

Trabajos, Comunicaciones y Conferencias

**Actas del Workshop Iberoamericano de
Estudios métricos de la actividad científica
orientada a temas locales/regionales**

Sandra Miguel
(coordinadora)

FaHCE
FACULTAD DE HUMANIDADES Y
CIENCIAS DE LA EDUCACIÓN



UNIVERSIDAD
NACIONAL
DE LA PLATA

**ACTAS DEL WORKSHOP
IBEROAMERICANO DE ESTUDIOS
MÉTRICOS DE LA ACTIVIDAD
CIENTÍFICA ORIENTADA A TEMAS
LOCALES/REGIONALES**
La Plata, 21 y 22 de agosto de 2018

Sandra Miguel
(coordinadora)

Diseño: D.C.V. Celeste Marzetti

Tapa: D.G.P. Daniela Nuesch

Editora por Prosecretaría de Gestión Editorial y Difusión: Leslie Bava

Queda hecho el depósito que marca la Ley 11.723

©2019 Universidad Nacional de La Plata

ISBN: 978-950-34-1742-3

Colección: Trabajos, comunicaciones y conferencias, 37

Cita sugerida: Miguel, S. (Coord.). (2019). *Actas del Workshop Iberoamericano de estudios métricos de la actividad científica orientada a temas locales/regionales* (2018 : La Plata). La Plata : Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. (Trabajos, comunicaciones y conferencias ; 37). Recuperado de <https://www.libros.fahce.unlp.edu.ar/index.php/libros/catalog/book/130>



Licencia Creative Commons 4.0.

Universidad Nacional de La Plata
Facultad de Humanidades y Ciencias de la Educación

Decana

Dra. Ana Julia Ramírez

Vicedecano

Dr. Mauricio Chama

Secretario de Asuntos Académicos

Prof. Hernán Sorgentini

Secretario de Posgrado

Dr. Fabio Espósito

Secretaria de Investigación

Prof. Laura Rovelli

Secretario de Extensión Universitaria

Dr. Jerónimo Pinedo

Prosecretario de Gestión Editorial y Difusión

Dr. Guillermo Banzato

Índice

Prólogo	9
EJE TEMÁTICO I: Presencia de temas orientados a lo local / regional en las políticas y agendas de investigación	11
Política científica y relevancia social de la investigación	
<i>Federico Vásen</i>	13
La investigación en áreas prioritarias y en temas locales/regionales. Presencia en las políticas y planes de los países latinoamericanos	
<i>Victoria Ugartemendía</i>	21
¿Responde la investigación a las necesidades de salud?	
<i>Ismael Rafols y Alfredo Yegros</i>	29
EJE TEMÁTICO II: Aproximaciones metodológicas para el abordaje cuantitativo de la producción en temas locales/regionales	37
Aspectos metodológicos para la construcción de categorías en temas específicos. El caso de la Nanociencia y la Nanotecnología	
<i>Zaida Chinchilla Rodríguez, Teresa Muñoz Ecija y Benjamín Vargas Quesada</i>	39
La recuperación de información por delimitadores geográficos y su aplicación en estudios bibliométricos de ciencia local	
<i>Claudia M. González, Gustavo Archuby y Sandra Miguel</i>	47

<u>El análisis y representación del contenido de la producción científica desde una perspectiva informétrica: aportes metodológicos</u>	
<i>Gustavo Liberatore</i>	55
<u>Aproximación metodológica para la extracción de temas de un corpus bibliográfico referencial a partir del lenguaje natural</u>	
<i>Sebastián Varela y Claudia M. González</i>	63
<u>EJE TEMÁTICO III: Estudios métricos sobre la producción científica en temas locales de países iberoamericanos.</u>	71
<u>Encontrar los temas locales en el CV de los investigadores uruguayos del área social</u>	
<i>Natalia Aguirre-Ligüera y Exequiel Fontans</i>	73
<u>Sesenta años de producción científica sobre Uruguay en la WOS: 1958-2017</u>	
<i>Exequiel Fontans y Natalia Aguirre-Ligüera</i>	83
<u>Argentina como tema o alcance geográfico de la investigación. Una mirada desde SciELO y Scopus</u>	
<i>Mónica Hidalgo, Lorena Caprile, Israel Jorquera Vidal y Sandra Miguel</i>	91
<u>Impacto de la investigación local mediante Altmetrics. El sector del vino en España</u>	
<i>Enrique Orduña Malea, Cristina Font y Adolfo Alonso-Arroyo</i>	99
<u>Indicadores bibliométricos de la producción científica sobre países latinoamericanos en perspectiva comparada</u>	
<i>Sandra Miguel, Claudia M. González y Claudia Boeris</i>	107
<u>Exploración de relaciones entre indicadores bibliométricos y otros indicadores del contexto económico, social y productivo</u>	
<i>Edgardo Ortiz Jaureguizar</i>	117

Prólogo

Este libro de actas reúne las intervenciones presentadas en el **Workshop Iberoamericano de estudios métricos de la actividad científica orientada a temas locales/regionales**, realizado en la ciudad de La Plata, Argentina, el 21 y 22 de agosto de 2018.

El evento estuvo organizado por el Instituto de Investigaciones en Humanidades y Ciencias Sociales (IdIHCS- UNLP/CONICET) en el marco del proyecto PICT 2015-2144 “La producción científica sobre los países latinoamericanos. Aproximación a su estudio desde una perspectiva bibliométrica y relación con indicadores del contexto económico y social”, acreditado por la Agencia Nacional de Promoción Científica y Tecnológica del Ministerio de Educación, Cultura y Ciencia y Tecnología de la República Argentina. Para la realización de la reunión se contó con el financiamiento de este proyecto y del subsidio para reuniones científicas RC 2017-0323 otorgado por la Agencia Nacional de Promoción Científica y Tecnológica.

El Workshop reunió investigadores con diferentes perfiles formativos y trayectorias, cuyas intervenciones permitieron generar un espacio compartido de debate y un intercambio de perspectivas teóricas y metodológicas relacionadas con las políticas científicas y con la obtención de métricas y visualizaciones derivadas de la producción y su relación con indicadores del contexto económico y social. Cabe destacar que aunque el interés en los estudios sobre América Latina y España no es nueva, sí es novedoso que desde las regiones periféricas se indague con una perspectiva bibliométrica la producción científica que se realiza sobre los países de la región, sea ésta producida en el propio país o llevada a cabo desde el extranjero.

Las exposiciones se desarrollaron en los siguientes ejes temáticos:

- Presencia de temas orientados a lo local / regional en las políticas y agendas de investigación
- Aproximaciones metodológicas para el abordaje cuantitativo de la producción en temas locales / regionales
- Estudios cuantitativos sobre la producción científica en temas locales de países iberoamericanos

Finalmente agradecer a todos los expositores y asistentes que hicieron posible el intercambio de conocimientos y experiencias en los temas abordados, y la concreción de esta publicación que recoge los principales resultados del encuentro.

Sandra Miguel
La Plata, 2019

EJE TEMÁTICO I

Presencia de temas orientados a lo local /
regional en las políticas y agendas de investigación

Política científica y relevancia social de la investigación

Federico Vasen¹

Introducción: relevancia, prioridades, agendas

El presente trabajo se propone recorrer la historia de la política científica haciendo énfasis en la forma en que se caracteriza en los distintos periodos la relevancia social de la investigación científica. Por concepciones de la relevancia entendemos el conjunto de juicios de valor acerca de las relaciones entre ciencia, tecnología y sociedad que permiten identificar las investigaciones que merecen promoción en el marco de una política científica particular. Estos sentidos pueden predicarse tanto de políticas científicas determinadas o generaciones de ellas, así como de las estrategias de resistencia frente a éstas, siempre y cuando no sean meramente reactivas y propongan una agenda alternativa (Vasen, 2011). Esta caracterización es la que luego permitirá puntualizar criterios de priorización de áreas o temáticas en el marco de instrumentos específicos.

En conexión con este concepto de relevancia, cabe distinguir a su vez el alcance de otras dos nociones asociadas: *prioridades* y *agendas de investigación*. Las concepciones de la relevancia deben comprenderse como un marco general para pensar las relaciones cómo la política científica interpreta y se posiciona en las relaciones ciencia/tecnología/sociedad. Las prioridades para investigación son elecciones de temáticas específicas de financiadores públicos o privados o de instituciones científicas en las cuales se quiere in-

¹ Universidad de Buenos Aires - CONICET, Ciudad Autónoma de Buenos Aires, Argentina.
fvasen@rec.uba.ar

centivar en forma prioritaria su desarrollo. Un ejemplo de ello es proponer a la nanotecnología o la biotecnología agraria como prioridades de un país o de una institución. Por otra parte, las agendas de investigación responden a la dinámica de la evolución de un campo científico particular, en el cual ciertos temas toman protagonismo por sobre otros. Se trata del grupo de temáticas al que un agente dentro del campo científico (investigador, grupo, comunidad disciplinar o subdisciplinar) dedica sus esfuerzos. Estas decisiones se basan no sólo en sus intereses sino también en las tradiciones del campo y los recursos teóricos, metodológicos y financieros disponibles.

En lo que resta de esta contribución nos ocupamos de la primera de las dimensiones mencionadas: la relativa a las concepciones de la relevancia. Para ello, realizamos un recorrido histórico de la política científica desde mediados del siglo XX, en el cual caracterizamos cuatro concepciones.

Sentidos de la relevancia

En esta sección presentamos cuatro concepciones de la relevancia de la investigación científica. La *relevancia diferida* puede considerarse el primer momento que se instaura a partir de la segunda posguerra y se identifica con el modelo lineal de investigación y desarrollo. La *relevancia nacional* emerge como una postura crítica ante el primer modelo en la década de 1960 y puede relacionarse a nivel local con el movimiento de Pensamiento Latinoamericano en Ciencia, Tecnología y Desarrollo. En tercer lugar, la *relevancia comercial-competitiva* adquiere protagonismo en la década de 1980 en conexión al modelo de los Sistemas de Innovación. Finalmente, la *relevancia poscompetitiva* puede entenderse como un reconocimiento de las limitaciones de la perspectiva económica vigente y se entronca con modelos de política como los de la innovación responsable o la innovación inclusiva. A continuación nos referimos brevemente a cada etapa.

Relevancia diferida

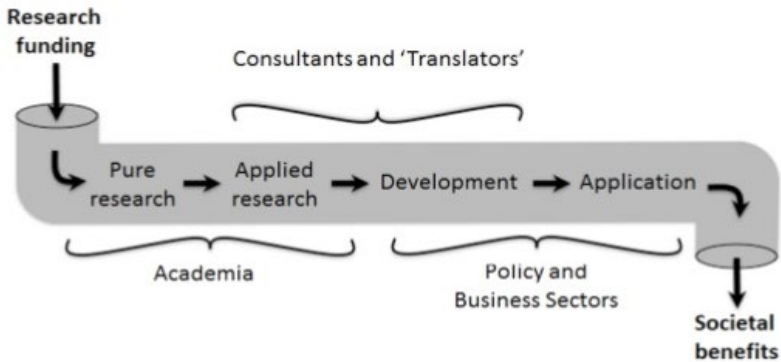
La publicación del informe Ciencia: la frontera sin fin por parte de Vannevar Bush en 1945 constituye un punto de inflexión en el financiamiento público de las actividades de ciencia y tecnología. A partir de esas ideas, se configurará lo que se dio en llamar el ‘modelo lineal de investigación y desarrollo’, que sostiene fundamentalmente los siguientes puntos:

- La importancia de la ciencia básica como motor de la investigación aplicada y el desarrollo tecnológico.
- La necesidad de financiamiento público para la ciencia básica ya que es una inversión muy riesgosa para el sector privado.
- La necesidad de resguardar la libertad de investigación en las etapas más básicas.
- La delegación en las comunidades académicas del juicio acerca de la pertinencia y la calidad de las propuestas de investigación que debían financiarse.

Cabe destacar que el modelo lineal brindó una suerte de “paraguas” conceptual para la justificación de la investigación que antes de denominaba “pura”, es decir, aquella que no tenía en vistas fines o aplicaciones tecnológicas específicas (Kaldewey y Schauz, 2017). Permitía asegurar el flujo de recursos públicos a la investigación básica sin la obligación de producir resultados aplicados tangibles. Es lo que David Guston (2000), un especialista en la historia de la política científica, llamó un contrato social de “autonomía relativa”, en el cual la sociedad le da recursos a la comunidad académica con el fin de que ella cumpla “las promesas de la ciencia” (ver figura 1). Pero no le explica cuál es la mejor forma de cumplirlas porque supone que son los científicos los más capacitados para escoger los temas y repartir internamente el financiamiento (Baldwin, 2018).

Hablamos de *relevancia diferida* en la medida en que los investigadores académicos que se involucran en las primeras etapas del proceso no están obligados a dar cuenta previamente a ser financiados la relevancia social de su trabajo. Esta sólo podrá mostrarse en forma diferida cuando ese conocimiento sea tomado, “traducido” y aplicado por actores del sector productivo e industrial.

Figura1. Modelo lineal de I+D



Fuente: <https://image.slidesharecdn.com/moseru54a-01engagementinfutureearthcredits-150304101943-conversion-gate01/95/engagement-in-future-earth-supporting-a-stepchange-in-global-sciencepolicy-interactions-19-638.jpg?cb=1425486042>

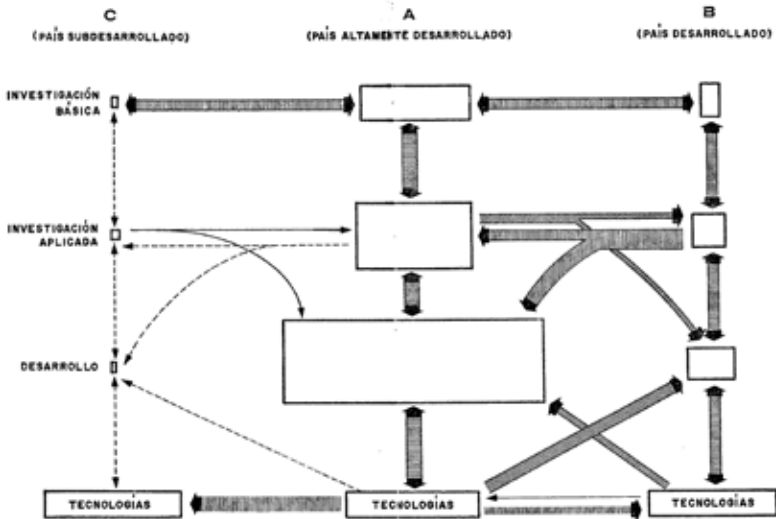
Relevancia nacional

A partir de mediados de la década de 1960, surge la preocupación por coordinar las políticas de los distintos organismos que realizan investigación científica. El debate se transporta entonces a un nivel nacional. Ya no se habla de la relevancia para una pequeña comunidad disciplinar sino de la relevancia en función de objetivos nacionales, como podría ser el prestigio del país o la capacidad de defensa ante ataques extranjeros. En los años '60, muchos científicos dejan de ser ciudadanos cosmopolitas de la República de la Ciencia, aquella comunidad idealizada de pares dedicadas al conocimiento por el conocimiento mismo, y comienzan a tomar nota de los intereses estratégicos que están detrás de quienes los financian.

En este marco surge en América Latina la reflexión del movimiento que se denominó Pensamiento Latinoamericano en Ciencia, Tecnología y Desarrollo Dagnino, Thomas y Davyt, 1996; Martínez Vidal y Marí, 2002; Galante et al., 2009). Sus autores más representativos -Amílcar Herrera, Jorge Sabato, Oscar Varsavsky- coincidían más allá de sus múltiples diferencias en criticar la forma en la que se había llevado adelante la política científica en la región hasta ese momento. Sabato, vinculado a las necesidades de la industria,

buscó articular espacios de transferencia de conocimientos entre el sector público y privado en el área de metalurgia, lejos de la ciencia académica básica que se producía en las universidades (Sabato, 2004). Herrera, un hombre de la universidad él mismo, se manifestaba en contra de la desarticulación entre las agendas de investigación perseguidas y las necesidades locales (Herrera, 1971). Finalmente, Varsavsky -el más radical- denunciaba el carácter ideológico del modo de producción de conocimiento científico vigente y trataba de sentar las bases de una ciencia que sirviera a la revolución socialista (1972). Más allá de las fuertes diferencias, puede notarse una coincidencia en lo que hace a la crítica al modelo ofertista y la búsqueda de una política científica que responda concientemente a un proyecto de país, pues se percibía que la política científica ofertista traía aparejada la consolidación de ciertos lazos de dependencia económica y cultural, que obturaban la posibilidad de pensar de forma autónoma cuál era el desarrollo científico-tecnológico más adecuado para cada país latinoamericano (ver figura 2). En este sentido es que hablamos de un sentido *nacional* de la relevancia donde los actores comienzan a pensar en el peso que tienen sus agendas de investigación para los objetivos nacionales y la política científica empieza públicamente a priorizar temáticas basadas en estos objetivos políticos.

Figura2. Señales de relevancia. Fuente: Herrera (1971)



Relevancia comercial-competitiva

Durante la década de los 90 se planteó la discusión acerca del surgimiento de un nuevo contrato social entre la ciencia y la tecnología y la sociedad. El contrato original de “autonomía relativa”, que habría quedado fijado desde 1945 junto al modelo lineal, podría caracterizarse como “el acuerdo de la comunidad política de proveer recursos a la comunidad científica y de permitirle a ésta retener sus mecanismos de decisión, y en contrapartida esperar beneficios tecnológicos futuros no especificados” (Guston, 2000 p. 62). El nuevo contrato, en cambio, estimularía un mayor vínculo entre científicos y promotores y un mayor control sobre el trabajo efectivo de investigación por parte de estos últimos a través de la inclusión de criterios externos a la comunidad científica en la evaluación de proyectos. Asimismo, los académicos son incentivados a trabajar en temas que puedan ser relevantes para aquellos con capacidad de financiar la investigación, en este sentido es que podemos hablar de una concepción de la relevancia ligada a lo comercial-competitivo.

Relevancia pos-competitiva

En los últimos años, se reconoce una tendencia que puede denominarse “giro poscompetitividad” en la política científica (Vasen, 2016). La idea es que se empiezan a desarrollar políticas científicas que no tienen como objetivo central fomentar a la industria para ganar competitividad económica en el contexto internacional. Se busca poner el énfasis en problemas nacionales relevantes para toda la sociedad, en generar un desarrollo socialmente responsable de las tecnologías emergentes o en impulsar tecnologías que tengan un fuerte valor social, incluso si económicamente no son sostenibles. Ejemplo de ello podrían ser los marcos de innovación responsable o innovación inclusiva (Vasen, 2017). Esta tendencia sin embargo no necesariamente es la que predomina hoy en las políticas. Dependiendo del país y la región puede ser corriente minoritaria o complementaria, en el marco de un discurso que sigue aún enfatizando la competitividad y los sistemas nacionales de innovación.

Conclusión

El trabajo precedente ha recorrido sucintamente las concepciones prevalecientes en lo que hace a los marcos sobre relevancia social de la investigación científica y su impacto en la política científica. A lo largo de la historia se evidencia una tensión entre los intereses de las comunidades académicas, los intereses de “la sociedad” o “el país” expresados en los objetivos políticos y los intereses de los actores económicos e industriales. La configuración resultante en cada etapa expresa el balance de fuerzas entre los distintos actores y sus intereses.

Referencias bibliográficas

- Baldwin, M. (2018). Scientific Autonomy, Public Accountability and the Rise of Peer Review in the Cold War United States. *Isis* 109(3), 538-558.
- Dagnino, R., Thomas, H. y Davyt, A. (1996). El pensamiento latinoamericano en Ciencia, Tecnología y Sociedad. Una interpretación política de su trayectoria, *Redes*, 3(7), 13-51.
- Galante, O., Marí, M., Carnota, R., Vasen, F. y Benso, O. (2009) *La Escuela Latinoamericana de Pensamiento en Ciencia, Tecnología y Desarrollo: proyecto de recuperación histórica y documental*. Trabajo presentado en XIII Encuentro ALTEC. Cartagena de Indias, Colombia.

- Guston, D. (2000). *Between Science and Politics*. Cambridge: Cambridge University Press.
- Herrera, A. O. (1971). *Ciencia y política en América Latina*. México: Siglo XXI.
- Kaldewey, D. y Schauz D. (2017). The ‘Politics of Pure Science’ Revisited. *Science and Public Policy*, 44(6), 883–886.
- Martínez Vidal, C. y Marí, M. (2002). La escuela latinoamericana de Pensamiento en Ciencia, Tecnología y Desarrollo. Notas de un proyecto de investigación. *Revista CTS+I (OEI)*, 4. Recuperado de [_Varsavsky, O. \(1972\). *Hacia una política científica nacional*. Buenos Aires: Periferia.](#)
- Vasen, F. (2011). Los sentidos de la relevancia en la política científica. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*. 7(19), 11-46.
- Vasen, F. (2016). ¿Estamos ante un giro poscompetitivo en la política de ciencia, tecnología e innovación? *Sociologias (Porto Alegre)*, 18(41), 242-268.
- Vasen, F. (2017). Responsible Innovation in Developing Countries: an Enlarged Agenda. En: Asveld, L., (Ed.), *Responsible Innovation 3 – An European Agenda?* (pp. 93-109). Cham: Springer.

La investigación en áreas prioritarias y en temas locales/regionales. Presencia en las políticas y planes de los países latinoamericanos

Victoria Ugartemendia¹

Introducción

A partir de la segunda mitad de los años noventa se sentaron las bases y se crearon las organizaciones que fueron consolidando los Sistemas Nacionales de Ciencia, Tecnología e Innovación (SNCTI) latinoamericanos. Durante esa década y las siguientes, las políticas de promoción de CTI dieron un giro importante con respecto a los patrones prevalecientes en las décadas previas de vigencia del modelo de industrialización sustitutiva de importaciones (Casas, 2014:3). Esto se expresó en la creación de planes de Ciencia, tecnología e innovación con la expectativa de que se instituyeran en instrumentos de la política pública.

En Argentina, por ejemplo, se elaboró el Plan Estratégico Nacional de Ciencia, Tecnología e Innovación “Bicentenario” (2006- 2010) (PB) y el Plan Nacional de Ciencia Tecnología e Innovación (2012-2015) (PNCTI). Brasil también desarrolla en esos años planes nacionales (si bien comienza con esta práctica en los años 1980). Aquí tomaremos para el análisis la Estrategia Nacional de Ciencia, Tecnología e Inovacao 2012- 2015 (ENCTI). En México la elaboración de planes nacionales viene desde antes- década de 1970-. En el año 2011 hizo el

¹ Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. vugartemendia@gmail.com

Programa Especial de Ciencia, Tecnología e Innovación 2008-2012 (PECTI) que tomaremos como referencia en este trabajo. Final del formulario

En este documento nos proponemos analizar las perspectivas de política científica que subyacen a los planes nacionales CTI de los países latinoamericanos. Nos concentraremos en Argentina, Brasil y México por ser los países con sistemas nacionales CTI más grandes de la región.

Metodología

La metodología de trabajo es cualitativa y las fuentes serán los planes de Ciencia, Tecnología e Innovación mencionados, y fuentes secundarias. Final del formulario.

Resultados y discusión

Perspectivas de política subyacentes a los planes nacionales

1-Una característica común a los planes nacionales analizados es que todos se plantean como objetivo que la I+D contribuya a lograr más competitividad económica en el marco de un Escenario de Desarrollo sustentable. Por ejemplo en los dos planes bajo análisis de Argentina se postula que la ciencia y la tecnología contribuyan significativamente al aumento de la productividad de la economía y a fortalecer el patrón de especialización productiva e inserción global del país. Se plantea además la búsqueda de la competitividad en sectores determinados y el “escalamiento” hacia posiciones más jerárquicas en las tramas globales de conocimiento e innovación (PNCTI, p. 22- 23).

En este sentido, prima aquí un sentido mercantil de la relevancia; es decir, lo que se define como importante en la política es lo que se espera que genere valor en el corto o en el mediano plazo. Como sostiene Vasen (2011), en los años 1980 emerge un nuevo contrato social entre ciencia, tecnología y sociedad, enmarcado en el documento de la OCDE Science and Technology Policy for the 80s. Se llamó a los países a emprender una política industrial activa y a buscar una mayor cooperación entre universidades e industrias.

2- En los Planes está presente la idea de que la meta a alcanzar es que la ciencia y la tecnología posibiliten la industrialización y el desarrollo sustentable. En el PB, por ejemplo, se menciona entre sus cuatro objetivos estratégicos se espera que el SNCTI contribuya a la apertura de senderos de desarrollo sustentable. En el ENCTI de Brasil se plantea algo similar: la

ampliación de las bases de la sustentabilidad ambiental y el desarrollo de una economía de bajo carbono.

Estas nociones de desarrollo no responden enteramente a las connotaciones que tuvieron en los años 1950 y 1960, tanto para los gobiernos caracterizados como desarrollistas, como para el pensamiento latinoamericano en CyT (PLACT), ya que a pesar de que incorpora la noción *aggiornada* de autonomía restringida y de vinculaciónismo, deja de lado las nociones muy caras al desarrollismo relacionadas con el ofertismo y la transferencia de tecnología. Además el desarrollo esperado se hace eco de nuevos aspectos y problemas al incorporar el término “sustentable”. Esta noción surge en la década de 1980 cuando emergen en el marco de distintas agencias de Naciones Unidas miradas que complejizaron la idea de que la industrialización de los países del Tercer Mundo, a partir de tomar conciencia sobre los problemas ecológicos y demográficos que trae aparejado cualquier proceso industrialización (Pakdaman N., 1999: 102).

