: corpus y objetivos

Como hemos observado esta investigación se cuestionó el porqué Wikipedia en español tenía un número tan bajo de artículos y cuál era la razón de que los usuarios activos y administradores fuesen en esta versión considerablemente menos que en las ediciones más amplias de la enciclopedia.

Huelga decir que plantear una investigación de estas características en un ámbito de más de quinientos millones de hablantes nativos o con competencias elevadas y un entorno geográfico y político de más de treinta países con uso prioritario de esta lengua, se trata de un trabajo que requiere varias fases y apoyos. Por ese motivo, planteamos este primer estudio, modesto, dentro de un espacio más fácilmente mensurable. Para ello se contó con el apoyo de docentes de tres universidades españolas: Universidad Complutense de Madrid (UCM), Universidad Rey Juan Carlos (URJC) y Universidad Antonio de Nebrija (Nebrija) y miembros de Wikimedia España. Por ese motivo, se optó en esta primera investigación, en estudiar en dos de las tres universidades citadas, dentro del ámbito geográfico de Madrid, España. Se espera que este estudio abra la posibilidad de solicitar una ayuda *Research Project Grant* de la Fundación Wikimedia para poder desarrollar una investigación que recoja al menos diez de los países centrales de habla hispana.

Para realizar estos experimentos se planteó junto a Wikimedia España un sondeo, basado en un taller de redacción de artículos en la enciclopedia online para estudiantes universitarios. Se seleccionaron tres asignaturas de dos universidades españolas: Rey Juan Carlos, de gestión pública, y Antonio de Nebrija, de gestión privada. En donde se formó a los docentes y a los discentes para crear un usuario y componer un artículo completo sobre cualquier aspecto no incluido en la Wikipedia en español o, en su defecto, la mejora sustancial, de más de las tres cuartas partes, de un artículo ya existente pero no desarrollado con profundidad.

Se optó por estas dos universidades porque al encontrarse en las cercanías de Madrid los estudiantes podrían trasladarse a charlas y encuentros de talleres impartidos por los diferentes grupos afines a la Fundación Wikimedia que operan en Madrid. Además, se buscó que las universidades representaran tanto el carácter público –Universidad Rey Juan Carlos– como el privado –Universidad Antonio de Nebrija–. Por último, los alumnos escogidos provenían de las disciplinas del área de las Ciencias Sociales, la Comunicación y el Periodismo.

Para calibrar todas las variables posibles se propuso realizar la mitad del experimento con carácter voluntario y la otra mitad como actividad reglada dentro de las distintas materias.

Además, a todos ellos se les envió una encuesta con carácter optativo para medir su satisfacción con el trabajo y su interés en Wikipedia.

Tabla 6. Datos del experimento. Fuente: elaboración propia.

Número de alumnos que realizaron el experimento	231
Alumnos con carácter voluntario	111
Alumnos con actividad reglada	120
Total de usuarios registrados para la actividad	140
Usuarios registrados (hombres)	41
Usuarías registradas (mujeres)	99
Lugar	- Universidad Rey Juan Carlos - Universidad Antonio de Nebrija
Período del experimento	Febrero marzo y abril de 2018
Fecha talleres	- Nebrija: 22/2/2018 - 08/03/2018 - 05/04/2018 - URJC: 16/3/2018 - 06/04/2018 - 20/04/2018- - URJC optativos: 13/3/2018 - 03/04/2018 -10/04/2018
Profesorado	Universidad Rey Juan Carlos, Universidad Antonio de Nebrija, Universidad Complutense de Madrid
Fecha de revisión de artículos	mayo y junio de 2018
Encuestas contestadas	29

Para que el experimento fuese eficaz se formó a todos los discentes con tres sesiones sobre creación de usuarios y la realización de talleres de creación. Además, se ofrecieron otros talleres públicos promovidos por Wikimedia España en la Embajada de Suecia en Madrid y en la Universidad Complutense.

