

El canon literario hispanoamericano en la era digital

CAROLINA FERRER
Université du Québec à Montréal
ferrer.carolina@uqam.ca

El concepto de canon es uno muy difícil de definir. Según algunos textos de referencia (Guillory, 1995; Lentricchia y McLaughlin, 1995), etimológicamente, la palabra proviene del griego *kanon* que significa regla o vara para medir. Inicialmente, el término es empleado para identificar aquellos libros del Antiguo y del Nuevo Testamentos que cuentan con la aprobación de las autoridades eclesiásticas en cuanto corresponden a la palabra divina, es decir aquellos textos que constituyen las Sagradas Escrituras. En las últimas décadas, en los estudios literarios se considera que la canonización de los textos literarios opera de manera análoga al proceso de constitución del canon bíblico.

En 1995, al comienzo de su libro *El canon occidental: los libros y la escuela de los años*, el crítico estadounidense Harold Bloom determina que:

Originalmente el Canon significaba la selección de libros en nuestras instituciones de enseñanza y, a pesar de las recientes políticas de multiculturalismo, la verdadera pregunta del Canon sigue siendo: ¿qué debería tratar de leer un individuo que desea leer a estas alturas de la historia? (1995: 15)¹.

Vista así, la pregunta fundamental del canon parece simple. Sin embargo, se trata de una noción significativamente compleja que ha suscitado polémicas considerables. Así, por ejemplo, en 1983, *Critical Inquiry* publica un volumen completo dedicado al concepto del canon. En su introducción, Robert von Hallberg establece que:

El interés en los cánones forma sin duda parte de una investigación más amplia dentro de las instituciones de estudio literario y de producción artística. 'Política', 'economía', 'social', 'autoridad', 'poder' –estos son algunos de los términos recurrentes a través de estos ensayos; nos mostramos muy curiosos ahora sobre aquellos puntos donde el arte parece ser menos privado que social» (1983: iii)².

¹ "Originally the Canon meant the choice of books in our teaching institutions, and despite the recent politics of multiculturalism, the Canon's true question remains: What shall the individual who still desires to read attempt to read, this late in history?" (Bloom, 1995: 15). Todas las traducciones son mías.

² "Interest in canons is surely part of a larger inquiry into the institutions of literary studies and artistic production. 'Politics', 'economics', 'social', 'authority', 'power' –these are some

En 1999, Nel van Djik analiza la formación del canon desde diferentes perspectivas: el nacionalismo, la literatura y el punto de vista institucional. Para él:

la lista de obras que constituye la herencia literaria de la sociedad occidental ya no es prescrita por la iglesia o el Estado, sino que por instituciones autorizadas tales como la crítica y la educación literarias. Asimismo, el desarrollo científico de las últimas décadas se ha traducido en una convicción generalizada de que el valor literario no es un valor intrínseco sino que una cualidad atribuida. Esta cualidad resulta del consenso que existe entre los miembros de una institución literaria en un determinado momento (121)³.

Según van Djik, los investigadores que trabajan sobre este tema se dividen en dos grupos: aquellos que lo abordan desde un punto de vista ideológico-hermenéutico y aquellos que optan por un enfoque sociológico.

Un año más tarde, Anderson y Zanetti (2000) determinan que el debate en torno al problema del canon se ha efectuado según dos polos opuestos. Por un lado, aquellos situados a la derecha del espectro político defienden un canon tradicional. Por otro lado, aquellos que poseen ideas de izquierda declaran que el canon es un artefacto obsoleto. Entre ambos, se sitúan una serie de posturas que abren la noción de canon con el propósito de incluir a las minorías o que promueven la existencia de múltiples cánones.

En el año 2003, Jeffrey Insko retoma el debate sobre este concepto, haciendo resurgir la tensión entre un canon imaginario y un canon pedagógico. Al año siguiente, Frank Kermode, en su libro *Placer y cambio: la estética del canon* (2004) deja de lado los aspectos ideológicos de la constitución del canon para introducir tres nuevas propiedades en dicho proceso: placer, cambio y azar.

