



Arte rupestre peruano: análisis textométrico

RUBÉN URBIZAGÁSTEGUI-ALVARADO

Resumen: El objetivo de este artículo es aplicar las estadísticas textométricas a la literatura publicada sobre arte rupestre peruano, en un esfuerzo por descubrir la representación social del discurso académico explicitada en la literatura publicada sobre este asunto hasta diciembre de 2020. Se utilizó el software Iramuteq como herramienta para el análisis estadístico del corpus textual. Todas las contribuciones incluidas y organizadas en un solo corpus, fue sometido a 6 tipos de análisis: análisis de léxico clásico, clasificación jerárquica descendiente (CHD), análisis de especificidades de grupos, análisis factorial de correspondencias (AFC), análisis de similitudes y nubes de palabras. El análisis de léxico clásico identificó 216 segmentos de textos formados por 8620 ocurrencias de unidades lingüísticas y de las cuales 1857 son lemas, 1776 palabras de formas activas; de este total de palabras activas, 414 tuvieron una frecuencia de 3 o mayor a tres usos en los textos; 81 palabras fueron formas suplementarias y 1043 palabras tuvieron una única aparición (56.17% de las formas) conocidas como hápax. El análisis de correspondencias identificó que los dos ejes principales de la coordenada cartesiana explican el 71.65% de la inercia en la distribución de los términos en clases. El análisis de similitud identificó cuatro clases léxicas que se muestran en un dendrograma. Finalmente se creó una nube de palabras con aquellos términos más utilizados resaltados en una figura.

Palabras Clave: Textometría, Bibliometría, Ciencometría; Arte rupestre; Quilcas; Perú.

Abstract: The objective of this paper is to apply textometric statistics to the published literature on Peruvian rock art, in an effort to discover the social representation of academic discourse made explicit in the published literature on this subject until December 2020. The Iramuteq software was used as a tool for the statistical analysis of the textual corpus. All contributions included and organized in a single corpus, were subjected to 6 types of analysis: classical lexicon analysis, descending hierarchical classification (CHD), group specificity analysis, factorial correspondence analysis (CFA), similarity analysis and cloud of words. The classical lexicon analysis identified 216 text segments made up of 8620 occurrences of linguistic units, from which 1857 are lemmas, 1776 words of active forms; of this total of active words, 414 had a frequency of 3 or greater than three uses in the texts; 81 words were supplementary forms and 1043 words had a single occurrence (56.17% of the forms) known as hapax. The correspondence analysis identified that the two main axes of the Cartesian coordinate explain 71.65% of the inertia in the distribution of the terms in classes. Similarity analysis identified four lexical classes that are displayed on a dendrogram. Finally, a word cloud was created with the most used terms highlighted in the figure.

Keywords: Textometry; Bibliometrics, Scietometrics, Rock art; Quilcas; Peru.

Introducción

Se afirma que cuando Colón sugirió que la tierra era redonda, lo ridiculizaron porque se sabía que la tierra era plana y cuando Galileo sostenía que los planetas giraban alrededor del sol, estuvo a punto de ser condenado a muerte por hereje. En ambos casos, la realidad social no estaba representada adecuadamente por las concepciones científicas de la época y la sociedad se resistía a realizar cambios en sus representaciones sociales de lo que hasta esas épocas se conocía. Los conocimientos sobre estos asuntos estaban influenciados por las ideas mitológicas y religiosas preexistentes en esas sociedades. Una disciplina que se interesa por este tipo de relaciones, por el conocimiento que emerge del sentido común como la representación de un conjunto de ideas de la vida cotidiana, que se construyen a través de las relaciones entre los individuos y los grupos sociales es la psicología social. Este entendimiento se basa en la teoría de la representación social (TRS) que se desarrolla a partir de los trabajos de Serge Moscovici en Francia durante los años sesenta y setenta. La representación social sería un conjunto de conceptos y explicaciones que se originan en la vida diaria en el curso de las interrelaciones entre los individuos. Serían equivalentes a los mitos y sistemas de creencias de las sociedades tradicionales (Moscovici 1981: 181). Por lo tanto, el intercambio de informaciones serían los fenómenos sociales que permiten la identificación y

el trabajo de las representaciones sociales de manera concreta, pues las representaciones se elaboran a partir del sentido común y el conocimiento popular presentes en las sociedades (Moscovici 1984: 61). Estas se forman cuando las personas se reúnen para intercambiar ideas o cuando se exponen a los medios de comunicación; es decir, las representaciones sociales son fenómenos que tiene sus raíces en la forma como el ser humano construye su identidad. Son fenómenos específicos que están relacionados con un modo particular de comprensión y comunicación, un modo que crea tanto la realidad como el sentido común. Las representaciones sociales son formas específicas de comunicar y comprender lo que ya sabemos y “ocupan, en efecto, una posición curiosa, en algún lugar entre los conceptos, que tienen como objetivo abstraer el significado del mundo, e introducirle orden, y las percepciones, que reproducen el mundo de una manera significativa” (Moscovici 1984: 17). Son sistemas cognitivos con una lógica y enunciado propio como los estereotipos, creencias, afirmaciones o valores que ayudan a las personas a orientarse y dominar su medio social.

De esa manera, las representaciones sociales (RS) proporcionan a los individuos un sistema común para clasificar y definir aspectos desconocidos de su mundo social derivados de sus historias individuales, grupales y generadas por la ciencia y la tecnología (Codol, 1974). Una hipótesis fundamental de la teoría de la representación social es que las representaciones cognitivas se



construyen, seleccionan y se impregnan en los grupos a través de la interacción social y la comunicación. Se constituyen a través de las relaciones que identifican al sujeto como un ser que acumula conocimientos y experiencias de lo vivido, creando representaciones de la realidad e interiorizando las costumbres de la sociedad, pasados a través de las generaciones y compartidos por los grupos sociales. Las relaciones entre el universo cognitivo y la realidad social se fortalecen en el juego dialéctico que envuelve la constitución de las representaciones sociales a través de los saberes y experiencias provenientes de diferentes orígenes. De hecho, Moscovici (1988) sugiere que las RS pueden definirse como patrones de creencias, valores y comportamientos compartidos colectivamente que evolucionan continuamente a lo largo del tiempo y el espacio dentro de los grupos y entre los grupos. Las RS funcionan para organizar la información y proporcionar un lenguaje mediante el cual los individuos pueden comunicarse en su mundo social y dominarlo y sería una consecuencia de la difusión y circulación del conocimiento perteneciente a grupos que, cuando se comparte con otros, dan como resultado la aparición de creencias, valores y normas que influyen en el pensamiento y el comportamiento. Las representaciones sociales a menudo se originan en conceptos relacionados con la ciencia y la tecnología. Jasper y Fraser (1984) sugieren además que las RS constituyen una realidad social en la que muchos individuos comparten una representación que a su vez influye en el comportamiento individual y que permiten conocimientos coherentes moldeados por un proceso de intercambio e interacción. Funcionan para permitir la comunicación y, por lo tanto, pueden considerarse como redes de conocimiento.

También en Francia y durante la década de los setentas, ya se desarrollaba investigaciones en estadística textual. Guiraud (1954, 1960) y Muller (1968, 1977) llamaban a estas investigaciones de textometría que aplicaban a los textos lingüísticos los métodos de análisis desarrollados por Benzécri (1973) quien desarrolló una serie de métodos, que denominaba de análisis de datos y que se puede considerar como el precursor de la minería de datos y el “big data”. El acento en los métodos de visualización es clave para comprender el éxito de la Escuela de Benzécri. “Lo que era un conjunto complejo de datos se organizó como un “espacio” para el beneficio del analista, y de repente la nube de datos se volvió accesible para la interpretación en su conjunto, con una estructura que podía explorarse, descubrirse, comentarse y mostrarse” (Beaudouin 2016: 59). Esas técnicas permitían la generación de mapas visuales de las palabras que se relacionan entre sí en un corpus textual. La textometría aplicaba así una amplia gama de cálculos lingüísticamente significativos y basados en las matemáticas para un análisis metódico de las colecciones de textos: asociaciones sintagmáticas y paradigmáticas, contrastes, caracterizaciones y evoluciones (Pincemin, 2008).

La textometría se desarrolló como una disciplina que admite una forma diferente de “leer” los textos ya que permite el estudio cuantitativo no lineal de los corpus textuales, combinando la investigación lexicométrica y estadística. Los resultados del análisis textométrico brindan una descripción general sintética, selectiva y sugerente del texto reorganizado, visto ahora a través de listas jerárquicas, mapas visuales, reagrupaciones y mejorías del texto. Una nueva forma de acceder a los

datos textuales versátiles y controlados y una nueva forma de “leer” el texto a partir de los datos que antes no estaban disponibles (Jaćimović, 2019). Estos métodos cuantitativos que son aplicados en la textometría también se utilizan en el campo de la minería de textos (MT). La combinación de las representaciones sociales y la textometría están en la base de los diversos softwares que han desarrollado una interfaz gráfica que permiten el análisis de datos textuales utilizando las funciones primarias de la textometría, como la visualización de concordancias, frecuencia de uso de las palabras o el análisis de correspondencias factoriales.