En Argentina y México, en continuidad con la tendencia iniciada en los años 1990, se abandona la idea de “cerrar la brecha” con los países desarrollados en términos de ciencia y tecnología. En Brasil en cambio figura el objetivo de “Reducción del desfase científico y tecnológico que aún separa a Brasil de las naciones más desarrolladas” (ENCTI, p. 33).

3-En los planes de Brasil y Argentina, en particular en el PB, aparece la preocupación porque la CyT contribuya a resolver la pobreza y la marginalidad. Este objetivo va perdiendo fuerza en el segundo Plan argentino y se plantea como especie de política de derrame en México.

4-En los casos de Brasil y Argentina, se espera que los planes sirvan para integrar los SIN, perspectiva que de la mano de economistas neoshumpeterianos creció y se difundió entre los investigadores y los hacedores de políticas científicas en los años '80 y desde los años 1990 en América Latina en particular. El Estado se corre del centro de la escena en la que lo pusieron las políticas latinoamericanas de CyT de los años 1950 y 1960, y entran a jugar un rol más relevante aquellos instrumentos de apoyo y fomento que consoliden las interacciones entre los diversos actores del sistema, teniendo en cuentas sus proyectos y programas y la manera en la que de acuerdo a cada una de sus particularidades, generan redes de apoyo y vínculos. De esta manera, se encuentran instrumentos dedicados a la generación de redes en todo el sistema. Por

ejemplo, uno de los programas a destacar es el llevado adelante por México denominado “Redes de Innovación”, que tiene como finalidad promover la articulación entre instituciones de investigación y empresas que al utilizar su sinergia incrementen la competitividad del sector productivo que les compete...” (CIECTI, 2015: 81 y 82).

Esto se encuadra en lo que Renato Dagnino (2006) denomina modelo de “Governanza en red”², en contraposición al modelo de “Modelo Jerárquico de Governanza” que dominó hasta los años 1980, y que se caracteriza por la coordinación descentralizada de la ciencia, por la que los sectores particulares tienden a ser formalmente autónomos y no están jerárquicamente relacionados y no existe ningún actor dominante o focal. En este sentido, el papel que se espera del Estado no es el de organizador de los intereses privados y de árbitro de los conflictos entre grupos, sino que los distintos actores interactúan entre sí y con los burócratas estatales en una base igualitaria (Dagnino, 2006:67³). Esto está presente en los planes: no se trata de poner al Estado en el centro de la escena, sino de convocarlo como mediador entre los componentes del sistema científico- tecnológico, y entre éstos y el resto de los actores sociales, entre los que se destacan las empresas.

Expresión de esta dinámica son procesos de consultas, conferencias, talleres y otras técnicas de participación a los que se convocaron a diversos actores de la sociedad civil para elaborar los planes, procesos de diseño de los PCTI caracterizados como “de abajo hacia arriba”. Por ejemplo, Casas et al (2014) sostiene que en el caso de México, en diferentes momentos de la historia de las PCTI, desde la década de 1970, se ha convocado a ejercicios participativos mismos que actualmente se han intensificado para la elaboración del nuevo Plan 2013-2037 (Casas et al, 2014: 12). En el caso del PNCTI de Argentina los actores participantes de la elaboración del Plan van desde integrantes de las Universidades Públicas, Universidades privadas, de Institutos de Investigación, Ministerios de la Nación, empresas privadas, cámaras empresarias, el COFECYT, hasta organismos internacionales como

² El término Governanza o *governaceo* “buen gobierno”, significa la gestión de los procesos a través de los cuales se crean las condiciones para gobernar: “... el “buen gobierno” se ocupa de suscitar las condiciones necesarias para un gobierno ordenado y una acción colectiva” (Stoker, G. y C. H. Alcántara, 1998:4).

³ La traducción es nuestra.

la CEPAL. Esta propuesta se hacen eco de lo planteado en la “Conferencia Mundial de la Ciencia” de Budapest de 1999, en la que se propuso realizar “... un nuevo “contrato social” basado en el supuesto de que la ciencia debe ser asunto de debate público y que sus prioridades (...) sean examinadas de manera democrática, no sólo dentro del recinto de los laboratorios y de los pasillos gubernamentales.” (Ávalos y Rengifo, 2001: 3).

Enfoque regional y local, y agenda de temas prioritarios

Casas et al (2014) sostiene que una de las características comunes en los países de América Latina es que “... se está transitando a planteamientos, argumentos y medidas específicas para políticas de CTI adecuadas a las regiones y las localidades...” (p. 13). Agregan que esta perspectiva ha sido dominante en las políticas europeas de CTI desde la década de los noventa (p. 14). Por otro lado en los planes aparece la definición de **áreas** o temas considerados estratégicos o prioritarios para los países, sobre las cuales los Estados se proponen desarrollar políticas diferenciadas. Esto aparece con más énfasis en los planes de Argentina y de Brasil que en los de México, que si bien define temas estratégicos, desarrolla más políticas de carácter horizontal.

En los cuatro planes analizados están presentes un conjunto de idénticos y similares temas considerados prioritarios, tales como Biotecnología, Tecnologías de la información y comunicación, Nanotecnología o Energía. Se han puesto en marcha múltiples instrumentos para canalizar este interés en temas prioritarios. Por ejemplo en Brasil y Argentina se hay constituido fondos sectoriales. En el primer caso varios de sus fondos sectoriales son financiados a través de tasas o impuestos a ciertos sectores económicos determinados. Uno de los más grandes, el CT - Petro, obtiene sus fondos de un porcentaje de las regalías que pagan las empresas petroleras para producir petróleo y gas. Esta masa de recursos luego se invierte en forma de subsidios, para estimular la innovación productiva en el sector petrolero y gasífero, mejorando la formación de recursos humanos, y financiando proyectos en conjunto entre empresas y universidades. Los fondos sectoriales tienen amplia difusión en la región, aunque solo en el caso brasileño cuentan con fondos suficientes como para generar un impacto significativo en el entramado productivo del país.” (CIECTI, 2015:81). En Argentina a partir del PNCTI, en el año 2010 se consolidaron los Fondos de Innovación Tecnológica Sectorial (FITS) y

los Fondos Tecnológicos Sectoriales (FTS) que destinan sus esfuerzos a financiar proyectos de I+D e innovación exclusivamente en los sectores de biotecnología, nanotecnología, tecnologías de la información y las comunicaciones y, los Fondos de Innovación Tecnológica Sectorial, por su parte, se destinan a financiar proyectos también de I+D e innovación exclusivamente en los sectores de Agroindustria, Energía, Salud, Desarrollo Social, Ambiente y Cambio climático (CIECTI, 2015: 106). Sin embargo, los temas prioritarios no nos dice mucho acerca de carácter regional o local de la política, aspectos que se suelen solapar y confundir. Es más pertinente analizar esta dimensión a partir de los instrumentos o del impacto de la política en regiones o localidades que desde los temas generales.

En este sentido existen programas y otro tipo de instrumentos directamente orientados a resolver problemas regionales o locales. Por ejemplo, en el caso de Argentina podemos mencionar al Programa de Reducción de los Desequilibrios Regionales y Provinciales inscripto en el PB, en el marco del cual se desarrolla el Proyectos Federales de Innovación Productiva (p. 22). (Convocatoria a PAE 2017, p. 9). Este tuvo la última convocatoria en 2017. Lo regional también aparece en el ENCTI de Brasil, donde se sostiene lo siguiente: “En lo relativo a las desigualdades regionales, el esfuerzo importante de desconcentración emprendido en el último gobierno alteró la distribución espacial de los programas de posgraduación, fortaleciendo las regiones Norte, Nordeste e Centro-Oeste en la formación de doctores. Algunos proyectos se desarrollaron en las regiones Norte e Nordeste, tales como la creación del Instituto Nacional do Semi-Arido (INSA) y el Centro de las Tecnologías Estratégicas do Nordeste (CETENE)...” (ENCTI, p. 38).

Referencias bibliográficas

- Argentina. Ministerio de Ciencia y Tecnología. SECyT. (2006). *Plan Estratégico Nacional de Ciencia, Tecnología e Innovación “Bicentenario” (2006- 2010)*.
- Argentina. Ministerio de Ciencia y Tecnología. (2012). *Plan Nacional de Ciencia Tecnología e Innovación (2012-2015)*. Recuperado de <https://www.argentina.gob.ar/plan-bicentenario-2006-2010>
- Argentina. Ministerio de Ciencia y Técnica. (2017). *Convocatoria a PAE 2017, Bases y condiciones del llamado a la presentación de PFIP 2017,*

- Proyectos Federales de Innovación Productiva*. COFECYT. Versión 02 (agosto/17). Recuperado de <https://www.argentina.gob.ar/sites/default/files/bases-de-la-convocatoria-pfip-mae-2016.pdf>
- Brasil. Ministério da Ciência, Tecnologia e Inovação. (2011). *Estrategia Nacional de Ciencia, Tecnologia e Inovacao 2012- 2015 (ENCTI)*. Recuperado de <http://livroaberto.ibict.br/218981.pdf>
- Casas R., Corona J. M. y Rivera R. (2014). Políticas de Ciencia, Tecnología e Innovación en América Latina: entre la competitividad y la inclusión social. *Conferencia Internacional LALICS 2013*, 11 y 12 de noviembre de 2013, Rio de Janeiro, Brasil. Recuperado de <http://www.redesist.ie.ufrj.br/lalics/papers/115 Politicas de Ciencia Tecnologia e Inovacion en America Latina entre la competitividad y la inclusion social.pdf>
- CIECTI (2015). Políticas CTI en países emergentes. Análisis comparado de experiencias heterogéneas y su aplicabilidad en Argentina. Informe final de proyecto-Noviembre. Recuperado de <http://www.ciecti.org.ar/wp-content/uploads/2016/09/CIECTI-Proyecto-Grupo-REDES.pdf>
- Dagnino, R. P. (2006). Enfoques empleados nos países avançados para a análise da política de C&T. *Redes*, 12 (24), 61-87.
- Mexico. Consejo Nacional de Ciencia y tecnología. (2008). *Programa Especial de Ciencia, Tecnología e Innovación 2008-2012 (PECTI)*

¿Responde la investigación a las necesidades de salud?¹

Ismael Ràfols² y Alfredo Yegros³

Introducción

La salud humana en el mundo ha mejorado enormemente durante los últimos cien años, principalmente gracias a la mejora de las condiciones de vida, la salubridad (por ejemplo, el acceso a agua no contaminada) y la atención sanitaria. Pero también, aunque quizá en menor medida, gracias a los descubrimientos científicos. Los rayos X y la penicilina son ejemplos de descubrimientos que han provocado grandes avances médicos, en tanto que las vacunas, que emanan de una investigación aplicada, han salvado, igualmente, millones de vidas.

Esos éxitos extraordinarios han inducido a pensar que si aumentara la investigación biomédica, la salud acabaría mejorando. Sin embargo, la realidad no es tan simple. Para que la investigación biomédica mejore la salud hace falta, primero, tener en consideración factores que, estrictamente, escapan del ámbito de la medicina. Antes que nada, se debe prestar más atención tanto a las necesidades humanas como a las causas de la mala salud. Teniendo todo esto en cuenta, es posible, entonces, establecer prioridades de I+D más efectivas.

¹ El texto original fue publicado por el Observatorio Social de La Caixa, disponible en <https://observatoriosociallacaixa.org/es/-/responde-la-investigacion-a-las-necesidades-de-salud>

² CSIC-UPV, Universitat Politècnica de València, Valencia, España. i.rafols@ingenio.upv.es

³ CWTS Leiden Universiteit, Leiden, Países Bajos. a.yegros@cwts.leidenuniv.nl

Muchos analistas aducen que los beneficios sanitarios de la investigación podrían mejorar sustancialmente si las prioridades se determinaran de un modo más sistemático. La asignación de fondos de investigación se basa en varios criterios, normalmente relacionados con la percepción que se tiene de la calidad científica de proyectos y equipos, de su potencial para generar avances científicos y de las demandas o necesidades sociales respecto a un problema concreto. Sin embargo, en la práctica, se insiste mucho en evaluar la calidad científica para fomentar la excelencia y se presta mucha menos atención a comprobar si la investigación responde a las necesidades sociales. Estudios recientes sugieren que hay un desajuste entre las prioridades de la investigación y las necesidades sanitarias en el nivel nacional y, todavía más, en el mundial (Gross *et al.*, 1999; Evans *et al.*, 2014).

En el presente artículo se analiza cómo mejorar la fijación de prioridades de la investigación sanitaria a la luz de si esta responde o no, y en qué medida, a necesidades sociales.

Cómo evaluar prioridades

El objetivo último de la investigación sanitaria es mejorar la salud y el bienestar de la humanidad, por lo que debería centrarse en alcanzar dicho objetivo. Cuando se establecen las políticas públicas y las prioridades de investigación, suele afirmarse que la investigación sanitaria debe orientarse a alcanzar «objetivos humanitarios y de solidaridad», los cuales deben «fundamentar todos los pasos del ciclo de la innovación, desde los descubrimientos de laboratorio y de investigación clínica hasta su aplicación en la atención sanitaria y la prevención». Pero, ¿cómo podemos saber realmente si esos objetivos se alcanzan a largo plazo?

Para responder a esta pregunta, en 1997 los Institutos Nacionales de la Salud (NIH) de EE. UU. crearon un grupo de trabajo sobre priorización de la investigación en el ámbito de la salud. Dos años después, Gross *et al.* (1999) publicaron un estudio que comparaba diversos indicadores de carga de enfermedad con iniciativas de investigación de los NIH, todas ellas referidas a una lista de enfermedades importantes. Las iniciativas se clasificaron teniendo en cuenta los fondos que se destinaban a cada proyecto y a cada enfermedad.

La “carga de enfermedad” da una pista sobre la importancia social de una dolencia. El valor de este indicador se establece teniendo en cuenta un con-

junto de parámetros; entre ellos, la prevalencia (el número de personas que sufre una enfermedad), la incidencia (la frecuencia de nuevos casos), los días de hospitalización, la mortalidad, los años de vida perdidos y los años de vida ajustados en función de la discapacidad (AVAD). Los AVAD indican cuántos años de vida sana se pierden a causa de una enfermedad, teniendo en cuenta tanto la mortalidad como la discapacidad que esa patología ocasiona. Dicho de otro modo, los AVAD expresan el número de *años de vida perdidos* a causa de una enfermedad, teniendo en cuenta tanto la esperanza de vida de la población como los años que no se han podido disfrutar plenamente, precisamente a causa de la enfermedad (*años perdidos por discapacidad*).

El trabajo de Gross y sus colegas (1999) puso de manifiesto que había una escasa correlación entre la financiación para el estudio de una enfermedad, por una parte, y la mortalidad y años de vida perdidos a causa de la misma, por la otra. Sin embargo, también observaron que había una fuerte correlación entre financiación y AVAD. Esto significaba, en definitiva, que la financiación que proporcionaban los NIH era, efectivamente, sensible a la carga de enfermedad. Es decir: las decisiones de qué investigar y cuánto esfuerzo dedicar a ello, tenían en cuenta la importancia, en términos sociales y personales, de las enfermedades.

Sin embargo, este cuidadoso análisis también evidenció la presencia de disparidades considerables en la cantidad de fondos que recibían enfermedades de igual carga de enfermedad. Patologías como el sida, el cáncer de mama o la diabetes recibían, proporcionalmente, mucha más financiación que la depresión, el cáncer de colon o las enfermedades perinatales. En consecuencia, había margen para alinear mejor la financiación a la carga de enfermedad.

Para relacionar la financiación de la investigación con la carga de enfermedad también se ha utilizado el número de publicaciones como indicador del esfuerzo de investigación (Evans *et al.*, 2014). Este enfoque bibliométrico tiene la ventaja de tener en consideración investigaciones que se realizan en universidades y hospitales por docentes y médicos y por tanto no están contabilizadas en el gasto en I+D de organismos de financiación.

Esfuerzo de investigación vs. carga de enfermedades

En el presente artículo, se emplean indicadores sobre publicaciones para vincular la carga de enfermedad (los AVAD) al esfuerzo de investigación. Se

combina información procedente de las bases de datos PubMed y Web of Science correspondiente al periodo 2009-2013. La relación entre las publicaciones y las enfermedades se basa en los descriptores de PubMed, que proporcionan expertos de la Biblioteca Nacional de Medicina de los NIH de EE.UU. Las cifras de carga de enfermedad mundial, finalmente, proceden de la OMS, con datos de 2012.

Si, para un conjunto de enfermedades, se compara el impacto de salud relativo de una de ellas, medido en el porcentaje de AVAD, con el volumen de publicaciones de investigación que ha generado, también medido en tanto por ciento sobre el total de publicaciones para el conjunto de enfermedades, entonces es posible hacerse una idea del ajuste o desajuste entre la carga de la enfermedad y el esfuerzo de investigación. Presentamos datos de todo el mundo y de España (cuyo perfil es similar al de muchos países desarrollados). La carga de enfermedad y las pautas de publicación evidencian disparidades sorprendentes en la investigación de diversos grupos de enfermedades. En el nivel mundial, el cáncer (los tumores malignos) representa más del 22% de las publicaciones médicas globales, aunque su carga de enfermedad no llega al 10% del total. Las enfermedades cardiovasculares, infecciosas y parasitarias representan más del 16% de la carga de enfermedad, pero menos del 10% de las publicaciones. Las patologías neonatales se han investigado poco si tenemos en cuenta su carga de enfermedad.

En España y otros países desarrollados, las cargas de enfermedad son distintas a las del conjunto del mundo. Por ejemplo, en España, el peso del cáncer en la carga de enfermedad es un poco mayor que la cuota de publicaciones sobre esta enfermedad. En cambio, la proporción de publicaciones sobre enfermedades infecciosas y parasitarias es mucho más elevada que la que correspondería a la carga de enfermedad de estas dolencias.

El gráfico 1 muestra la relación entre la carga de enfermedad relativa (porcentaje respecto al total) y el vertical, la proporción de publicaciones (también en relación con el total de publicaciones sobre enfermedades). En las patologías que se sitúan por encima de la línea de 45°, las publicaciones superan relativamente su carga de enfermedad; es decir, se podrían calificar como “sobre-estudiadas”. Las que aparecen por debajo de esa línea tienen, en cambio, un nivel proporcionalmente menor de publicaciones, por lo que podrían calificarse de “infra-estudiadas”.

La comparación entre las pautas mundiales y las españolas pone de manifiesto que hay algunas enfermedades que presentan cargas de enfermedad similares en los países desarrollados (por ejemplo, las cardiovasculares o el cáncer), y algunas otras que presentan una carga mucho mayor en países de renta baja o media que en países de renta alta. Algunas de esas últimas patologías siguen estando presentes en países de renta alta (por ejemplo, el sida/VIH o la tuberculosis), en tanto que otras (como la malaria) solo suponen una gran carga para los países pobres.

El hecho de que el 90% de la investigación del mundo se produzca en países desarrollados hace que las iniciativas de investigación mundiales encajen más con las necesidades sanitarias de las sociedades ricas que con las del conjunto del mundo (Evans *et al.*, 2014). En consecuencia, hay buenas razones para transformar la investigación en salud pública con el fin de que se ocupe, local y globalmente, de necesidades sanitarias desatendidas. Esta orientación cuestiona tanto la relativa infra-inversión, local y global, en enfermedades como la depresión o la obstrucción pulmonar crónica, como el excesivo esfuerzo en investigar prioridades sanitarias locales (como la diabetes) y no desafíos sanitarios globales (como la malaria).

Conclusiones

Este trabajo plantea cómo un enfoque comparado puede fomentar un mejor ajuste entre las iniciativas de I+D y las necesidades sanitarias. Los datos sugieren que algunas enfermedades prevalentes en países de renta baja y media, como la malaria o la tuberculosis, merecen más inversión en I+D, pero lo mismo puede decirse de patologías como la depresión o el ictus, de elevada carga de enfermedad en todo el mundo.

El análisis puede enriquecerse profundizando en el diseño de fórmulas más eficientes de distribución de las inversiones en distintos tipos de investigaciones; es decir, considerando varios enfoques científicos para combatir una enfermedad. La enorme incertidumbre que rodea el resultado de una investigación hace que el estudio de una determinada patología deba incluir diversas estrategias, que vayan desde la biología básica a la medicina, pasando por los estudios sociales y las consideraciones de salud pública.

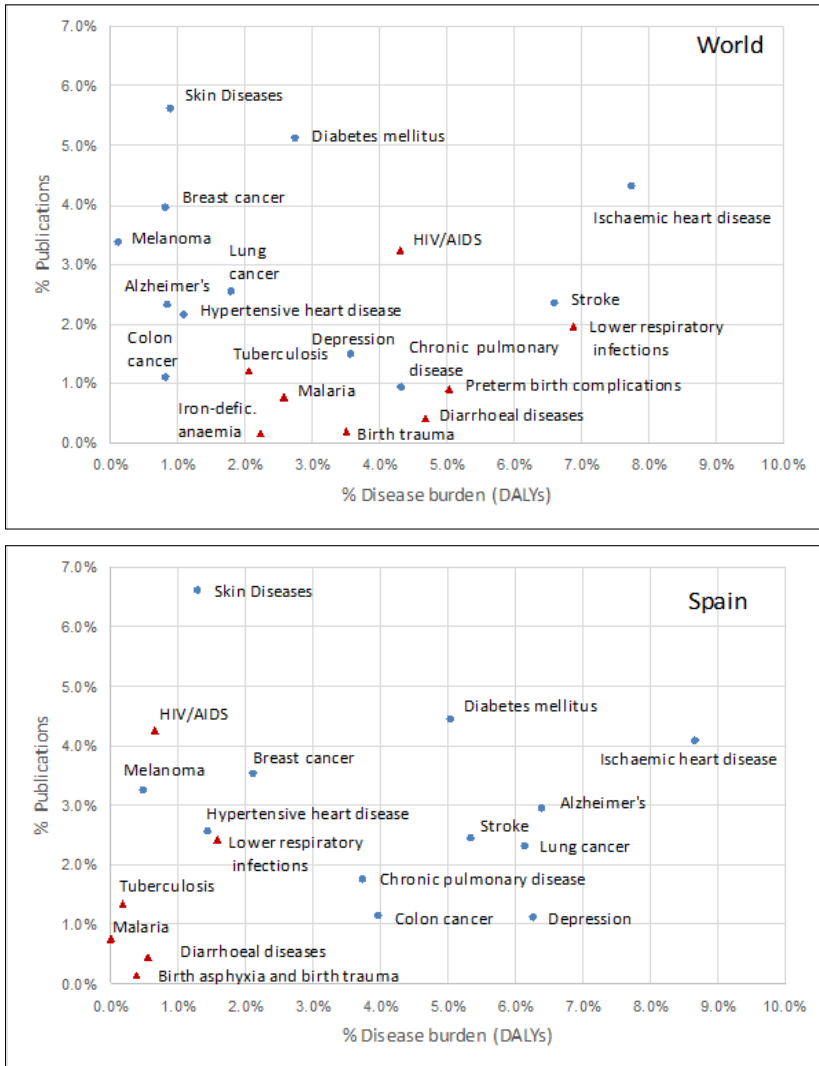
Con el fin de que la I+D sanitaria sea más sensible a las necesidades sociales, es preciso mejorar el proceso de selección de prioridades, incorpo-

rando información sobre necesidades sanitarias y dialogando con pacientes y otros sectores afectados. De acuerdo con los presupuestos de la ciencia abierta y la investigación y la innovación responsables, la priorización de las agendas de investigación debería basarse en dos estrategias principales: en primer lugar, la recogida de los datos disponibles sobre las inversiones en investigación realizadas para abordar un determinado problema sanitario; en segundo lugar, la evaluación de las preferencias de los ciudadanos respecto a los problemas que hay que afrontar y las mejores estrategias para resolverlos.

Referencias bibliográficas

- Cassi, L., Lahatte A., Rafols I., Sautier P. y de Turckheim É. (2017). Improving fitness: mapping research priorities against societal needs on obesity. *Journal of Informetrics* 11(4). Recuperado de <https://www.sciencedirect.com/science/article/abs/pii/S1751157717301542>
- Evans, J.A., Shim J. M. y Ioannidis J. P. A. (2014). Attention to local health burden and the global disparity of health research. *PLOS One*, 9(4). Recuperado de <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090147>
- Gross, C.P., Anderson G. F. y Powe N. R. (1999). The relation between funding by the National Institutes of Health and the burden of disease. *New England Journal of Medicine*, 340(24). Recuperado de <https://www.nejm.org/doi/full/10.1056/NEJM199906173402406>

Gráfico 1. Relación entre la carga de enfermedad, expresada en años de vida ajustados en función de la discapacidad (AVAD), y el porcentaje de publicaciones sobre determinadas enfermedades en el mundo (arriba) y en España (abajo).



EJE TEMÁTICO II

Aproximaciones metodológicas para el abordaje
cienciométrico de la producción en temas locales/regionales

Aspectos metodológicos para la construcción de categorías en temas específicos. El caso de la Nanociencia y la Nanotecnología

*Zaida Chinchilla Rodríguez¹, Teresa Muñoz Ecija²
y Benjamín Vargas Quesada³*

Introducción

La delimitación de campos consiste en la asignación de revistas o publicaciones a campos científicos determinados (Zhao, 2009). Es considerada un área de investigación esencial porque permite responder a las necesidades de información de los investigadores y estudiar la estructura intelectual y cognitiva de los campos científicos, además de sus dinámicas. Por tanto, la calidad de los estudios cientométricos de un campo científico depende de su correcta delimitación.