Mi intención, al mencionar estos diversos estudios sobre el concepto del canon, no es continuar este debate, sino mostrar que la constitución de un corpus canónico es un proceso altamente complejo que no deja indiferente a nadie. Asimismo, me propongo introducir una metodología alternativa para identificar aquellas obras que son más frecuentemente analizadas al interior de una literatura nacional⁴. Es decir, en vez de presentar un canon literario, mi objeto será trazar una cartografía literaria. Específicamente, mi análisis se centra en las letras de Hispanoamérica.

ENFOQUES TEÓRICO Y METODOLÓGICO

Con el propósito de desarrollar este método experimental, baso mi investigación en la articulación de varias nociones esenciales. Desde el punto de vista

of the terms that recur throughout these essays; we are most curious now about those points where art seems less private than social" (von Halberg, 1983: iii).

³ "The list of works that count as our western society's literary inheritance is no longer prescribed by the church and the state, but by authorized institutions such as literary criticism and literary education. In addition, scientific developments in the past decades have resulted in the widespread conviction that literary value is not an intrinsic but an attributed quality. This quality results from the consensus that exists between the members of a literary institution at a certain moment" (van Djik, 1999: 121).

⁴ En su libro *Dialogues with/and Great Books* (2012), David Fishelov utiliza indicadores bibliométricos y Webométricos no con el propósito de proponer un método alternativo a la constitución del canon, sino que para confirmar la importancia de las listas canónicas ya existentes.

teórico, fundo mi trabajo, por un lado, en el concepto de campo literario establecido por Pierre Bourdieu (1992), y, por otro lado, en la cienciometría (Price, 1963; Garfield, 1980; Leydesdorff, 1998). Metodológicamente, dos enfoques resultan medulares: en primer lugar, el análisis de campo del conocimiento (Albrechtsen, 1997; Hjørland, 2001; Hjørland y Albrechtsen, 1995) y, en segundo lugar, las teorías de las citas (Kaplan, 1965; Moravcsik y Murugesan, 1975; Gilbert, 1977; Small, 1978, 1998; Leydesdorff, 1998; Enger, 2009).

Si bien varios cientometristas se muestran reacios a la utilización de técnicas cuantitativas para analizar el comportamiento de las humanidades debido a las diferencias en cuanto a las formas de citar a través de las disciplinas (Cole, 1983; Cozzens, 1985; Larivière *et al.*, 2006; Nederhof *et al.*, 1989), recientemente han aparecido numerosos análisis bibliométricos relativos a las humanidades, en general, y a los estudios literarios, en particular (Ardanuy *et al.*, 2009; Hammarfelt, 2011; Hérubel y Goedeken, 2000; Osca-Lluch y Haba, 2005). Estas últimas publicaciones indican que, a pesar de las diferencias antes mencionadas, los estudios cientométricos constituyen un enfoque pertinente para aumentar nuestra comprensión del campo literario.

Por ello, en los últimos años, he llevado a cabo diversos estudios cuantitativos relativos al área de las humanidades que me han permitido, entre otras cosas, crear un atlas de los recursos bibliográficos digitales, así como una cartografía de los estudios sobre la literatura hispanoamericana (Ferrer, 2012).

En el presente estudio, la base de datos utilizada es la Modern Language International Bibliography, que de ahora en adelante llamaré MLAIB. Esta bibliografía electrónica, mundialmente reconocida como la más importante en el área de los estudios literarios, contiene más de 2.200.000 referencias e incluye aproximadamente 4.400 revistas académicas. Además de los artículos, la base MLAIB contiene las referencias de libros, capítulos de libros y tesis⁵. En cuanto a su cobertura cronológica, ésta se extiende de 1896 al presente.

A través de las técnicas de *data mining* (Han *et al.*, 2012) y de palabras clave (Callon *et al.*, 1993), obtengo, en primer lugar, una muestra de la bibliografía crítica sobre la literatura hispanoamericana. La fecha límite utilizada es el año 2010.

