
APLICACIONES LEXICOMÉTRICAS AL ANÁLISIS DISCURSIVO Y TEXTUAL EN CIENCIAS HUMANAS Y SOCIALES

LEXICOMETRIC APPLICATIONS TO DISCURSIVE AND TEXTUAL ANALYSIS IN HUMAN AND SOCIAL SCIENCES

Juan Marcelo Columba-Fernández¹

Cómo citar: Columba-Fernández, J. M. (2019). Aplicaciones lexicométricas al análisis discursivo y textual en ciencias humanas y sociales. *Página y Signos: Revista de Lingüística y Literatura*, 15, 75-88. <https://doi.org/10.5281/zenodo.8373333>

RESUMEN

Las ciencias humanas y sociales en el siglo XXI enfrentan una serie de transformaciones teóricas y metodológicas que exhiben, a la vez, nuevas y enriquecedoras perspectivas analíticas en torno a sus objetos de estudio. El análisis cuantitativo del léxico, conocido también bajo la denominación de lexicometría, forma parte de las posibilidades metodológicas orientadas al análisis de la materia prima de la actividad humana y social. Esta manera de aproximarse al sentido –en la que interactúan conocimientos lingüísticos, estadísticos e informáticos– hace posible la puesta en marcha de una hermenéutica interdisciplinaria y una lectura deslinearizada que enriquece y objetiviza la interpretación en torno al quehacer del hombre y de la sociedad. El presente trabajo presenta una perspectiva panorámica, que gravita en torno a algunas pautas metodológicas de estudio cuantitativo y empírico del vocabulario contenido en textos y discursos producidos en el contexto local.

Palabras clave: Discurso, ciencias sociales, humanidades, lexicometría, texto

¹ Máster en Ciencias del Lenguaje por la Universidad de Franche-Comté y Licenciado en Lingüística Castellana por la Universidad Mayor de San Andrés. Correo electrónico: suxta@hotmail.com.

ABSTRACT

Humanities and social sciences in the 21st century are facing several theoretical and methodological transformations that manifest, in parallel, new and enriching analytical perspectives on their objects of study. The quantitative analysis of the lexicon, also known as lexicometry, constitutes a methodological possibility to analyze the essential substance of human and social activity. This way of approaching the meaning –in which linguistic, statistical and computer sciences interact– enables an interdisciplinary hermeneutic and a non-linearised type of reading that enriches and objectifies the interpretation of human endeavor and societal dynamics. The article presents a panoramic perspective and some methodological guidelines of quantitative and empirical study of the vocabulary contained in texts and discourses in the local context.

Keywords: Discourse, humanities, lexicometry, social sciences, text

1. Introducción

Desde una perspectiva histórica, el sintagma “*Discours Analysis*” (Análisis del Discurso) fue introducido por el lingüista Z. S. Harris en un artículo del mismo nombre publicado el año 1952 en el número 28 de la revista anglófona *Language*. En ese uso, el término “*discourse*” designaba una unidad lingüística constituida por un conjunto de oraciones y, desde una perspectiva estructuralista, se concebía el “*analysis*” según la etimología del vocablo como la descomposición de un todo en sus partes. El año 1969, en Francia, el número 13 de la revista “*Langages*” consagra un número especial al entonces novedoso dominio del “*Analyse du discours*”. Dicha irrupción es contemporánea a la publicación del libro “*Analyse automatique du discours*” (Análisis automático del discurso) de M. Pêcheux y a la circulación de la influyente “*Archeologie du savoir*” (Arqueología del saber) de M. Foucault quien también sitúa la noción de discurso al centro de su reflexión (Maingueneau, 2014, pp. 10, 13).