En la década de los ochenta, inspirado por el análisis de datos, Reinert (1979) diseña la metodología de ALCESTE (Analyse des Lexèmes Cooccurrents dans les Énoncés Simples d'un Texte). Las preocupaciones de Reinert tomaron una orientación particular que considera un corpus como una secuencia de declaraciones producidas por un sujeto-enunciador. Así, el texto se modela en una tabla que contiene declaraciones en filas, con la marca del sujeto-emisor, y las palabras o lexemas en columnas, que se refieren a objetos en el mundo (sin preconcepciones sobre la realidad de estos objetos). El objetivo es entonces sacar a relucir “mundos léxicos” (Beaudouin 2016: 18). Sin embargo, estas herramientas de estadística textual se desarrollaron con lenguajes de programación que se fueron convirtiendo en obsoletos y actualizarlos requería que los códigos se escriban de nuevo. Por ejemplo, el software Alceste de Reinert (1979) fue totalmente reprogramado por Ratinaud (2009) y renombrado como “Iramuteq”, con una interfaz más moderna y con la capacidad de procesar mayores volúmenes de datos. Este software Iramuteq (2020) últimamente está siendo utilizado en una serie de estudios calificados de “cualitativos” y “exploratorios” y será usado también en esta investigación sobre arte rupestre peruano.

¿Porque estudiar a los profesionales de arte rupestre en el Perú? Motivado por el deseo de capturar las prácticas sociales tangibles y las interacciones que constituyen gran parte de las representaciones sociales del arte rupestre peruano, este artículo se centra en la comprensión basada en la práctica de los profesionales del arte rupestre peruano. El campo académico que lidia con el arte rupestre se construye socialmente en la sala de clases, las conferencias llevadas a cabo sobre este asunto y en las publicaciones difundidas por los especialistas. Un estudio de las representaciones sociales del arte rupestre que tienen estos profesionales, de los que forman parte estas prácticas cotidianas, se relaciona directamente con las vivencias reales de las personas cuyo contacto con los objetos rupestres los define socialmente como especialistas en arte rupestre. Los profesionales del arte rupestre peruano (ARP) desempeñan un papel excepcional en la construcción cambiante de los sistemas de creencias de las comunidades de prácticas del arte rupestre. Ocupan una posición distintiva en la interacción entre el conocimiento experto y el conocimiento del novicio, sistemas de creencias cuyas suposiciones y agencias básicas pueden no corresponder necesariamente. Es a través de la práctica diaria de los profesionales que la investigación y las publicaciones que difunden se implementan, se integran en la realidad social y en los sistemas de creencias de la sociedad. Este artículo parte del supuesto de que las formas en que se organiza la información, la documentación y la forma en que se comunican a los interesados reflejan y constituyen la forma



en que estos problemas son representados socialmente. Por lo tanto, al estudiar lo que los profesionales del ARP creen y practican en relación con el “arte rupestre”, y luego publican, se puede obtener conocimientos sobre las formas de representación social concretizadas en el lenguaje usado en las publicaciones.

Actualmente no existe una metodología textométrica para estudiar el significado discursivo ni el análisis de la producción de textos académicos en arte rupestre; por esa razón el objetivo de este artículo es aplicar las estadísticas textométricas a la literatura publicada sobre arte rupestre peruano, en un esfuerzo por descubrir la representación social del discurso académico explicitada en la literatura publicada sobre este asunto en este campo. Se busca identificar patrones de coincidencias temáticas en la literatura académica publicada en este campo, respondiendo preguntas como ¿cuáles son las palabras y categorías gramaticales predominantes en esta literatura publicada? ¿es posible verificar una proximidad léxica en los documentos que conforman el corpus textual? ¿existe asociación léxica en los documentos que conforman el corpus de análisis?

Para lograr el objetivo propuesto, este artículo está organizado en seis partes. En la primera se presenta una introducción al tema, se detalla el problema y se formulan las preguntas de investigación. En la segunda se revisa la literatura publicada sobre este asunto. En la tercera se describe la metodología empleada con énfasis en la forma de recolección de los datos y la forma de medición de los datos. En la cuarta parte se exponen los resultados. En la quinta se proponen las conclusiones y discusiones de los resultados obtenidos. Por último, se presenta la bibliografía revisada para la elaboración de este documento.

Marco teórico y revisión de la literatura

Todavía no existen estudios en la arqueología, sociología o ciencias sociales en general que hayan estudiado la literatura publicada en esos asuntos usando el software Iramuteq. En el campo de la enfermería, Acauan, et. al (2020) evaluaron el uso del software Iramuteq como herramienta de apoyo al análisis de datos textuales y verificaron que este software es una buena herramienta auxiliar para la comprensión de lo que llaman investigación cualitativa ya que este software “permite la integración de métodos estadísticos con análisis cualitativo subjetivo con la ayuda de representaciones gráficas y relativamente simple de entender. Además, permite el uso de bases de datos de diferentes tamaños, fortaleciendo la importancia del investigador en todas las fases de la investigación, ya que requiere de conocimientos teóricos y experiencia significativos para el armado del corpus textual que será procesado” (Acauan, et. al, 2020: 4). Ya Arditi, et al (2020) aplicaron una encuesta transversal multicéntrica a pacientes que padecían de cáncer invitándolos a calificar su atención en los hospitales y a agregar comentarios de texto libre sobre sus experiencias relacionadas con esa enfermedad. Realizaron un análisis textual utilizando el software Iramuteq sobre los comentarios proporcionados por los encuestados e identificaron cinco clases temáticas principales, dos de las cuales consisten en una descripción detallada de las “camino para la atención del cáncer”. Las tres clases restantes estaban relacionadas con la “atención médica”, la “gratitud y elogio” y la forma en que vivían los pacientes con cáncer (“el cáncer y yo”).

Un análisis más detallado de esta última clase mostró que los comentarios de los pacientes estaban relacionados con los siguientes temas: “shock inicial”, “soledad”, “comprensión y aceptación”, “repercusiones del cáncer” e “información y comunicación”. Hoffmann, et. al (2020) buscan acercarse a un enfoque de las Humanidades Digitales, utilizando como datos la producción de conocimiento en la Historia de la educación matemática (HEM) en Brasil, en particular la producción de tesis y disertaciones en los años 2012 y 2017, utilizando como herramienta para el análisis de los datos, el software Iramuteq. Para el análisis elaboraron dos corpus textuales, uno referente a los resúmenes del año 2012 y otro al año 2017. Lamentablemente los resultados se limitan a ofrecer el número de textos analizados, el número de ocurrencias de las palabras, número de formas y número de hapax encontrados.

La transición de la denominada Bibliotecología basada en evidencias (EBL) a la Bibliotecología basada en narrativas (NBL), en línea con el “giro narrativo” que ocurre en las ciencias sociales y las humanidades, también se está produciendo en Italia. Partiendo del reconocimiento de que las encuestas del campo que utilizan métodos narrativos están aumentando en las bibliotecas públicas italianas, el artículo de Faggiolani e Galluzzi (2019) presenta la metodología y los primeros resultados de una meta-investigación construida sobre un corpus que abarca los datos textuales producidos mediante técnicas de investigación cualitativa. Aplicando un software para el análisis automático de textos (Iramuteq) y utilizando un enfoque tanto inductivo como deductivo, los autores analizan el léxico detrás de las narrativas de los usuarios, no usuarios, líderes de opinión y miembros del personal. El resultado es un mapa de preguntas y respuestas que saca a la luz patrones de investigación emergentes e impulsa consideraciones adicionales sobre el enfoque metodológico elegido. Ya Ferreira e Correa (2018) estudiaron el asunto “Biblioteca digital” en la literatura científica brasileña en Ciencias de la Información, analizando los términos presentes en los campos del título, resumen y palabras clave, de los trabajos publicados en artículos de revistas e indexados en la Base de datos de Periódicos em Ciência da Informação (BRAPCI) publicados en el Brasil, visualizando los términos utilizando el software Iramuteq. Según los autores, la propuesta metodológica implica el relevamiento y visualización de términos en forma de gráficos generados a partir del análisis de frecuencias (ley de Zipf) y co-ocurrencias de las palabras clave identificadas. Como resultado de la aplicación, el método permite la visualización de los temas que involucran el tema “Biblioteca digital” en la Ciencia de la Información brasileña en el período de 2001 a 2017.