CARTOGRAFÍA LITERARIA

Los resultados de esta exploración arrojaron una muestra de más de 75.000 documentos. En la figura 1, podemos observar dicha bibliografía clasificada en las diecinueve literaturas nacionales que constituyen Hispanoamérica. Adicionalmente a las referencias compiladas en el gráfico, hay un subconjunto de más de 11.000 referencias sobre las letras hispanoamericanas sin mencionar país alguno. Puesto que en este estudio me concentro en las literaturas nacionales, he omitido dichas referencias. Como podemos apreciar, la importancia relativa de los diferentes países es muy variada. Argentina y México representan cada uno 22% de la muestra y sus literaturas se destacan claramente como las más estudiadas según la base MLAIB. Luego, Cuba, Chile, Perú y Colombia muestran una bibliografía relativamente importante con muestras que fluctúan entre 12%

⁵ En este estudio, no considero las tesis, puesto que la MLAIB incluye sólo las tesis de las universidades norteamericanas. Ver Fitz-Enz (2008).

Bibliografía por literatura nacional MLAIB 1896-2010

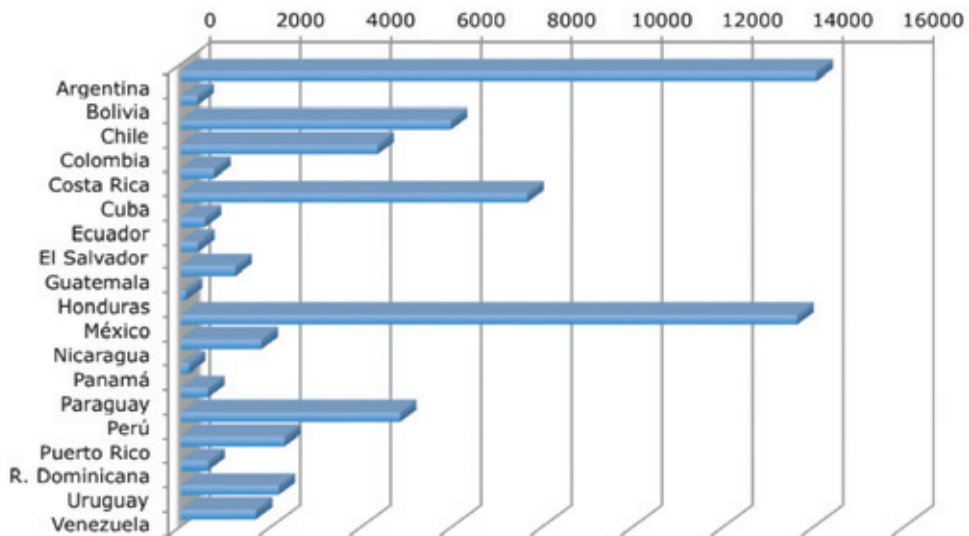


Fig. 1. Bibliografía crítica de la literatura hispanoamericana por literatura nacional.

y 7% del total. Las bibliografías sobre las literaturas de Puerto Rico, Uruguay, Nicaragua, Venezuela y Guatemala acumulan muestras pequeñas que se sitúan entre 2.300 y 1.200 referencias. Por último, Costa Rica, República Dominicana, Paraguay, Ecuador, El Salvador, Bolivia, Panamá y Honduras reflejan una situación bastante embrionaria con bibliografías que oscilan entre 800 y 100 referencias por país.

ESCRITORES Y OBRAS

Ahora bien, con el propósito de identificar las obras y los autores más estudiados de Hispanoamérica, interrogué nuevamente la base MLAIB. En la figura 2, represento, por país, el número de autores y el número de obras que reúnen 10 o más referencias dentro de la base de datos. Llegamos así a identificar 717 autores y 451 obras. Desde luego, Argentina y México siguen ubicándose en primera fila según ambos indicadores.