Entre estas tres vertientes, el Análisis Automático del Discurso (AAD) constituye, a nivel metodológico, un cuestionamiento a la lectura intuitiva y propone un trabajo de interpretación crítica apoyado en procedimientos analíticos informatizados. En este sentido, la constitución de un corpus cerrado que permita confrontar efectos de sentido con otras totalidades equivalentes resulta una tarea central del AD. En la misma línea de automatización en el tratamiento de datos textuales, hacia los años setenta, la denominación de “*lexicométrie*” (lexicometría) vino a designar una metodología de estudio del discurso que pretende exhaustividad, sistematicidad y computarización en la medición de unidades léxicas. En vista de la realización de comparaciones cuantitativas, la lexicometría efectúa operaciones como la segmentación y elección de unidades analizables en la cadena textual, la constitución de un corpus finito y su fraccionamiento interno, además de la comparación de constataciones cifradas de las unidades consideradas. La realización de dichas operaciones implica el respeto de principios como la invariabilidad de la unidad de conteo, las cantidades significativas y equilibradas de los usos de la unidad, ello sin obviar la posibilidad de comparación e interpretación de las constataciones efectuadas (Charaudeau y Maingueneau, 2002, pp. 38, 342).

Estas operaciones buscan una “*désubjectivation*” (de-subjetivación) de la lectura mediante las herramientas propias a los estudios de una lingüística cuantitativa que pretende proporcionar, mediante la sistematicidad y exhaustividad de los tratamientos automatizados, una garantía de objetividad y cientificidad alejada de propensiones político-ideológicas (Mazière, 2010, p. 19).

Esta “*statistique lexicale*” (estadística léxica), concebida como la aplicación de métodos estadísticos a la descripción del vocabulario (Dubois, 2007, p. 441), constituye una perspectiva que nace del encuentro entre disciplinas interesadas, por una parte, en el estudio del lenguaje y, por otra, en la cuantificación y tratamiento de la información. La utilización de las herramientas informáticas para la gestión de importantes conjuntos textuales proporciona un servicio de clasificación, descripción y comparación, en tareas relacionadas con encuestas socio-económicas, entrevistas, investigaciones literarias, archivos históricos u otro tipo de bases documentales (Lebart y Salem, 1994, p.7).

En este marco, el presente trabajo tiene por objeto aproximarse, desde una perspectiva al interior de las ciencias del lenguaje, a las posibilidades de aplicación de la metodología lexicométrica en conjuntos verbales propios al dominio de las ciencias humanas y sociales en el contexto local. Para llevar a cabo dicho objetivo, el presente texto se organiza en apartados que dan cuenta de exploraciones realizadas en diferentes *corpora* constituidos por textos de tipo político, filosófico, literario, histórico, legal e institucional. En cada sección planteada, se realiza una brevísima descripción del corpus, una constatación estadística de datos verbales y una interpretación de los mismos que, si bien adolece de un retorno al texto analizado, intenta adaptarse a las características espaciales de la literatura gris. En relación a cuestiones terminológicas esenciales, cabe mencionar que el presente trabajo utiliza el término “discurso” para referirse, fundamentalmente, a la materialización verbal de un fenómeno lingüístico supra-oracional –resultado de sus condiciones externas de producción– mientras que reservamos el uso del vocablo “texto” a la cristalización del discurso en formas escriturales y, por esa razón, analizables mediante la metodología propuesta.

2. Aplicaciones lexicométricas en el análisis de textos políticos y filosóficos

El discurso político ha venido siendo el objeto de estudio privilegiado de disciplinas como la retórica o el análisis argumentativo del discurso. Estudios precedentes en torno a las renuncias presidenciales del año 2005 en Bolivia (Columba, 2009, 2015), desde una perspectiva argumentativa, muestran los mecanismos lingüísticos y figurativos desplegados en torno a los participantes de la interacción comunicativa en el ámbito político. Una aproximación lexicométrica a las frecuencias de utilización de elementos pronominales, que designan al locutor y al interlocutor del discurso político (yo/nosotros/usted/ustedes) en un corpus constituido por más de ciento sesenta alocuciones presidenciales, producidas durante los siete trimestres de gobierno de C. Mesa, muestran una dinámica donde el uso enfático de las unidades consideradas varía significativamente, sobre todo, en momentos de alta tensión política.