El principal problema al que nos enfrentamos cuando queremos delimitar un dominio científico es cómo recuperar el conjunto de datos que mejor representa un campo, de manera que este conjunto de datos debe mostrar un alto nivel en sus valores de precisión y exhaustividad (Glänzel, 2015).

Existen diferentes metodologías para la delimitación de campos temáticos. La primera de ellas consiste en la asignación de revistas a un campo temático. La segunda utiliza una estrategia de búsqueda basada en palabras

¹ Instituto de Políticas y Bienes Públicos (IPP), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, España. zaida.chinchilla@csic.es

² Facultad de Documentación y Comunicación, Universidad de Granada, España; teresamunozecija@gmail.com

³ Facultad de Documentación y Comunicación, Universidad de Granada, España; benjamin@ugr.es

clave. La tercera radica en el uso de las relaciones de citación: citación directa, co-citación o emparejamiento bibliográfico. La cuarta opción consiste en la combinación de alguna o todas las metodologías anteriormente descritas.

La Nanociencia y la Nanotecnología (N&N) es una ciencia relativamente nueva, con un marcado carácter interdisciplinar y con un gran potencial económico y social. En N&N, el método más usado para su delimitación ha sido la combinación de diferentes estrategias de búsqueda (utilizando el prefijo nano más palabras claves), junto al análisis de citación. Estas delimitaciones han servido de base para muchos de los estudios cuantitativos en el campo de la N&N, aun mostrando debilidades en los resultados de la delimitación (Huang, Notten, & Rasters, 2011).

El objetivo principal de este estudio es poner de manifiesto la problemática de la delimitación de campos emergentes a través de la delimitación del dominio N&N, con el fin de que pueda servir a otros investigadores que trabajan en la delimitación de campos.

Metodología

Diferentes fueron los materiales utilizados en este estudio. Por un lado, usamos las bases de datos Scopus y Web of Science (WoS). Por otro lado, también utilizamos los índices de revistas Journal Citation Reports (JCR) y SCImago Journal Rank (SJR), en su versión de 2016, ya que ambos incluían la categoría científica Nanociencia y Nanotecnología. En relación al período de tiempo estudiado, se utilizaron dos períodos diferentes. El primer período fue de 2008-2015, debido a la estabilidad que presentó la producción científica de N&N durante esos años (Chinchilla-Rodríguez, Ocaña-Rosa, & Vargas-Quesada, 2016; Muñoz-Écija, Vargas-Quesada, & Chinchilla-Rodríguez, 2017). El segundo período comprendió desde 2000 hasta 2016 y estuvo condicionado por la tercera fuente de datos con la que se trabajó: el sistema de clasificación creado por el Centre for Science and Technology Studies (CWTS). Esta clasificación sólo cubría ese período temporal y utilizaba las publicaciones indexadas en la Web of Science asignándolas a tres niveles jerárquicos: macroclusters, mesoclusters y microclusters, empleando el método de conteo fraccionado (Waltman & van Eck, 2012).

Se emplearon tres aproximaciones para la delimitación de campos: 1) aproximación a nivel de categoría (A_1); 2) aproximación a nivel de publicación (A_2); y 3) aproximación a nivel de palabras clave (A_3).

Aproximación a nivel de categoría (A_1)

Para delimitar a nivel de categoría, seleccionamos las revistas incluidas en la categoría Nanoscience and Nanotechnology de la WoS. Se identificaron 88 revistas. Solo pudimos realizar la delimitación a este nivel en WoS porque Scopus no contenía esta categoría.

Aproximación a nivel de publicación (A_2)

En esta aproximación utilizamos la versión de la WoS del CWTS. Para ello, diseñamos una estrategia de búsqueda basada en palabras clave y limitamos la consulta solo a los artículos y las revisiones publicadas entre 2000 y 2016. Se recuperaron 1.005.801 documentos. Posteriormente, calculamos el porcentaje de solapamiento de los datos recuperados entre el total de publicaciones que conforman cada uno de los microclusters propuestos en el sistema de clasificación del CWTS. Se eligieron los microclusters con un porcentaje de solapamiento mayor o igual al 60%, un total de 35 microclusters. A continuación, calculamos el porcentaje de publicaciones en revistas, dividiendo el número de publicaciones solapadas en las revistas de los 35 microclusters por el número total de publicaciones en revistas de esos microclusters. Finalmente, extrajimos las revistas con un porcentaje de publicaciones superior al 20% y que contuviesen más de 50 publicaciones. En total seleccionamos 76 revistas.

Aproximación a nivel de palabras clave (A_3)

Varios pasos fueron aplicados paralelamente al conjunto de datos de WoS y de Scopus en esta aproximación: 1) recuperación de datos a través de la estrategia de búsqueda (711.464 y 902.082 documentos recuperados de 2008 a 2015); 2) análisis de citación directa de las revistas citadas por la colección y posterior selección de las 100 revistas más citadas; 3) aplicación del método estadístico *scree test* combinado con la regresión de mínimos cuadrados ordinarios, con el fin de elegir el mejor punto de corte para seleccionar las revistas más importantes de la variable; 4) búsqueda en los campos *source/publication* de WoS y Scopus, respectivamente, utilizando el prefijo nano*; 5) recopilación de revistas en las categorías N&N en JCR y SJR, más las revistas del paso 4, eliminando las revistas duplicadas o previamente identificadas en el paso 3; 6) análisis de citación directa de las revistas citadas por las revistas recopiladas en el paso 5; 7) aplicación, nuevamente, del método estadístico *scree test* combinado con la

regresión de mínimos cuadrados ordinarios, para volver a elegir las revistas más importantes. Por último, fusionamos las revistas de los pasos 3, 4 y 7, y la suma resultante de esta fusión fueron consideradas como las revistas más importantes de N&N, 113 revistas en total.

Precisión y exhaustividad

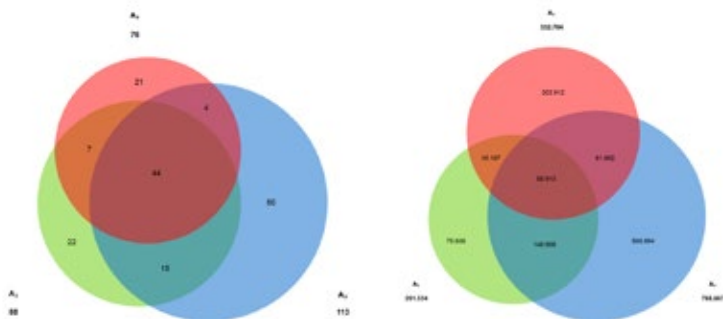
La precisión a nivel de revistas se calculó estudiando la sección *Aim & Scope* de cada una de las revistas identificadas y asignando un valor entre 0 y 1 a cada una de las revistas. La precisión a nivel de publicaciones se calculó con la ayuda de expertos en N&N. Para ello, se diseñó un cuestionario de preguntas cerradas compuesto de una muestra aleatoria de 50 publicaciones, en el que solicitábamos a los expertos que valorasen los documentos como relevantes o no relevantes, atendiendo a la definición de N&N propuesta por el *National Nanotechnology Initiative*. Para el envío de los cuestionarios se utilizó el software Qualtrics.

La exhaustividad se calculó considerando la exhaustividad total al conjunto de revistas identificadas en las 3 aproximaciones. En el caso de las publicaciones, se consideró la exhaustividad total al total de las publicaciones de las revistas identificadas por las 3 aproximaciones.

Resultados y discusión

La Figura 1 muestra que un total de 44 revistas fueron coincidentes en las 3 aproximaciones. La aproximación a nivel de palabras clave es la que identificó un número mayor de revistas. Posiblemente esto se debe a la mayor cobertura de revistas que presenta la base de datos Scopus. El solapamiento de las publicaciones entre las tres aproximaciones fue menor que el solapamiento entre las publicaciones de la aproximación a nivel de categorías y la aproximación a nivel de palabras clave.

Figura1. Distribución del total de revistas y publicaciones identificadas por cada una de las aproximaciones y los valores de solapamiento entre ellas



Atendiendo a los niveles de precisión y exhaustividad (Tabla 1), pudimos observar como no existen grandes diferencias en los valores obtenidos. A nivel de revistas la precisión es más alta que a nivel de publicaciones. Igualmente ocurre con la exhaustividad. La tendencia aquí fue que la aproximación a nivel de palabras clave presentó los valores más altos, excepto en la precisión a nivel de publicaciones de la aproximación 2.

Tabla 1. Precisión y exhaustividad a nivel de revistas (R) y de publicaciones (P)

	Precisión (R)	Precisión (P)	Exhaustividad (R)	Exhaustividad (P)
A ₁	0,87	0,47	0,49	0,34
A ₂	0,86	0,62	0,44	0,34
A ₃	0,89	0,56	0,64	0,46

Analizando los cuestionarios, pudimos observar un alto nivel de desacuerdo entre los expertos. Para calcular el nivel de acuerdo utilizamos el indicador estadístico *fleiss kappa*, cuyo valor oscila entre 0 y 1. Este valor fue de 0,27, lo que indica un ajustado nivel de acuerdo entre los expertos, siendo los nanotecnólogos especializados en física los que mostraron un nivel de acuerdo mayor.

Conclusión

Con este trabajo hemos intentado sintetizar la dificultad de la delimitación de campos. Confirmamos que existe una problemática a nivel cuantitativo cuando queremos delimitar campos científicos. A pesar de la aplicación de

diferentes métodos para delimitar campos científicos, la utilización de diferentes fuentes de datos, el análisis de las relaciones de citación y medidas como la precisión y la exhaustividad, y validación de resultados por parte de expertos, sigue estando latente que la delimitación de campos es problemática y seguirá siéndolo. No existe «el mejor y único método» que nos permita obtener los mejores resultados, por ejemplo, si atendemos a las medidas de precisión y exhaustividad para evaluar la pertinencia de la delimitación. Además, es difícil obtener una delimitación precisa, incluso cuando se aplican mejoras o se combinan metodologías.

Tampoco podemos olvidar que existe la problemática a nivel cualitativo, pues cuando validamos los resultados con la ayuda de expertos en la materia, el nivel de acuerdo es muy bajo. La percepción de los expertos tiende a ser subjetiva y, dependiendo de su especialidad, van a definir un documento como relevante o como no relevante.

En resumen, este estudio ha intentado mostrar varias aproximaciones para la delimitación temática. No podemos proponer ninguno de ellos como el más adecuado. Cada uno de los métodos a utilizar será más o menos adecuado, dependiendo del campo a determinar y/o del objetivo que se pretenda conseguir.

Referencias bibliográficas

- Chinchilla-Rodríguez, Z., Ocaña-Rosa, K., y Vargas-Quesada, B. (2016). How to Combine Research Guarantor and Collaboration Patterns to Measure Scientific Performance of Countries in Scientific Fields: Nanoscience and Nanotechnology as a Case Study. *Frontiers in Research Metrics and Analytics*, 1, 2. <https://doi.org/10.3389/frma.2016.00002>
- Glänzel, W. (2015). Bibliometrics-aided retrieval: where information retrieval meets scientometrics. *Scientometrics*, 102(3), 2215–2222. <https://doi.org/10.1007/s11192-014-1480-7>
- Huang, C., Notten, A., & Rasters, N. (2011). Nanoscience and technology publications and patents: a review of social science studies and search strategies. *The Journal of Technology Transfer*, 36(2), 145–172. <https://doi.org/10.1007/s10961-009-9149-8>
- Muñoz-Écija, T., Vargas-Quesada, B., y Chinchilla-Rodríguez, Z. (2017). Identification and visualization of the intellectual structure and the main research lines in nanoscience and nanotechnology at the worldwide level.

Journal of Nanoparticle Research, 19(2), 62. <https://doi.org/10.1007/s11051-016-3732-3>

Waltman, L., y van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science: A New Methodology for Constructing a Publication-Level Classification System of Science. Journal of the American Society for Information Science and Technology, 63(12), 2378–2392. <https://doi.org/10.1002/asi.22748>

Zhao, D. (2009). Mapping library and information science: Does field delineation matter? Proceedings of the American Society for Information Science and Technology, 46(1), 1–11.

La recuperación de información por delimitadores geográficos y su aplicación en estudios bibliométricos de ciencia local

Claudia M. González¹, Gustavo Archuby² y Sandra Miguel³

Introducción

La vinculación entre lenguaje y espacio geográfico tiene su complejidad. El espacio físico de la tierra poblada tomada en el contexto del cosmos, puede referenciarse de diferentes maneras. A veces se usa la geografía física, otras la geografía política, pero también hay geografía en un sinnúmero de situaciones tales como la mención de áreas sanitarias, distritos de votación, zonas inundables; además de declaraciones explícitas que pueden tomar la forma de topónimos reconocidos, coordenadas geográficas, códigos postales o prefijos telefónicos.

En el discurso científico de tipo referencial -entendiendo como tal a los registros bibliográficos con resumen de las producciones científicas- se encuentran expresiones de lo geográfico muy variadas, que dependiendo del área disciplinar pueden presentar diferente grado de ambigüedad o vaguedad. Sin embargo, el esfuerzo puesto en la identificación de dichas expresiones

¹ Instituto de Investigaciones en Humanidades y Cs Sociales (UNLP/CONICET), La Plata, Argentina. cgonzalez@fahce.unlp.edu.ar

² Facultad de Humanidades y Cs de la Educación (UNLP), La Plata, Argentina. gustavo@fahce.unlp.edu.ar

³ Instituto de Investigaciones en Humanidades y Cs Sociales (UNLP/CONICET), La Plata, Argentina. smiguel@fahce.unlp.edu.ar

puede verse retribuido si a partir de ellas encontramos nuevas maneras de visualizar los resultados de la actividad científica, permitiendo demostrar el esfuerzo que un colectivo humano de investigación realiza para solucionar los problemas del propio territorio.

Para abordar esta problemática, el trabajar con los topónimos es un primer paso. En esta ponencia presentamos algunos avances a los que hemos arribado en relación a su uso para la selección de la producción bibliográfica pertinente y luego a la manera en que dicha producción puede representarse sobre un mapa.

Metodología

Los topónimos y la selección de la producción

La metodología bibliométrica centra su atención en el análisis del tamaño, crecimiento y distribución de la bibliografía científica en diferentes tipos y niveles de agregación temáticos, institucionales, geográficos, etc. (Okubo, 1997) por una parte, y en el estudio de la estructura social de los grupos que la producen y la utilizan por otra (López Piñero, 1972). Como punto de partida siempre se considera que la producción se encuentra almacenada en alguna fuente de datos que posee cierto nivel de normalización y que cuenta con interfaces de acceso a dicha información ya sea en forma de motor de búsqueda o de servicio de tipo Interfaz de Programación de Aplicaciones (*API*). La elección de las fuentes a utilizar suele ser un asunto controvertido ya que no existe una sola fuente que sea totalmente comprensiva. Las más usadas son *Web of Science* y *Scopus*, a nivel multidisciplinar mundial, *Google Scholar* y actualmente *Dimensions* para los estudios webmétricos. Para estudios relacionados con Latinoamérica, se ha propuesto el uso de *Scielo* y *Redalyc* como las más representativas de la región.

La estructura básica de los registros bibliográficos que ofrecen las fuentes nos brindan cuatro campos de datos con capacidad de contener topónimos: el título, el resumen, las palabras claves y la afiliación de los autores. Mientras la selección de registros basada en la mención de un topónimo (por ejemplo el nombre de un país) en el campo afiliación implica delimitar la producción científica “del país”, es decir la que producen sus autores; seleccionar los registros de acuerdo a la misma mención en cualquiera de los otros 3 campos de datos, implica estar seleccionando la producción de investigación

que se realiza “sobre un país”. Luego, la combinación de diferentes formas de estas estrategias nos permite obtener, por ejemplo, la producción del país que menciona al país, la producción del país que no menciona al país, la producción que menciona al país escrita por extranjeros (Miguel, González y Chinchilla-Rodríguez, 2012).

La selección del/los topónimos a utilizar para la recuperación de registros tiene al menos dos aspectos problemáticos. Por un lado se debe decidir hasta que nivel de subagregados se utilizarán. Así, si el nivel de agregación elegido es claramente administrativo (caso país), es fácil de resolver ya que se piensa en términos de los subagregados que lo integran: provincias, estados, etc., solo conlleva decidir hasta que nivel de profundidad se desea llegar. El otro aspecto es que hay que ser cuidadoso al momento de elegir la forma completa o truncada del topónimo garantizando la mayor exhaustividad sin pérdida de precisión, ya que las designaciones ambiguas producen recuperación de registros no deseados que luego deberán revisarse manualmente para su descarte. (Miguel, González e Hidalgo, 2013)

Reconocimiento automático de topónimos

Una vez que se ha realizado la selección de los registros de la fuente de acuerdo a alguno de los criterios presentados en el apartado anterior, tenemos conformado un corpus textual susceptible de ser trabajado con técnicas de Procesamiento del Lenguaje Natural (NPL). Se pueden usar para identificar los temas de investigación, o para el caso que nos interesa aquí que consiste en la extracción de información que hace referencia al espacio o territorio.

Dentro de las técnicas que se pueden utilizar están las denominadas NERC (Named Entity Recognition and Classification), que se emplean para reconocer automáticamente en textos los nombres de entidades, entendiendo como tales los nombres propios de organizaciones, personas, localizaciones, y entidades numéricas tales como fechas, tiempo, montos, expresiones porcentuales. Existen diferentes aproximaciones heurísticas para su implementación. Algunas formas se encuadran dentro de lo que se denomina Aprendizaje Supervisado, tales como los Modelos Ocultos de Markov (HMM), los Árboles de Decisión, Modelos de Máxima Entropía, Máquinas de Vectores de Soporte (SVM) y Campo Aleatorio Condicional (CRF) (Nadeau y Sekine, 2007). Son todas variaciones que consisten en leer un corpus anotado, memorizar las listas

de entidades y crear reglas de desambiguación basadas en las características discriminativas. Un método de referencia que a menudo se propone consiste en etiquetar palabras de un corpus de prueba cuando están anotadas como entidades en el corpus de entrenamiento.

En este trabajo se realizó una prueba de evaluación utilizando una aplicación desarrollada por el Grupo de procesamiento del Lenguaje Natural de la Universidad de Stanford denominada StanfordNE (Stanford NLP, 2018), que es una implementación en Java basada en CRF. La aplicación cuenta con tres modelos que fueron entrenados con una mezcla de los corpus anotados para nombres de entidades denominados CoNLL, MUC-6, MUC-7 y ACE. El modelo Clase 3 reconoce localizaciones, personas y organizaciones; el modelo clase 4 reconoce localizaciones, personas, organizaciones y misceláneas y el modelo Clase 7 reconoce localizaciones, personas, organizaciones, monedas, porcentajes, fechas y tiempos. El objetivo en este trabajo es evaluar el desempeño de los tres modelos sobre un corpus muestral de registros, de tal manera que nos permita decidir cuál de los tres utilizar para la etapa de entrenamiento.

Las medidas utilizadas para la evaluación son la Exhaustividad (*Recall*) y Precisión, medidas clásicas utilizadas para la evaluación de la recuperación de información desde el pionero Proyecto Cranfield (Cleverdon, 1966). También utilizamos la medida conocida como F1, o score F1, que permite dar cuenta del balance que existe entre las otras dos:

Precisión: $\text{Verdaderos positivos} / \text{Verdaderos Positivos} + \text{Falsos positivos}$

Exhaustividad (*Recall*): $\text{Verdaderos positivos} / \text{Verdaderos positivos} + \text{Falsos negativos}$

F1: $2 * (\text{Precisión} * \text{Exhaustividad} / \text{Precisión} + \text{Exhaustividad})$

De esta manera se puede estimar de manera comparativa el desempeño de cada uno de los modelos en la identificación y anotación automática de los nombres de localidades.

Resolución de topónimos y elaboración de mapas

Una vez que se tienen marcados los topónimos, la función de resolución es mapear esos topónimos del documento con los topónimos contenidos en otro documento de referencia (Gazetteer), que además tiene resueltas las coordenadas geográficas. El sistema de coordenadas es prácticamente universal,

donde un punto $P = \varphi, \lambda$, en el que φ es la latitud, es decir el ángulo entre el punto y el ecuador, medido en grados entre -90 y + 90; y λ es la longitud, el ángulo este oeste del punto 0 por acuerdo internacional que es el meridiano de Grenweech, medido en grados entre -180 + 180. Si bien la latitud y la longitud son siempre relativas a un sistema geodésico de referencia, es decir a un modelo matemático que aproxima a la forma de la tierra (como esfera, elipsoide o geode), el sistema WG84 (World Geodesic System) es el más usado.

El problema que se presenta al tratar de realizar mapas basados en topónimos y en información bibliométrica -que cuenta la cantidad de publicaciones- es resolver visualmente de manera óptima el anidamiento geográfico y la densidad de publicación por áreas temáticas.

Resultados y discusión

Los topónimos y la selección de la producción

Se muestran los resultados de aplicar diferentes estrategias de búsqueda para el caso de la producción Argentina que contiene como topónimo el nombre del país, de alguna de sus provincias o de una región (Bonaerense, Cuyo, Patagónica, Pampeana) en el título, resumen o palabras claves para el periodo 2007-2016.

Tabla 1. Resultados de diferentes estrategias de búsqueda en Scopus

1	argentin*	27938	91,31 %
	[Provs] NOT argentin*	1549	5,06 %
	[Region] NOT argentin*	1222	3,99 %
		30709	100 %
2	argentin* AND [Provs]	13026	42,42 %
	argentin* AND [Region]	3710	12,08 %
3	argentin* OR [Provs]	29548	96,22 %
4	argentin* OR [Provs] OR [Region]	30598	99,64 %

Cuando se piensan las estrategias de búsqueda lo importante es poner en consideración que los nombres de los topónimos sean únicos de la región. Por ejemplo, las provincias como Córdoba o Santa Cruz pueden recuperar registros que tienen que ver con otras regiones del mundo, España y Bolivia, en este caso. Cuando se tiene conocimiento de casos tan evidentes, será conveniente ajustar la expresión excluyendo dichos países.

Reconocimiento automático de topónimos

Luego de aplicar la estrategia de búsqueda 4 de la tabla 1, la muestra de registros que se utilizó para la evaluación de los 3 modelos de StanfordNER sin entrenamiento fueron 372 casos de un total de 12379 correspondientes al área de Agricultura y Ciencias Biológicas.

Después de ejecutar los 3 algoritmos se revisaron manualmente los registros de la muestra para realizar la anotación y el conteo de los casos positivos y negativos.

Tabla2. Resultado de las medidas de evaluación para los 3 modelos

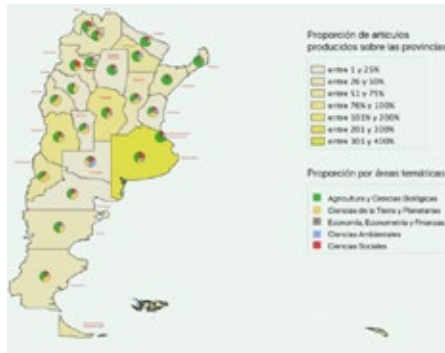
	VP	FP	FN	PRECISION	RECALL	F1
Mod 4	843	808	232	0,5106	0,7842	0,6185
Mod 3	766	257	326	0,7488	0,7015	0,7243
Mod 7	733	63	394	0,9209	0,6504	0,7624

Como se puede observar, el mejor desempeño se obtuvo con el Modelo 7 (F1=0,7624). Recordemos que cuanto más precisos se vuelven los sistemas, por lo general pierden exhaustividad y a la inversa. En este caso, se observa que los modelos 4 y 7 tienen valores menos armónicos que el modelo 3. Mientras el Modelo 4 es más exhaustivo y menos preciso, el Modelo 7 es más preciso y menos exhaustivo.

Resolución de topónimos

Se muestra aquí un mapa que muestra la producción científica obtenida con una estrategia similar a la de la búsqueda 4 (ver Tabla 1) para el periodo 2008-2012.

Figura 1. Producción asociada a temas de investigación locales (Argentina 2008-2012)



En el mismo se puede observar el resultado de asignar a la capa de las provincias Argentinas los atributos de datos totales de producción y los mismos discriminados por áreas temáticas. Los primeros se manifiestan en la diferente intensidad de colores que toma cada provincia en el mapa, los segundos en los gráficos circulares. Así, tenemos provincias con gráficos con preponderancia del área Agricultura y Ciencias Biológicas, como es el caso de Misiones y Corrientes (Esteros del Iberá), o un gráfico con preponderancia de las Cs de la Tierra y Planetarias como es el caso de la zona petrolera de Neuquén.

Conclusiones

Si bien la frase “ciencia local” puede presentar interpretaciones ambiguas, en este trabajo se le ha dado la connotación de ser la ciencia que produce un país sobre cuestiones que de alguna manera involucran a su propio territorio. Si bien los estudios de producción científica realizados con este recorte no cubren la totalidad de la ciencia de interés local, se considera que es una aproximación a su delimitación.

Trabajar con la representación espacial que aparece mencionada en el discurso científico se presenta como una oportunidad poco explorada en los estudios Bibliométricos. La capacidad de aportar información valiosa implica encontrar técnicas automáticas o semi-automáticas de tratamiento del lenguaje natural que permitan explotar corpus referenciales grandes.