En el cuadro 1, presento una lista con los 20 escritores hispanoamericanos que registran el mayor número individual de referencias en la base MLAIB. Nuevamente, observamos que Argentina y México se destacan cada uno con 4 autores, seguidos muy de cerca por Cuba y Perú cada uno con 3 autores. Cabe mencionar que los seis escritores hispanoamericanos galardonados con el Premio Nobel de literatura se encuentran presentes. Sin embargo, el líder absoluto de este indicador es Jorge Luis Borges, quien nunca fuera reconocido por la Academia sueca. Asimismo, apreciamos la coexistencia de autores de distintas épocas, de Juana Inés de la Cruz a Mario Vargas Llosa. Obviamente, aquellos

**Número de autores y de obras con 10 o más referencias
MLAIB 1896-2010**

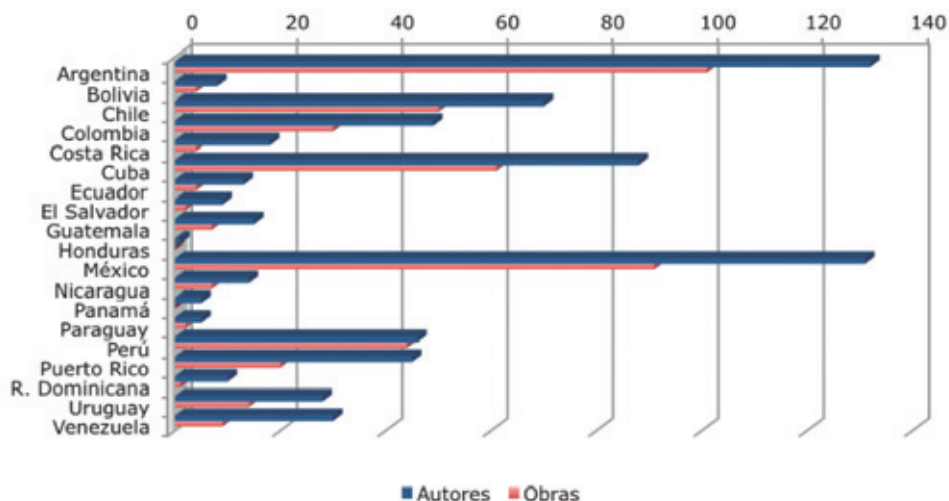


Fig. 2. Número de autores y de obras por literatura nacional.

Cuadro 1. Escritores hispanoamericanos con 100 o más referencias

País	Autor
Argentina	Borges, Jorge Luis (1899-1986)
Colombia	García Márquez, Gabriel (1928-)
Argentina	Cortázar, Julio (1914-1984)
México	Fuentes, Carlos (1928-2012)
Nicaragua	Darío, Rubén (1867-1916)
Chile	Neruda, Pablo (1904-1973)
México	Paz, Octavio (1914-1998)
Cuba	Martí, José (1853-1895)
Cuba	Carpentier, Alejo (1904-1980)
Perú	Vargas Llosa, Mario (1936-)
México	Juana Inés de la Cruz (1648-1695)
México	Rulfo, Juan (1918-1986)
Perú	Vallejo, César (1892-1938)
Guatemala	Asturias, Miguel Ángel (1899-1974)
Argentina	Puig, Manuel (1932-1990)
Perú	Arguedas, José María (1911-1969)
Chile	Mistral, Gabriela (1889-1957)
Argentina	Sarmiento, Domingo Faustino (1811-1888)
Uruguay	Onetti, Juan Carlos (1909-1994)
Cuba	Lezama Lima, José (1912-1976)

autores más recientes tienen una probabilidad relativamente baja de encontrarse en esta lista, puesto que no han alcanzado aún a acumular un número significativo de referencias. De hecho, sólo dos de los escritores de esta lista se encuentran aún en vida: Gabriel García Márquez y Mario Vargas Llosa. Por último, entre estos 20 escritores, observamos la presencia de sólo dos mujeres: Juana Inés de la Cruz, mexicana, y Gabriela Mistral, chilena.