3.1. Relación de objetivos del experimento

A) Mensurar la capacidad de la creación de contenidos para Wikipedia en el ámbito universitario.

B) Monitorizar la actividad y comprobar la asiduidad de colaboración de los usuarios creados, es decir, si se mantuvieron activos una vez concluidos los artículos o si abandonaron la creación de contenidos.

C) Comprobar la existencia de los sesgos intrínsecos de Wikipedia y su importancia en la redacción de contenidos en la versión en español.

D) Valorar la calidad del contenido de los artículos y el rechazo o la aprobación de los mismos según los administradores y la comunidad de la enciclopedia online.

E) Contabilizar el número de ediciones, versiones, cambios e intercambios creados por los usuarios y su interrelación con la comunidad.

F) Categorizar los contenidos creados según los temas y aspectos tratados en cada artículo.

G) Cuantificar las aportaciones visuales o infográficas realizadas por los usuarios.

4. Resultados y discusión

Para mostrar los resultados del experimento seguimos la estrategia propuesta de división de la muestra en tres grupos: el primero realizado en la URJC y con carácter de tarea reglada, el segundo planteado en la misma universidad pública y con carácter optativo y el tercero desarrollado en la Nebrija y actividad reglada. Se ofreció a cien alumnos en el primer caso, 111 en el segundo y a veinte en el tercero. Lógicamente la creación de usuarios fue dispar, aquellos que participaron como tarea reglada concluyeron el artículo en un 89% en la URJC y la totalidad en la Nebrija. Sin embargo, cuando la creación del texto era una opción voluntaria solo la concluyeron el 28,18%. Este dato nos ofrece ya un resultado relevante y es el poco interés que despertó entre el alumno de carácter voluntario. La enciclopedia online se observa como algo complejo y alejado y que requiere un gran esfuerzo, comentarios como estos se escucharon en los talleres y quedaron reflejados en el *verbatimum* final de la encuesta. En la tabla 7 presentamos los datos básicos de la muestra y el cómputo de los artículos realizados:

Tabla 7. Datos básicos de la muestra y los artículos realizados por grupos y totales. Fuente: elaboración propia en base a los datos del experimento.

	Pública actividad reglada		Pública optativa		Privada reglada		Totales	
	Nº	%	Nº	%	Nº	%	Nº	%
Alumnos susceptibles de hacer la actividad	100		111		20		231	
Usuarios totales registrados en Wikipedia	89	89%	31	28,18%	20	100%	140	
Artículos publicados en Wikipedia	89		31		20		140	
Artículos rechazados	13	14,61%	8	25,81%	1	5,00%	22	15,71%
Artículos validados por la comunidad (2 meses)	76	85,39%	23	73,03%	19	95,00%	118	84,29%
Ediciones en taller promedio por alumno	8		10		5		7,63	
Total de ediciones en taller	677		314		103		1094	
Promedio de ediciones sobre el artículo, por alumno, una vez publicado	1,38		1,29		3,85		2,17	
Total de ediciones, por los alumnos, sobre los artículos después de publicados	123		40		77		240	
Artículos sin imagen	65	73,03%	18	58,06%	12	60,00%	95	67,86%
Artículos con imagen	24	26,97%	13	41,94%	8	40,00%	45	32,14%
Artículos con imágenes tomadas de WMC	21	23,60%	9	29,03%	6	30,00%	36	25,71%

Artículos con imágenes originales aportadas por los alumnos	3	3,37%	4	12,90%	2	10,00%	9	6,43%
Usuarios activos después de 1 mes de la actividad	2	2,25%	0	0,00%	1	5,00%	3	2,14%

Como se observa en la tabla 7, se lograron publicar un total de 140 artículos nuevos (o con cambios superior al 75% de la versión previa existente). De estos solo veintidós (15,71%) fueron rechazados por los administradores y la comunidad de usuarios de la Wikipedia, que consideraron que los textos propuestos carecían del nivel suficiente o no resultaban de carácter enciclopédico. Atendiendo a este resultado, se trata de un éxito importante ya que 118 artículos se encuentran disponibles en la Red y verificados. Estas nuevas entradas se han mantenido tras la publicación en el mes de abril de 2018 y han superado los dos meses de prueba.