Como se puede ver por la literatura revisada, todavía no existen investigaciones que hayan tomado a la literatura diseminada en el campo del arte rupestre, o a la arqueología o a algunos campos de las ciencias sociales como objeto de análisis utilizando el paquete estadístico Iramuteq. Entonces, es pues propicio el momento para un análisis de este tipo en la literatura publicada sobre arte rupestre en el Perú.

Metodología

Este trabajo analiza la literatura producida sobre arte rupestre en el Perú, por lo tanto, como unidades de



análisis fueron tomados cada uno de los libros, capítulos de libros, artículos, trabajos presentados en congresos y otras formas de comunicación escritas publicadas sobre este asunto desde que se tuvieron las primeras noticias sobre su existencia hasta diciembre de 2020. Para recolectar los documentos publicados sobre este asunto se hicieron búsquedas con múltiples combinaciones booleanas usando las palabras claves listadas en el Anexo A, en las bases de datos bibliográficas listadas en el Anexo B. Por ejemplo, combinaciones booleanas con las palabras clave: Arte rupestre AND Perú; Geoglifos AND Perú; etc. En las búsquedas se utilizaron preferentemente los idiomas español e inglés. También se hicieron búsquedas en repositorios de acceso abierto, así como en los catálogos disponibles en internet para consulta online de las bibliotecas y sitios listadas en el Anexo C. Las referencias bibliográficas que se encontraron se exportaron a una base de datos bibliográfica diseñada en EndNote (Versión X8). Cada documento que se encontró se leyó y revisó cuidadosamente para verificar que tratara sobre el asunto de esta investigación. Se revisaron las referencias bibliográficas listadas en cada documento para identificar los trabajos referenciados en esos documentos pertinentes al asunto “arte rupestre en el Perú” pero que no se encontraron en las búsquedas realizadas en las bases de datos. Se normalizaron los nombres de los autores, las palabras clave, los títulos de las revistas, y los nombres de los congresos, mesas redondas y eventos similares, con el fin de eliminar las referencias bibliográficas duplicadas. Solo se retuvieron documentos únicos y no duplicados.

En el proceso de búsqueda se encontraron también bibliografías levantadas sobre el arte rupestre peruano y sudamericano. Esas bibliografías fueron cuidadosamente analizadas y las referencias pertinentes ingresadas en la base de datos creada en Endnote (Versión X8). Esos aportes proceden de las bibliográficas elaboradas por Monteverde Sotil (2008); Narváez Luna (s.d.); Guffroy (1999), Marymor (2020), Hostnig (2021), y otros.

Con esta revisión y selección minuciosa se produjo una base de datos bibliográfica especializada sobre arte rupestre en el Perú totalizando 1815 referencias bibliográficas únicas y no repetidas conteniendo libros, capítulos de libros, artículos de revistas, comunicaciones presentadas en congresos, y tesis producidos por diferentes autores hasta diciembre del 2020. De estos textos se seleccionaron solo 1 texto de cada año para con este texto hacer la prueba del análisis de textos con el paquete tm (Feinerer & Hornik, 2020) de R que es simplemente un entorno computacional que puede realizar cálculos numéricos, procesamiento de cadenas con caracteres y mucho más. Para seleccionar cada texto se usó la estrategia de muestra randómica aleatoria sin reemplazos, pero usando los algoritmos pertinentes de R (R Core Team, 2017). Este algoritmo enumera los documentos de 1 a n. Por ejemplo, si en un determinado año se publicaron 50 documentos, el algoritmo enumera los documentos de 1 a 50; luego de estos 50 documentos, se le solicita a R que genere 1 número aleatorio. Digamos que R genera como número aleatorio el 33, entonces el documento 33 es aleatoriamente seleccionado como muestra. De cada uno de estos documentos seleccionados aleatoriamente se copiaron los resúmenes y se llevaron a Notepad. Para guardar cada resumen se usó el apellido y en ciertos casos abreviando el apellido del autor seguido del año de publicación del documento. Cada documento se guardó en el formato xx.csv, por ejemplo: Azevedo-2020.

csv.

El uso de recursos informáticos y estadísticos en el análisis de textos ayuda a potencializar el alcance de las investigaciones. Aunque el software es solo una herramienta que auxilia en el análisis, la familiaridad del investigador con el asunto es esencial. El uso de softwares de análisis lexical permite el análisis de bases de datos complejas, proporcionando más detalles para explorar, describir y comparar los datos. Por esa razón se utiliza el software Iramuteq (Interface for R for Les Multidimensional Analyzes de Textes et Questionnaires, 2020) como herramienta para el análisis estadístico del corpus textual. Se eligió este software porque es una herramienta de acceso libre y gratuito que permite varios tipos de análisis estadísticos. Cuando se propone la utilización de softwares es con el objetivo de facilitar y volver más eficiente el proceso de tratamiento de los datos. Tratando los datos de un corpus textual de manera estadística, su potencialidad para el análisis de datos está ligado a que no requiere del investigador un conocimiento específico de la estadística ni siquiera un lenguaje de programación informática. La creación de representaciones gráficas mediante el tratamiento de los datos permite al investigador un análisis e interpretación más precisos atenuando la pérdida de información.

Todas las contribuciones fueron incluidas y organizadas en un solo corpus, el cual fue sometido a 5 tipos de análisis: análisis de léxico clásico, clasificación jerárquica descendiente (CHD), análisis de especificidades de grupos, análisis factorial de correspondencias (AFC), análisis de similitudes y finalmente se trazó una nube de palabras.

Resultados

El aprendizaje de las ciencias puede verse como el desarrollo de un repertorio de prácticas discursivas con las que se participa en la adquisición de conocimiento en las prácticas cotidianas de los grupos sociales. Para que este aprendizaje se afiance el lenguaje y la comunicación son elementos esenciales. Los grupos sociales pueden representar compañeros de clases, disciplinas científicas profesionales u organizaciones ciudadanas. A través de la participación y la comunicación, los participantes transforman estas comunidades y transforman el conocimiento. La importancia de la comunicación para el aprendizaje de las ciencias se puede entender de varias maneras. La enseñanza y el aprendizaje se construyen a través del lenguaje y los procesos sociales. A través de los procesos del discurso, las comunidades construyen colectivamente normas y expectativas, definen y legitiman el conocimiento, construyen afiliaciones y posicionan a los participantes a través de la interacción social. El acceso de los aprendices se logra mediante la participación de estos en los espacios sociales y simbólicos de las comunidades especializadas. Aprender a hacer ciencia consiste en convertirse en miembro de una comunidad discursiva especializada y, por lo tanto, comprender cómo emplear el conocimiento y como usarla en la práctica. Participar en un conjunto de prácticas discursivas implica no solo el uso del lenguaje, sino también un conjunto de valores, creencias, actitudes y formas de estar en el mundo. El significado se construye socialmente a través del lenguaje; los iniciantes aprenden a usar y comprender términos científicos. La ciencia es conocimiento social y el aprendizaje implica comprender los significados reconocidos por las comunidades epistémicas. Por



lo tanto, el conocimiento disciplinario se construye, enmarca, comunica y evalúa a través de las prácticas discursivas. En el campo del arte rupestre en el Perú, estas prácticas discursivas se manifiestan como “uno de los aspectos que caracteriza y define el mundo académico es la construcción de comunidades científicas cuyos miembros propician la producción y la circulación del conocimiento, debemos reconocer el papel fundamental de la formación escolar y del lenguaje en tal proceso. Es por ello que este universo de saberes se conforma y transforma a través de la ciencia y la alfabetización disciplinar y, de igual manera, por medio de la creación y el uso de estrategias discursivas particulares” (Londoño Zapata, 2015, p. 335) como aquellas difundidas a través de las publicaciones académicas.

La función del análisis textométrico es develar la constitución lingüística y discursiva de esta comunidad académica, utilizando el corpus textual compuesto de una muestra de 42 resúmenes de la literatura publicada sobre arte rupestre peruano por miembros de esta comunidad. En estos 42 resúmenes se encontraron 216 segmentos de texto (con una media de 39.9 segmentos por texto) que están formados por 8620 ocurrencias de unidades lingüísticas y de las cuales 1857 son lemas y 1776 palabras de formas activas (verbos, adjetivos, adverbios, sustantivos); de este total de palabras activas, 414 tuvieron una frecuencia de 3 o mayor a tres usos en los textos; 81 palabras fueron formas suplementarias (artículos, proposiciones, conjunciones, etc.) y del total de palabras 1043 tuvieron una única aparición (56.17% de las formas) conocidas como hápax. Las palabras con formas activas con el mayor número de frecuencias (mayor a 10) se presentan en la Tabla 1. Estas cifras son auto explicativas; sin embargo, se pueden ejecutar una serie de combinaciones para darles sentido; por ejemplo: Sitio rupestre; sitio con geoglifos; Sitio arqueológico; sitio con

petroglifos, y muchas otras.