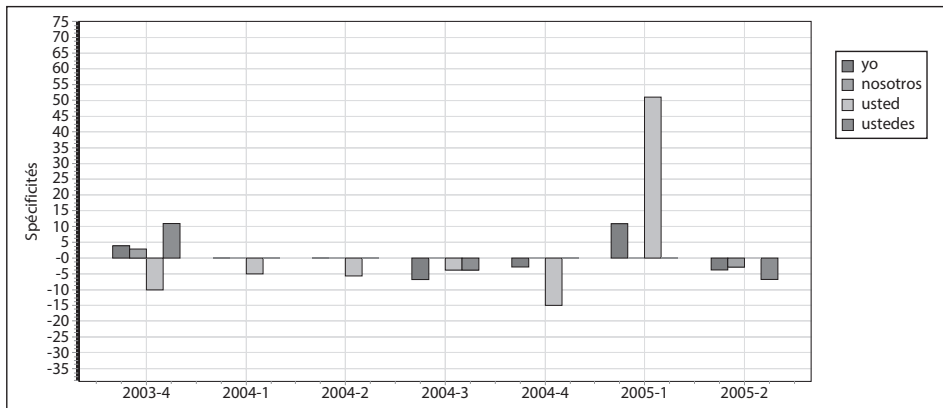


Gráfico 1. Especificidades pronominales en los discursos presidenciales por trimestres (2003-2005).

Fuente: Elaboración propia.

El gráfico 1 evidencia un uso específico significativo de los pronombres “yo”, “usted” y “ustedes” en dos momentos de crisis política en Bolivia (octubre 2003 y marzo de 2005). Así, durante el trimestre en el que se produjo la presentación de la renuncia del nuevo mandatario, es posible constatar el uso enfático del

pronombre “usted”, que alcanza proporciones significativas (183 apariciones) en relación a la totalidad de empleos de la misma forma (353 a.) en los discursos estudiados. Así, secuencias como “el Congreso Nacional que lo representa a **usted**, ciudadana y ciudadano de Bolivia definirá el tema, yo quiero expresarle a **usted**, que es ante **usted**, ante la consideración suya, que presento mi renuncia a la Presidencia de Bolivia” (el énfasis es nuestro) interpelan frontalmente al interlocutor y elaboran una imagen discursiva que apela a la importancia del rol reflexivo y decisivo del ciudadano ante la posibilidad del alejamiento del Jefe de Estado de sus funciones ejecutivas.

Por otra parte, las aplicaciones de la estadística léxica en el campo filosófico pueden hacerse presentes al comparar los sistemas de pensamiento de connotados intelectuales, a través del vocabulario que utilizan para expresar su razonamiento. Así, escritos pertenecientes a grandes pensadores bolivianos que desarrollan temas como la “raza”, tal el caso de A. Arguedas en *Pueblo enfermo* (1909), y F. Tamayo, en *La creación de la pedagogía nacional* (1910), se caracterizan a partir de la utilización de vocablos singulares al punto de vista de cada autor, cuando examinan dicho aspecto que es considerado como un elemento central en la concepción societal boliviana (Roca, 1979, p. 16).

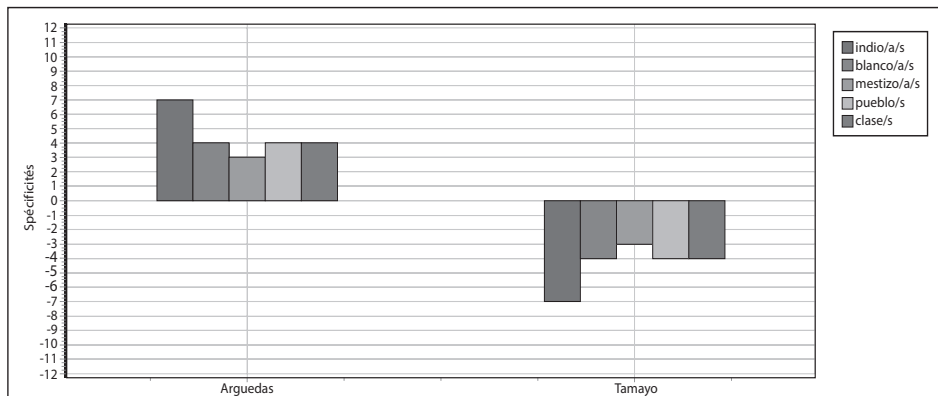


Gráfico 2. Vocabulario específico de Arguedas en torno al tema de la raza.