Un dato interesante de este trabajo es que los artículos fueron elaborados en la mayoría de los casos en talleres online antes de ser publicados de forma definitiva. Esto permitió que los discentes pudieran mejorar y reeditar sus textos antes de la redacción publicable. Cada texto final tuvo de media más de 7,5 ediciones. Lo que implica que los nuevos artículos nacían ya con cierta profundidad.

Sin embargo, el experimento mostró una carencia en el uso de imágenes y/o infografías. Sesenta y cinco artículos (73,03%) carecían de cualquier soporte visual. Aunque las fotografías no garantizan la mayor calidad de una entrada sí miden la implicación y el cuidado en la misma. Además, de la baja inclusión de imágenes y/o infografías, se observa que cuando se usan estas, casi en su totalidad eran enlazadas desde WMC (Wikimedia Commons: repositorio audiovisual que provee de material a los proyectos de la fundación, entre otros a todas las Wikipedias) y muy pocas veces eran obras inéditas subidas ex profeso para el proyecto enciclopédico.

Por último, un aspecto clave del experimento era saber si los usuarios, una vez formados, mantendrían su actividad. El resultado es muy claro ya que tras dos meses únicamente continuaban en activo tres alumnos lo que supone tan solo el 2,14% de la muestra seleccionada. Esto confirmada claramente el comportamiento habitual de los internautas de habla hispana en Wikipedia que, si bien crean usuarios y participan en la redacción de la enciclopedia, su continuidad y sus colaboraciones reales son muy breves en el tiempo y escasas en la generación de número de artículos nuevos.

Para entender el comportamiento de los usuarios y, sobre todo, para analizar su desafección con la Wikipedia, se realizó una encuesta a los estudiantes. Esta se propuso dos meses después de terminadas las actividades y en ella se medía su interés como lectores y también como usuarios de la enciclopedia online (Tabla 8). Además, en esta encuesta, se les ofrecieron diversas preguntas abiertas para que pudieran responder con libertad. Las respuestas nos generaron un *verbatimum* con algunos comentarios muy relevantes.

Todos los alumnos reconocían que eran lectores de Wikipedia y más del sesenta por cien la consultaban diaria o semanalmente. Aún más interesante es que ninguno indicó que no fuese lector de la misma. Del mismo modo casi el noventa por cien entendía que la enciclopedia online era muy útil o práctica para sus vidas, es decir, un porcentaje ampliamente mayoritario reconocía su uso y señalaba su relevancia.

Tabla 8. Resultados de las encuestas realizadas pasados 2 meses de terminada la actividad con los alumnos. Fuente: elaboración propia en base a los datos obtenidos de la encuesta.

Número total de encuestas contestadas	29				
Sexo	Hombre	Mujer			
	37,93%	62,07%			
1. ¿Con qué frecuencia utilizas Wikipedia? Hablamos de lectura/consulta:	A diario	Cada dos o tres días	Cada semana	Sólo ocasionalmente	Nunca
	34,48%	27,59%	13,79%	24,14%	0,00%
2. ¿Es útil y práctica Wikipedia?	Muy Poco	Poco	Normal	Algo	Mucho
	0,00%	0,00%	10,34%	27,59%	62,07%
3. Hay pocos artículos en Wikipedia en español porque el entorno para nuevos editores es hostil:	Nada Hostil	Un poco hostil	Normal	Hostil	Muy hostil
	3,45%	34,48%	44,83%	17,24%	0,00%
4. Hay pocos artículos en Wikipedia en español porque la interfaz para editores no es amigable:	Nada Amigable	Algo amigable	Normal	Amigable	Muy amigable
	6,90%	20,69%	41,38%	24,14%	6,90%
5. Hay pocos artículos en Wikipedia en español por falta de interés y desidia:	Poco de acuerdo	Algo de acuerdo	Normal	De acuerdo	Muy de acuerdo
	10,34%	10,34%	31,03%	20,69%	27,59%
6. Hay pocos artículos en Wikipedia en español por desconocimiento para editar:	Poco de acuerdo	Algo de acuerdo	Normal	De acuerdo	Muy de acuerdo
	3,45%	10,34%	17,24%	27,59%	41,38%
7. Hay pocos artículos en Wikipedia en español porque en otras Wikipedias hay más campañas para creación de artículos	Poco de acuerdo	Algo de acuerdo	Normal	De acuerdo	Muy de acuerdo
	10,34%	10,34%	27,59%	44,83%	6,90%