En el cuadro 2, encontramos las obras hispanoamericanas que poseen 100 o más referencias en la base MLAIB. *Cien años de soledad*, del escritor colombiano Gabriel García Márquez, se destaca claramente como la obra preferida por los académicos. Como podemos constatar, la lista incluye obras desde el siglo XVII, tal como *Comentarios reales de los incas* de Garcilaso de la Vega, hasta los años ochenta. Al igual que en el caso de los autores, debido al efecto acumulativo de la base bibliográfica, las obras más recientes se encuentran difícilmente presentes en esta lista. En este caso, hay tres obras de escritoras: Juana Inés de la Cruz, mexicana, Isabel Allende, chilena, y Rigoberta Menchú, guatemalteca galardonada en 1992 con el premio Nobel de la Paz.

Cuadro 2. Obras hispanoamericanas con 100 o más referencias

<i>País</i>	<i>Autor</i>	<i>Obra</i>	<i>Ref.</i>
Colombia	García Márquez, Gabriel	<i>Cien años de soledad</i> (1967)	412
México	Rulfo, Juan	<i>Pedro Páramo</i> (1955)	270
Argentina	Cortázar, Julio	<i>Rayuela</i> (1963)	184
Argentina	Puig, Manuel	<i>El beso de la mujer araña</i> (1976)	164
Paraguay	Roa Bastos, Augusto	<i>Yo el Supremo</i> (1974)	152
Argentina	Sarmiento, Domingo Faustino	<i>Facundo</i> (1845)	150
Guatemala	Menchú, Rigoberta	<i>Me llamo Rigoberta Menchú</i> (1985)	149
Perú	Garcilaso de la Vega, el Inca	<i>Comentarios reales de los incas</i> (1609)	147
Colombia	García Márquez, Gabriel	<i>Crónica de una muerte anunciada</i> (1981)	137
Chile	Allende, Isabel	<i>La casa de los espíritus</i> (1982)	132
Perú	Poma de Ayala, Felipe Huamán	<i>El primer nueva corónica y buen gobierno</i> (1615-16)	124
Cuba	Carpentier, Alejo	<i>Los pasos perdidos</i> (1953)	118
Cuba	Cabrera Infante, Guillermo	<i>Tres tristes tigres</i> (1967)	116
México	Juana Inés de la Cruz	<i>Primero sueño</i> (1680-81)	116
México	Fuentes, Carlos	<i>Terra Nostra</i> (1975)	115
Cuba	Lezama Lima, José	<i>Paradiso</i> (1966)	113
Cuba	Carpentier, Alejo	<i>El reino de este mundo</i> (1949)	109
México	Fuentes, Carlos	<i>La muerte de Artemio Cruz</i> (1962)	103
Argentina	Hernández, José	<i>Martín Fierro</i> (1872, 1879)	102

El cuadro 3 corresponde a los escritores que acumulan el mayor número de obras con 10 o más referencias en la base MLAIB. A nivel de país, México presenta 6 autores, seguido por Chile con 5 autores. Individualmente, el líder es nuevamente Jorge Luis Borges. Al igual que en los otros casos, el período cubierto se extiende del siglo XVII, con Juana Inés de la Cruz, hasta el presente. En este caso, el número de escritoras alcanza a 6: las chilenas Isabel Allende y Diamela Eltit, las mexicanas Juana Inés de la Cruz, Elena Poniatowska y Carmen Boullosa y la portorriqueña Rosario Ferré.

Cuadro 3. Escritores con el mayor número de obras con 10 o más referencias

<i>País</i>	<i>Autor</i>	<i>Nº de obras</i>
Argentina	Borges, Jorge Luis	24
Perú	Vargas Llosa, Mario	18
México	Fuentes, Carlos	16
Colombia	García Márquez, Gabriel	15
Argentina	Cortázar, Julio	14
Cuba	Carpentier, Alejo	13
Chile	Eltit, Diamela	8
Chile	Neruda, Pablo	8
Chile	Allende, Isabel	7
Cuba	Martí, José	7
Uruguay	Onetti, Juan Carlos	7
Argentina	Puig, Manuel	7
México	Juana Inés de la Cruz	6
Cuba	Arenas, Reinaldo	5
Perú	Arguedas, José María	5
Chile	Bolaño, Roberto	5
México	Boullosa, Carmen	5
Chile	Donoso, José	5
Puerto Rico	Ferré, Rosario	5
México	Paz, Octavio	5
México	Poniatowska, Elena	5
México	Rulfo, Juan	5
Cuba	Sarduy, Severo	5
Perú	Vallejo, César	5