Después de la reducción del vocabulario a sus raíces, el 85% de las unidades de contextos elementales (u.c.e.) fueron retenidas en el corpus (122), que se distribuyeron en dos subgrupos, separando la clase 4 de las demás clases. Luego, se subdividió este segundo subcorpus, originando la clase 2 en oposición a las clases 1 y 3. La función semántica expresada en cada división se aleja después de cada subdivisión al “descender” en el árbol de jerarquía de las palabras representadas en cada una de las clases identificadas en el corpus. Los textos (corpus) fueron segmentados y clasificados en cuatro clases temáticas que se proporcionan en un dendrograma que muestra la agrupación jerárquica de las clases, los detalles de las clases con las palabras típicas y los extractos léxicos. Para facilitar el análisis de similitud, el análisis léxico usa la clasificación jerárquica descendente (Reinert, 1983, 1990) que identificó cuatro clases léxicas que se muestran en el dendrograma de la Figura 1.

La clasificación jerárquica descendente (Método de Reinert) realiza subdivisiones del corpus hasta lograr clases específicas de contenido léxico que conforman las unidades de contexto elementales (u.c.e.). En el análisis se consideraron palabras con frecuencia igual o mayor a la media de 3 y con chi-cuadrado $\times 2$ mayor a 3.84 (g.l. = 1, $p \geq 0.05$). En efecto, las clases que componen el dendrograma están compuestas por las palabras más significativas o frecuentes y sus correspondientes asociaciones con cada clase (chi-cuadrado). En el dendrograma (Tabla 2) se presenta cada clase indicando el número de u.c.e., nombres, atributos y variables relacionados, además de las palabras que las componen indicando además los valores de chi-cuadrado correspondientes.

Este dendrograma muestra dos grandes registros léxicos que se subdividen en, por un lado, la clase 4, y por el otro lado, las clases 1, 3 y la clase 2 (Ver Figura 1). La

Tabla 1. Palabras con frecuencias mayores a 10

Palabras	Frecuencia	Palabras	Frecuencia	Palabras	Frecuencia
sitio	48	realizar	17	imágenes	11
rupestre	47	articular	16	palpa	11
geoglifo	34	evidenciar	16	peruano	11
arqueológico	33	linear	16	presente	11
arte	31	Nasca	16	camélidos	10
petroglifo	31	trabajo	16	carácter	10
valle	30	desarrollo	15	características	10
parte	29	medio	15	cerro	10
investigación	28	paisaje	15	complejo	10
cultural	26	grabar	14	conjunto	10
grande	26	primero	14	cuenca	10
rio	24	asociar	13	dato	10
encontrar	21	estilo	13	departamento	10
estudio	20	grupo	13	distrito	10
perú	19	provincia	13	espacio	10
ubicar	19	quebrar	13	hacer	10
alto	18	tiempo	13	proceso	10
andino	18	año	12	registro	10
periodo	18	lugar	12	relación	10
presentar	18	motivo	12	representación	10
roca	18	sociedad	12	resultado	10
dos	17	zona	12	proyecto	10

Elaborado por el autor

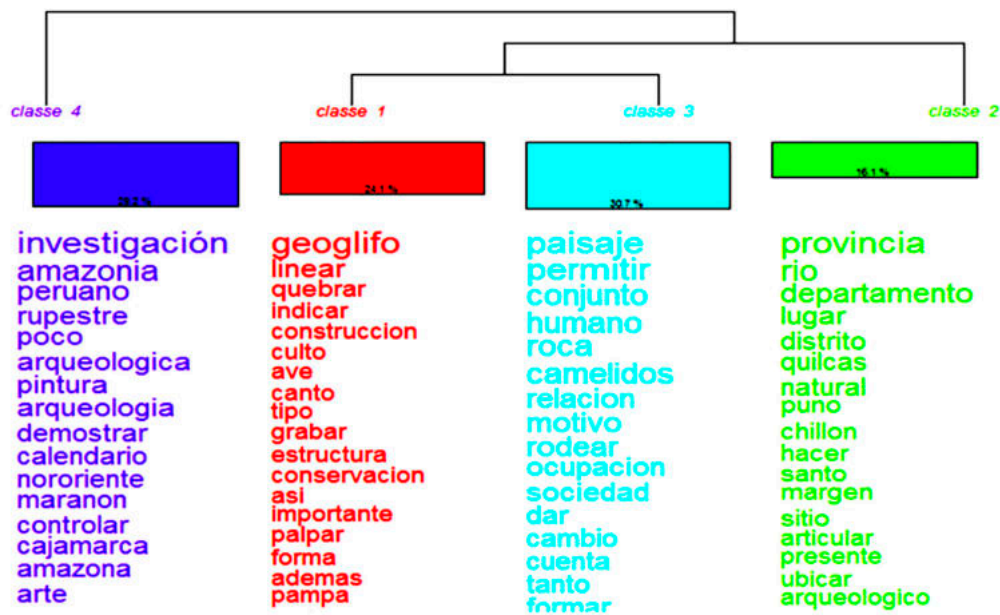


Figura 1. Dendrograma de las clases léxicas.

clasificación jerárquica descendente empleada permite identificar conjuntos de palabras cuya independencia se caracteriza por formas específicas correlacionadas entre sí. Construidas sobre la distribución de las principales palabras del léxico, las clases léxicas identificadas permiten describir los “mundos léxicos” o universos del discurso. Estos “mundos léxicos”, como los caracteriza Reinert (2007: 195-196), serían producto de los diferentes “ambientes mentales” invertidos por los profesionales del arte rupestre peruano. La primera categoría, que incluye solo a la clase 4, representa el 29.2% de los segmentos léxicos que se refieren a las investigaciones sobre arte rupestre peruano en la Amazonía y en el nororiente peruano, así como en las regiones aledañas a la Amazonía como, por ejemplo, el Departamento de Cajamarca. Ya la clase 2 (16.1%) se refiere a un léxico tomado de las exploraciones rupestres siguiendo los ríos en los departamentos de Puno y Lima respectivamente. Las categorías léxicas que reúnen las clases 1 (24.1% de los segmentos de textos) y la clase 3 (30.7% de los segmentos de textos), enfrenta a dos clases, una de las cuales postula el análisis de los geoglifos como indicación de la construcción de cultos, y la necesidad de grabar y conservar esas estructuras; mientras la otra (clase 3), invoca al paisaje como la posibilidad de permitir el análisis de un conjunto humano en la ocupación de esos espacios arqueológicos.

La Tabla 2 muestra la robustez de estas categorizaciones según la indicación del chi-cuadrado. Por cuestiones de espacio solo se muestran las 20 primeras palabras o categorías léxicas.

Obsérvese que Iramuteq no es muy eficiente al construir los segmentos textuales para análisis pues no reconoce algunas categorías gramaticales como la conjunción “así” y el adverbio “además”. También la lematización de la ciudad de Palpa es reconvertida al verbo “palpar”. Sin embargo, es bueno reconocer que a pesar de estas pequeñas deficiencias la segmentación opera adecuadamente.

Para explorar las cuatro clases léxicas del arte rupestre

peruano y representarla gráficamente, se utilizó el análisis de correspondencia múltiple (ACM) que toma como entrada las tablas que contienen variables categóricas nominales, estima las distancias entre las variables y luego transpone esas distancias a una gráfica bidimensional. Cada fila y cada columna de la tabla se representa, así como un punto en un espacio euclidiano. Si todos las clases léxicas tuvieran el mismo perfil, se agruparían en el centro del trazado, donde se cruzan los ejes horizontal y vertical. Este no es el caso, como lo demuestra la cantidad de palabras que se distribuyen por el trazado. Muchos de los términos se agrupan en la esquina superior derecha del trazado, mientras que otros términos se agrupan en la esquina inferior derecha (Ver Figura 2).