De un tiempo a esta parte, la cienciometría ha puesto empeño en encontrar formas de visualización con mayor potencia explicativa. Los Sistemas de Información Geográfica pueden ser un aporte para reflejar distintos aspectos de la relación ciencia/territorio. En el caso particular de los temas de investigación, un primer paso para realizar su georeferenciamiento proviene de la identificación de los topónimos mencionados en el texto.

Referencias bibliográficas

- Cleverdon, C.W., Mills, J.Y Keen, E.M. (1966). *ASLIB Cranfield project: Factors Determining the performance indexing Systems*. Cranfield: ASLIB.
- López Piñero, J. M. (1972). *El análisis estadístico y sociométrico de la literatura científica*. Valencia: Facultad de Medicina. Centro de Docum. e Informática Médica.

- Miguel, S, González, C.M. y Chinchilla-Rodríguez, Z. (2012). Lo local y lo global en la producción científica argentina con visibilidad en Scopus, 2008-2012. *Información, Cultura y Sociedad*, 32, 59-78. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/1375/1352>
- Miguel, S, González, C.M. e Hidalgo, M. (2013). Argentina como objeto de investigación. Reflexiones conceptuales y aproximaciones metodológicas para el abordaje bibliométrico de la producción científica sobre temas de alcance nacional. Trabajo presentado en Actas de las 3ras Jornadas de Intercambio y Reflexiones de la Investigación en Bibliotecología. La Plata: UNLP. Recuperado de http://jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2013/actas-2013/miguel_gonzalez_hidalgo.pdf/view
- Nadeau, D. y Sekine, S. (2007). A survey of named entity recognition and classification. *Linguistic e Investigations*, 30(1), 3-26.
- Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*. (Report No. STI Working Papers 1997/1). Paris: OECD. <https://doi.org/10.1787/18151965>
- Stanford N.L.P. (2018). Recuperado de <https://nlp.stanford.edu/software/CRF-NER.shtml>.

El análisis y representación del contenido de la producción científica desde una perspectiva informétrica: aportes metodológicos

Gustavo Liberatore¹

Introducción

El “paradigma del análisis de dominio” es un enfoque teórico de la Ciencia de la Información que establece que la mejor forma de comprender la organización de los campos de conocimiento es delimitando y estudiando las distintas “comunidades discursivas” que los componen. Dichas comunidades deben abordarse desde una perspectiva holística, a partir de los contextos históricos, sociales y culturales en que se constituyen. Desde esta perspectiva, el conocimiento puede ser organizado y representado dinámicamente desde múltiples variables: la organización social del trabajo, los patrones de cooperación, el lenguaje, las formas de comunicación, los sistemas de información y los criterios subjetivos de relevancia. (Hjørland y Albrechtsen, 1995; Hjørland, 2002).

Dentro de este marco se proponen un conjunto de métodos y técnicas empíricas orientadas al estudio y descripción de los campos de conocimiento a partir del análisis de las relaciones que los estructuran (conceptos, entidades) y de sus vinculaciones con otros campos (Smiraglia, 2015). De entre ellos se destaca el método bibliométrico como el de mayor aporte en la delimitación y estudio de los campos de conocimiento tomando como unidades de análisis a los autores, los contenidos y las citas. Complementariamente, distintas

¹ Depto. Ciencia de la Información. Fac. Humanidades. Universidad Nacional de Mar del Plata, Mar del Plata, Argentina. gliberat@mdp.edu.ar

técnicas de visualización de información y de dominios científicos proporcionan formas de representación de la información complejas que aportan claves de interpretación ocultas para las técnicas y métodos más tradicionales (Díaz-Pérez, Moya-Anegón y Carrillo-Calvet, 2017).

Teniendo en cuenta esta perspectiva se pretende abordar una problemática específica de las investigaciones empíricas realizadas con base en la producción científica que se relaciona con la captura, análisis y representación de la investigación científica aplicada a entornos geográficos específicos. El desafío aquí es delimitar y representar de manera combinada dos dominios diferentes: el geográfico y el disciplinar o científico.

Bajo este marco se tomó como objeto de estudio la producción científica indizada en la corriente principal sobre medio ambiente aplicada a la Argentina, en donde al menos uno de los autores de las publicaciones reportara una afiliación en dicho país. A partir de esta delimitación se fijaron una serie de interrogantes que guiaron esta investigación: ¿Cuáles son las áreas geográficas más estudiadas en la Argentina en relación al medio ambiente? ¿Cuáles son las temáticas mediambientales investigadas? ¿Cómo se distribuyen las temáticas en función de las áreas geográficas?

Materiales y Metodología

El corpus bibliográfico fue seleccionado de la base de datos SCOPUS tomando en consideración los artículos publicados en el periodo 2007-2016. Se recuperaron 3672 registros en base a la definición de la siguiente ecuación de búsqueda²:

```
(TITLE-ABS-KEY (argentin*) AND AFFILCOUNTRY (argentina))  
AND PUBYEAR >2006 AND PUBYEAR <2017 AND (LIMIT-TO(-  
SUBJAREA,"ENVI"))
```

Luego de la exportación de los resultados en formato *BibTex* se procedió a analizar el contenido tomando en consideración el campo de palabras clave de autor (*authorkeywords*) a partir del cual se construyó un índice de 9689 entradas o expresiones. Con este resultado se aplicaron una serie de cálculos con vista a la representación de las variables temática y geográfica.

2 El área temática de búsqueda es la proporcionada por la clasificación interna de SCOPUS denominada "EnvironmentalScience".

Dentro del conjunto seleccionado de palabras clave se contabilizaron las frecuencias absolutas y luego se aplicó el cálculo de co-ocurrencias (*co-word analysis*) con el objetivo de generar una matriz de datos para la representación de las relaciones correspondientes.

Para analizar y representar las áreas geográficas involucradas de Argentina se procedió a una detección manual de todas aquellas expresiones que remitiesen a una ubicación geográfica específica. Se aislaron 514 palabras clave con este perfil. Para su organización se realizó una lematización geográfica agrupando a las distintas expresiones en regiones y accidentes geográficos y provincias. Como resultado se obtuvo una taxonomía *ad hoc* constituida por 14 regiones y accidentes geográficos y 22 provincias.

Para el cálculo de frecuencias y análisis de co-ocurrencias se utilizó el software BibExcel (Persson) en su versión 2016-02-20. En la tabulación de datos y generación de algunas gráficas se usó el paquete estadístico Excel en su versión 2013. Las matrices de co-ocurrencia fueron representadas por grafos derivados de la aplicación del análisis de redes sociales (ARS). Para la valoración de los componentes de la red se utilizaron medidas de centralidad (grado e intermediación) y de modularidad (análisis de comunidades o clusters). Para la visualización de datos, se ha utilizado el programa Gephi en su versión 0.9.2 (2017), seleccionando para la distribución el algoritmo ForceAtlas2.

Resultados y discusión

El diseño de una taxonomía que agrupara los lugares geográficos hacia donde estaban enfocadas las investigaciones permitió la representación, en primer lugar, de las áreas y sitios de mayor interés en el campo disciplinar estudiado. Así, se pudo verificar que en el grupo *regiones y accidentes geográficos* la Patagonia (18%) y la región Pampeana (7%) concentran la mayor cantidad de publicaciones (Figura 1). En el grupo *Provincias* la Ciudad Autónoma de Buenos Aires³ (11%) y Córdoba (3%) aparecen como los espacios geográficos con mayor presencia.

³ Desde 1994 la Ciudad de Buenos Aires goza de un estatus administrativo similar al de las provincias que constituyen el territorio argentino por lo que, en esta división, se la considera como tal.

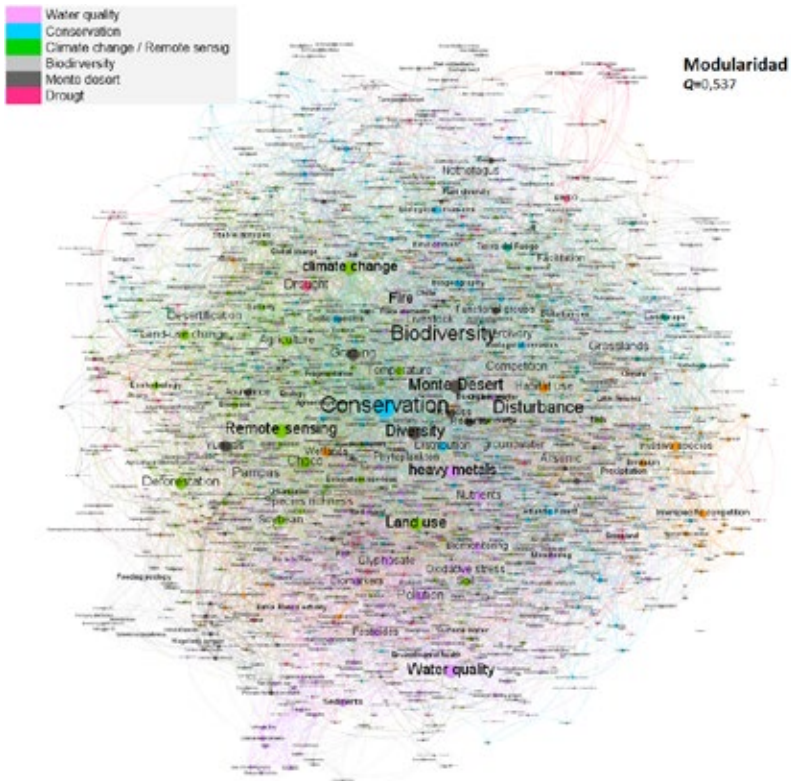
Figura 1. Distribución porcentual de las regiones y accidentes geográficos estudiados en relación al total de publicaciones observadas (n=3672).



Luego de circunscribir el dominio geográfico el análisis se enfocó en el contenido temático de las publicaciones. A partir del cálculo de co-ocurrencias de las palabras clave se aplicó el análisis de redes sociales con la concurrencia de dos medidas específicas: centralidad de grado de los nodos y modularidad de red. Previamente, en la generación del primer grafo se procedió a la poda del nodo *Argentina* para evitar una representación egocéntrica y poco útil a los fines de esta investigación. Luego de equilibrar la red, la primera medida permitió observar cuáles eran las principales temáticas globales del corpus bibliográfico determinadas por el nivel de participación de los nodos (número de enlaces) dentro de la estructura de relaciones. En un segundo momento se aplicó el cálculo de modularidad que, como medida estructural de la red, permite observar si existen módulos o divisiones conformadas por nodos densamente interconectados entre sí y con conexiones más débiles con el resto de los módulos (Figura 2). Esto permitió comprobar que los nodos

con mayor peso o grado calculados previamente definen las principales comunidades o clusters temáticos.

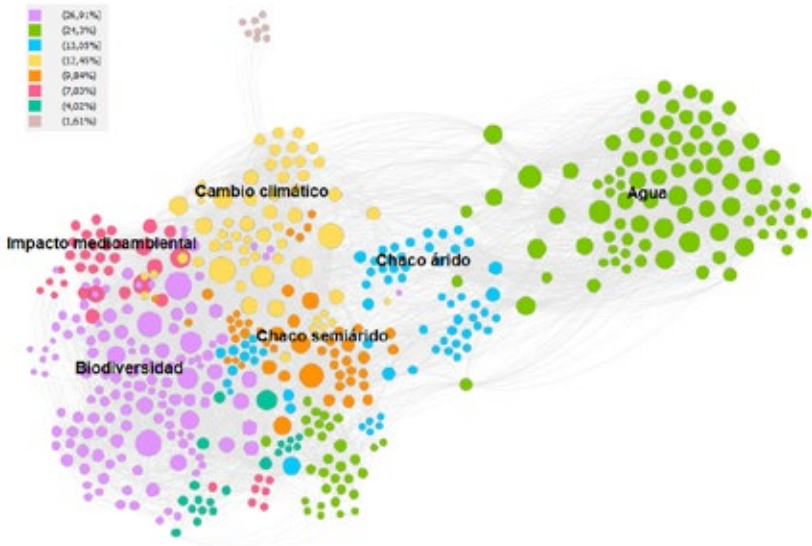
Figura 2. Principales clusters temáticos observados.



A partir de esta delimitación general en lo geográfico y temático se buscó cruzar ambos dominios con el fin de observar la distribución de los tópicos investigados en función de un área geográfica específica. Se tomó en consideración la *Región del Chaco*, a modo de ejemplo, para lo cual se procedió a aislar los registros correspondientes a esta área y realizar un cálculo de co-ocurrencia de las palabras claves y su representación mediante un grafo de red social (Figura 3). A la red resultante se le aplicó nuevamente el cálculo de modularidad a partir del cual pudo discriminarse los principales temas que

concentran la atención en el campo medioambiental. Para una mejor visualización el grafo fue editado solamente con las principales etiquetas⁴ (temas) que representan a las comunidades o clusters observados. Así, la biodiversidad, el cambio climático, el agua y el impacto mediambiental constituyen los principales frentes de investigación, mientras que aparecen en foco dos sub-regiones, el Chaco árido y el semiárido, como áreas geográficas de interés.

Figura 3. Principales temáticas estudiadas en la Región del Chaco.



Conclusión

Dada la naturaleza metodológica de esta breve presentación respecto a las posibilidades o variantes para el análisis y representación de la investigación aplicada a espacios geográficos se postulan tres reflexiones sobre la experiencia presentada:

- La faceta geográfica en un análisis de contenido debe estar sujeta a modelos de representación estandarizados, tanto desde el punto de vista de la captura de datos (búsqueda) cuanto de sus modos de organización (taxonomías).

⁴ Para una mejor comprensión han sido traducidas al español.

- Por el contrario, las temáticas vinculadas a los espacios geográficos se encuentran sujetas a formas discursivas asociadas a los distintos abordajes disciplinares y a la manera en que están relacionadas.
- Finalmente, son los interrogantes (intereses) fijados a priori los que demandarán las formas de representación y visualización más adecuadas para poder aportar respuestas concretas.

Referencias bibliográficas

- Díaz-Pérez, M., Moya-Anegón, F. y Carrillo-Calvet, H. A. (2017). Técnicas para la visualización de dominios científicos y tecnológicos. *Investigación Bibliotecológica*, (Núm. especial de Bibliometría), 17-42. doi: <http://dx.doi.org/10.22201/iibi.24488321xe.2017.nesp1.57884>
- Hjorland, B. (2002) Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.
- Hjorland, B. y Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal Of The American Society For Information Science*, 46(6), 400-425.
- Smiraglia, R. P. (2015). *Domain Analysis for Knowledge Organization: tools for ontology extraction*. Amsterdam: Elsevier-Chandos Publishing.

Aproximación metodológica para la extracción de temas de un corpus bibliográfico referencial a partir del lenguaje natural

Sebastián Varela¹ y Claudia M. González²

Introducción

Se presenta un ejercicio exploratorio de text mining en el que se busca explorar y comparar la producción científica de Argentina y México en el área de las ciencias ambientales. Para ello se utilizaron un conjunto de registros bibliográficos extraídos de la base de datos Scopus en el periodo 2007-2016 en cuyos títulos, resúmenes o palabras claves aparece mencionado el nombre del país.

Metodología

En la base de datos Scopus se encontraron en dicho periodos 5.943 artículos mexicanos y 3.190 argentinos. Los registros fueron descargados en formato BibTeX y convertidos en *data frames* del lenguaje de programación R utilizando el paquete *bibliometrix*.

La unidad de análisis es el resumen o abstract de artículo. Se procedió a desmenuzar el texto en tokens³ individuales -términos en este caso- y a transformarlo en una matriz ordenada de datos (Silge y Robinson, 2017).

¹ varela.sebastian@gmail.com

² Instituto de Investigaciones en Humanidades y Cs Sociales (UNLP/CONICET), La Plata, Argentina. cgonzalez@fahce.unlp.edu.ar

³ Un token es una unidad significativa de texto, y el proceso de tokenización es el proceso de división de un corpus textual en tokens. En este trabajo usamos como tokens en primer lugar terminos, y posteriormente bigramas (grupos de dos palabras).

La matriz de datos resultante tiene 1.931.639 filas (términos), de los cuales 1.233.227 provienen de abstracts mexicanos y 698.412 de argentinos. Posteriormente se aplican diccionarios de stopwords⁴: los diccionarios estándar del paquete *tidytext* y uno propio para eliminar números y otros caracteres inapropiados. Como resultado quedan un total de 73.666 términos, 44.240 procedentes de resúmenes mexicanos y 29.426 de resúmenes argentinos.

Resultados y discusión

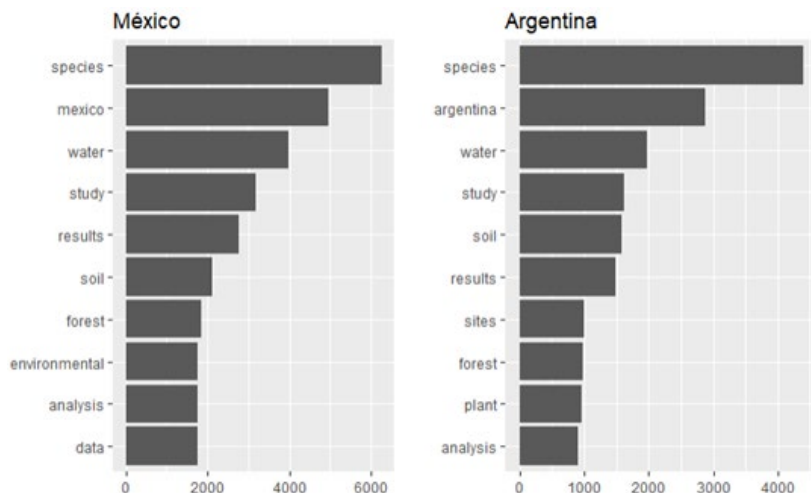
En la Figura 1 podemos observar con gráficos de barras el top ten de los términos más frecuentemente utilizados en la producción académica de ambos países. Llama la atención que no aparezca el término *grassland* (pastizal, pradera), lo que invita a pensar que hay un descuido en la agenda de investigación en relación a ese ecosistema, habida cuenta de que es el ecosistema predominante en Argentina, y muy relevante en México (aunque en este país los bosques y los ambientes áridos tienen mayor proporción)⁵. Sería interesante indagar en qué medida los artículos con las palabras *soil* y *plant* en sus abstracts refieren a estudios sobre *grasslands*.

Los términos *results*, *study*, *data*, *analysis* son palabras típicas del lenguaje científico que no aportan desde el punto de vista disciplinar, y se podrían eliminar del análisis. El examen global de estos gráficos permite inferir que no hay diferencias importantes entre las agendas de investigación de ambos países. Se encuentran los siguientes términos en común: *species*, *water*, *soil*, *forest*. En cambio, el término *environmental* aparece sólo en el corpus mexicano, mientras que *plant* entra en el top ten de Argentina y no en el de México.

⁴ Las stopwords son palabras comunes que son inútiles para el objetivo analítico. Algunos ejemplos en inglés son: “*the*”, “*of*”, “*to*”, “*from*” etc.

⁵ En Argentina *grassland/grassland* -singular y plural sumados- tiene 517 menciones, y en México 193.

Figura 1. Términos más frecuentes



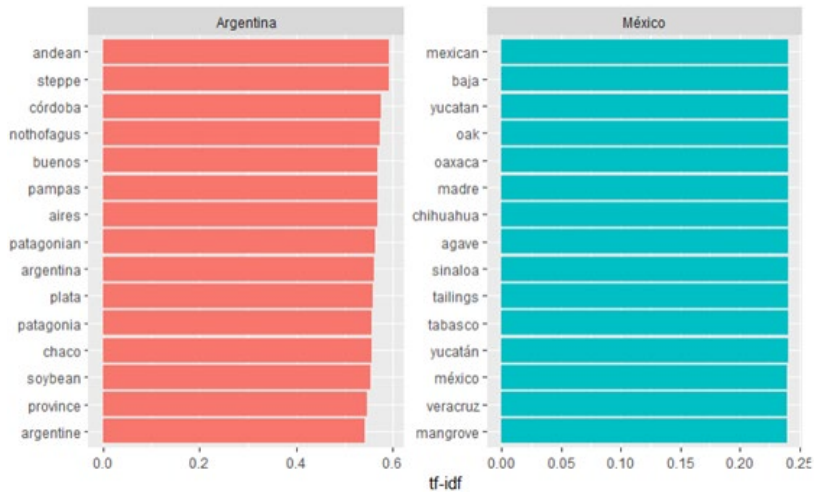
A continuación, en la Figura 2, están las word clouds (nubes de palabras) para cada país. Las nubes de palabras son listados en los que la importancia de los términos se pondera según el tamaño y color de letra. Respecto de estas nubes, cabe decir algo similar a lo dicho respecto de la Figura 1: no se observan mediante diferencias relevantes entre ambos corpus.

Figura 2. Nubes de palabras



Para observar comparativamente las características de cada colección, el paso siguiente sería utilizar el estadístico *tf-idf* (frecuencia de término – frecuencia inversa de documento), un estadístico usado para identificar términos que son especialmente relevantes para un documento (o colección de documentos) particular⁶. Se observan en la Figura 3 los bigramas más frecuentes para cada corpus: en el caso de Argentina aparecen las referencias geográficas más relevantes para la producción de conocimiento medioambiental⁷: *andean*, *cordoba*, *buenos*, *aires*, *patagonian*, *patagonia*, *plata* (referencia al río o cuenca de La Plata), *pampas*, *chaco*. Por otro lado hay referencias a un ecosistema: *steppe*(referencia mayormente a la Patagonia); y finalmente a especies vegetales: *nothofagus* (planta patagónica) y *soybean* (soja), este último término relacionado a la actividad agrícola.

Figura 3. TF-IDF



En el caso de México, las referencias geográficas son *baja* (California), *yucatan* (que aparece con y sin acento), *oaxaca*, *chihuahua*, *sinaloa*, *madre* (Sierra), *tabasco*, y *veracruz*. Por otro lado hay referencias a especies

⁶ Para más detalles sobre este estadístico véase Silge y Robinson (2017).

⁷ Esto puede deberse en parte a que en dichas regiones hay instituciones de investigación importantes dedicadas a la producción de conocimiento en ciencias ambientales.

vegetales: *oak*, *mangrove* y *agave*. Finalmente aparece el término *tailings*, referido a desechos tóxicos de la actividad minera.

Algo llamativo cuando se analizan estos gráficos es que no aparecen las disciplinas del campo ambiental: *physiology*, *ecology*, *restoration*, *conservation*. Se requiere mayor elucidación sobre este punto aunque se puede conjeturar que no aparecen porque en los resúmenes se tiende a mencionar las diversas subdisciplinas y especialidades del campo ambiental.

Continuando el proceso de exploración, se cambia el proceso de tokenización de palabra a bigrama (conjunto de dos palabras), lo cual puede ser más interesante para la detección de temas de investigación. Para facilitar la interpretación se utilizan gráficos de redes de palabras (generados con los paquetes de R *widyr* y *ggraph*). Estas redes tienen la ventaja de permitir visualizar las relaciones entre las palabras más frecuentes de manera simultánea. A partir de la interpretación de los nodos resultantes se pueden inferir temáticas. El orden de aparición de las palabras -o direccionalidad- se indica con una flecha.

En primer lugar cabe mencionar referencias geográficas: de izquierda a derecha aparece una pequeña red de bigramas con referencia geográfica: *south* → *america/south* → *american*, y luego un cluster de bigramas probablemente explicado por la producción de grupos de investigación con referencia geográfica en la provincia de Buenos Aires y la Patagonia norte, aunque esta conjetura requiere una indagación más profunda⁸. Otras referencias geográficas son Tierra del Fuego, que a pesar de ser una provincia pequeña aparece como una provincia relevante para las ciencias ambientales⁹; y Bahía Blanca que también aparece como una referencia geográfica destacada¹⁰.

⁸ Cabe señalar que una parte de dichos trabajos se deben a la producción del Instituto de Investigaciones Fisiológicas y Ecológicas vinculadas a la Agricultura (IFEVA), de la Facultad de Agronomía (UBA).

⁹ En dicha provincia se encuentra el Centro Austral de Investigaciones Científicas (CADIC-CO-NICET).

¹⁰ En dicha ciudad está la Universidad Nacional del Sur e importantes institutos de investigación del CONICET orientados a investigación ambiental y oceánica.

Figura 4. Bigramas corpus argentino



Hay luego pequeños clusters de bigramas que dan cuenta de referencias técnicas: análisis espacial y de mapas (*spatial*, *distribution*, *scales*) y *environmental* → *variables*. Finalmente aparecen los agrupamientos relacionados con *temáticas específicas* de investigación: en primer lugar dos nodos numerosos, uno referido a estudios sobre ecología de comunidades vegetales (*species*, *richness*, *composition*, *plant*, *exotic*, etc.); y otro sobre limnología (*water*, *quality*, *surface*, *drinking*, etc.). También una serie de bigramas de fácil interpretación: contaminación con metales pesados (*heavy* → *metals*); cambio climático (*climate* → *change*); estudios de ecosistemas semi áridos (*semi* → *arid*), estudios del crecimiento de las plantas (*growing* → *season*); estudios sobre montes desérticos (*monte* → *desert*); estudios sobre materia orgánica (*organic* → *matter*); cambio climático (*climate* → *change*), y estudios sobre los parques nacionales (*national* → *park*).