Sin embargo, cuando se les preguntó a los estudiantes por su relación como usuarios y redactores en la Wikipedia las respuestas fueron menos entusiastas. Aunque un porcentaje elevado (38 %) entendía que en el entorno de creación no era hostil, otros (17%) percibía que sí lo era. Además, la mayoría no sabía decantarse por si el entorno era o no hostil con los nuevos usuarios. Aún más, cuando se les preguntó sobre si la enciclopedia online resultaba amigable en su interfaz y modo de redacción, los discentes que indicaban que era poco o muy poco amigable alcanzaban casi el 30 % de la encuesta.

En el *verbatimum* se reflejó con claridad la desafección con la Wikipedia como usuarios y redactores por la mayoría de los estudiantes, que escribieron y comentaron cosas como: “Porque no se me ha ocurrido de que/quien hacerlo” (E4) y “No he encontrado una temática que me empujase a ello” (E16). Pero a la vez muchos de los alumnos reconocían que si bien no colaboraban, sí observaban que la fiabilidad y el valor de la enciclopedia online: “nunca se me ocurre, pero sé que es útil hacerlo” (E23) y “No me interesa escribir ni contribuir en ella aunque si utilizo Wikipedia” (E12).

Solo un porcentaje bajo expresaba claramente su intención de ser un usuario activo y continuado o, al menos, de colaborar de nuevo en el futuro: “No he tenido mucho tiempo para hacerlo, pero me gustaría hacer alguno más. El que hice lo enseñé de forma orgullosa, me parece muy interesante que algo que he escrito yo esté ahí como primer resultado de búsqueda” (E24) y “Un mundo apasionante, que en mi caso siempre quise descubrir y que pienso que tiene mucho que ver en el periodismo y la comunicación multimedia. El único problema que tiene es que la sociedad aún no lo concibe como una nueva forma de comunicar, de transmitir conocimiento de manera honesta y fiable” (E2).

4.1. Resultado y discusión de los sesgos

En nuestra investigación nos propusimos medir no solo el comportamiento de los nuevos usuarios en Wikipedia en español, sino también ver si la construcción de nuevos artículos repetía los sesgos intrínsecos internacionales y domésticos de la enciclopedia online en español.

Para ello analizamos todos los trabajos realizados, aceptados o no por la comunidad, y los clasificamos según sus contenidos y sus temas para poder someterlos a los cinco sesgos intrínsecos. La mayoría de ellos fueron fácilmente clasificables, aunque se dieron seis ejemplos de artículos que no se podían analizar en todos los sesgos al tratarse de conceptos o definiciones. En la Tabla 9 se ofrecen los resultados del experimento.

Tabla 9. Sesgos intrínsecos presentes en el experimento realizado. Fuente: elaboración propia partir de los datos recabados en el experimento.