Fuente: Elaboración propia.

El gráfico 2 permite observar el vocabulario mediante el cual Arguedas trata el tema de la raza. Así es posible evidenciar, entre las principales palabras con sentido pleno, los vocablos “indio/a/s” (28 veces) e “indígena/s” (11 v.), que constituyen el sujeto principal sobre el cual discurre el autor en torno al tema racial. Asimismo, la frecuencia de las asociaciones léxicas de los vocablos señalados con la voz “pampa” (4 v.) deja entrever una concepción racial del ser, íntimamente vinculada a su medio geográfico en fragmentos como “La **pampa** y el **indio** no forman sino una sola entidad. No se comprende la **pampa** sin el **indio**, así como éste sentiría nostalgia en otra región que no fuera la **pampa**” (el énfasis es nuestro).

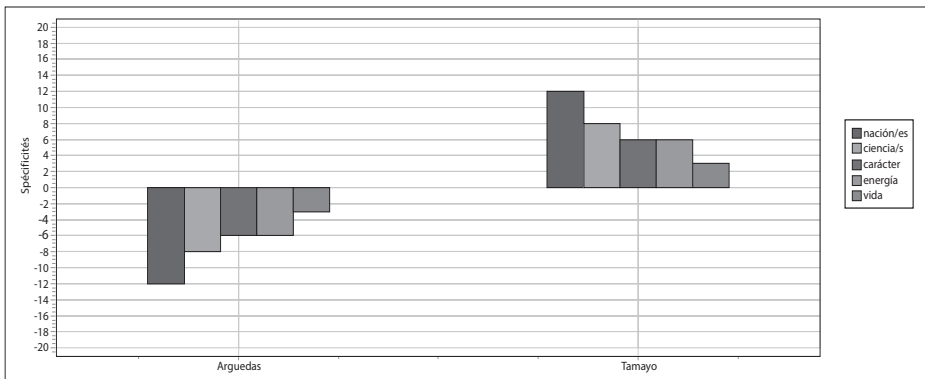


Gráfico 3. Vocabulario específico de Tamayo en torno al tema de la raza.

Fuente: Elaboración propia.

A diferencia de Arguedas, Tamayo no incluye en su reflexión dichos vocablos (indio/indígena), y prefiere referirse a un sujeto colectivo designado mediante el sustantivo “nación/es” (29 apariciones), tal como se observa en el gráfico 3. Entre los vocablos más frecuentemente asociados a este último, encontramos el sustantivo “historia” (6 veces), expresando una inscripción de la nación (biológicamente racializada en Tamayo) en una linealidad histórica. Dicho aspecto puede comprobarse en segmentos como “Desde que hay **nación**, esto es, desde que hay un grupo humano que permanece en la **historia** y genera en la naturaleza, dentro de un marco de condiciones especiales, propias y permanentes, entonces hay raza, y entonces, hay carácter nacional” (el énfasis es nuestro).

3. Aplicaciones lexicométricas en el análisis de textos literarios e históricos

La clasificación literaria clásica encuentra en la poesía la expresión por excelencia del género lírico, pues esta constituye la manifestación estética del sentimiento humano mediante la palabra. Destaca entre la poesía modernista latinoamericana el boliviano Ricardo Jaimes Freyre, cuya obra poética expresa una tensión epicúrea cuando enuncia sentimientos de alegría y tristeza (Coello 1971, p. 629). Una dicotomía que bien podría corresponder al tránsito entre un universo semántico “luminoso” y otro “oscuro” –en sus respectivas acepciones de alegre y triste– está presente en los textos del poeta.