A continuación (Figura 5) se examina la red de bigramas de frecuencias absolutas resultante de los abstracts mexicanos:

Figura 5. Bigramas corpus mexicano



Encontramos en este caso algunos bigramas que sugieren temáticas en común con la producción argentina: *heavy* → *metal*; *organic* → *matter*; *environmental* → *conditions*; *climate* → *change*. Algunas redes de bigramas también son similares sugiriendo temáticas comunes: ecología de comunidades vegetales (*species*, *richness*, *plant*, etc.), y limnología (*water*, *quality*, *drinking* -aunque en este caso se agrega el bigrama *air* → *quality*).

Por otro lado las siguientes son temáticas distintivas de México: diversidad genética (*genetic* → *diversity*); servicios de ecosistemas (*ecosystem* → *services*); cobertura del suelo (*land* → *cover*); bosques de neblina o altura (*cloud* → *forest*) y biosfera de reserva (*biosphere* → *reserve*)

Finalmente las referencias geográficas más frecuentes son *sierra* → *madre*; *yucatan* → *peninsula*; y el cluster que menciona Baja California, México central, Veracruz México y México City. Se ve a continuación una comparación de los corpus interpretando los bigramas.

Figura 6. Comparación temática de los corpus



Conclusión

El ejercicio propuesto permite explorar la producción científica de ambos países una disciplina específica, permitiendo identificar similitudes y diferencias.

El uso de bigramas (y eventualmente otros n-gramas) resulta de potencial utilidad para la identificación de temas de investigación.

Luego de este primer ejercicio exploratorio, resulta de interés propiciar estudios entre la producción científica Argentina y la de países más cercanos (Brasil o Chile), focalizando sobre tópicos menos generales y más acotados, restringiendo el foco a ecosistemas específicos (por ejemplo ambiente acuático o estepa), o bien en temáticas particulares como la contaminación.

La identificación de temas o tópicos en grandes volúmenes de datos sigue siendo un desafío para las ciencias de la información y de la computación. El uso de algoritmos y técnicas estadísticas cada vez más sofisticados parece ser un camino posible.

Referencias bibliográficas

- Aria, M. y Cuccurullo, C. (2017) *bibliometrix: An R-tool for comprehensive science mapping analysis*, Journal of Informetrics, 11(4), pp 959-975, Elsevier.
- Pedersen, T. L. (2017). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. Recuperado de <https://cran.r-project.org/package=ggraph>.
- Silge J, R. D (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *JOSS*, 1(3). Recuperado de <http://dx.doi.org/10.21105/joss.00037>.
- Silge, J. and Robinson, D. (2017). *Text mining with R*. New York: O'Reilly.

EJE TEMÁTICO III

Estudios métricos sobre la producción científica
en temas locales de países iberoamericanos

Encontrar los temas locales en el CV de los investigadores uruguayos del área social

Natalia Aguirre-Ligüera¹ y Exequiel Fontans²

Introducción

Identificar temas locales en la producción científica de los investigadores de los denominados países no centrales resulta clave para comprender algunas dinámicas en la generación de conocimiento a nivel nacional y presenta desafíos metodológicos a salvar.

Como es sabido, en la literatura se plantea a menudo la importancia de los temas locales en la construcción de las agendas de investigación en general y de las ciencias sociales en particular, cuyo objeto de estudio es esa sociedad en la que se instalan problemáticas que urgen ser abordadas (Hicks, 1999, 2004; Larivière, Archambault, Gingras, & Vignola-Gagné, 2006; Moed, Luwel, & Nederhof, 2002) namely: the centrality of books in social science literature and their high citation rate; and the national orientation of social science literatures. The paper then looks at reasons why social science bibliometrics holds increasing promise, namely: increasing internationalization; and good coverage of scholarly journals.”,”author”:[{“dropping-particle”:.”,”family”:”Hicks”,”given”:”Diana”,”non-dropping-particle”:.”,”parse-names”:-false,”suffix”:.””}],”container-title”:”Scientometrics”,”id”:”ITEM-1”,”issue”:”2”,”issued”:{“date-parts”:[["1999"]]},”page”:”193-215”,”title”:”The

¹ Departamento Tratamiento y Transferencia de información, Facultad de Información y Comunicación. Universidad de la República, Montevideo, Uruguay natalia.aguirre@fic.edu.uy

² Departamento Tratamiento y Transferencia de información, Facultad de Información y Comunicación. Universidad de la República, Montevideo, Uruguay. exequiel.fontans@fic.edu.uy

difficulty of achieving full coverage of international social science literature and the bibliometric consequences”, ”type”: ”article-journal”, ”volume”: ”44”}, ”uris”: [”http://www.mendeley.com/documents/?uuiid=424b917c-439e-4f6c-8e80-e28328a21326”]], {”id”: ”ITEM-2”, ”itemData”: {”DOI”: ”10.1007/1-4020-2755-9”, ”ISBN”: ”978-1-4020-2755-0”, ”abstract”: ”This chapter reviews bibliometric studies of the social sciences and humanities. SSCI bibliometrics will work reasonably well in economics and psychology, whose literatures share many characteristics with science, and less well in sociology, characterised by a typical social science literature. The premise of the chapter is that quantitative evaluation of research output faces severe methodological difficulties in fields whose literature differs in nature from scientific literature. Bibliometric evaluations are based on international journal literature indexed in the SSCI, but social scientists also publish books, write for national journals and for the non-scholarly press. These literatures form distinct, yet partially overlapping worlds, each serving a different purpose. For example, national journals communicate with a local scholarly community, and the non-scholarly press represents research in interaction with contexts of application. Each literature is more trans-disciplinary than its scientific counterpart, which itself poses methodological challenges. The nature and role of each of the literatures will be explored here, and the chapter will argue that by ignoring the three other literatures of social science fields.”}, ”author”: [{”dropping-particle”: ””, ”family”: ”Hicks”, ”given”: ”-Diana”, ”non-dropping-particle”: ””, ”parse-names”: false, ”suffix”: ””}], ”container-title”: ”Handbook of Quantitative Science and Technology Research / ed. Henk Moed, Kluwer Academic”, ”id”: ”ITEM-2”, ”issue”: ”2001”, ”issued”: {”date-parts”: [[”2004”]]}, ”page”: ”473-496”, ”title”: ”The four literatures of social science”, ”type”: ”chapter”}, ”uris”: [”http://www.mendeley.com/documents/?uuiid=05db1267-afed-40a8-a4cd-7fd3ac-70f972”]], {”id”: ”ITEM-3”, ”itemData”: {”author”: [{”dropping-particle”: ””, ”family”: ”Moed”, ”given”: ”Henk F.”, ”non-dropping-particle”: ””, ”parse-names”: false, ”suffix”: ””}], {”dropping-particle”: ””, ”family”: ”Luwel”, ”given”: ”Marc”, ”non-dropping-particle”: ””, ”parse-names”: false, ”suffix”: ””}], {”dropping-particle”: ””, ”family”: ”Nederhof”, ”given”: ”Anton J.”, ”non-dropping-particle”: ””, ”parse-names”: false, ”suffix”: ””}], ”-

container-title": "Library Trends", "id": "ITEM-3", "issue": "3", "issued": {"date-parts": [{"2002"}]}, "page": "498-520", "title": "Towards research performance in the Humanities", "type": "article-journal", "volume": "50", "uris": [{"http://www.mendeley.com/documents/?uuiid=a34bd08e-5496-499c-8208-0ce04879ff07"}], {"id": "ITEM-4", "itemData": {"DOI": "10.1002/asi.20349", "author": [{"dropping-particle": "", "family": "Larivière", "given": "Vincent", "non-dropping-particle": "", "parse-names": false, "suffix": ""}, {"dropping-particle": "", "family": "Archambault", "given": "Éric", "non-dropping-particle": "", "parse-names": false, "suffix": ""}, {"dropping-particle": "", "family": "Gingras", "given": "Yves", "non-dropping-particle": "", "parse-names": false, "suffix": ""}, {"dropping-particle": "", "family": "Vignola-Gagné", "given": "Étienne", "non-dropping-particle": "", "parse-names": false, "suffix": ""}], "container-title": "Journal of the American Society for Science and Technology", "id": "ITEM-4", "issue": "8", "issued": {"date-parts": [{"2006"}]}, "page": "997-1004", "title": "Measuring use and creation of open educational resources in higher education", "type": "article-journal", "volume": "57", "uris": [{"http://www.mendeley.com/documents/?uuiid=533b1d8f-5d89-44aa-9a19-3bd97f6c23ca"}]}, "mendeley": {"formattedCitation": "(Hicks, 1999, 2004; Larivière, Archambault, Gingras, & Vignola-Gagné, 2006; Moed, Luwel, & Nederhof, 2002. Ciertas demandas en torno a temas sensibles ingresan a las universidades empujadas por la sociedad civil, los estados, los organismos financiadores, los medios de comunicación y otros actores no académicos. Algunas de estas temáticas revisten un carácter muy restringido al ámbito local, -en el caso de Uruguay podría pensarse en el Plan Ceibal o el PANES (Plan de Asistencia Nacional a la Emergencia Social)-, mientras que otras son más amplias, y aunque concitan atención nacional, tienen interés para la región e incluso para el mundo -por ejemplo: la instalación de plantas de celulosa; la utilización de agrotóxicos para la producción agropecuaria-. Estos temas demandan la intervención de la academia y la generación de conocimiento genuino que permita dar respuesta o alternativas a estos asuntos de interés social.

Este conocimiento no debería necesariamente provocar escaso interés a nivel global, dependiendo de cuán local sea la temática y también de cómo se comunique. Sin embargo, es frecuente que estos trabajos "locales" sean

más difíciles de publicar en el ámbito internacional. Esto produce una tensión entre los sistemas de evaluación de la investigación y los instrumentos de incentivo a la investigación. Los primeros tienden a premiar las publicaciones en revistas de corriente principal y a estimar poco las contribuciones nacionales, mientras que entre los segundos destacan los que estimulan la producción de conocimiento sobre temas de interés nacional. Entre éstos destaca el Programa Vinculación Universidad-Sociedad y Producción³, o Proyectos Orientados a la Inclusión Social⁴, financiados por la Comisión Sectorial de Investigación Científica de la Universidad de la República; o los instrumentos de apoyo de la Agencia Nacional de Investigación e Innovación orientados a problemáticas de interés nacional⁵, por ejemplo: Equidad Territorial; Descentralización; Educación; Salud; Energía; Agro, entre otros.

Identificar los temas locales en la producción científica es un desafío mayor, ya que no se han desarrollado aún formas automáticas o semiautomáticas para esta clasificación. Miguel, González y Chinchilla(2015)proponen una metodología a través de la detección de topónimos en los títulos, resúmenes y palabras claves de artículos indizados en SCOPUS. En este trabajo se emula esta metodología y se la aplica al currículum vitae (CV) estandarizado de los investigadores uruguayos del área social.

El uso de esta fuente de datos se justifica en que las investigaciones sociales se vehiculizan a partir de otros tipos de documentos y en gran medida en idiomas nacionales (Hicks, 2004; Moed, 2005; Nederhof, 1989)and focuses on the contribution of scholarly work to the advancement of scholarly knowledge. Its principal question is: how can citation analysis be used properly as a tool in the assessment of such a contribution?Citation analysis involves the construction and application of a series of indicators of the impact', influence' or quality' of scholarly work, derived from references cited in footnotes or bibliographies of scholarly research publications. It describes primarily the use of data extracted from the Science Citation Index and the Web of Science, published by the Institute for Scientific Information (ISI). El CV comprende todo lo publicado por los investigadores, en este sentido,

³ Disponible: <http://www.csic.edu.uy/renderPage/index/pageId/125> [Consulta: 01/09/2018]

⁴ Disponible: <http://www.csic.edu.uy/renderPage/index/pageId/139> [Consulta: 01/09/2018]

⁵ Disponible: <http://www.anii.org.uy/apoyos/investigacion/> [Consulta: 01/09/2018]

supera las limitaciones de las bases de datos de artículos, con fuertes sesgos idiomáticos y contempla la variada tipología documental utilizada por los investigadores sociales y no angloparlantes.

Objetivos

Identificar temáticas locales en la producción científica de las ciencias sociales uruguayas.

Observar la incidencia de las variables: tipo de publicación, idioma, lugar de publicación y subárea de conocimiento en la producción de temas locales.

Poner a prueba la metodología propuesta por Miguel et al. (2015) para reconocer temas locales-globales, y adaptar su uso a otra fuente de datos.

Metodología

La fuente de datos es el CV de los investigadores activos de las CCSS en el SNI, a octubre de 2014. La unidad de análisis son las publicaciones reportadas por los investigadores, el período temporal es 2004-2013.

Se clasifican los registros de acuerdo a su temática: local o no-local, tomando como referencia la metodología propuesta por Miguel et al. (2015), adaptada a la fuente de datos y al caso de Uruguay. Se buscan los topónimos Uruguay, y los correspondientes a los diecinueve departamentos (división política del país), las capitales y principales ciudades⁶, en los campos título, título de libro en caso de capítulos de libro, y palabras clave. Se clasifican las publicaciones como locales o no-locales en función de que aparezca al menos uno de estos topónimos en algunos de los campos considerados.

Se descartan los registros sin lugar de publicación y los que se asignan a más de una subárea de conocimiento. Se parte de 7242 registros de publicaciones, sin duplicados, reportadas por 260 investigadores. Se quitan: 400 registros sin lugar de publicación asignado y 163 adscriptos a 2 o más subáreas de conocimiento.

Las herramientas informáticas usadas son: Exel, SPSS, OpenRefine, Pajek, Gephy.

Resultados y discusión

Al observar la distribución porcentual de acuerdo a temática de los documentos, la tabla 1 muestra que entre los artículos hay casi 10% más trabajos

⁶ Disponible: <http://www.ine.gub.uy/web/guest/codificadores-basedatos> [Consulta: 1/9/2018]

no-locales, mientras que en los capítulos la situación se revierte encontrando un 13% más publicaciones locales. En las otras dos categorías la diferencia porcentual es prácticamente despreciable. Por lo tanto, es posible afirmar que existe una relación débil entre tipo de documento y temática de la publicación, particularmente que los trabajos sobre temáticas no-locales tienden a publicarse más en artículos y las investigaciones sobre temas locales en capítulos.

En la tabla 2 se aprecia que hay un 6% más de publicaciones locales si el idioma es el español y un 5% menos si el idioma es el inglés. Con lo cual, ambos idiomas están levemente relacionados con las temáticas de los trabajos y es más probable publicar sobre temas locales en español y no en inglés.

Tabla 1-Tipo de doc. por temática.

		Temática			
			No local	Local	Total
Tipo de documento	Artículo	Recuento	1461	644	2105
		% de Temática	35,1%	25,5%	31,5%
	Capítulo	Recuento	988	924	1912
		% de Temática	23,8%	36,6%	28,6%
	Libro	Recuento	473	260	733
		% de Temática	11,4%	10,3%	11,0%
	Trabajo evento	Recuento	1235	694	1929
		% de Temática	29,7%	27,5%	28,9%
Total	Recuento	4157	2522	6679	
	% de Temática	100,0%	100,0%	100,0%	

Tabla 2-Idioma por temática.

		Temática			
			No local	Local	Total
Idioma	Español	Recuento	3376	2196	5572
		% de Temática	81,2%	87,1%	83,4%
	Inglés	Recuento	643	264	907
		% de Temática	15,5%	10,5%	13,6%
	Otro	Recuento	138	62	200
		% de Temática	3,3%	2,5%	3,0%
Total	Recuento	4157	2522	6679	
	% de Temática	100,0%	100,0%	100,0%	

Tabla 3- Lugar de publicación por temática.

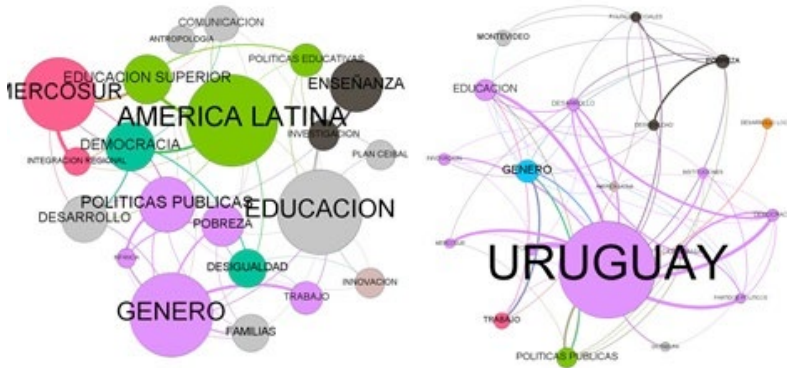
Lugar		Temática		
		No local	Local	Total
Uruguay	Recuento	1671	1290	2961
	% de Temática	40,2%	51,1%	44,3%
ALC	Recuento	1453	773	2226
	% de Temática	35,0%	30,7%	33,3%
Europa	Recuento	768	357	1125
	% de Temática	18,5%	14,2%	16,8%
América del N.	Recuento	218	92	310
	% de Temática	5,2%	3,6%	4,6%
Otras regiones	Recuento	47	10	57
	% de Temática	1,1%	0,4%	0,9%
Total	Recuento	4157	2522	6679
	% de Temática	100,0%	100,0%	100,0%

Al observar la tabla 3 surge que entre los trabajos publicados en Uruguay hay un 11% más que se ocupan de temas locales, mientras que entre los publicados en ALC y en Europa hay un 4% más trabajos no locales, en ambas regiones. Hay entonces una relación leve entre lugar de publicación y temática, que se vuelve más intensa entre la publicación de temas locales en Uruguay y bastante más suave entre publicar trabajos sobre temas no-locales en ALC y Europa.

La distribución de las publicaciones locales por subárea de conocimiento se sitúa en un rango que va de 11% en Ciencias de la Educación a un 52% en Ciencias Políticas. En el medio se sitúan Comunicación y Medios; Derecho; Economía y Negocios; y Geografía Económico Social, subáreas en las que los temas locales representan alrededor de 1/3 de lo que publican. Se entiende que deberían estudiarse con mayor profundidad las subáreas, observando su comportamiento en función del resto de las variables.

Las figuras que siguen muestran la red de co-ocurrencia de las 20 palabras clave con mayor grado del componente gigante, usadas por los autores para describir el contenido de sus trabajos. La figura 1 muestra las palabras más utilizadas en las publicaciones clasificadas como no-locales. La figura 2 permite visualizar las palabras que describen el contenido de las publicaciones locales. Varias de las palabras se repiten como: género, educación, políticas públicas, innovación, desarrollo, democracia. Es sugerente la repetición de las palabras en ambos mapas, parece indicar que en ambos grupos de publicaciones según temática local o no-local se abordan los mismos tópicos. Llama la atención que el Plan ceibal, tema local, aparezca en el mapa de las publicaciones no-locales.

Figuras 1 y 2 (izq. y der.)-20 palabras clave con mayor grado en publicaciones no-locales (izq.) y locales (der.).



Consideraciones finales

Parece haber una relación leve entre tipo de publicación y temática local o no-local. Más intensa entre la producción de capítulos sobre temas locales y artículos sobre temas no-locales.

- Habría una relación más leve aún entre idioma y temática.
- Se identifica una cierta relación entre lugar de publicación y temas, donde es más frecuente publicar sobre temas locales en Uruguay.
- Las subáreas de conocimiento se comportan de manera muy dispar en función de las temáticas y habría que replicar los cruces con las otras variables aplicadas al conjunto.
- Las palabras clave del componente gigante de cada agrupación de trabajos de acuerdo a la temática no parecen mostrar diferencias importantes sobre los temas abordados. Habría que afinar las visualizaciones y profundizar el análisis a ver si es posible explicar esta semejanza. Una explicación posible es que en el escenario internacional los investigadores deben marcar la aproximación a temas locales en los títulos y resúmenes a partir de los nombres geográficos que circunscriben sus trabajos. Sin embargo, en ámbitos de publicación domésticos tal vez no resulte tan necesario delimitar el alcance de un trabajo mediante la utilización de topónimos en los títulos.

Surge la pregunta sobre si la metodología testeada funciona mejor en la clasificación de trabajos en las bases de datos internacionales como SCOPUS que cuando se la aplica al CV.

En suma, este es un trabajo exploratorio del que surgen algunas interrogantes que sirven de punto de partida para continuar testeando esta metodología y eventualmente complementarla incorporando otros abordajes.

Referencias bibliográficas

- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2), 193–215. Recuperado de <https://doi.org/10.1007/BF02457380>
- Hicks, D. (2004). The four literatures of social science. In *Handbook of Quantitative Science and Technology Research* / ed. Henk Moed, Kluwer Academic (pp. 473–496). Recuperado de <https://doi.org/10.1007/1-4020-2755-9>
- Larivière, V., Archambault, É., Gingras, Y., y Vignola-Gagné, É. (2006). Measuring use and creation of open educational resources in higher education. *Journal of the American Society for Science and Technology*, 57(8), 997–1004. Recuperado de <https://doi.org/10.1002/asi.20349>
- Miguel, S., González, C., y Chinchilla-Rodríguez, Z. (2015). Lo local y lo global en la producción científica argentina con visibilidad en Scopus, 2008-2012. Dimensiones nacionales e internacionales de la investigación. *Información, Cultura y Sociedad*, 32(junio), 59–78.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Recuperado de <https://doi.org/10.1007/1-4020-3714-7>
- Moed, H. F., Luwel, M., y Nederhof, A. J. (2002). Towards research performance in the Humanities. *Library Trends*, 50(3), 498–520.
- Nederhof, A. J. (1989). Books and chapters are not to be neglected in measuring research productivity. *American Psychologist*, 44(4), 734–735.

Sesenta años de producción científica sobre Uruguay en laWOS: 1958-2017

Exequiel Fontans¹ y Natalia Aguirre-Ligüera²

Introducción

En materia de política científica uno de los tópicos en el que la comunidad académica se ha centrado desde las últimas décadas del siglo XX, es el establecimiento de las agendas de investigación (Arellano Hernández y Kreimer, 2011). En este marco han surgido algunos trabajos que exploran el impacto de lo local y lo global en las agendas de investigación.

Por su parte en Argentina, si bien es profusa la producción académica sobre la temática desde el campo Ciencia, Tecnología y Sociedad, desde la perspectiva de los Estudios Métricos de la Información se ha comenzado a tratar mucho más recientemente (Miguel, S., González, C. y Chinchilla-Rodríguez, Z., 2015; Miguel, S., González, C. y Hidalgo, M., 2013; Miguel, S. y Ugartemendia, V., 2015).

Este trabajo tiene el objetivo de brindar insumos para contribuir a la comprensión de la producción que se realiza sobre Uruguay en la literatura de corriente principal. En este marco, las preguntas que guían este estudio son:

1. ¿Cuánto se produce sobre Uruguay y cómo evoluciona esa producción?
2. ¿Sobre qué temáticas se produce?
 - ¿Qué produce Uruguay sobre sí mismo?

¹ Departamento Tratamiento y Transferencia de información, Facultad de Información y Comunicación. Universidad de la República, Montevideo, Uruguay exequiel.fontans@fic.edu.uy

² Departamento Tratamiento y Transferencia de información, Facultad de Información y Comunicación. Universidad de la República, Montevideo, Uruguay; natalia.aguirre@fic.edu.uy

- ¿Varían las temáticas de la producción cuando Uruguay colabora con otros países?
 - ¿Sobre qué temáticas se interesan los países que escriben sobre Uruguay cuando lo hacen sin participación de Uruguay?
3. ¿Quién lidera la investigación cuando hay colaboración con países extranjeros?

Metodología

Se realizó una búsqueda el 13/08/2018, en la Colección Principal³ de la *Web ofScience* de *ClarivateAnalytics* con la expresión “uruguay*” en el campo tema; el periodo de tiempo fue 1958-2017. Se obtuvieron 8935 registros con la expresión buscada en el título, resumen o palabras clave.

Los registros se descargaron y luego se procesaron con BibExcel (Persson, O., Danell, R. y Wiborg Schneider, J., 2009). Se extrajo:

- Afiliación institucional a nivel de país (campo C1)
- Liderazgo (campo RP, dirección de ReprintAddress) (Miguel, S., González, C. y Chinchilla-Rodríguez, Z., 2015);
- Temáticas (campo WC (*WosCategory*), se asignan a los títulos de revista, es una tematización de “granularidad gruesa”.

Se normalizaron con Open Refine (Verborgh, R. y Wilde, M. D., 2013) y se elaboraron los cuadros y gráficos con Excel.

Se clasificaron los registros en tres grupos según los datos de filiación disponibles en el campo C1: (1) con firma solo extranjera, (2) en colaboración con Uruguay y (3) con firma solo de Uruguay.

En cuanto a las firmas, extraídas del campo C1, el dato no está reportado en una importante cantidad de registros de las cuatro primeras décadas (83%, 39%, 78% y 65% respectivamente), por lo que esas décadas no se consideran cuando se cruza este dato con las temáticas. De todas maneras, cabe resaltar que, como la ausencia de datos se da en los periodos con menos producción, para todo el corpus el dato falta solamente en el 15% de los registros.

Para el análisis de temática, se realizó un top 10 de las WC de mayor frecuencia por categoría de filiación. Ello dio por resultado 15 temáticas, que

³ Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

fueron categorizadas en cuatro grupos a saber: (a) solo interés del grupo, temáticas que aparecen en el top 10 de un grupo y no aparecen en ninguno de los otros dos, busca responder a las preguntas 2 a y 2 c; (b) interés del grupo y de las colaboraciones, busca responder a la pregunta 2 b; (c) independiente del grupo considerado, las temáticas que aparecen en los tres top 10; (d) sin interés en las colaboraciones.