	U. pública actividad reglada		U. pública optativa		U. privada reglada		Totales	
	Nº	%	Nº	%	Nº	%	Nº	%
Alumnos susceptibles de hacer la actividad	100		111		20		231	
Usuarios totales registrados en Wikipedia	89		31		20		140	
Artículos publicados en Wikipedia	89		31		20		140	
Artículos rechazados	13	14,61%	8	25,81%	1	5,00%	22	15,71%
Artículos validados por la comunidad (2 meses)	76	85,39%	23	73,03%	19	95,00%	118	84,29%
Sesgos								
Tema hombre	60	67,42%	11	35,48%	8	40,00%	79	56,43%
Tema mujer	27	30,34%	6	19,35%	0	0,00%	33	23,57%
Tema neutro	2	2,25%	14	45,16%	12	60,00%	28	20,00%
Urbano	85	95,51%	25	80,65%	16	80,00%	126	90,00%
Rural	4	4,49%	4	12,90%	4	20,00%	12	8,57%
No clasificable en Urbano/Rural	0	0,00%	2	6,45%	0	0,00%	2	1,43%
Oriente	5	4,49%	0	0,00%	0	0,00%	5	3,57%
Occidente	84	94,38%	29	93,55%	19	95,00%	132	94,29%
No clasificable en Oriente/Occidente	0	0,00%	2	6,45%	1	5,00%	3	2,14%
Hispanosfera	48	53,93%	24	77,42%	7	35,00%	79	56,43%

No Hispanosfera	41	46,07%	5	16,13%	12	60,00%	58	41,43%
No clasificable	0	0,00%	2	6,45%	1	5,00%	3	2,14%
Deportes	3	3,37%	4	12,90%	0	0,00%	7	5,00%
No deporte	86	96,63%	27	87,10%	20	100,00%	133	95,00%

Al analizar los sesgos internacionales se observa que la elección entre temas occidentales (América, Europa y Oceanía) frente a otras propuestas, se decantan mayoritariamente por la primero. Así, 132 artículos (94,29%) se inspiraban en aspectos relacionados con occidente. Este sesgo (S.I.2), que como se indicó contiene una connotación racista en su propia definición, se confirma claramente en el experimento. Y, además, fue uno de los temas centrales en los encuentros pre conferencia y conferencia de Wikimanía los días 18, 19 y 20 de julio de 2018 en Ciudad del Cabo, Suráfrica, en los encuentros denominados *Descolonizando Internet* (Colaboradores de Wikimanía 2018a, 2018b).

El sesgo (S.I.1), la variable urbano/rural, reproduce la carencia habitual de la enciclopedia en este debate, es decir, el noventa por cien de los estudiantes decidió escribir sobre temas y actividades que se enmarcan dentro de lo urbano. Lo rural quedaba muy relegado y es casi inexistente. Es cierto que el taller se realizó en centros urbanos en las proximidades de una gran ciudad. También en las conferencias de Wikimnia de 2018 se trató esta deficiencia en las propuestas denominadas *Documenting rural áreas by WikiTakes activities* (Colaboradores de Wikimanía 2018c).

El tercer sesgo intrínseco (S.I.3) de carácter internacional es el que indaga sobre la presencia de la mujer en la enciclopedia. La Wikipedia ha observado cómo los artículos sobre hombres, actividades masculinas o temas relaciones con los varones superan en creces a los de temática femenina o relacionados con mujeres. Nuestro estudio muestra cómo los nuevos estudiantes también reproducen este patrón. Aunque la mayoría de los discentes eran mujeres y durante marzo de 2018 hubo una importante manifestación feminista en España (Seis millones de trabajadores secundan la huelga feminista 2018) lo cierto es que la mayoría de los artículos son sobre temas asociados a varones.

Ante el sesgo intrínseco de género, en España se vienen realizando diversas actividades interesantes. La primera es la creación de varios grupos de trabajo de temática feminista como Wikisfera, Wikimujeres, Cuarto propio y Wikimedia España, y la segunda es la formación y realización de talleres ex profeso para generar contenidos sobre mujeres clave en la historia, ciencia o sociedad.

Al analizar los sesgos intrínsecos propios o domésticos de Wikipedia en español se percibe que efectivamente los temas de las hispanoesferas (S.P.1) son mayoritarios (56,43%) pero no llegan a ser tan problemáticos como la dicotomía urbano/rural u occidental/oriental. Gran parte de este resultado se debe a que la Wikipedia en español padece una fuerte influencia de la versión inglesa. La relación con la misma implica que los temas estadounidenses y de las comunidades inglesas sean muy frecuentes y moderen el impacto de la hispanoesfera.