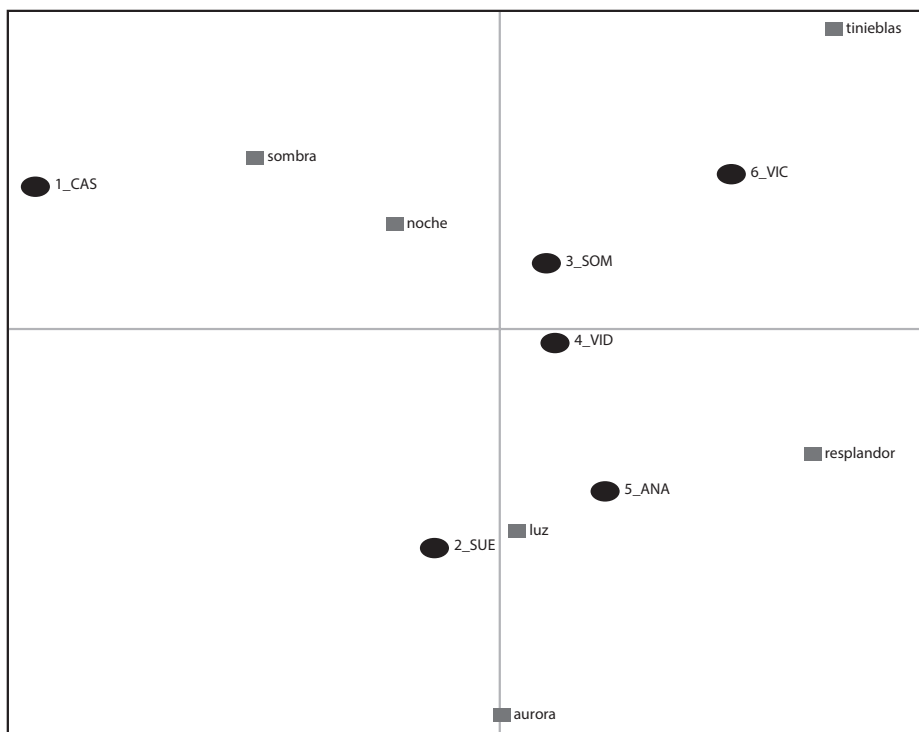


Gráfico 4. Distancias y oposiciones léxicas en la obra poética de R. Jaimes Freyre.

Fuente: Elaboración propia.

El estudio cuantitativo del vocabulario específico y las distancias léxicas presentes en los poemarios de Jaimes Freyre representan, en el gráfico 4, una oposición entre un campo semántico referente a la claridad conformado por vocablos como “luz” (31 apariciones), “aurora” (10 a.) y “resplandor” (5 a.), todos ellos más característicos de los poemarios *Los sueños son vida* (1917), *Anadiomena* (1917) y *País de Sueño* (1918). Asimismo, la semanticidad de otros conjuntos poéticos como *Castalia Bárbara* (1899), *Las Víctimas* (1917) y *País de Sombra* (1918) se encuentra orientada hacia un universo de sentido más lúgubre, ello se pone de manifiesto a través del uso frecuente de formas específicas como “noche” (49 a.), “sombra” (30 a.) y “tinieblas” (8 a.). Se evidencia así un distanciamiento entre dos bloques poéticos a partir de las similitudes y diferencias que los mismos presentan en el uso léxico.

En cuanto al estudio cuantitativo de textos históricos, este permite una objetivación interpretativa en relación al trabajo sobre documentos de archivo que registran la memoria de un conjunto social. Así, el análisis de coapariciones léxicas en documentos como el “Memorándum dirigido al H. Congreso y a la Nación sobre las ventajas del Ferrocarril Oriental” (1904), considerado un hito en la historia del desarrollo autónomo regional (Dory, 2009, p. 94), muestra, a través de una representación arbórea, las principales agrupaciones temáticas y las relaciones terminológicas establecidas en el texto.