REFLEXIONES FINALES

Al inicio de este artículo, evoqué diversos aspectos que, según los académicos, dificultan el proceso de identificación de las obras literarias canónicas. Luego, introduje una metodología alternativa que se basa, esencialmente, en la minería

de datos. De esta forma, logré obtener un corpus de obras que revela el interés que, desde hace más de cien años, la crítica académica ha manifestado por la literatura hispanoamericana. A mi modo de ver, esta nueva metodología permite sobrepasar las diferentes tensiones relativas al proceso de constitución del canon y, al mismo tiempo, responde tanto a lo planteado por Nel van Dijk, cuanto a la interrogante de Harold Bloom.

Por un lado, según van Dijk, el canon es prescrito por "instituciones autorizadas tales como la crítica y la educación literarias" (1999: 121)⁶. En este sentido, el método aquí propuesto es justamente una forma de considerar una parte muy significativa de la actividad crítica académica. Al utilizar la base MLAIB y explotarla con métodos cuantitativos, la bibliografía obtenida no está sujeta a criterios de selección subjetivos, sino que se basa en la ley de los grandes números. En efecto, en el caso de Hispanoamérica, los resultados aquí compilados son el reflejo de más de 75.000 referencias obtenidas con técnicas cuantitativas reproducibles y verificables.

Por otro lado, la cartografía presentada responde a la interrogante de Harold Bloom. Es decir, adecuándola a la zona geográfica en cuestión, "¿qué debería tratar de leer un individuo que desea leer [la literatura de Hispanoamérica] a estas alturas de la historia?" (1995: 15)⁷. En este sentido, me gustaría hacer notar que los veintiocho escritores hispanoamericanos que Bloom incluye en su libro sobre el canon occidental están presentes en esta nueva cartografía.

Sin embargo, considero que la bibliografía obtenida es claramente superior a la lista canónica de Bloom, puesto que, en primer lugar, estos resultados pueden ser clasificados por número de citas y ello según diferentes parámetros: geográfico, cronológico y lingüístico, entre otros. En segundo lugar, esta cartografía es dinámica, ya que es fácilmente actualizable mediante la interrogación periódica de la base MLAIB. Por último, el corpus obtenido es claramente más rico y diverso que la lista canónica propuesta por Bloom que, en el caso de Hispanoamérica, se compone de sólo treinta y seis obras y diecisiete autores pertenecientes a ocho países. Dicha lista, que data de 1995, es claramente más restringida que las cuatrocientos cincuenta y una obras, los setecientos diecisiete autores y los diecinueve países que constituyen la cartografía propuesta. Cabe mencionar que la célebre lista de Bloom se concentra en el siglo XX y sólo considera una obra de fines del siglo XIX. Por lo demás, dicho canon no incluye a ninguna escritora, puesto que Bloom incluye a la incontestablemente sobresaliente Juana Inés de la Cruz en el canon literario de España de lo que él denomina "La era aristocrática" (1995: 503).

A mi parecer, este estudio es una demostración más de la pertinencia y la importancia de introducir métodos cuantitativos en los estudios literarios como una manera de incrementar y de profundizar el conocimiento relativo a nuestra disciplina. Por ello, considero fundamental el desarrollo de las humanidades digitales.

⁶ "by authorized institutions such as literary criticism and literary education" (van Dijk, 1999: 121).

⁷ "What shall the individual who still desires to read [Spanish American literature] attempt to read, this late in history?" (Bloom, 1995: 15).

AGRADECIMIENTOS

Esta investigación contó con el apoyo del Consejo de Investigación en Ciencias Sociales y Humanidades de Canadá (Social Sciences and Humanities Research Council of Canada).