Resultados y discusión

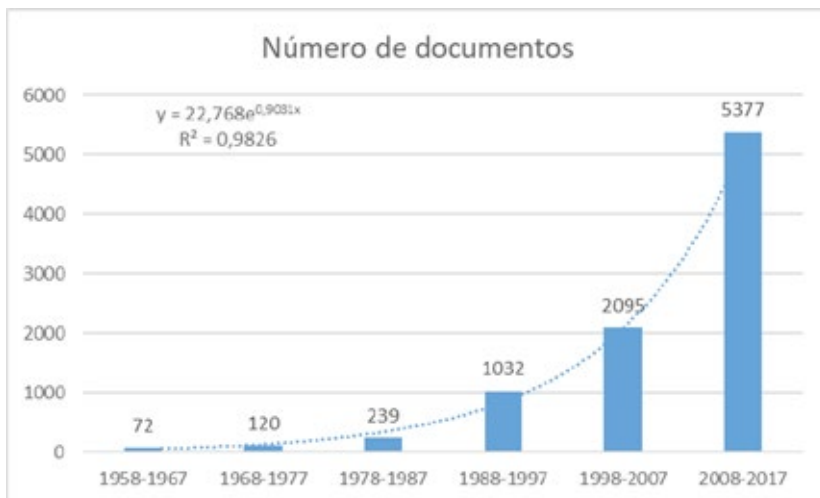
¿Cuánto se produce sobre Uruguay y cómo evoluciona esa producción?

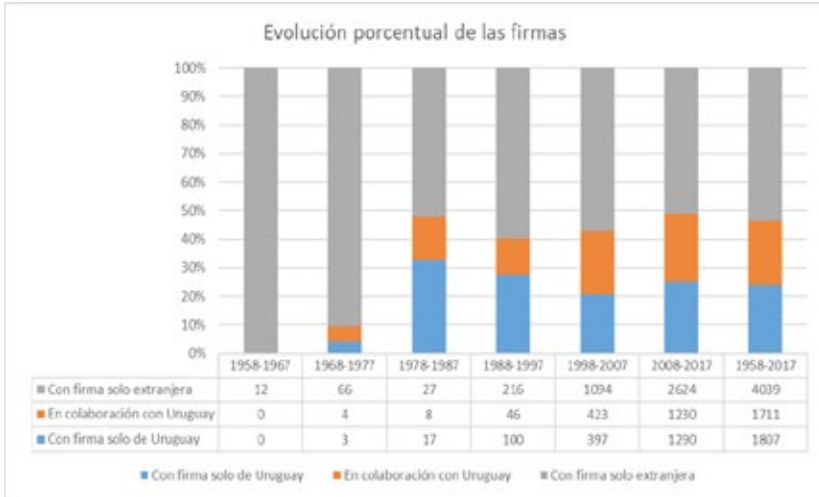
El crecimiento de la producción sobre Uruguay, observado por décadas, es exponencial. El 60% de la producción se concentra en la última década (2008-2017) (Figura 1).

Para los registros que tienen el dato, el 50% de la producción sobre Uruguay está firmada por extranjeros sin colaboración de investigadores de Uruguay.

En cuanto a las firmas por países, para todo el periodo, Uruguay encabeza la producción con un 39% (n=3518), USA 17% (n=1534), Argentina 16% (n=1450) y Brasil 16% (n=1434). Otros países que integran el top 10 son, España, Reino Unido, Chile, Francia, México y Alemania.

Figura 1. Evolución porcentual de las firmas (izquierda) y evolución del número de documentos (derecha).





¿Sobre qué temáticas se produce?

Cuando se establece un top 10 con las temáticas de mayor frecuencia en cada categoría de colaboración, se observa un solapamiento importante en las mismas, dado que la producción se concentra en solo 15 WC (Figura 2).

a. ¿Qué produce Uruguay sobre sí mismo?

“Medicina, General e Interna” es la temática que aparece en el top 10 de la producción de Uruguay, y lejos del top 10 para las otras dos categorías, puesto 47 para colaboración y puesto 49 para extranjeros.

b. ¿Varían las temáticas de la producción cuando Uruguay colabora con otros países?

“Salud Pública, Ambiental y Ocupacional” y “Ciencias de la Tierra, Multidisciplinar” son las dos temáticas que figuran en el top 10 para Uruguay en colaboración con extranjeros y solo extranjeros. Por su parte, cuatro temáticas, “Agronomía”, “Ciencias ambientales”, “Ciencia y tecnología de los alimentos” y “Oncología”, están en el top 10 de las publicaciones de Uruguay y también cuando publica en colaboración con extranjeros (Figura 2 y 3).

c. ¿Sobre qué temáticas se interesan los países que escriben sobre Uruguay cuando lo hacen sin participación de Uruguay?

“Ciencia política”, “Historia” y “Relaciones internacionales”, son las tres temáticas de interés de los extranjeros cuando publican sobre Uruguay sin participación de autores uruguayos (Figura 2 y 3).

Finalmente, encontramos que hay una importante producción sobre “Economía” de autores uruguayos sin colaboración extranjera (está en el lugar 3), y, curiosamente, también esta temática resulta de interés para los extranjeros sin la colaboración de Uruguay, estando incluso en el lugar uno.

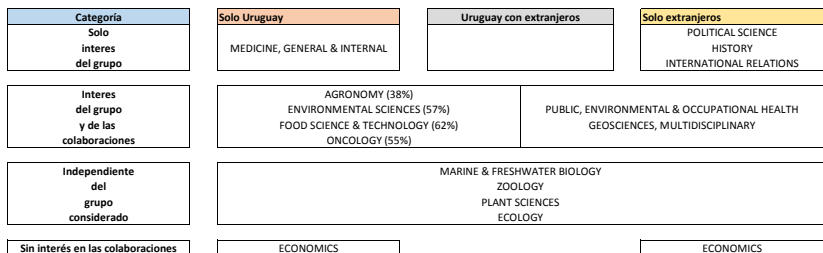
¿Quién lidera la investigación cuando hay colaboración con países extranjeros?

Uruguay lidera el 46% de la producción cuando colabora con firmas extranjeras, le sigue Argentina (11%), Estados Unidos (10%), Brasil (7%) y España (5%). El porcentaje acumulado de estos cinco, explica el 79% de los liderazgos.

Figura 2. Top 10 de las WC más frecuentes por categoría de filiación.

Wos Category	Solo Uruguay		Uruguay con extranjeros		Solo extranjeros	
	Rank	Documentos % / 1807	Rank	Documentos % / 1711	Rank	Documentos % / 4039
AGRONOMY	1	106, 0,06	10	66, 0,04	25	77, 0,02
MARINE & FRESHWATER BIOLOGY	2	105, 0,06	1	130, 0,08	3	233, 0,06
ECONOMICS	3	93, 0,05	16	53, 0,03	1	360, 0,09
ZOOLOGY	4	84, 0,05	6	93, 0,05	2	350, 0,09
PLANT SCIENCES	5	80, 0,04	3	100, 0,06	5	182, 0,05
MEDICINE, GENERAL & INTERNAL	6	70, 0,04	47	16, 0,01	49	35, 0,01
ECOLOGY	7	68, 0,04	7	93, 0,05	10	137, 0,03
ENVIRONMENTAL SCIENCES	8	60, 0,03	8	80, 0,05	14	113, 0,03
ONCOLOGY	9	58, 0,03	5	96, 0,06	24	79, 0,02
FOOD SCIENCE & TECHNOLOGY	10	55, 0,03	9	67, 0,04	20	92, 0,02
PUBLIC, ENVIRONMENTAL & OCCU	15	45, 0,02	2	107, 0,06	9	138, 0,03
GEOSCIENCES, MULTIDISCIPLINARY	12	48, 0,03	4	98, 0,06	6	180, 0,04
POLITICAL SCIENCE	26	35, 0,02	38	19, 0,01	4	188, 0,05
HISTORY	25	37, 0,02	48	16, 0,01	7	166, 0,04
INTERNATIONAL RELATIONS	76	10, 0,01	73	8, 0,00	8	165, 0,04

Figura 3. Categorización de las temáticas según interés de los grupos.



Conclusión

A continuación, se listan algunas conclusiones y preguntas abiertas de este trabajo:

- La producción con firma solo de Uruguay está en torno al 25%.

- El 50% de lo que se escribe sobre Uruguay lo hacen extranjeros, sin colaboración de uruguayos, lo que representa un contraste importante con los resultados de (Miguel, Ugartemendia, 2015), en donde los valores eran aproximadamente 40% producción argentina, 30% extranjera y 30% en colaboración, aunque hay que tener en cuenta la fuente (Scopus) y el periodo (2008-2012).
- Uruguay lidera el 46% de los trabajos en colaboración. Esto incluye 4 temáticas que están en el top 10 de su producción individual.

Por último, algunas reflexiones y cuestiones que quedan planteadas para futuros estudios:

- La estrategia de búsqueda, “uruguay*”, para nuestro país, ¿es suficiente? ¿No es necesario agregar a la expresión de regiones u otros topónimos? Aparentemente sí, aunque esto no se ha testeado en este trabajo. Uruguay es un país pequeño que no tiene eco regiones que destaquen a nivel internacional, como podría ser el caso de la Pampa Húmeda para Argentina, o el Amazonas para Brasil.
- ¿Qué pasa con los trabajos sobre temas globales? Por ejemplo, cambio climático; epidemias globales, en dónde la aparición de los nombres de las regiones no indica un tratamiento local del tema. Este aspecto ameritaría un análisis en profundidad en futuros trabajos; algunos temas, como los mencionados, podría decirse que son “supranacionales”, y el país y las localidades no tienen un peso específico sobre la temática del documento analizado.
- ¿Qué pasa con lo “no hecho” (Frickel, S. et al, 2010), las áreas que no aparecen en este estudio? ¿Por qué no están en las agendas de investigación?
- ¿Los extranjeros que miran a Uruguay, lo hacen desde las Ciencias Sociales? La literatura, en general, postula que los temas sociales son de más interés local que internacional; este aspecto debería ser explorado con más detalle.

Referencias bibliográficas

ANII, Unidad de evaluación y monitoreo. (2017). Boletín ANII: Indicadores de Ciencia, Tecnología e Innovación en Uruguay 2017. Boletín ANII: Vol. 5. Montevideo: ANII.

- Arellano Hernández, A. y Kreimer, P. (2011). *Biblioteca sociedad y tecnociencia. Colección Estudios Sociales de Tecnociencia desde América Latina. Estudio social de la ciencia y la tecnología desde América Latina*. Bogotá: Siglo del hombre editores.
- Frickel, S., Gibbon, S., Howard, J., Kempner, J., Ottinger, G. y Hess, D. J. (2010). Undone Science: Charting social movement and civil society challenges to research agenda setting. *Science, Technology, & Human Values*, 35(4), 444–473. Recuperado de <https://doi.org/10.1177/0162243909345836>
- Miguel, S., González, C. y Chinchilla-Rodríguez, Z. (2015). Lo local y lo global en la producción científica argentina con visibilidad en Scopus, 2008-2012. Dimensiones nacionales e internacionales de la investigación. *Información, cultura y sociedad*, (32), 59–78.
- Miguel, S., González, C. y Hidalgo, M. (2013). Argentina como objeto de investigación. Reflexiones conceptuales y aproximaciones metodológicas para el abordaje bibliométrico de la producción científica sobre temas de alcance nacional. En *Terceras Jornadas de Intercambios y Reflexiones acerca de la Investigación en Bibliotecología*, La Plata. Recuperado de <http://jornadabibliotecologia.fahce.unlp.edu.ar>
- Miguel, S. y Ugartemendia, V. (2015). Extranjeros que miran a nuestro país: la producción científica sobre Argentina entre 2008 - 2012. En UBA (Ed.), *XI Jornadas de Sociología - UBA*. Buenos Aires: UBA.
- Persson, O., Danell, R. y Wiborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. En F. Aström, R. Danell, B. Larsen y J. Wiborg Schneider (Eds.), *Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday* (Vol. 5, pp. 9–24). Leuven, Belgium: International Society for Scientometrics and Informetrics.
- Verborgh, R. y Wilde, M. D. (2013). *Using OpenRefine: The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web*. Birmingham, England: Packt Publishing.

Argentina como tema o alcance geográfico de la investigación. Una mirada desde SciELO y Scopus

*Mónica Hidalgo¹, Lorena Caprile²,
Israel Jorquera Vidal³ y Sandra Miguel⁴*

Introducción

Existen varios estudios abocados al análisis bibliométrico de la producción científica de un país, que aportan un valioso conocimiento sobre las características del desarrollo de la ciencia de ese país, objeto de estudio. Sin embargo son escasos los trabajos que, como éste, estudian la producción científica sobre un país o región objeto de investigación. Esta investigación centra su atención en el estudio de la producción científica sobre Argentina, delimitado por el conjunto de publicaciones científicas en los que el país (Argentina) es el tema o el alcance geográfico de las investigaciones. A este segmento se lo ha denominado “producción científica sobre temas de alcance nacional”, y abarca tanto los trabajos donde Argentina, o alguna de sus provincias o algunas de sus regiones geográficas estén comprendidas en las investigaciones. También comprende aquellos trabajos en los que Argentina sea único país estudiado, o sea parte de

¹ Departamento de Bibliotecología. Facultad de Humanidades y Ciencias de la Educación. Universidad Nacional de La Plata. La Plata, Argentina. monicahidalgorubio@gmail.com

² mlorenacaprile@gmail.com

³ Biblioteca. Facultad de Ciencias Naturales y Museo. Universidad Nacional de La Plata. La Plata, Argentina. jorqueravidal@gmail.com

⁴ Instituto de Investigaciones en Humanidades y Ciencias Sociales - IdIHCS (UNLP-CONICET), La Plata, Argentina. smiguel@fahce.unlp.edu.ar

un estudio de comparación con otros países, e investigaciones de personalidades, eventos, instituciones, objetos o símbolos argentinos.

Los objetivos que esta investigación se plantea son:

- Determinar y comparar la presencia de estudios sobre Argentina en la producción científica 2008-2012 incluídas en SciELO y Scopus.
- Identificar las áreas disciplinares en las que hay mayor presencia de temas relacionados con el país objeto de estudio en ambas fuentes.
- Identificar las procedencias geográficas de los autores y comparar los patrones de colaboración a nivel de países.
- Analizar y describir la distribución de la producción según las regiones geográficas de Argentina, y su relación con las regiones de las instituciones de procedencia de los autores.

Metodología

Fuentes de datos

Como fuentes de datos se seleccionaron las bases de datos de SciELO y Scopus. La primera, por constituir una fuente de referencia de la literatura científica en el ámbito regional de América Latina y el Caribe. La segunda, Scopus, de Elsevier, por su amplia cobertura de geográfica y temática (Moya-Anegón y otros, 2007) y por ser una de las bases de datos más comprensivas en la difusión de la literatura científica de circulación internacional. Asimismo su elección se fundamenta en que su cobertura en la producción científica procedente de países latinoamericanos es incluso mayor que Web of Science (WoS) de Thomson Reuters (Collazo-Reyes, 2014) y que SciELO y Redalyc (Miguel, 2011).

Estrategia de búsqueda

Búsquedas en el campo: Título, Resumen y Palabras clave.

Términos utilizados: “Argentina”; nombre de cada una de las 23 provincias; nombre de las 5 regiones socioeconómicas (Cuyo, Noreste, Noroeste, Pampeana, Patagonia). Período de la producción científica: 2008-2012. Total de registros recuperados: 1.551 SciELO y 16.301 Scopus.

Preparación y depuración de los datos

Los registros recuperados en ambas fuentes se exportaron a dos bases de datos ad-hoc. Se revisaron uno a uno determinando su pertinencia o no al objeto de estudio.

Los registros resultantes se agruparon en 6 áreas disciplinares (Ciencias Agrícolas, Ciencias Médicas y de la Salud, Ciencias Naturales y Exactas, Ciencias Sociales e Ingenierías y Tecnologías). La clasificación por regiones consistió en asignar a cada registro el nombre de la/s regiones estudiadas (Cuyo, Noreste, Noroeste, Pampeana, Patagonia). Se clasificó como “Argentina” a aquellos registros abocados al estudio del país en su totalidad.

Indicadores calculados

Se calcularon los siguientes indicadores: a) Tamaño de la producción a partir del recuento de documentos (Tabla 1); b) Distribución disciplinar y regional de la producción según las clasificaciones previamente establecidas (Tablas 2-3); c) Tasa de colaboración nacional e internacional (Tabla 4); d) Relación entre la región de procedencia de los autores argentinos y las regiones geográficas de la Argentina estudiadas (Tabla 5).

Resultados

Se presentan a continuación los resultados del análisis.

Tabla 1. Volumen y evolución de la producción científica sobre temas de alcance nacional en SciELO y Scopus, 2008-2012

Año	SciELO		Scopus	
	ndoc	% ndoc	ndoc	% ndoc
2008	251	16,2	2.811	17,2
2009	370	23,9	3.052	18,7
2010	352	22,7	3.198	19,6
2011	346	22,3	3.512	21,5
2012	232	15,0	3.728	22,9
Totales	1.551	100,0	16.301	100,0
Tasa de Crecimiento en el quinquenio (%)	-7,6		32,6	

Tabla 2. Distribución de la producción científica sobre temas de alcance nacional por áreas temáticas en SciELO y Scopus, 2008-2012

Área temática MINCyT	SciELO		Scopus	
	ndoc	% ndoc	ndoc	% ndoc
Ciencias Agrícolas	144	9,3	6.365	39,0
Ciencias Médicas y de la Salud	153	9,9	5.711	35,0
Ciencias Naturales y Exactas	674	43,5	5.041	30,9
Ciencias Sociales	402	25,9	3.186	19,5
Ingenierías y Tecnologías	10	0,6	917	5,6
Humanidades	149	9,6	817	5,0

Los % no suman 100% porque hay solapamiento entre las distintas áreas temáticas

Tabla 3. Distribución de la producción científica por región geográfica estudiada 2008-2012

SciELO			Scopus		
Región estudiada	ndoc	% ndoc	Región estudiada	ndoc	% ndoc
Argentina	75	4,8	Argentina	6.745	41,4
Cuyo	221	14,2	Cuyo	891	5,5
Noreste	279	18,0	Noreste	952	5,8
Noroeste	447	28,8	Noroeste	1.105	6,8
Pampeana	320	20,6	Pampeana	4.081	25,0
Patagonia	244	15,7	Patagonia	2.926	17,9
Otros	18	1,2	Otros	59	0,4

Tabla 4. Distribución de la producción científica sobre temas de alcance nacional según procedencia geográfica de los autores, SciELO y Scopus, 2008-2012

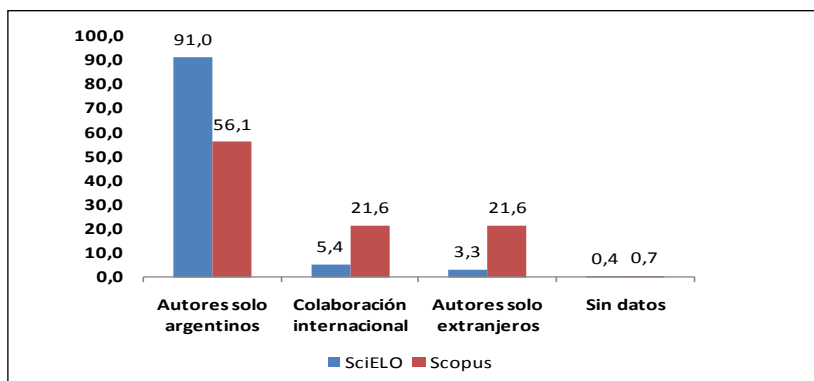


Tabla 5. Región estudiada vs. región de procedencia del autor en SciELO, 2008-2012

		% ndoc SciELO				
		Región de procedencia del autor				
		Cuyo	Noreste	Noroeste	Pampeana	Patagonia
Región estudiada	Argentina	2,4	0,8	2,3	7,0	1,0
	Cuyo	72,9	3,3	1,1	9,3	6,0
	Noreste	1,2	86,2	14,3	13,9	3,0
	Noroeste	6,0	3,3	79,4	19,1	6,0
	Pampeana	4,2	4,9	1,1	33,2	2,0
	Patagonia	12,0	1,6	1,4	16,2	81,0
	Otros	1,2	0,0	0,3	1,2	1,0

		% ndoc Scopus				
		Región de procedencia del autor				
		Cuyo	Noreste	Noroeste	Pampeana	Patagonia
Región estudiada	Argentina	24,5	20,1	25,3	37,3	11,9
	Cuyo	41,3	2,7	3,0	3,5	2,3
	Noreste	2,5	55,9	10,1	5,2	0,8
	Noroeste	8,9	8,5	47,5	4,9	2,1
	Pampeana	10,5	10,2	9,0	35,6	6,5
	Patagonia	12,4	2,6	5,0	13,3	76,2
	Otros	0,0	0,0	0,1	0,2	0,1

Conclusión

- El volumen de la producción científica sobre temas de alcance nacional es mayor en Scopus (16.301 docs.) que en SciELO (1.551 docs.)
- La cobertura regional de revistas de la base de datos SciELO, no está relacionada a un mayor volumen de producción sobre temas regionales, al menos no para trabajos sobre Argentina.
- Una mayor presencia en Scopus, demuestra que los temas de alcance nacional son también publicados en revistas extrarregionales (Miguel y otros, 2015).
- SciELO posee mayor producción del área de Ciencias Naturales y Exactas, mientras que en Scopus hay mayor presencia de documentos

- de Ciencias Agrícolas, Ciencias Médicas y Ciencias Naturales. Las Ciencias Humanas y Sociales están levemente mejor representadas en SciELO que en Scopus.
- En SciELO hay un claro predominio de firmas de autores argentinos exclusivamente.
- En Scopus, cerca de la mitad de las firmas corresponden solo a autores argentinos, y la otra mitad se distribuye entre trabajos firmados en colaboración internacional y trabajos firmados solo por autores extranjeros.
- Se puede afirmar que el lugar geográfico en el que desempeñan su labor los investigadores argentinos se vincula, aunque no de manera excluyente, con el contexto geográfico en el que se sitúan las investigaciones.
- En Scopus, más del 50% de la producción refiere a Argentina como tema o territorio, y el resto a regiones específicas hacia el interior del país.
- En SciELO, el 95% de la producción está circunscripta a alguna de las 5 regiones socioeconómicas.
- Se requieren más estudios que permitan profundizar en la relación entre temáticas y regiones, así como determinar las instituciones y grupos de investigación más comprometidos con los estudios regionales, estimar el impacto de las investigaciones, además de favorecer a la revisión y perfeccionamiento de la estrategia metodológica empleada.

Referencias bibliográficas

- Collazo Reyes, F. (2014). Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. *Scientometrics*, 98(1). 197-209.
- Miguel, S. (2011). Revistas y producción científica de América Latina y el Caribe: su visibilidad en SciELO, RedALyC y SCOPUS. *Revista Interamericana de Bibliotecología*, 34(2), 187-199. Recuperado de <http://www.redalyc.org/articulo.oa?id=179022554006>
- Miguel, S., González, C. y Chinchilla Rodríguez, Z. (2015). Lo local y lo global en la producción científica argentina con visibilidad en Scopus, 2008-2012. Dimensiones nacionales e internacionales de la investigación. *Información, Cultura y Sociedad*, 32, 59-78. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/1375/1352>

Moya-Anegón, F. de; Chinchilla-Rodríguez, Z.; Vargas-Quesada, B.; Corera-Álvarez, E.; Muñoz-Fernández, F.J.; González-Molina, A. y Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78.

Impacto de la investigación local mediante Altmetrics. El sector del vino en España

*Enrique Orduña-Malea¹, Cristina Font-Julian²
y Adolfo Alonso-Arroy³*

Introducción

La ‘Declaration of Research Assessment’ (DORA)⁴ explicita el malestar de la comunidad científica por las prácticas extendidas en los procesos de evaluación de la investigación científica, proponiendo un amplio listado de buenas prácticas (firmado actualmente por 505 instituciones y 12.492 individuos), dirigidas a los distintos actores involucrados en los procesos de publicación, difusión y evaluación de la ciencia. En esa misma línea *Nature* publica en 2015 el conocido como ‘Manifiesto de Leiden sobre Indicadores de Investigación’ (Hicks et al., 2015), un decálogo de buenas prácticas relacionados específicamente con el buen uso de las técnicas bibliométricas en las tareas de evaluación de los resultados de investigación. En concreto, el tercer punto del Manifiesto indica explícitamente que “La excelencia en investigación de relevancia local debe ser protegida”.

Los países y regiones del mundo poseen unas características - sociales, económicas, culturales, climatológicas -tales que determinan sus principales sectores de actividad industrial y, por tanto, las áreas con mayor financiación

¹ Universitat Politècnica de València, Valencia, España enorma@upv.es

² Universitat Politècnica de València, Valencia, España crifonju@upv.es

³ Universitat de València, Valencia, España. adolfo.alonso@uv.es

⁴ <https://sfdora.org>

por parte de gobiernos y entidades financiadoras. El florecimiento de ciertos sectores industriales y culturales conlleva de forma paralela un desarrollo científico para dar respuesta a las necesidades locales. Aunque existen disciplinas y campos científicos de naturaleza internacional, otras muchas áreas (especialmente algunas ramas de la Ingeniería, las Ciencias Sociales y Humanidades) se caracterizan por tratar problemas que difícilmente pueden reproducirse en otros lugares y, por tanto, a captar atención internacional.