Por último, el sesgo intrínseco propio de la sobre abundancia de los temas deportivos (S.P.2) no parece confirmarse en nuestra investigación. Es cierto que se escriben un cinco por cien de artículos referidos a este concepto, pero es un porcentaje relativamente bajo para ser considerando un sesgo fundamental o demérito estructural de la enciclopedia. Además, el tema futbolístico es tratado

directamente en solo un artículo. Aunque estos datos se pueden deber a que la investigación se realizó en un ámbito universitario, su porcentaje se aleja completamente de lo planteado por otros estudios (Yasseri et al. 2013, 25-49). Por ello, de los cinco sesgos estudiados es el único que no se confirma.

5. Conclusiones

Nuestra investigación nace como un primer paso para entender el porqué Wikipedia en español no alcanza el volumen de artículos y de usuarios activos que le correspondería a dicho idioma. Es, como se plantea, un estudio germinal ya que la lengua española tiene más de quinientos millones de hablantes nativos o con competencias desarrolladas suficientes para escribir en la enciclopedia online y, además, esta se consulta de forma continuada en más de treinta estados del mundo. Por eso, nuestro trabajo se ofrece a la comunidad como un paso inicial y modesto al que se espera se puedan sumar otras investigaciones. Las conclusiones de nuestro análisis se enmarcan dentro de la comunidad de wikipedistas de España, donde se encuentra un gran número de enciclopedistas online pero menos el diez por ciento de los hablantes de esta lengua. Se contempla continuar este estudio con una *Research Project Grant* de Wikimedia Foundation, pero también se ofrecerán las conclusiones (y los resultados en abierto) para que otros investigadores puedan realizar indagaciones análogas en distintos países de la hispanosfera.

Nuestro estudio evidencia que la comunidad hispana, de hablantes nativos o con competencia desarrolladas en español, no produce el mismo volumen de artículos y contenidos relevantes en Wikipedia que las otras lenguas de su entorno: inglés, francés, alemán e italiano. El porcentaje y la actividad por usuario creado es menor que en todos los idiomas citados y de otros como el portugués, el japonés. Para entender estos datos hemos investigado el comportamiento de nuevos usuarios en español en el entorno de Madrid. Nuestros resultados nos indican que los sujetos de estudio cumplen algunas características comunes.

El usuario medio del experimento, nivel educativo universitario no diplomado, posee las herramientas básicas para elaborar artículos relevantes y de calidad. Tras el experimento se constata que casi el 85% de las entradas fueron aprobadas por la comunidad y por los administradores reconocidos de la Wikipedia. Se trata de un porcentaje elevado que nos permite indicar que con un nivel educativo universitario es suficiente para elaborar artículos solventes.

El experimento muestra como la tasa de participación en Wikipedia de forma voluntaria es muy baja. Entre el grupo de control que se ofreció la propuesta como una actividad no reglada la mayoría del alumnado no participó, aunque tuvieran los recursos y los conocimientos para hacerlo. Un dato positivo, es que sin embargo un elevado número de estudiantes sí reconoce el valor de la enciclopedia y la consultan a diario o semanalmente como lectores. Además, muchos estudiantes recurrieron al uso de talleres online, participaron en debates y acudieron a charlas y encuentros presenciales sobre la enciclopedia virtual.

Las entradas escritas por los alumnos del experimento, desgraciadamente, cumplían en conjunto los tres sesgos intrínsecos internacionales de la Wikipedia y uno de los sesgos intrínsecos propio de la versión en español. Es decir, se observa

como la mayoría de los artículos trataban temas urbanos, masculinos (o de creaciones realizadas por hombres), occidentales y situados en la hispanosfera. Sin embargo, no se comprobó el sesgo deportivo, ya que los artículos dedicados a este tema solo representaban el cinco por cien de la producción total.