El análisis de correspondencia factorial, a partir de cálculos de la inercia de las palabras del corpus, resalta los paralelos y oposiciones entre las formas y donde se plasman las distintas clases obtenidas y su interdependencia, en un plano cartesiano. Luego, la pertenencia a una clase las superpone en una distribución determinada por los factores propios de la matriz de inercia informada en los dos ejes. Es decir, el análisis factorial de correspondencias léxicas asigna a cada uno de los “términos” del corpus una coordenada y en base a los valores obtenidos al calcular la distancia entre cada uno de los “términos” del corpus, los coloca en un espacio bidimensional. La nube de palabras así creada revela las repulsiones y atracciones de los conjuntos léxicos del corpus. En el espacio de representación, las áreas alrededor de las cuales giran las formas, así como las líneas normativas ascendentes de los ejes, se intenta identificar si la distribución de las formas léxicas están articuladas por un eje vertical que representa un continuo polarizado entre la problematización de los temas. El eje horizontal que permite discernir una división entre, por un lado, axiologías y repertorios argumentativos, y por otro, actores, dispositivos o eventos. Debido a que el análisis de correspondencias visualiza las relaciones entre los elementos que componen los datos como distancias en un gráfico, a menudo se pueden descubrir patrones basados



Tabla 2. Las cuatro clases léxicas del arte rupestre peruano

Clase 1 (24.15%) 33 u.c.e.	x ²	Clase 2 (16.1%) 22 u.c.e.	x ²	Clase 3 (30.7%) 42 u.c.e.	x ²	Clase 4 (29.2%) 40 u.c.e.	x ²
geoglifo	51.51	provincia	50.35	paisaje	14.73	investigación	26.16
linear	29.21	rio	36.3	permitir	14.19	amazonia	20.6
quebrar	12.44	departamento	32.72	conjunto	11.74	peruano	17.89
indicar	9.67	lugar	23.29	humano	11.74	rupestre	14.52
construcción	9.67	distrito	21.9	roca	10.52	poco	13.97
culto	9.67	quilcas	21.54	camélidos	10.52	arqueológica	13.45
ave	9.67	natural	21.07	relación	10.06	pintura	12.58
canto	9.67	puno	16.03	motivo	9.32	arqueología	12.58
tipo	9.04	chillón	16.03	rodear	9.32	demostrar	9.99
grabar	8.87	hacer	15.74	ocupación	8.19	calendario	9.99
estructura	8.87	santo	15.74	sociedad	7.85	nororient	9.99
conservación	8.87	margen	15.74	dar	6.94	marañón	9.99
así	8.87	sitio	13.06	cambio	6.94	controlar	9.99
importante	8.87	articular	9.24	cuenta	6.94	cajamarca	9.99
palpar	7.61	presente	9.22	tanto	6.94	amazona	9.99
forma	6.86	ubicar	9.08	formar	5.94	arte	9.72
además	6.22	arqueológico	7.67	existir	5.94	contexto	8.9
pampa	6.22	izquierda	7.43	través	5.77	constituir	8.9
serie	6.22	producción	5.83	cultural	4.74	andino	8.69
figura	5.84	norte	5.83	diferente	4.05	lograr	7.44

Elaborado por el autor

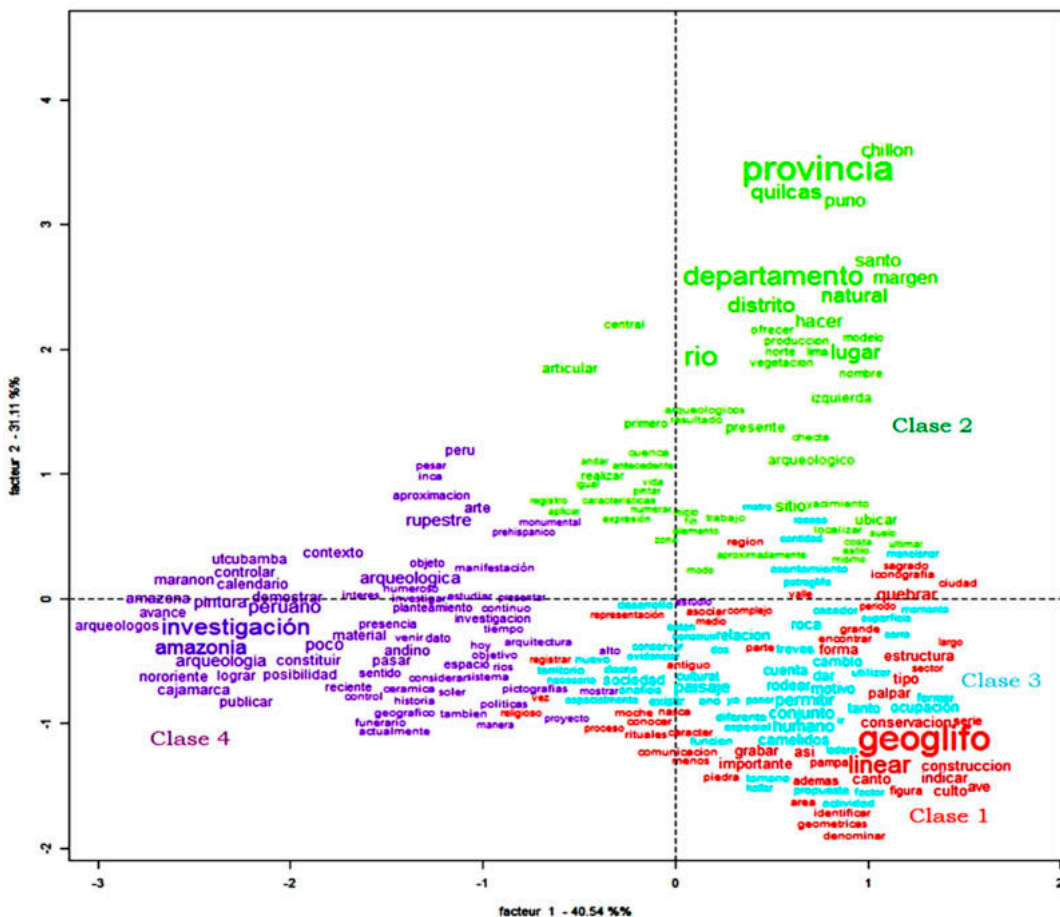


Figura 2. Distribución de los términos y las clases.



en qué elementos de una categoría aparecen cerca de los elementos de la otra. Por lo tanto, el análisis de correspondencias sirve para filtrar los patrones principales de un gran conjunto de datos. Por ejemplo, los dos ejes de esta coordenada explican el 71.65% de la inercia en la distribución de los términos en clases. Claramente la Clase 1 está mucho más cerca de la Clase 3, con la que comparten muchos términos cercanos. Por lo tanto, los asuntos estudiados por los autores que estarían presentes en estas dos clases parecen estar estudiando los mismos asuntos y las mismas regiones. Sin embargo, estas dos clases comparten también algunas preocupaciones con los asuntos estudiados por la clase 2, por eso se posicionan en el mismo eje vertical con valores positivos de la coordenada. Ya los asuntos de la clase 4 tienen sus propias características e intereses y por eso se posicionan próximo al eje central, pero con valores positivos y negativos (Ver Figura 2).

La Figura 3 muestra a los autores de los textos y las clases en las cuales fueron categorizados. Los autores de las clases 1 y 3 están localizados muy próximos en la coordenada cartesiana significando que probablemente están abordando los mismos asuntos y por eso están muy próximos unos de los otros. Estas dos clases también están próximas de los autores clasificados como pertenecientes a la clase 2. Sin embargo, en esta clase dos se nota una polarización en dos subgrupos de autores. Ya los autores de la clase 4 son más homogéneos y no están polarizados. En esta etapa del análisis, se busca comprender en qué

se diferencia cada clase en relación con el campo común de las representaciones sociales. El software Iramuteq proporciona un gráfico de las clases en forma de árbol resultante del análisis de semejanza y aproximación entre palabras. El punto de partida del análisis de similitud es una matriz construida sobre las formas plenas del corpus basada en la idea de la relación proporcional a la frecuencia, pero introduce las relaciones de coocurrencia entre las formas en los segmentos del texto y permite describir clases léxicas, perfiles de especificidad o incluso cuerpos completos (Marchand y Ratinaud, 2012).

El objetivo de este método es el de estudiar la proximidad y las relaciones entre las formas que integran el corpus, representándolo de una manera arbórea. Cuanto más grueso es el borde, más fuerte será el vínculo entre estas las palabras. El lugar que ocupan las palabras entre sí también es importante, al igual que los enlaces representados. El interés está centrado en restaurar la parte común, representando las relaciones entre las formas léxicas en un corpus no dividido por variables exógenas, pero solo en unidades contextuales (Marchand y Ratinaud, 2012). Este es el árbol más simple que se puede obtener, pero también el más pesado en términos de información que permite leer las relaciones lexicales más fuertes entre las formas y las más frecuentes de un corpus.