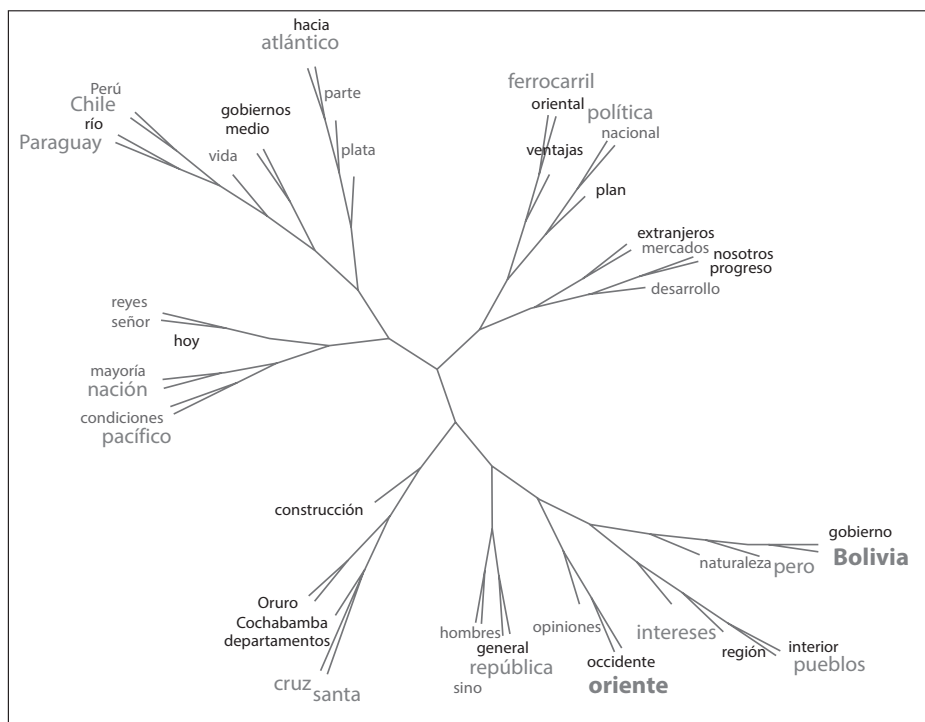


Gráfico 5. Asociaciones y frecuencias léxicas en el memorándum cruceño de 1940.

Fuente: Elaboración propia.

La representación arborescente presente en el gráfico 5, resultante del análisis estadístico de coocurrencias o apariciones simultáneas en segmentos del documento considerado, permite vislumbrar la vocación de integración nacional del “Oriente” (24 apariciones) en su relación con “Bolivia” (28 a.), formulada por los representantes cruceños a inicios del siglo XX. Asimismo, resulta posible observar la importancia otorgada al proyecto del “ferrocarril” (16 a.) vinculado al “progreso” (6 a.) y al “desarrollo” (5 a.), además de resaltar las posibilidades de acceso al océano “Atlántico” (9 a.) por el río “Paraguay” (10 a.) ante la pérdida territorial sufrida como consecuencia de la Guerra del Pacífico de 1879.

4. Aplicaciones lexicométricas en el análisis de textos de ley e institucionales

El estudio de las normas y los textos de ley constituye un ámbito fértil para la aplicación de la estadística léxica pues, además de proporcionar un corpus estabilizado y sistematizado en el ámbito jurídico, expresa los cambios sociales y las políticas públicas implementadas a lo largo de la historia del país. Dichos cambios, en el ámbito de la educación pública, no han sido pocos y han manifestado, prácticamente desde la fundación de la República, una permanente inquietud estatal sobre la planificación relacionada a la problemática educativa. Un análisis longitudinal de los usos verbales presentes en la normativa y las reformas efectuadas en el ámbito educativo boliviano hace posible apreciar las “herencias” que recibe la legislación educativa vigente (Talavera, 2011, p. 263).