La aplicación de ciertas técnicas bibliométricas puede no sólo evidenciar el carácter local de ciertas líneas de investigación sino potenciarlo. Por ello, la utilización de bases de datos selectivas, como Web of Science (WoS) y Scopus, basadas principalmente en un corpus de artículos de revistas en inglés, es desaconsejada a la hora de evaluar el impacto de ciertas disciplinas locales. Contra este problema destacan actualmente dos líneas complementarias: a) el uso de bases de datos no selectivas; y b) la aplicación de métricas alternativas (Altmetrics). En el primer caso, destaca el uso de Google Scholar (Orduña-Malea et al, 2016), aunque sus limitaciones para poder extraer datos de forma automática, su falta de control de calidad y su facilidad de manipulación provocan cierto rechazo en la comunidad bibliométrica. En el segundo caso, destacan las métricas procedentes de plataformas de redes sociales (Facebook, Twitter, Mendeley, etc.), que reflejan evidencias alternativas de impacto (Priem y Hemminger, 2010), aunque el significado de estas métricas y su uso en procesos evaluativos están en constante discusión (Haustein, 2016). Ambos procedimientos (bases de datos no selectivas y Altmetrics) podrían ayudar a mostrar evidencias de un mayor nivel de internacionalización o difusión de la investigación local.

Este trabajo pretende centrar el foco en el sector vitivinícola, de gran importancia en España. Aunque el nivel de consumo de vino presenta un descenso importante durante las últimas décadas⁵, España posee, a fecha de 2017, la mayor cantidad de superficie de viñedo en Europa (967 mha). Así mismo, es la tercera potencia del mundo (tras Italia y Francia) en producción de vino (32.1 mill. hl), y la séptima en consumo (10.3 mill. hl), según los datos de la Organización Internacional de la Viña y el Vino (OIV)⁶. Además, las exportaciones de vino español en 2017 supusieron más de 2.800 millones de euros a las arcas del

⁵ <http://www.oiv.int/public/medias/4710/oiv-noteconjmars2016-en.pdf>

⁶ <http://www.oiv.int/public/medias/5960/aspectos-de-la-coyuntura-mundial-oiv-abril-2018.pdf>

Estadosegún datos del Observatorio Español del Mercado del Vino (OEMV)⁷, dando buena muestra de la importancia del sector en el país.

El principal objetivo de este trabajo es determinar el grado de internacionalización de la producción científica española sobre vino, y comprobar si las métricas alternativas muestran o reflejan un impacto internacional diferente al proporcionado por las métricas tradicionales.

Metodología

Dado que el sector del vino aglutina a una amplia variedad de disciplinas científicas (Ciencias Agroalimentarias, Química, Economía, Turismo, etc.), se decidió restringir la búsqueda a aquellos trabajos relacionados con su industria a través de la siguiente estrategia de búsqueda: [(TITLE-ABS-KEY (“wine* industr*” OR “wine* sector” OR “wine* business”) AND AFFILCOUNTRY (spain)) AND (EXCLUDE (PUBYEAR, 2018))]. El corpus de producción bibliográfica española se confinó al conjunto de trabajos con al menos un/a autor/a con afiliación a una institución española. La búsqueda se realizó en julio de 2018 en Scopus, por ser esta base de datos más extensa que WoS y permitir la exportación automática de los resultados con un alto control de calidad.

De forma paralela, se diseñó un script mediante Phyton para generar consultas automáticas a la API de Altmetric.com⁸. De este modo se extrajeron los datos de menciones a los documentos recuperados mediante la búsqueda anterior en Twitter, Mendeley y en blogs. Así mismo, se recopilaron los datos geográficos de los Tweets y Mendeley (éstos se descartaron al no estar disponibles para la mayoría de registros). Este procedimiento se realizó durante la primera semana de agosto de 2018. Finalmente, todos los datos obtenidos fueron volcados a una hoja de cálculo para su análisis estadístico. Para ello se utilizó la aplicación XLSTAT.

Resultados

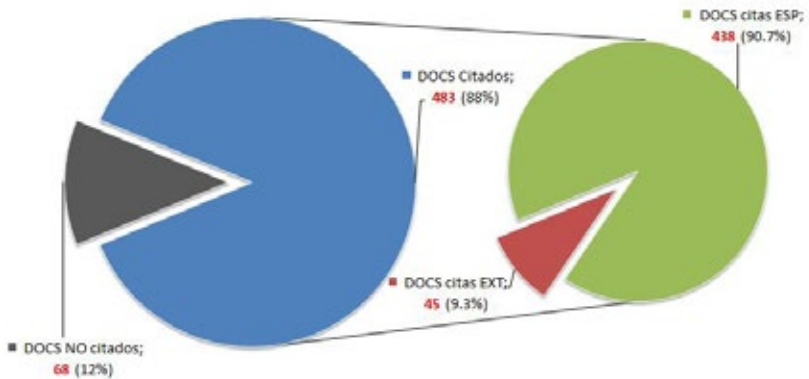
La estrategia de búsqueda arroja un total de 551 registros (17.5% del total sin la restricción geográfica de afiliación española), siendo la amplia mayoría artículos de revista (79%), *conferencepapers* (9.4%) y capítulos de

⁷ <http://www.oemv.es/>

⁸ <https://api.altmetric.com/>

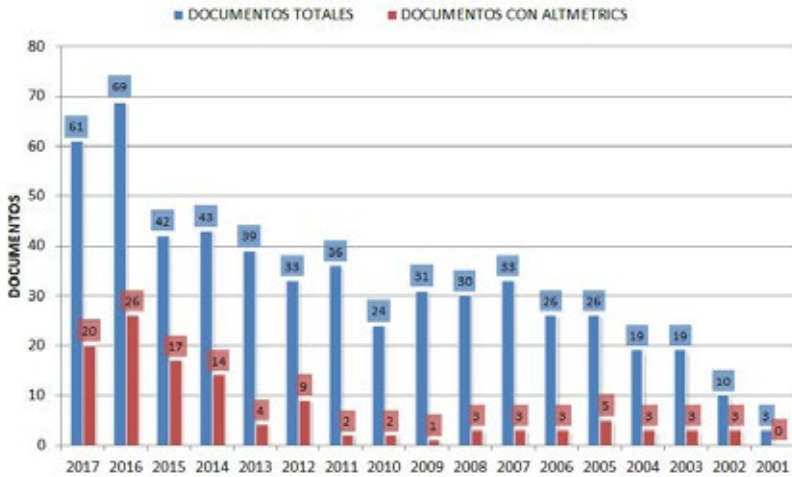
libro (5.1%). El 88% de este corpus (483 documentos) recibe al menos una cita. De este subconjunto, el 90.7% recibe al menos una cita procedente de un documento con un coautor cuya afiliación institucional es española (Fig.1). No obstante, el total de citas recibidas acumuladas por los 483 documentos citados asciende a 2.637, mientras que solamente el 32.2% (848) provienen de contribuciones con autores locales.

Figura1. Producción en industria del vino con coautoría española (Scopus)



De los 551 registros recuperados, un total de 475 incorporan un DOI. De éstos, únicamente 118(25% de los documentos con DOI) reciben alguna mención. El corpus altmétrico queda formado por 429 menciones en posts, 385 Tweets y 4.674 readers. En la Figura 2 se puede observar el conjunto de documentos publicados (2001-2017) y la cantidad de documentos que presentan al menos una mención en Altmetrics.com (Posts, Tweets o Readers).

Figura2. Documentos por año de publicación y cobertura en Altmetrics



En lo que respecta a la localización geográfica de los Tweets, se localizan menciones de hasta 29 países diferentes, mientras que en 86 Tweets (22.3%) la localización es desconocida. A nivel de penetración (cantidad de trabajos con Tweets) sobresalen Estados Unidos (30), Francia (27), España (21) y Reino Unido (21). A nivel de menciones (cantidad total de Tweets), destacan estos mismos países. Estados Unidos ocupa nuevamente el primer lugar (71), seguido en este caso de España (39 Tweets), Reino Unido (31) y Francia (30) (Tabla 1). El porcentaje de Tweets desde España respecto del total es del 12.7%.

En la Tabla 1 se muestra la correlación (Spearman; $\alpha > 0.1$) entre todas las métricas. Para compensar el efecto del año de publicación (documentos más antiguos han tenido más tiempo para recibir citas y menciones), las menciones se han normalizado a través del número de años transcurridos desde la publicación de los documentos.

Tabla1. Coeficiente de correlación de Spearman (n= 119)

Variables	Posts	Tweets	Tweets (ES)	Readers	Citas	Citas (ES)	Citas (EXT)
Posts	1						
Tweets	0.88	1					
Tweets (ES)	0.46	0.43	1				
Readers	0.39	0.43	0.14	1			
Citas	-0.16	-0.09	-0.11	0.47	1		
Citas (ES)	-0.16	-0.12	-0.18	0.32	0.74	1	
Citas (EXT)	-0.17	-0.11	-0.08	0.44	0.94	0.50	1

El número de Readers correlaciona de forma moderada-baja con el número de citas totales (.47). El número de menciones en Posts y en Tweets está fuertemente correlacionado (.88). Finalmente, no se identifica correlación alguna entre el número de citas recibidas y el número de menciones en Twitter.

Discusión y conclusión

El análisis de un sector industrial complejo en lugar de una disciplina o campo científico provoca que el corpus bibliográfico obtenido dependa en gran medida de la estrategia de búsqueda, mezclando además documentos de áreas distintas (con patrones distintos de citación). Existe por otra parte una dependencia en las fuentes usadas (Scopus y Altmetrics.com). Google Scholar podría proporcionar un corpus bibliográfico más preciso (documentos y citas), mientras que el uso de otras fuentes alométricas como PlumAnalytics obtendrían una cobertura de Twitter y Mendeley diferente (Zahedi y Costas, 2018).

Considerando el carácter exploratorio de este trabajo, se concluye que el nivel de internacionalización del impacto de la producción española sobre vino en Scopus es bajo (solamente el 9.3% de los documentos citantes no tiene un co-autor con filiación española). Twitter muestra un nivel mayor de internacionalización, donde España representa sólo el 12.7% del total de Tweets con menciones a alguno de los documentos producidos.

Agradecimientos

Este trabajo ha recibido financiación del Proyecto Nacional eMarketwine (CS02016-78775-R), del Ministerio de Economía, Industria y Competitividad.

Cristina I. Font Julián disfruta de un contrato FPI (BES-2017-079741) financiado por el Ministerio de Ciencia, Innovación y Universidades.

Referencias bibliográficas

- Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413-423. Recuperado de <https://link.springer.com/article/10.1007/s11192-016-1910-9>
- Hicks, D., Wouters, P., Waltman, L., y Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431. Recuperado de <http://www.leidenmanifesto.org/>
- Orduña-Malea, E., Martín-Martín, A., Ayllón, Juan M. y Delgado López-Cózar, E. (2016). *La revolución Google Scholar: destapando la caja de Pandora Académica*. Granada: UNE.
- Priem, J., y Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *FirstMonday*, 15(7). Recuperado de <https://firstmonday.org/article/view/2874/2570>
- Zahedi, Z., y Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PloS one*, 13(5), e0197326. Recuperado de <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197326>

Indicadores bibliométricos de la producción científica sobre países latinoamericanos en perspectiva comparada

*Sandra Miguel¹, Claudia M. González²
y Claudia Boeris³*

Introducción

La definición de áreas prioritarias y temas de interés científico y tecnológico constituyen aspectos de especial relevancia en el desarrollo de las agendas de investigación en la mayoría de los países del mundo. América Latina no es una excepción. En un contexto de creciente incorporación del acontecer local y nacional en contextos globales (Bodemer, 2010), el impulso de políticas que orientan la investigación hacia temas locales y a la resolución de problemas sociales, productivos y medioambientales plantea nuevos desafíos en la valoración de los resultados e impacto que esa investigación produce, así como también de la transferencia e innovación (Lugones, Porta y Codner, 2013; Codner, Miguel, 2014).

Si bien existen numerosas investigaciones que desde una aproximación bibliométrica estudian la producción de los países latinoamericanos, son escasos e incipientes los trabajos que enfocan la atención en la producción sobre

¹ Instituto de Investigaciones en Humanidades y Ciencias Sociales - IdIHCS (UNLP-CONICET), La Plata, Argentina smiguel@fahce.unlp.edu.ar

² Instituto de Investigaciones en Humanidades y Ciencias Sociales - IdIHCS (UNLP-CONICET), La Plata, Argentina. cgonzalez@fahce.unlp.edu.ar

³ Departamento de Bibliotecología. Facultad de Humanidades y Ciencias de la Educación. Universidad Nacional de La Plata, La Plata, Argentina clauboe@gmail.com

los países, es decir, donde el país es tema o alcance territorial de las investigaciones, independientemente de la afiliación institucional y procedencia geográfica de los autores firmantes de las publicaciones.

Algunos trabajos previos enfocados en Argentina revelan que para el período 2008-2012 un 25% de la producción científica de ese país, con visibilidad internacional en Scopus, es abocada a estudios sobre el propio país. Difiere según las disciplinas alcanzando en algunas un porcentaje cercano al 50%, como es el caso de Ciencias sociales, Economía, Ciencias agrícolas y biológicas, Ciencias de la Tierra y relacionadas con el medioambiente. Solo el 28% de los trabajos se comunica en revistas argentinas o latinoamericanas, siendo el mayor porcentaje publicado en revistas extra-regionales. Predominan los trabajos sin colaboración y en colaboración nacional, y es alto el porcentaje de liderazgo de los investigadores argentinos en la producción firmada en colaboración internacional. Desde las comunidades extranjeras existe un mayor interés en los estudios sobre Argentina por parte de Estados Unidos y países europeos como España, Reino Unido, Alemania, Francia e Italia. Aunque se observa también la presencia incipiente de países de América Latina, especialmente Brasil, Chile y México (Miguel, González y Chinchilla-Rodríguez, 2015; Miguel y Ugartemendía, 2015). Sobre el impacto científico de la producción en temas locales no se tiene mayor conocimiento, excepto en el área de las Humanidades y Ciencias Sociales donde Chinchilla-Rodríguez, Miguel, S. y Moya-Anegón, 2015 encuentran que las publicaciones de Argentina del período 2003-2012 incluidas en Scopus alcanzan menores niveles de impacto que las de tópicos de alcance global, predominando los trabajos en español y de autoría única.

En otras áreas temáticas y aunque sin distinción en el alcance local / global de los temas, Meneghini et al (2008) encuentran que los trabajos firmados sin colaboración internacional por autores de Argentina, Brasil, Chile y México reciben menos citas que los firmados por autores de fuera de la región. Para esos mismos países y para el área de ecología Rau et al (2017) observan que mientras el volumen de artículos en revistas internacionales de alto impacto registra un crecimiento exponencial desde 1975 a 2015, a partir de 2000 el número de citas por artículo disminuye marcadamente, a diferencia de lo que ocurre en la misma área en trabajos de autores de países centrales.

Este escenario abre numerosos interrogantes en torno a los patrones de publicación, colaboración e impacto de la producción de y sobre los países latinoamericanos. En este trabajo se realiza una primera aproximación a su estudio, presentando indicadores de la producción científica sobre países latinoamericanos en perspectiva comparada. Interesa conocer: ¿Cómo está representada en el contexto internacional de la ciencia la investigación sobre los países latinoamericanos? ¿Cómo se distribuye la producción en relación a la procedencia geográfica de los autores? ¿Cómo es la relación del impacto de la producción según las prácticas de coautoría y la procedencia geográfica de los autores y de las revistas?

Metodología

Se estudia la producción científica de circulación internacional incluida en Scopus, sobre diez países latinoamericanos: Brasil, México, Argentina, Chile, Colombia, Venezuela, Cuba, Perú, Ecuador, Uruguay, abarcando el período 2007-2016. Para ello, se realizaron búsquedas a partir del campo país de afiliación de los autores, obteniendo el volumen de la producción científica total de cada país registrada en la base de datos. Por otro lado y para cada uno de los países, se realizaron consultas por el nombre del país (con truncamiento) en los campos de título, palabras clave y resumen, sin distinción en la afiliación de autores. Sobre este conjunto de registros que se ha llamado “temas locales” se conformaron dos subconjuntos: uno, conteniendo los artículos firmados por al menos un autor del país estudiado (autores del país), y otro, con los artículos firmados exclusivamente por autores externos al país estudiado (autores del extranjero). Se calculó el porcentaje de documentos en temas locales sobre el total de la producción de cada país y luego, la distribución de documentos en temas locales según la procedencia de los autores. Se seleccionaron dos áreas temáticas bien diferenciadas en los volúmenes de producción y en los patrones de colaboración y citación, y países que tuvieran más del 40% de producción en temas locales en dichas áreas: Agricultura y ciencias biológicas (8 países) y Ciencias sociales (9 países). Se calcularon los siguientes indicadores bibliométricos: Porcentaje de documentos (% ndoc), Citas x documento (citas x doc), Autores x documento (autores x doc), comparando los resultados según la procedencia geográfica de los autores y de las revistas.

Resultados

El porcentaje de documentos en temas locales respecto del total de la producción de cada país varía entre el 21,7% en Uruguay hasta el 45,5% en Perú (Fig. 1). En Argentina, Brasil, Chile y Colombia la mayoría de los documentos que estudian al país son producidos por autores del mismo país estudiado; mientras que en el resto de los países se da una relación inversa, siendo el mayor porcentaje de la producción firmada solo por autores del extranjero (Fig. 2). Al comparar la producción en las dos áreas temáticas seleccionadas se observan diferencias. En AGR, en seis de ocho países los temas locales son mayormente estudiados por autores del país (Fig. 3); en SOC, en cambio, esto se da solo en Argentina, Brasil y Chile (Fig. 4).

El número medio de autores por documento es mayor cuando los trabajos son firmados por autores del país que cuando firman solo autores del extranjero; patrón observado en todos los países y en las dos áreas temáticas (Fig. 5).

Tanto en AGR como en SOC la producción firmada por autores del país estudiado recibe menos citas por documento que la producida por autores del extranjero. Una excepción se da para el caso de Ecuador que presenta una paridad en este indicador en AGR, y es más elevada para autores del país en SOC (Fig. 6).

En AGR en la mayoría de los casos, los trabajos que reciben más citas por documento son los firmados por autores del extranjero, independientemente de la región de procedencia de las revistas donde publican. En SOC la situación es más dispar y varía según los distintos países y regiones. (Figs. 7 y 8)

Fig. 1 Presencia de temas locales en la producción científica de países latino-americanos (2007-2016)

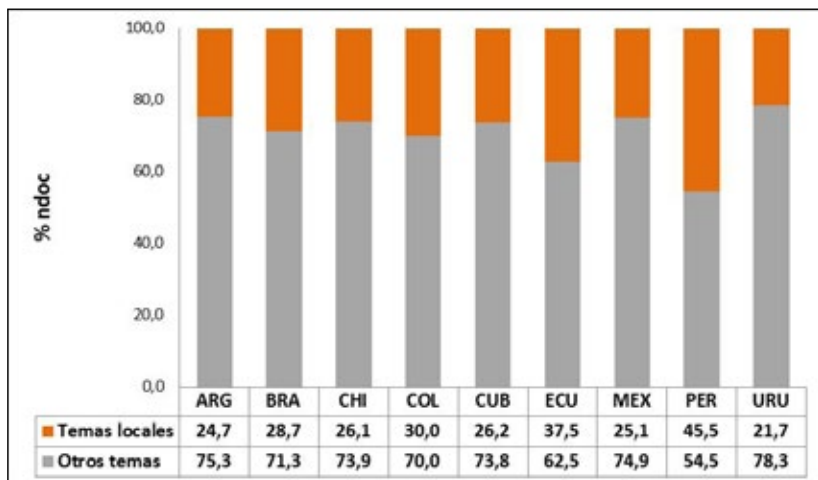


Fig. 2 Producción en temas locales según la procedencia de los autores (2007-2016)

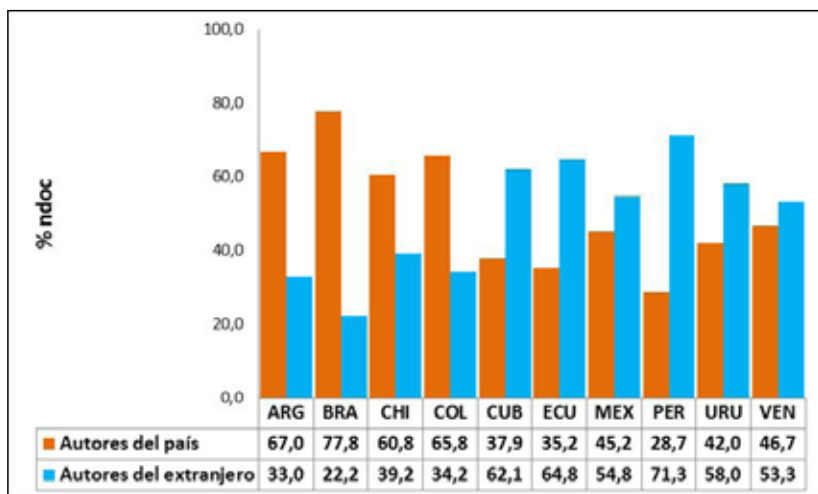


Fig. 3 Producción en temas locales según la procedencia de los autores, AGR (2007-2016)

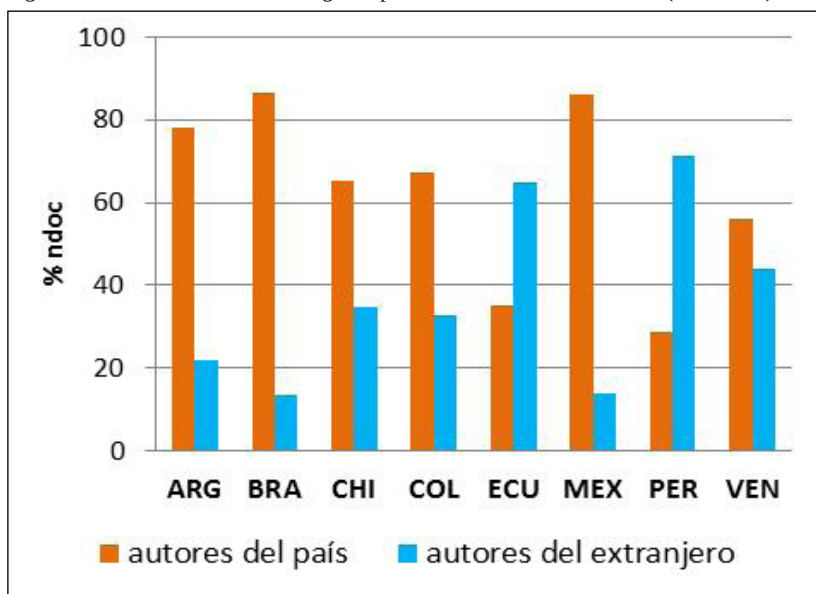


Fig. 4 Producción en temas locales según la procedencia de los autores, SOC (2007-2016)

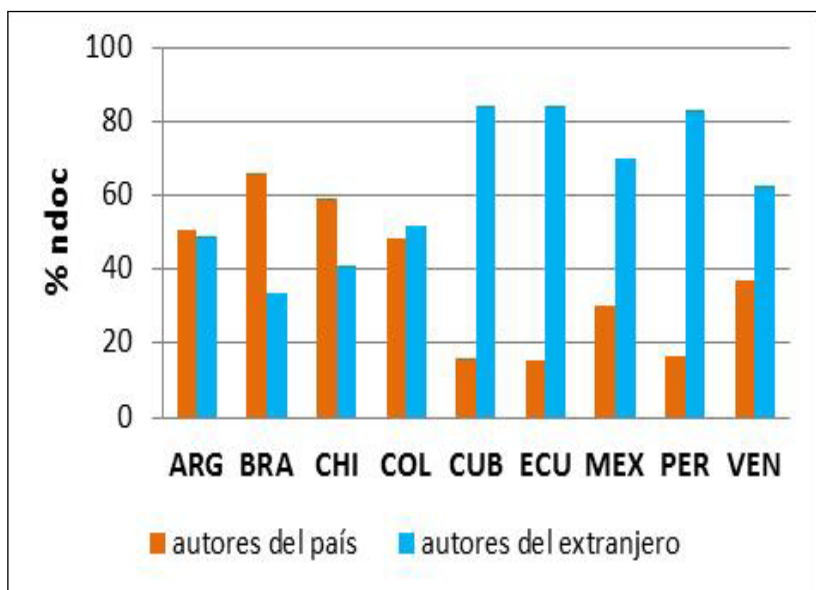


Fig. 5 Autores por documento en la producción en temas locales según procedencia de autores, AGR y SOC (2007-2016).

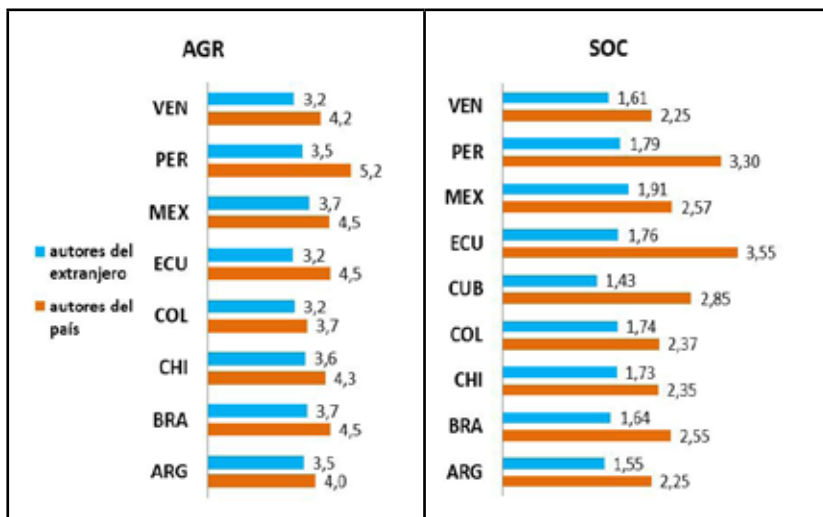


Fig. 6 Citas por documento en la producción en temas locales según procedencia de autores, AGR y SOC (2007-2016).