El análisis de similitudes muestra un grafo que representa las ligaciones entre las palabras del corpus textual a partir de los cuales se pueden inferir la estructura del corpus, así como identificar los temas

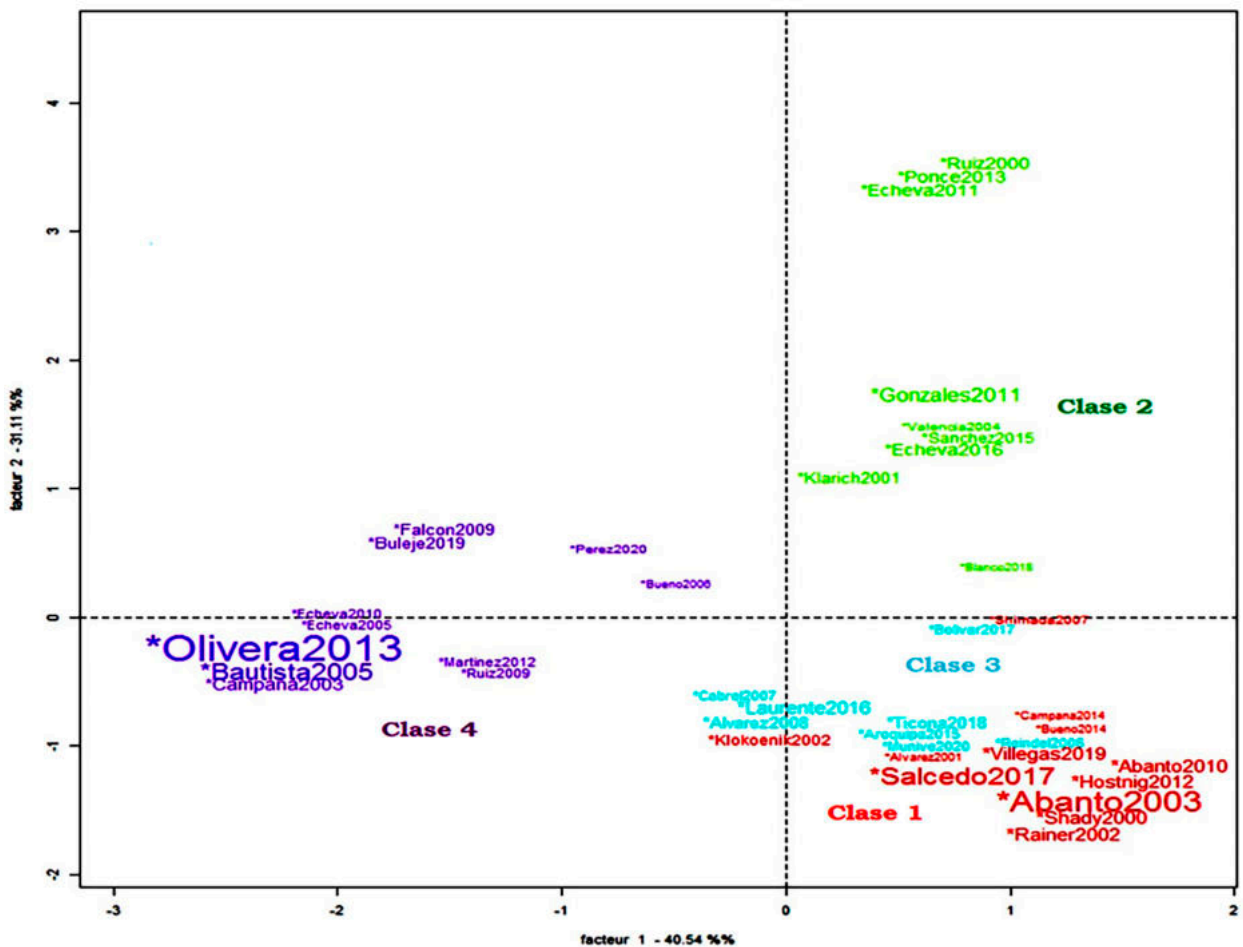


Figura 3. Distribución de los autores según las clases.

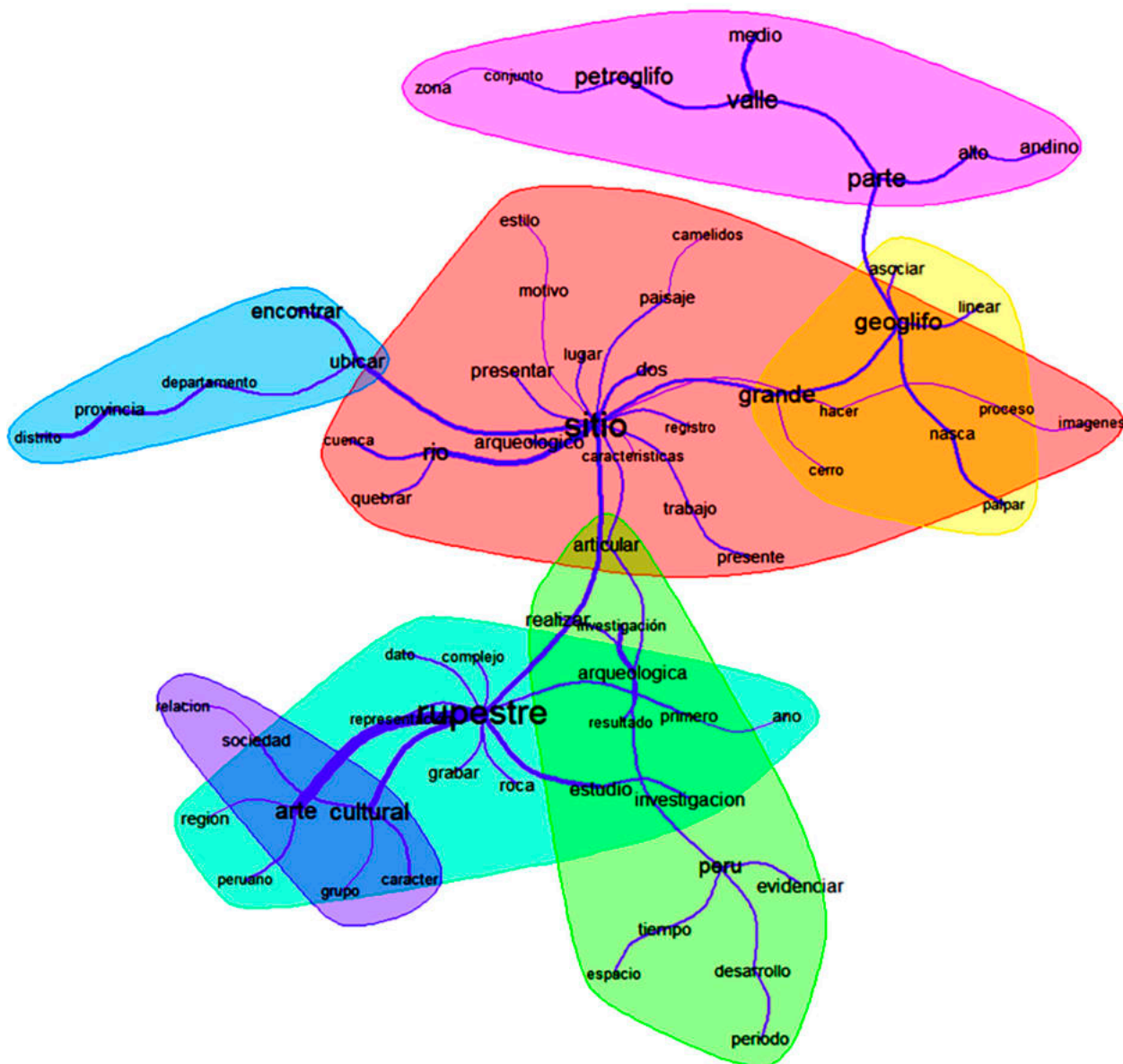


Figura 5. Grafo de similitud del corpus textual con agrupamientos.

el tiempo y espacio del periodo de desarrollo.

La Figura 6, muestra la nube de palabras formadas a partir del corpus. Su representación consiste en el cálculo de la frecuencia de ocurrencias de las palabras y donde aparecen las palabras con mayor representatividad con el tipo de letras en tamaño mayor y las menos representativas en un tamaño menor. Por ejemplo, “sitio” y “rupestre”, aparecen con mayor tamaño por que tuvieron mayor frecuencia de apareamiento en los textos analizados. Igualmente “geoglifo” y “petroglifo” aparecen en destaque por la misma razón, y así sucesivamente las palabras que componen el mundo léxico de las preocupaciones de los investigadores en arte rupestre en el Perú. Básicamente, estas son las mismas que ya se analizaron anteriormente.