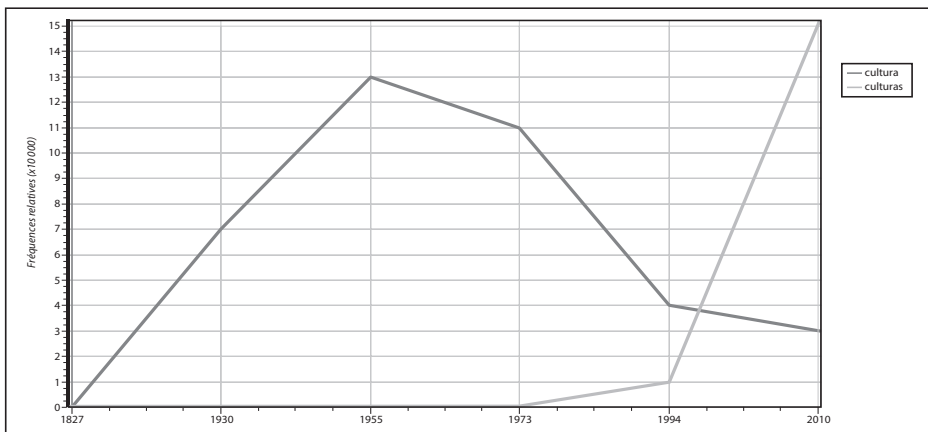


Gráfico 6. Evolución de frecuencias de “cultura” y “culturas” en la legislación educativa boliviana.

Fuente: Elaboración propia.

La temática de la diversidad cultural y la interculturalidad en la normativa educativa boliviana, como permite observar el gráfico 6, se inscribe en transformaciones léxicas del vocablo “cultura” (de 25 apariciones en 1955 a 4 en 2010) y su plural “culturas” (de una primera aparición en 1994 a 18 en 2010).

Ello hace evidente una singular ruptura que comienza en los años posteriores a la aprobación del Código de Educación de 1955 y se consolida en la Reforma Educativa de 1994. Tal quiebre manifiesta la irrupción de la diversidad cultural en la concepción de las políticas referentes a la educación pública en Bolivia.

En relación al estudio estadístico de documentos institucionales, el análisis de las memorias de los congresos universitarios bolivianos revela las tendencias y las políticas asumidas en relación a la educación superior en el país. Como órgano nacional de gobierno, el Congreso Nacional de Universidades constituye la máxima autoridad del sistema que, reuniéndose cada cuatro años, define las políticas y estrategias, aprueba los planes de desarrollo y fija las políticas económicas y financieras universitarias, entre otras funciones de suma importancia atribuidas a esta instancia de decisión (CEUB, 2009, p.150).

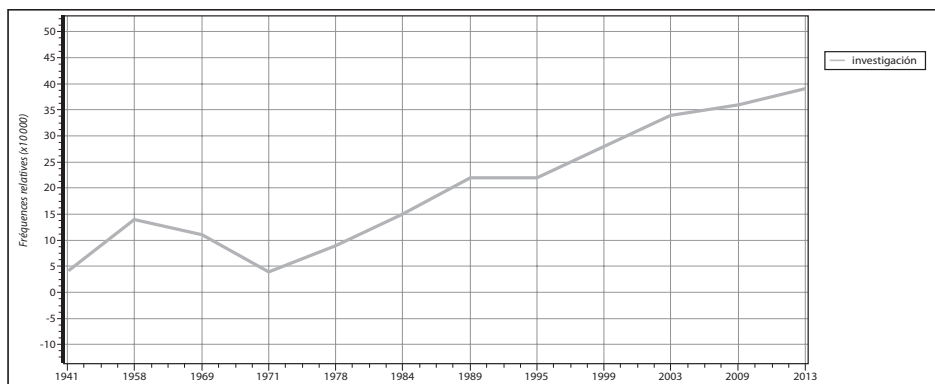


Gráfico 7. Evolución de usos de “investigación/es” en los documentos de congresos universitarios bolivianos.

Fuente: Elaboración propia.