El dato más decepcionante del experimento es que de 231 alumnos que fueron sujetos de la muestra solo tres continuaron siendo usuarios activos de la Wikipedia tras el período de observación y control. Este dato tan bajo, confirma el desinterés que se observa en la comunidad hispana en la creación de contenidos originales en la enciclopedia online.

Sin embargo, un resultado valioso del experimento es que todos los usuarios reconocieron la valía cultural de la Wikipedia y el prestigio social de participar en la elaboración de creación de contenidos de forma colaborativa. Como confirmó una estudiante tras el cierre del experimento: “El [artículo de Wikipedia] que hice lo enseñó de forma orgullosa, me parece muy interesante que algo que he escrito yo esté ahí como primer resultado de búsqueda”.

6. Referencias bibliográficas

- Alexa (2018). *The top 500 sites on the web*. Alexa International Inc.
<https://www.alexa.com/topsites>
- Alvar, Manuel (1961). “Hacia los conceptos de lengua, dialecto y hablas”. *Nueva Revista De Filología Hispánica* 15, 51-60.
- Clark, Justin; Faris, Robert; and Jones, Rebekah Heacock (2017). Analyzing Accessibility of Wikipedia Projects Around the World. Berkman Klein Center Research Publication Series. <https://ssrn.com/abstract=2951312>
- Colaboradores de Meta (2018). List of Wikipedias by speakers per article. *Meta*
https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias_by_speakers_per_article&oldid=18152604
- Colaboradores de Wikimanía (2005). “Main Page”. *Wikimania 2005*.
https://wikimania2005.wikimedia.org/w/index.php?title=Main_Page&oldid=8281
- Colaboradores de Wikimanía (2018a). “Preconference”. *Wikimania 2018*.
<https://wikimania2018.wikimedia.org/w/index.php?title=Preconference&oldid=20322>
- Colaboradores de Wikimanía (2018b). “Program/Centering Knowledge from the Margins: A Whose Knowledge?”. *Wikimania 2018*.
https://wikimania2018.wikimedia.org/w/index.php?title=Program/Centering_Knowledge_from_the_Margins:_A_Whose_Knowledge%3F_discussion&oldid=21067
- Colaboradores de Wikimanía (2018c). “Program/Documenting rural areas by WikiTakes activities”. *Wikimania 2018*.
https://wikimania2018.wikimedia.org/w/index.php?title=Program/Documenting_rural_areas_by_WikiTakes_activities&oldid=21115
- Colaboradores de Wikipedia (2018). “Anexo:Wikipedias”. Wikipedia, La enciclopedia libre, 2018.
<https://es.wikipedia.org/w/index.php?title=Anexo:Wikipedias&oldid=109596488>
- Comrie, Bernard, (ed., 2009). *The World's Major Languages*. New York: Routledge.