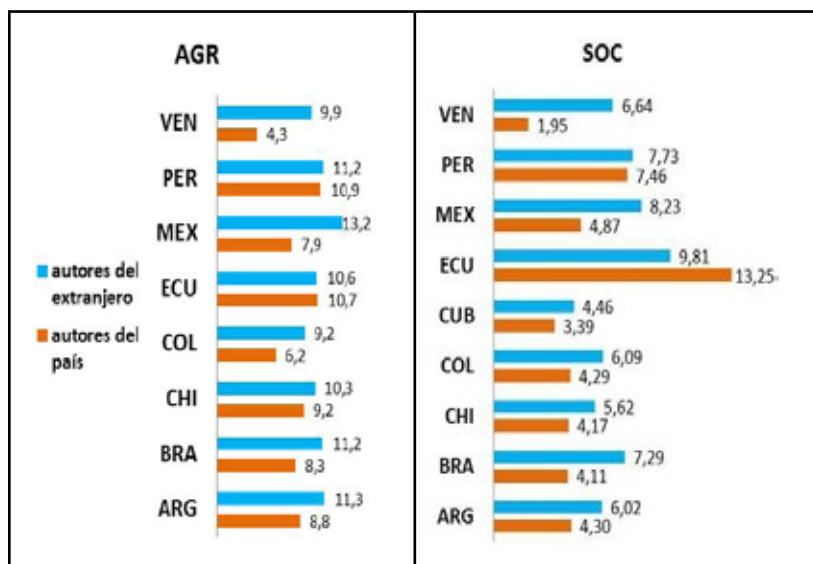


Fig. 7. Citas por documento según procedencia de revistas donde publican y de autores, AGR (2007-2016)

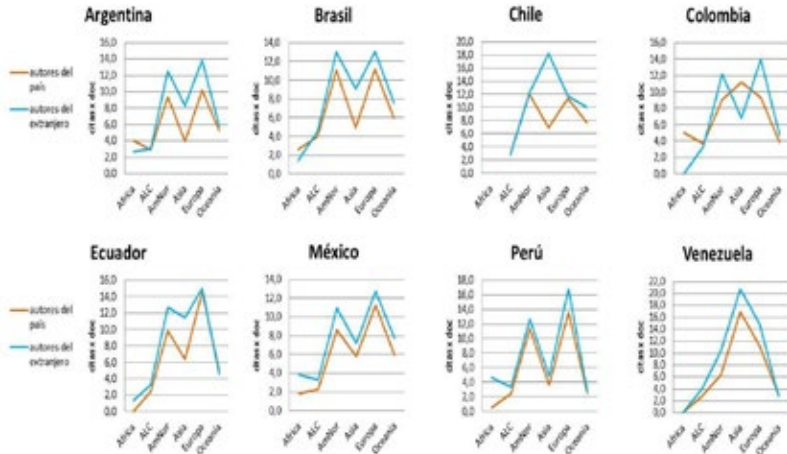
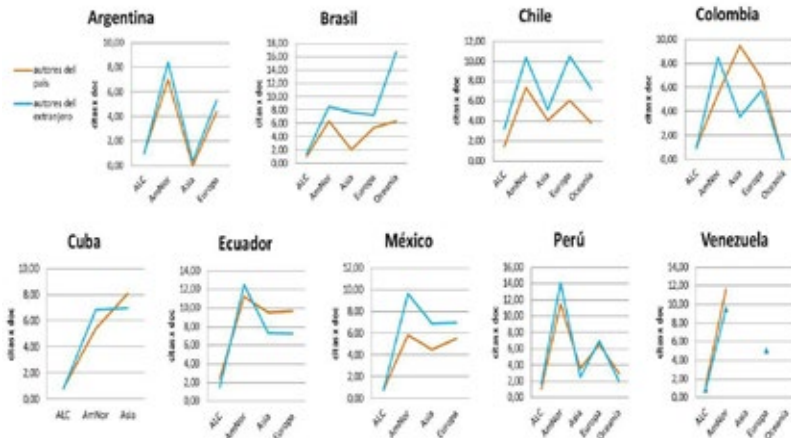


Fig. 8. Citas por documento según procedencia de revistas donde publican y de autores, SOC (2007-2016)



Discusión y conclusiones

La primera conclusión a la que arriba este estudio es que las investigaciones sobre América Latina son de interés tanto para los propios países de la región como para países externos. No obstante, en términos relativos al

total de la producción de cada país, el porcentaje de trabajos donde el país es tema o alcance territorial de la investigación varía considerablemente. Asimismo, llama la atención que en varios países el mayor volumen de producción en temas locales sea firmada por autores del extranjero. Este fenómeno, mayormente evidenciado en SOC, podría ser un indicio de la mayor o menor importancia que tienen esos estudios en las agendas de investigación, como también deberse a que tengan menos visibilidad en el contexto internacional de la ciencia, estando menos representados en la fuente de datos utilizada.

La segunda conclusión es que excepto el caso particular de Ecuador, los trabajos firmados por autores del país alcanzan menor impacto que los firmados por autores del extranjero. Este fenómeno no parece estar vinculado con la colaboración entre autores, desvaneciendo la idea de que a más número de autores mayor es el número de citas por documento recibidas. Tampoco parece guardar relación con la región de procedencia de las revistas donde publican, ya que aunque los trabajos en revistas de América del Norte y Europa son, en general, los que alcanzan mayor impacto; en promedio, los trabajos de extranjeros reciben más citas por documento que los de autores del país, incluso cuando publican en revistas latinoamericanas. Este fenómeno es más evidente en AGR que en SOC, donde se observan diferencias entre países. Si bien para el caso argentino en Ciencias Sociales y Humanidades, Chinchilla-Rodríguez et al. (2015) encontraron que los estudios en temas locales reciben menos citas que los de tópicos globales, estos resultados preliminares dan indicios de que más allá de los temas o las revistas elegidas para publicar, los autores latinoamericanos reciben, en general, menos citas que los autores de los países centrales, tal como afirman los trabajos de Meneghini et al. (2008) y Rau et al. (2017). Para confirmar esta hipótesis se requiere un estudio más exhaustivo y en más áreas temáticas.

Agradecimientos

Este trabajo ha sido realizado en el marco de los proyectos PICT 2015-2744 y H827 acreditados por Ministerio de Educación, Cultura y Ciencia y Tecnología de la República Argentina.

Referencias bibliográficas

Bodemer, K. (2010). Los estudios latinoamericanistas en Europa: legados históricos, situación actual y perspectivas. En *Anuario Americanista*

- Europeo*, 8, 1-20. Recuperado de <http://www.red-redial.net/revista/anuario-americanista-europeo/article/viewFile/140/110>
- Chinchilla-Rodríguez, Z., Miguel, S. y de Moya-Anegón, F. (2015). What factors affect the visibility of Argentinian publication in Humanities and Social Sciences in Scopus? Some evidence beyond the geographic realm of research. *Scientometrics*, 102(1), 789-810. Recuperado de <http://doi.org/10.1007/s11192-014-1414-4>
- Codner, D., Miguel, S. y Dutrenit, G. (2014). Midiendo el impacto de las políticas públicas a partir de indicadores bibliométricos. En *Taller de Indicadores de Ciencia y Tecnología en Latinoamérica* (pp. 117-126). México: Foro Consultivo Científico y Tecnológico.
- Lugones, G., Porta F. y Codner D. (2013). Perspectiva sobre el impacto del Programa de Modernización Tecnológica del BID en la política de CTI de Argentina. En Crespi, G. y Dutrenit G. (Ed.), *Políticas de ciencia, tecnología e innovación para el desarrollo. La experiencia latinoamericana*. México: Foro Consultivo Científico y Tecnológico. Recuperado de http://www.foroconsultivo.org.mx/libros_editados/politicas_de_cti.pdf
- Meneghini, R., Packer, A. y Nassi-Caló, L. (2008). Articles by Latin American Authors in Prestigious Journals have fewer citations. *PlosOne*, 3(1), e3804. Recuperado de <https://doi.org/10.1371/journal.pone.0003804>
- Miguel, S., González, C. y Chinchilla-Rodríguez, Z. (2015). Lo local y lo global en la producción científica argentina con visibilidad en Scopus, 2008-2012. Dimensiones nacionales e internacionales de la investigación. *Información, Cultura y Sociedad*, 32, 59- 78. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/1375/1352>
- Miguel, S. y Ugartemendía, V. (2015). Extranjeros que miran a nuestro país: la producción científica sobre Argentina entre 2008- 2012. *XI Jornadas de Sociología UBA*, Buenos Aires. Argentina.
- Rau, J.R., Monjeau, A., Cristobal-Pizarro, J. y Anderson, C.B. (2017). Cuanto más publicamos, menos nos citan. *Ecología Austral*, 27, 385-391. Recuperado de <https://doi.org/10.25260/EA.17.27.3.0.453>

Exploración de relaciones entre indicadores bibliométricos y otros indicadores del contexto económico, social y productivo

Edgardo Ortiz-Jaureguizar¹

Introducción

En la literatura bibliométrica existen diversos estudios que abordan la relación existente entre diferentes indicadores bibliométricos (*e.g.*, van Raan, 2006; Falagaset *al.*, 2008; García-Pachón y Arencibia-Jorge, 2014; Ortiz-Jaureguizar *et al.*, 2015). Sin embargo, a nivel de países, no existen estudios en los cuales se analice la información común que contienen los indicadores bibliométricos y otros indicadores, tales como aquellos provistos por el Banco Mundial o la Red de Indicadores de Ciencia y Tecnología -Iberoamericana e Interamericana- (RICYT).

El objetivo de este estudio es determinar si existe información común entre los indicadores bibliométricos de *ScimagoJournal and Country Rank* (SCIJCR) y los indicadores de contexto, graduación superior, insumo y bibliométricos provistos por la RICYT, tomando como unidades de estudio un conjunto de países iberoamericanos y europeos. Para ello se utilizará un enfoque empleado en la taxonomía biológica, tendiente a determinar el grado de correspondencia existente entre dos o más clasificaciones de los mismos objetos de estudio por medio de técnicas de análisis multivariado (Crisci y López Armengol, 1983).

¹ LASBE, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata. CONICET, La Plata, Argentina. eortiz@fcnym.unlp.edu.ar

Metodología

Las unidades de estudio fueron nueve países iberoamericanos [Brasil (BRA), México (MEX), Argentina (ARG), Chile (CHI), Colombia (COL), Venezuela (VEN), Cuba (CUB), Perú (PER) y Uruguay (URU)] a los que se agregaron España (ESP) y Portugal (POR). Como variables se utilizaron 11 indicadores bibliométricos tomados de *ScimagoJournal& Country Rank* (SCIJCR) (Documentos, Documentos citables, Documentos no citables, Citaciones, Autocitaciones, Citaciones externas, Documentos citados, Documentos no citados, Citaciones por documento, Índice h, y Porcentaje de colaboración internacional) y 23 indicadores tomados de la RICYT, de los cuales cuatro fueron de contexto [Población, Población económicamente activa (PEA), Producto bruto interno en dólares (PBI) y PBI en dólares expresados en Paridad de Poder de Compra (PPC)], dos de graduación superior (Titulados de grado y Doctores), cinco bibliométricos [(Publicaciones en SCOPUS por habitante, Publicaciones en SCOPUS en relación al PBI, Publicaciones en SCOPUS en relación al gasto en Investigación y Desarrollo (I+D), Publicaciones en SCOPUS cada 100 investigadores expresados en Personas Físicas (PF), y Publicaciones en SCOPUS cada 100 investigadores expresados en Equivalente a Jornada Completa (EJC)] y 12 de insumo [(Gasto en I+D en dólares, Gasto en I+D en PPC, Gasto en I+D en relación al PBI, Gasto en I+D por habitante en dólares, Gasto en I+D por habitante en dólares expresados en PPC, Gasto en I+D por investigador en miles de dólares (PF), Gasto en I+D por investigador en miles de dólares (EJC), Gasto en I+D por investigador en miles de dólares PPC (PF), Gasto en I+D por investigador en miles de dólares PPC (EJC), Personal (investigadores) en I+D (PF), Investigadores cada 1000 habitantes de la PEA (PF), y Personal (investigadores) en I+D (EJC)].

Tanto para los indicadores de SCIJCR como para aquellos de la RICYT, los datos fueron los del periodo 2006-2015, por ser este el período en el cual podían tomarse datos para los mismos años en ambas fuentes de información. De tal modo, para cada país el valor de cada uno de los indicadores representó el promedio de los valores de dicho indicador para los años comprendidos en el periodo arriba mencionado.

Para establecer las relaciones de similitud entre los países, se construyeron dos matrices básicas de datos, una de 11 países por los 11 indicadores bibliométricos de SCIJCR y otra de 11 países por los 23 indicadores

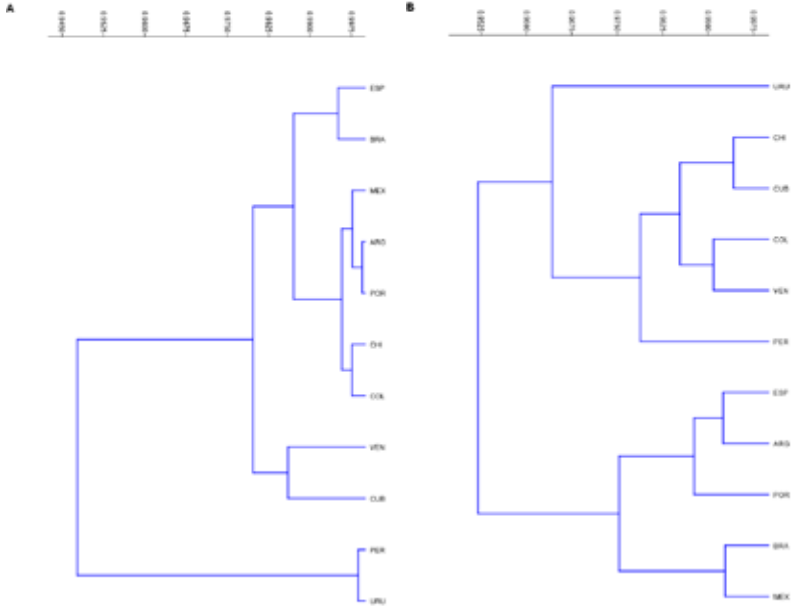
de RICYT. Ambas matrices fueron estandarizadas, para disminuir el impacto de aquellas variables donde la diferencia entre los valores máximos y mínimos era muy acentuada. Para obtener y graficar las relaciones de similitud entre los países, las matrices fueron analizadas por medio de dos técnicas de análisis multivariado: fenogramas y análisis de los componentes principales (ACP). La distorsión de los fenogramas se estimó por medio del coeficiente de correlación cofenética (CCC), en tanto que las distorsiones en las distancias entre países que pudiesen surgir en el gráfico del ACP fueron evaluadas mediante un árbol de distancias mínimas (ADM). Para más detalles acerca de estas técnicas, véase Sneath y Sokal (1973), Crisci y López Armengol (1983), Sokal (1986) y Cuadras (2007).

Todos los cálculos vinculados con el análisis multivariado, así como la representación gráfica del mismo, se realizaron con el programa PAST 3.20 (Hammeret *al.*, 2001).

Resultados y discusión

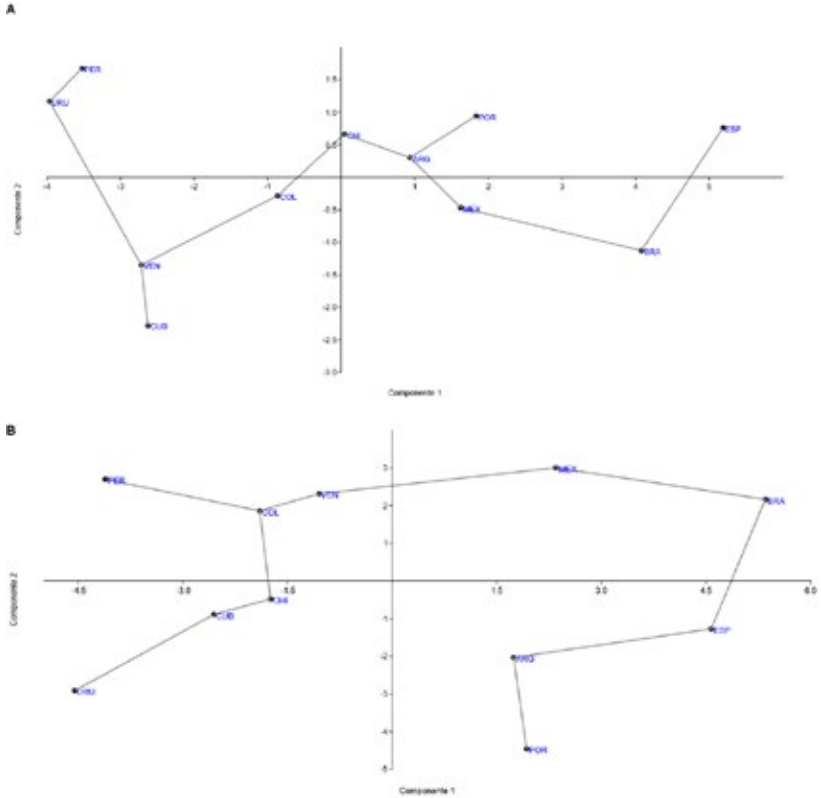
En el caso de los fenogramas, el basado en las variables de SCIJCR (Figura 1A) muestra nueve grupos: 1) Perú, Uruguay; 2) España, Brasil, México, Argentina, Portugal, Chile, Colombia, Venezuela, Cuba; 3) España, Brasil, México, Argentina, Portugal, Chile, Colombia; 4) Venezuela, Cuba; 5) España, Brasil; 6) México, Argentina, Portugal, Chile, Colombia; 7) México, Argentina, Portugal; 8) Chile, Colombia; y 9) Argentina, Portugal. Por su parte, el fenograma basado en las variables de la RICYT (Figura 1B) también muestra nueve grupos: 1') España, Argentina, Portugal, Brasil, México; 2') Uruguay, Chile, Cuba, Colombia, Venezuela, Perú; 3') España, Argentina, Portugal; 4') Brasil, México; 5') España, Argentina; 6') Chile, Cuba, Colombia, Venezuela, Perú; 7) Chile, Cuba, Colombia, Venezuela; 8') Chile, Cuba; y 9) Colombia, Venezuela. No existen grupos comunes en ambos fenogramas.

Figura1.Comparación de los fenogramas de similitud entre países (coeficiente de correlación de Pearson-ligamientos promedios no ponderados). A. Basado en las 11 variables de SCIJR (CCC: 0,81); B: basado en las 23 variables de la RICYT (CCC: 0,65). Para los acrónimos, véase Metodología.



En el caso de los análisis de los componentes principales, aquel basado en las variables de SCIJR (Figura 2A) muestra cuatro grupos: 1) Perú, Uruguay; 2), Venezuela, Cuba; 3) Colombia, Chile, Argentina, México, Portugal; y 4) Brasil, España. Por su parte, aquel basado en las variables de la RICYT (Figura 2B) muestra cuatro grupos: (1') Argentina, Portugal; 2') Brasil, España; 3') Venezuela, Colombia; y 4') Chile, Cuba. En este caso, solo se observa un grupo común entre ambos gráficos (Brasil, España) y, aunque de manera más difusa, también se observa que, por un lado, Brasil, España, Portugal, Argentina y México y, por el otro, Venezuela, Colombia, Uruguay, Chile y Cuba, aparecen en ambos gráficos del lado de los mayores y menores valores del primer componente principal, respectivamente. Asimismo, existe cierta relación entre algunos países en ambos gráficos (*i.e.*, Argentina-Portugal, Brasil-México, Chile-Colombia, y Colombia-Venezuela) aunque no puedan reconocerse como grupos en los dos gráficos.

Figura 2. Comparación de los gráficos bidimensionales (componentes 1 y 2) del análisis de componentes principales entre países. A. Basado en las 11 variables de SCIJCR; B: basado en las 23 variables de la RICYT. En ambas figuras, las líneas que conectan los países corresponden al ADM. Para los acrónimos, véase Metodología.



Los resultados obtenidos sugieren que existe poca o ninguna relación entre ambos conjuntos de variables para el grupo de países estudiados. Los fenogramas descartan la existencia de información común, aunque debe señalarse que esto puede obedecer a la propia forma de construir los agrupamientos. La naturaleza jerárquica de los fenogramas empleados puede forzar la separación de objetos con valores de similitud similares. Por otra parte, el hecho de que los métodos de ordenación producen agrupamientos que carecen de estructura jerárquica, puede explicar el reconocimiento de agrupamientos reunidos por criterios menos restrictivos (James y McCulloch, 1990).

Con la finalidad de profundizar este estudio, se realizará un análisis similar al aquí presentado, con el agregado de reunir en una tercera matriz de datos a las 11 variables de SCIJCR con las 23 de la RICYT. Este estudio conjunto permitirá evaluarse la combinación de las variables produce una clasificación diferente, o es similar a alguna de las dos obtenidas en esta presentación. Asimismo, al correlacionar ambos grupos de variables en un mismo análisis, será posible poner a prueba la hipótesis de que la relación existente entre ambos conjuntos de variables es nula o escasa para el grupo de países estudiados.

Conclusión

Para el conjunto de los 11 países estudiados, las 11 variables de SCIJCR y las 23 variables de RICYT no producen clasificaciones congruentes al emplearse fenogramas, mientras que solo se recupera un único grupo al utilizarse el análisis de componentes principales. Esto permite inferir que ambos conjuntos de variables poseen muy poca o ninguna información común, siendo dicha información fundamentalmente complementaria.

Se realizará un análisis conjunto de las 34 variables para los mismos 11 países y se compararán los resultados con los aquí obtenidos, a fin de ver si la combinación de las variables produce una clasificación diferente, o es similar a alguna de las dos vistas en esta presentación. Asimismo, al correlacionar ambos grupos de variables en un mismo análisis, será posible testear la hipótesis de que la información común es nula o escasa

Referencias bibliográficas

- Crisci, J.V., & **López Armengol, M.F.** (1983). *Introducción a la teoría y práctica de la Taxonomía Numérica*. Washington: Monografías de la Organización de los Estados Americanos.
- Cuadras, C.M. (2007). *Nuevos métodos de análisis multivariante*. Barcelona: CMC Editions.
- Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., y Karageorgopoulos, D.E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal*, 22,2623-2628. doi: 10.1096/fj.08-107938
- García-Pachón, E., y Arencibia-Jorge, E. (2014). Comparación del factor de impacto y el índice SCImagoJournal Rank en las revistas del sistema respiratorio. *Archives of Bronconeumology*, 50(7), 308-309. Recuperado

- de <http://www.archbronconeumol.org/es-comparacion-del-factor-impacto-el-articulo-resumen-S030028961300313X>
- Hammer, Ø., Harper, D.A.T., y Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1), 1-9.
- James, F.C., y McCulloch, C.F. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review in Ecology and Systematics*, 21, 129-166.
- Ortiz Jaureguizar, E., Miguel, S., y Posadas, P. (2015). Relaciones de similitud y valor discriminativo de los indicadores bibliométricos: los indicadores de ScimagoJournal & Country Rank en las revistas generalistas de Paleontología. En *Actas de las 4as Jornadas de intercambio y reflexión acerca de la investigación en Bibliotecología*. Recuperado de http://jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2015/archivos-pdf/jjrib2015_ORTIZ_texto.pdf
- Sneath, P.H.A., y Sokal, R.E. (1973). *Numerical taxonomy*. San Francisco: W.H. Freeman Co.
- Sokal, R.R. (1986). Phenetic taxonomy: theory and methods. *Annual Review of Ecology and Systematics*, 17, 423-442.
- van Raan, A.F.J. (2006). Comparisons of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502.

En un contexto de creciente incorporación del acontecer local y nacional en contextos globales, el impulso de políticas que orientan la investigación hacia temas locales y a la resolución de problemas sociales, productivos y medioambientales, plantea nuevos desafíos en la valoración de los resultados e impacto que esa investigación produce, así como también de la transferencia e innovación. Este libro de actas reúne las ponencias presentadas en el Workshop Iberoamericano de estudios métricos de la actividad científica y tecnológica en temas locales/regionales, con aportes de autores de diferentes perfiles formativos y trayectorias, en un intento por contribuir a los debates teóricos y metodológicos para la obtención de métricas y visualizaciones derivadas de la producción científica de países iberoamericanos y la relación con indicadores del contexto económico y social. Los contenidos están organizados en tres ejes temáticos. El primero se enfoca en cuestiones relativas a las políticas y agendas de investigación; el segundo presenta diferentes aproximaciones metodológicas para el abordaje cuantitativo de la producción científica en temas locales/regionales, y el tercero recoge estudios de caso de un grupo de países de la región.

**Trabajos, Comunicaciones
y Conferencias, 37**

ISBN 978-950-34-1742-3