El gráfico 7 muestra la evolución del uso del término “investigación/es” que, en el más reciente congreso universitario, alcanza un total de 706 actualizaciones; tal empleo del término prácticamente duplica su uso en el documento precedente (382 a. en 2009). Esta tendencia, cuya progresión no cesa desde inicios de

los años setenta, refleja la política investigativa que, junto a otras temáticas como la “acreditación” (404 a.) y la educación “virtual” (222 a.), forman parte de las más recientes preocupaciones universitarias. Este tipo de constataciones demuestra su importancia al momento de constituirse en insumos decisionales para la planificación estratégica de la actividad universitaria en el país, más aún ante la natural evolución institucional y la consecuente necesidad de una reforma institucional de las casas de estudios superiores.

5. A manera de cierre

El presente trabajo pretende contribuir, mediante una somera exploración realizada en torno a conjuntos documentales propios al ámbito de las ciencias humanas y sociales, a la presentación de pistas de investigación y posibilidades metodológicas para el estudio estadístico y automatizado de diferentes tipos de material lingüístico.

El énfasis hecho en una perspectiva metodológica que gravita en torno al estudio cuantitativo del léxico pretende mostrar una forma de complementación, o incluso de optimización, de las aproximaciones cualitativas características de un ámbito de estudio donde el componente verbal y escritural constituye la materia prima del trabajo de campo y de archivo.

Nuestra contribución se enmarca en una reflexión en torno a lo que ha venido a denominarse “*digital humanities*” o humanidades numéricas, una noción que considera inherentemente el tratamiento automatizado de la información documental, ello ante la colosal irrupción del “*big data*” y sus posibilidades de análisis como corpus.

En este contexto, la perspectiva interdisciplinaria resulta invaluable y fructífera para la comunidad local interesada en el estudio del hombre y de la sociedad. Se proyecta, así, una mirada renovadora e integradora que plantea nuevos desafíos intelectuales y académicos a las humanidades y las ciencias sociales en el siglo XXI.

Referencias

Charaudeau, P. y Maingueneau, D. (2002). *Dictionnaire d'analyse du discours*. Paris: Seuil.

Coello, C. (1971). Subliminar: síntesis de una concepción de la vida. En *Thesavrus. Boletín del Instituto Caro y Cuervo* (1), Tomo XXVI. Consultado el 28 de julio de 2016 en http://cvc.cervantes.es/lengua/thesaurus/pdf/26/TH_26_003_146_0.pdf

Columba, M. (2009). *Palabras del presidente. Análisis argumentativo de los discursos del 6, 8 y 10 de marzo de 2005*. La Paz: Instituto de Estudios Bolivianos.

Columba, M. (2015). L'argumentation dans le discours politique bolivien. Schèmes argumentatifs et dynamique figurative lors des démissions présidentielles de Carlos Mesa. Tesis de Maestría. Université de Franche Comté.

Comité Ejecutivo de la Universidad Boliviana-CEUB (2009). Estatuto Orgánico de la Universidad Boliviana. Consultado el 28 de julio de 2016 en http://www.ceub.edu.bo/index.php?option=com_content&view=article&id=80&Itemid=88

Dory, D. (2009). *Las raíces históricas de la autonomía cruceña. Una interpretación política*. Santa Cruz de la Sierra: El País.

Dubois, J. et al. (2007). *Grand dictionnaire de linguistique et sciences du langage*. Paris: Larousse.

Lebart, L. y Salem, A. (1994). *Statistique textuelle*. Paris: Dunod.

Maingueneau, D. (2014). *Discours et analyse du discours*. Paris: Armand Colin.

Mazière, F. (2010). *L'analyse du discours. Histoire et pratiques*. Paris: Presses Universitaires de France.

Roca, J. (1979). Bolivia en Arguedas y Tamayo. En *Latinoamérica* (79). México D.F.: UNAM.

Talavera, M. (2011). Herencias que recibe la ley Avelino Siñani – Elizardo Pérez. En *Temas Sociales* (31). La Paz: Carrera de Sociología – UMSA.