Conclusión

Este artículo tuvo como objetivo verificar el conocimiento desarrollado y transmitido por los medios

de difusión académicos (libros, revistas, congresos, etc.), sobre arte rupestre en el Perú. Si bien la comunidad científica es considerada como un grupo social peculiar de producción de conocimiento, este universo es producido por especialistas, teniendo en cuenta el rigor y los métodos propios de la ciencia, que no pueden equipararse a la dinámica presente en el proceso de formación de representaciones sociales del sentido común. Identificar la cosmovisión y los conocimientos construidos y compartidos que orientan las prácticas socioculturales de los científicos sociales revela una nueva dinámica de interacción social que se superpone a las características objetivas de la información, dilucidando parte de los conceptos sobre los que se articulan las representaciones sociales y exponen los elementos consistentes que explican las estrategias de confrontamiento que utilizan los científicos que lidian con el arte rupestre en el Perú. “El objetivo del método propuesto es el de poner en evidencia, a través del análisis de un conjunto de textos, los ‘mundos lexicales usuales’ evocados por los enunciadores. La estadística



Figura 5. Nube de palabras.

encuentra todo su interés en ello: es la redundancia de las huellas lexicales [o sucesiones de palabras] lo que permite localizar los mundos más frecuentes. Un mundo lexical es entonces la huella estadística de un lugar en el vocabulario, lugar frecuentemente ‘habitado’ por el enunciador” (Reinert, 1997: 274).

Se encontró que ese mundo lexical se representa socialmente teniendo como centro de exploración a las palabras “sitio”, “rupestre”, “geoglifo”, “petroglifo”, “arte” y “arqueológico” que se agrupan en cuatro clases léxicas que dan estructura a los estudios del arte rupestre peruano. Sin embargo, la categoría lexical indígena de “quilca”, debería ser considerada para darle validez epistemológica al “estudio de las manifestaciones gráficas en roca o tierra, como son los pictogramas, petroglifos, geoglifos, o arte mobiliario con tradición rupestre” (Echevarría López, 2016: 11). Aunque esta categoría lexical todavía no es utilizada ampliamente por quienes configuran la representación social del conocimiento sobre quilcas en el Perú, es necesario incentivar su uso como una forma de recuperar y relocalizar el término original en el centro de las categorías lexicales de los investigadores comprometidos con este campo.

La proximidad léxica identificó cuatro clases temáticas que se proporcionan en un dendrograma que muestra la agrupación jerárquica de las clases. La asociación entre las palabras es mostrada con el análisis de similitudes que permitió identificar núcleos semánticos representados por las palabras “rupestre” y “arte” que aparecen de forma mucho más destacada que las demás; sin embargo, ambas palabras se relacionan con la palabra “sitio”. Pero también con palabras como “espacio”, “tiempo”, “periodo” y “desarrollo” opuestos a “evidenciar” pero enlazados por la palabra “Perú”.

La Teoría de las Representaciones Sociales como aporte teórico para la investigación realizada con la literatura académica publicada sobre arte rupestre

peruano reafirma que esta reproducción es el resultado de la objetivación textual y terminológica; es decir, a través de la objetivación se tiene conocimientos transferidos de conceptos a significados concretos, que al seguir siendo utilizados en los contextos sociales y académicos en los que se insertan, se convierten en representaciones de la realidad, representaciones que se inscriben en un marco contextual del pensamiento preexistente; por lo tanto, dependiente de sistemas de creencias anclados en valores, tradiciones e imágenes del mundo y la existencia. Los sujetos que las producen están anclados en una red de significados en torno al objeto de estudio (el arte rupestre peruano), en consecuencia, a través de la objetivación, esta red de significados organizados en un proceso comprensible y adecuado a la realidad convierte los conceptos producidos en una imagen lingüística, para comprenderlos, producirlos y reproducirlos.

Según Moscovici (2003), la representación social simplemente “puede ser analizada en términos de una trayectoria icónica y lingüística, ascendiendo a una fuente (las ‘ideas fuente’) y al mismo tiempo buscando estandarizar en la dirección descendente en forma de campos semánticos y esquemas demostrados, de fácil transmisión” (Moscovici 2003: 249). Así, al permitir la organización y clasificación de segmentos de texto según similitudes entre ellos, el software iramuteq permite la indicación de posibles representaciones sociales que organizan y configuran el pensamiento y el conocimiento social del arte rupestre peruano. Los medios de comunicación y difusión de la ciencia producida por la academia emergen en las reflexiones como elementos contribuyentes a la afirmación del arte rupestre en el Perú y sirven como anclas de las ideas referentes a las representaciones sociales lexicales del arte rupestre peruano. Los medios de comunicación académicos influyen pues en la construcción de conceptos socialmente compartidos y orientan las prácticas sociales



relacionadas con un patrón hegemónico de un discurso oficial y oficializado del discurso académico. Por lo tanto, contribuyen también a su reproducción y perseverancia.

Rubén Urbizagástegui-Alvarado
 Doctor en Ciencia de la Información
 Universidad de California en Riverside
 Riverside, California
 ruben@ucr.edu
 Orcid.org/0000-0001-5014-801X

REFERENCIAS

- ACAUAN, L. V., C. Vianna ABRANTES, M. Andrade Conceição STIPP, L. Amorim Corrêa TROTTE, G. Oroski PAES & A. B. Azevedo QUEIROZ 2020. Utilização do software iramuteq para análise de dados qualitativos na enfermagem: um ensaio reflexivo. *Revista Mineira de Enfermagem* 24:1-5. <http://www.reme.org.br/artigo/detalhes/1480>
- ARDITI, Ch., D. WALTHER, I. GILLES; S. LESAGE, A.-C. GRIESSER, C. BIENVENU; M. EICHER, and I. PEYREMAN-BRIDEVAUX 2020. Computer-assisted textual analysis of free-text comments in the Swiss Cancer Patient Experiences (SCAPE) survey. *BMC Health Services Research* 20(1): 1-12. <https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-020-05873-4>
- BEAUDOUIN, V. 2016. Statistical Analysis of Textual Data: Benzécri and the French School of Data Analysis. *Glottometrics*, 33: 56-72, .
- BENZECRI, J.-P. 1973. *L'analyse des données*. Dunod, Paris.
- CODOL, J. P. 1974. On the system of representations in a group situation. *European Journal of Social Psychology* 4(3): 343-365.
- ECHAVARRIA LOPEZ, G.-T. 2016. "Quilca" y "arte rupestre", disquisiciones en el contexto del arte, la arqueología y la ciencia peruana. *Arqueología y Sociedad*, 31: 11-22.
- FAGGIOLANI, Ch., & A. GALLUZZI 2019. Storytelling and the Italian public libraries. Emerging research patterns. *Qualitative & Quantitative Methods in Libraries* 8(4): 437-455. https://iris.uniroma1.it/retrieve/handle/11573/1344014/1317153/Faggiolani_Storytelling_2019.pdf
- FEINERER I., & K. HORNIK. 2020. tm: Text Mining Package. R package version 0.7-8, <https://CRAN.R-project.org/package=tm>. Viena.
- FERREIRA, M. H. W. & R. Fernandes CORRÊA 2018. Estudo métrico sobre biblioteca digital: uso do software iramuteq. *Encontro Nacional de Pesquisa em Ciência da Informação*, n. XIX ENANCIB. Londrina, Paraná. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102876>.
- GUFFROY, J. 1999. Bibliografía. En. *El arte rupestre del antiguo Perú*, pp. 139- 147. IFEA, Lima, Perú.
- GUIRAUD, P. 1954. *Les Caractères statistiques du vocabulaire*. PUF, Paris.
- GUIRAUD, P. 1960. *Problèmes et méthodes de la statistique linguistique*. Larousse, Paris.
- HOFFMANN, Y. T., E. BISSET ALVAREZ, & Y. MARTÍ-LAHERA. 2020. Análise textual com IRaMuTeQ de pesquisas recentes em História da educação matemática no Brasil: um exemplo de Humanidades Digitais. *Investigación bibliotecológica*, v. 34, no. 84:103-133. Mexico, D.F.
- HOSTNIG, R. 2021. Bibliografía sobre Arte Rupestre del Perú. [Cusco, Perú : Rainer Hostnig]. https://www.academia.edu/41737737/BIBLIOGRAF%C3%8DA_SOBRE_ARTE_RUPESTRE_DEL_PER%C3%9A_Mayo_2021_PERUVIAN_ROCK_ART_BIBLIOGRAPHY_May_2021_Compilado_por_Rainer_Hostnig
- IRAMUTEQ. 2020. Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres. Francia. <http://www.iramuteq.org/>
- JACIMOVIĆ, J. 2019. Textometric methods and the TXM platform for corpus analysis and visual presentation. *Infoteka* 19(1): 30-54.
- JASPARS, J. & C. FRASER. 1984. Attitudes and Social Representations. En Farr, R.M. & Moscovici, S. (eds), *Social Representations*, pp.101-124. Cambridge University Press, Cambridge.
- LONDOÑO ZAPATA, O. I. 2015. La ciencia como discurso: Estudios del discurso científico y académico Entrevista a Anamaria Harvey Arellano. *Literatura y lingüística* 31: 335-358.
- MARCHAND, P. & P. RATINAUD. 2011. L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles*, pp. 687-699. Francia.
- MARYMOR, L. 2020. Peruvian rock art bibliography as extracted from the rock art studies bibliographic database for the years 1886 to 2019 : first approximation = bibliografía peruana sobre quilcas o arte rupestre extraída de la base de datos de estudios bibliográficos para los años 1886 al 2019: primera aproximación. *Quellca Rumi*, 2(2): 75-154.
- MOSCOVICI, S. 1981. On social representations. En: J. Forgas (ed.), *Social Cognition: Perspectives on Everyday Understanding*, pp. 181-210. Academic Press, New York.
- MOSCOVICI, S. 1984. The phenomenon of social representations. En. R.M. Farr & S. Moscovici (eds), *Social Representations*, p. 3-69. Cambridge University Press, Cambridge/Paris.
- MOSCOVICI, S. 1988. Notes towards a description of social representations. *European journal of social psychology* 18(3), 211-250.
- MOSCOVICI, S. 2003. *Representações sociais: investigações em psicologia social*. Editora Vozes, Petrópolis.
- MULLER, Ch. 1968. *Initiation à la statistique linguistique*. Larousse, Paris.
- MULLER, Ch. 1977. *Statistique linguistique: exercices et applications*. Centre de Philologie et Littératures Romanes, Université des Sciences Humaines de Strasbourg, Strasbourg.
- MONTEVERDE SOTIL, L. R. 2008. *Bibliografía Básica de Arte Rupestre del Perú en los catálogos de las páginas web de la biblioteca del Instituto Francés de Estudios Andinos (Ifea) de Lima y de la Pontificia Universidad Católica del Perú (PUCP)*. En Rupestreweb, <http://www.rupestreweb.info/bibliografiarupestre.html>.
- NARVÁEZ Luna, J. J. s.f. *Archaeological bibliography of Amazonia and the Eastern Peruvian Andes*. <https://faculty.ucr.edu/~legnerref/biados/texts/bibliosa.htm>
- PINCEMIN, B. & S. HEIDEN 2008. "What is textometry? Introduction", Textometry project website, <http://textometrie.ens-lyon.fr/spip.php?rubrique80&lang=en>.
- R CORE TEAM. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Viena.
- RATEAU, P., P. MOLINER, C. GUIMELLI, and J. Claude ABRIC. 2011. Social representation theory. *Handbook of theories of social psychology* 2:4 77-497.
- Ratinaud, P. 2009. *IRAMUTEQ: Interface de R pour les analyses multidimensionnelles de textes et de questionnaires (computer software) [Internet]*. [cited 2015 Oct 15]. Available from: <http://www.iramuteq.org>.
- REINERT, M. 1997. Les mondes lexicaux des six numéros de la revue "le surréalisme au service de la révolution". *Cahiers du centre de recherche sur le surréalisme (Mélusine)*. 16: 270-302.
- REINERT, M. 1990. ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. *Bulletin de méthodologie sociologique*, 28: 24-54.
- REINERT, M. 1979. Analyse de deux corpus verbaux et présentation d'un programme de classification descendante hiérarchique. PhD diss., Thèse de 3ème cycle, (Université Pierre et Marie