- Council For Cultural Cooperation Education Committee (2002). Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Madrid: Secretaría General Técnica del MECED, Subdirección General de Información y Publicaciones.
- Enciclopedia Contributors (2018). Enciclopedia Libre Universal en Español. http://enciclopedia.us.es/index.php/Enciclopedia:Sobre_la_Enciclopedia_Libre
- Fundación Princesa de Asturias (2015). Wikipedia Premio Princesa de Asturias de Cooperación Internacional 2015. <http://www.fpa.es/es/premios-princesa-de-asturias/premiados/2015-wikipedia.html?texto=acta&especifica=0>
- Garayzábal, Elena y Zhang, Jianfeng (2006). “China y España, lengua y dialecto: mismos términos, significados diferentes”. *Revista de investigación lingüística* 9, 233-249
- Han, Byung Chul (2013). *La sociedad de la transparencia*. Barcelona, Herder Editorial S.L.
- Instituto Cervantes (2017). El español: una lengua viva. Informe 2017. https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2017.pdf
- Internet World Stats (2108). Internet world users by language. <https://www.internetworldstats.com/stats7.htm>
- Ippolita (2012). *El acuario de Facebook*. Madrid, Enclave de libros.
- Jones, Sam (2010). “Cuba opens new online frontline in war of words: Castropedia”. *The Guardian*, 15 de diciembre, <https://www.theguardian.com/world/2010/dec/15/cuba-usa>
- Koch, Tommaso (2018). Wikipedia cierra en protesta por la reforma de la ley de derechos de autor en la Unión Europea. *El País*. 7 de abril. https://elpais.com/cultura/2018/07/04/actualidad/1530694176_811190.html
- La Vanguardia* (2018). “Seis millones de trabajadores secundan la huelga feminista, según los sindicatos”. Barcelona, 9 de marzo. <https://www.lavanguardia.com/economia/20180308/441354667931/dia-mujer-8m-huelga-feminista-manifestaciones.html>
- Simons, Gary F. and Fenning, Charles D. (2018). *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas, SIL International. <http://www.ethnologue.com>
- Tkacz, Nathaniel (2011). The Spanish Fork: Wikipedia's ad-fuelled mutiny. *Wired*. <https://www.wired.co.uk/article/wikipedia-spanish-fork>
- United Nations, Department of Economic and Social Affairs, Population Division (2017a). World Population Prospects: The 2017 Revision, World Population 2017 Wallchart. ST/ESA/SER.A/398. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_Wallchart.pdf
- United Nations, Department of Economic and Social Affairs, Population Division (2017b). World Population Prospects: The 2017 Revision, Methodology of the United Nations Population Estimates and Projections, Working Paper No. ESA/P/WP.250. New York: United Nations. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_Methodology.pdf
- Web Technology Surveys (2018). Historical yearly trends in the usage of content languages for websites. https://w3techs.com/technologies/history_overview/content_language/ms/y
- Wikistats (2018). List of MediaWikis. <http://wikistats.wmflabs.org/>
- Yasseri, Taha; Spoerri, Anselm; Graham, Mark; and Kertész, János (2013). “The most controversial topics in Wikipedia: A multilingual and geographical analysis”. En:

Global Wikipedia: International and cross-cultural issues in online collaboration. Editado por Fichman, Pnina y Hara, Noriko, 25-49. Maryland, Rowmand & Littlefield. Zafra, Remedios (2017). *El entusiasmo*. Barcelona, Anagrama.

7. Agradecimientos y ayudas

Esta investigación ha sido posible gracias a la ayuda de un proyecto de investigación *PERLAD, Pensamiento y representación literaria y artística digital ante la crisis de Europa y el Mediterráneo* y la ayuda de Wikimedia España. Sin el apoyo de estas instituciones hubiera sido imposible la realización del trabajo. Agradecemos también el necesario apoyo de la Universidad Antonio de Nebrija, de la Universidad Rey Juan Carlos y de la Universidad Complutense de Madrid.

Florencia Claes es Profesora asociada en la Universidad Rey Juan Carlos de Madrid con varias publicaciones en el ámbito de las redes sociales y comunicación digital. Ha impartido clases en la Universidad Complutense y en la Antonio de Nebrija. Su campo de investigación son las redes sociales, los líderes de opinión online y las comunidades virtuales. Desde hace 8 años es colaboradora activa en los proyectos de la Fundación Wikimedia.

Luis Deltell es Profesor contratado doctor de la Universidad Complutense de Madrid. Investigador Principal del Proyecto de Investigación UCM-Santander. Referencia: PR26/16-6B-1, titulado: *PERLAD, Pensamiento y representación literaria y artística digital ante la crisis de Europa y el Mediterráneo* y codirector del grupo complutense de investigación ESCINE. Su campo de investigación es la historia de la imagen, el cine y el análisis de las nuevas tecnologías de las ciencias de la información. Ha sido visiting scholar en Stanford University, Berkeley University of California y profesor invitado en una veintena de universidades europeas y en Cairo University. Además, es director de cortometrajes y documentales.