BIBLIOGRAFÍA

- Albrechtsen, Hanne, "Knowledge organization in the humanities", *Knowledge Organization*, 24, 2 (1997), pp. 61-63.
- Anderson, Earl R. y Zanetti, Gianfrancesco, "Comparative Semantic Approaches to the Idea of a Literary Canon", *The Journal of Aesthetics and Art Criticism*, 58, 4 (2000), pp. 341-360.
- Ardanuy, Jordi; Urbano, Cristóbal y Quintana, Lluís, "A Citation Analysis of Catalan Literary Studies (1974-2003): Towards a Bibliometrics of Humanities Studies in Minority Languages", *Scientometrics*, 8, 12 (2009), pp. 347-366.
- Bloom, Harold, *The Western Canon: the Books and School of the Ages*, New York, Riverhead, 1995.
- Bourdieu, Pierre, *Les règles de l'art. Genèse et structure du champ littéraire*, París, Seuil, 1992.
- Callon, Michel; Courtial, Jean-Pierre y Penan, Hervé, *La Scientométrie, Que Sais-Je?*, París, Presses Universitaires de France, 1993.
- Cole, Stephen, "The Hierarchy of the Sciences", *American Journal of Sociology*, 89, 1 (1983), pp. 111-139.
- Cozzens, Susan E., "Using the Archive – Price, Derek Theory of Differences among the Sciences", *Scientometrics*, 7, 3-6 (1985), pp. 431-441.
- Enger, K. Brock, "Using citation analysis to develop core book collections in academic libraries", *Library & Information Science Research*, 31, 2 (2009), pp. 107-112.
- Ferrer, Carolina, "Los estudios literarios hispanoamericanos y las bases de datos digitales en la era de la globalización", en *América Latina, globalidad e integración*, Madrid, Ediciones del Orto, 2012, pp. 635-646.
- Fishelov, David, *Dialogue with/and Great Books. The Dynamics of Canon Formation*, Brighton, Sussex Academic Press, 2010.
- Fitz-Enz, Daniel, *MLA International Bibliography Database Guide*, CSA Illumina, 2008, <www.csa.com> [26/06/2013].
- Garfield, Eugene, "Is Information-Retrieval in the Arts and Humanities Inherently Different from that in Science – Effect that ISIS-Citation-Index-for-the-Arts-and-Humanities is Expected to Have on Future Scholarship", *Library Quarterly*, 50, 1 (1980), pp. 40-57.
- Gilbert, G. Nigel, "Referencing as Persuasion", *Social Studies of Science* 7, 1 (1977), pp. 113-122.
- Guillory, John, "Canon", en Joseph Childers y Gary Hentzi (eds.), *Columbia Dictionary of Modern Literary and Cultural Criticism*, New York, Columbia University Press, 1995, pp. 233-249.
- Hammarfelt, Bjorn, "Interdisciplinarity and the Intellectual Base of Literature Studies: Citation Analysis of Highly Cited Monographs", *Scientometrics*, 86, 3 (2011), pp. 705-725.
- Han, Jiawei; Kamber, Micheline y Pei, Jian, *Data Mining. Concepts and Techniques*, Waltham, Morgan Kaufmann, 2012.
- Hérubel, Jean-Pierre V.M. y Goedeken, Edward A., "Metadisciplinarity, Belles Lettres, and André Malraux: A bibliometric exploration of knowledge formation", *Serials Librarian*, 37, 4 (2000), pp. 51-68.
- Hjorland, Birger y Albrechtsen, Hanne, "Toward a New Horizon in Information-Science – Domain-Analysis", *Journal of the American Society for Information Science*, 46, 6 (1995), pp. 400-425.