Curie, Paris VI, 1979). Francia.

REINERT, M. 2007 Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours. *Langage et société* 21(122): 189-202.

REINERT, M. 1983. Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données* 8(2):187-198.

Anexo A: Palabras claves usadas en las búsquedas

Arte rupestre AND Perú
 Geoglifos AND Perú
 Arte en las rocas AND Perú
 Pictógrafos AND Perú
 Petroglifos AND Perú
 Quilcas AND Perú
 Sitios de arte rupestre AND Perú
 Formas precolombinas AND Perú
 Manifestaciones rupestres AND Perú
 Arte primitivo AND Perú
 Rock art AND Perú
 Art on the rocks AND Perú
 Pictorials AND Perú
 Petroglyphs AND Perú
 Rock art sites AND Perú
 Pre-Columbian forms AND Perú
 Cave Manifestations AND Perú
 Primitive art AND Perú
 Otras

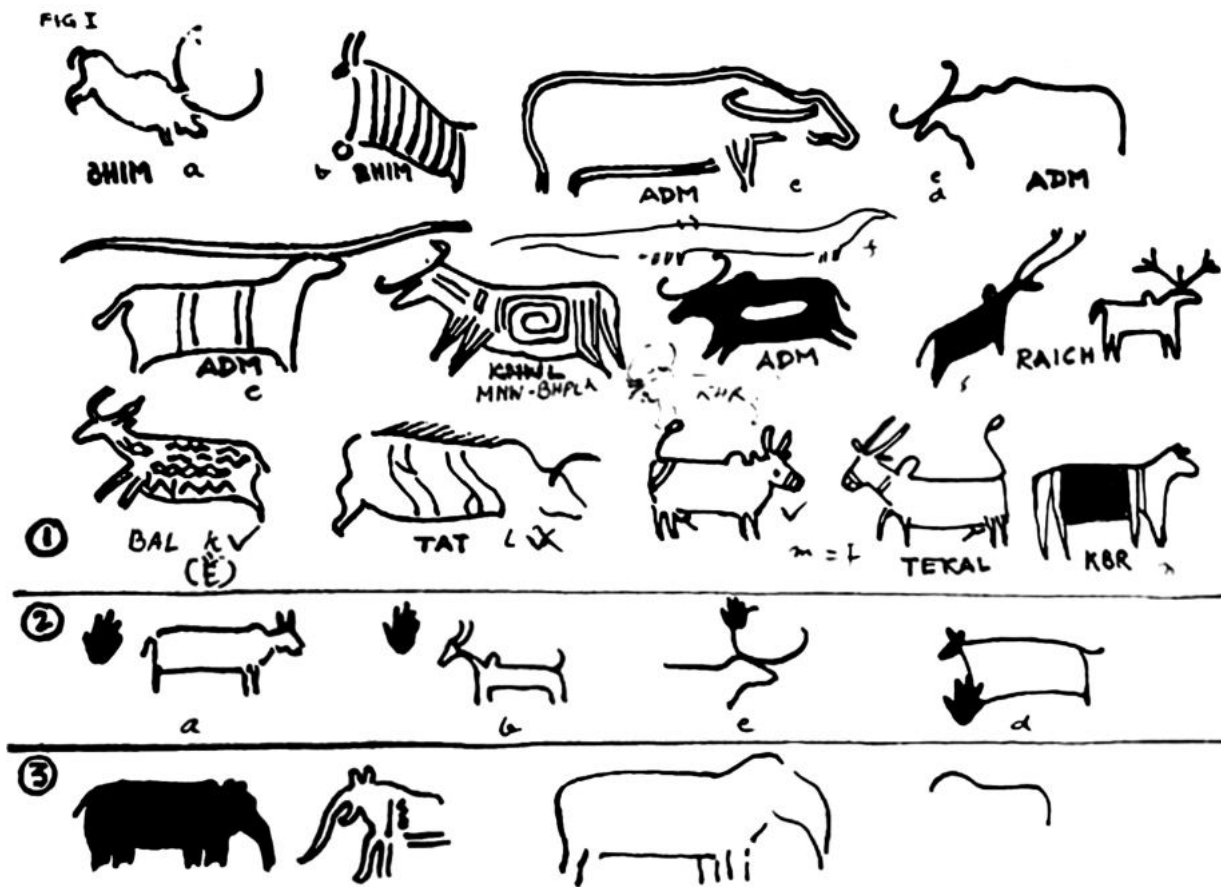
Anexo B: Bases de Datos Bibliográficas

Academic Search Complete
 Anthropology Online
 Anthropology Plus
 AnthroSource

Archaeological Bibliography (Archäologische Bibliographie)
 ArticleFirst
 Arts and Humanities Citation Index (Web of Science)
 Google Scholar
 Handbook of Latin American Studies
 HAPI (Hispanic American Periodical Index)
 JSTOR: The Scholarly Journal Archive
 Latin America Data Base
 Melvyl Catalog
 Social Sciences Citation Index (Web of Science)
 Scopus
 Web of Science
 WorldCat
 Scielo-Peru
 Redalyc
 Otros

Anexo C: Repositorios y Catálogos online

Biblioteca Nacional del Peru
 Biblioteca Central de la Universidad Nacional de San Marcos
 Biblioteca Central de la Universidad Católica del Perú
 Biblioteca del Instituto de Estudios Peruanos
 Biblioteca del Centro Francés de Estudios Andinos
 Biblioteca del Centro de Investigaciones Bartolomé de Las Casas
 Biblioteca del IFEA
 Persée: CNRS, l'Université de Lyon
 Sociedad de Investigación del Arte Rupestre de Bolivia (SIARB)
 Museum of Northern Arizona
 IECTA (Instituto para el Estudio de la Cultura y Tecnología Andina, Arica, Chile)
 Otros.



Dibujos y notas de investigación del Dr. Vishnu Shridhar Wakankar.