- Hjorland, Birger, "Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance", *Journal of the American Society for Information Science and Technology*, 52, 9 (2001), pp. 774-778.
- Insko, Jeffrey, "Generational Canons", *Pedagogy*, 3, 3 (2003), pp. 341-358.
- Kaplan, Nancy R., "The norms of citation behavior – Prolegomena to the footnote", *American Documentation* 16, 3 (1965), pp. 179-184.
- Kermode, Frank, *Pleasure and Change. The Aesthetics of Canon*, Oxford, Oxford University Press, 2004.
- Larivière, Vincent; Archambault, Éric; Gingras, Yves y Vignola-Gagné, Étienne, "The Place of Serials in Referencing Practices: Comparing Natural Sciences and Engineering with social Sciences and Humanities", *Journal of the American Society for Information Science and Technology*, 57, 8 (2006), pp. 997-1004.
- Lentricchia, Frank y McLaughlin, Thomas (eds.), *Critical Terms for Literary Studies*, Chicago, The University of Chicago Press, 1995.
- Leydesdorff, Loet, "Theories of Citation?", *Scientometrics*, 43, 1 (1998), pp. 5-25.
- Modern Language Association International Bibliography, <www.mla.org>.
- Moravcsik, Michael J. y Murugesan, Ponraj, "Some Results on Function and Quality of Citations", *Social Studies of Science* 5, 1 (1975), pp. 86-92.
- Nederhof, Anton J.; Zwan, Rolf A.; De Bruin, Renger E. y Dekker, P.J., "Assessing the Usefulness of Bibliometric Indicators for the Humanities and the Social and Behavioral-Sciences – A Comparative Study", *Scientometrics*, 15, 5-6 (1989), pp. 423-435.
- Oscá-Lluch, Julia y Haba, Julia, "Dissemination of Spanish social sciences and humanities journals", *Journal of Information Science*, 31, 3 (2005), pp. 230-237.
- Price, Derek De Solla, *Little Science, Big Science*, New York, Columbia University Press, 1963.
- Small, Henry, "Citations and consilience in science – Comments on Theories of citation?", *Scientometrics* 43, 1 (1998), pp. 143-148.
- Small, Henry, "Cited documents as concept symbols", *Social Studies of Science*, 8 (1978), pp. 327-340.
- Van Dijk, Nel, "Research into canon formation: nationalism, literature, and an institutional point of view", *Poetics Today*, 20, 1 (1999), pp. 121-132.
- Von Halberg, Robert, "Editor's introduction", *Critical Inquiry*, 10, 1 (1983), pp. iii-vi.



RESUMEN

En este artículo, propongo un método alternativo al tradicional para la constitución del canon literario. En vez de basar la determinación del canon en diferentes valores y criterios, extraigo de la base de datos *Modern Language Association International Bibliography* los autores y las obras con el mayor número de referencias. Específicamente, mi objeto de estudio es la literatura hispanoamericana. A través de la minería de datos, obtengo una muestra de más de 75.000 referencias que nos permite observar la bibliografía crítica de las diecinueve literaturas nacionales de Hispanoamérica. Este método cuantitativo arroja un corpus de 451 obras y 717 autores que son citados más de 100 veces en la base de datos analizada. Consecuentemente, esta bibliografía no es el resultado de criterios subjetivos de selección, sino que se basa en la ley de los grandes números. Asimismo, este estudio es una demostración de la pertinencia de utilizar métodos cuantitativos para analizar las bases de datos bibliográficas digitales con el propósito de profundizar nuestros conocimientos del campo literario.

Palabras clave: Canon literario, literatura hispanoamericana, minería de datos, bibliometría, métodos cuantitativos.

ABSTRACT

In this article, I propose an alternative technique to the traditional method of constitution of the literary canon. Instead of basing the determination of the canon on different values and criteria, I scrutinize the *Modern Language Association International Bibliography* database in order to determine the most cited authors and literary works. Specifically, I study Spanish American literature. Thus, through the process of data mining, I obtain a sample of over 75,000 references that allows us to observe the critical bibliography about the nineteen national literatures of the subcontinent. This quantitative technique yields a corpus of 451 titles and 717 writers that are cited more than 100 times in the database. Consequently, this bibliography is not the result of subjective selection criteria, but is based on the law of large numbers. Furthermore, this study shows that the quantitative analysis of bibliographic digital databases is an effective way to bring new light to the field of literary studies.

Keywords: Literary canon, Spanish American literature, data mining, bibliometrics, big data.