

Modelado de tópicos (LDA) aplicado a una controversia ambiental

Isotopías del discurso
ambiental sobre el fuego en
islas del Paraná, Argentina

80

***Francisco José Preiti y
Francisco Javier Alomar***

franciskojo03@gmail.com - Escuela de Antropología

franciscojalomar@gmail.com - Escuela de Antropología

La Trama de la Comunicación - Volumen 28 Número 1 - enero a junio de 2024

ISSN 2314-2634 (en línea) / P. 080-117

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina

Francisco José Preiti y Francisco Javier Alomar

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina.

Resumen:

El artículo aborda el uso de la metodología cuantitativa modelado de tópicos, Latent Dirichlet Allocation (LDA), en el estudio de la controversia ambiental que se configuró en relación a los incendios en islas del río Paraná entre 2020 y 2022. Se enfoca en cómo las metodologías cualitativas y cuantitativas se complementan para enriquecer la investigación en antropología, ecología política y análisis del discurso. A través de LDA, se explora la construcción de isotopías como agrupaciones semánticas emergentes de entrevistas y una asamblea realizadas durante las movilizaciones ambientales en la ciudad de Rosario, Argentina. Los resultados revelan métricas, tópicos e interpretaciones que resultan fundamentales para la definición de isotopías, vinculadas a la organización e identidad política de las agrupaciones, la tensión entre conservación y desarrollo, y la interacción entre discursos globales y locales. De este modo, se caracteriza al discurso ambiental en un caso particular, narrado por sujetos que participaron en la dinámica de conflictos. El escrito se propone como integración metodológica para un análisis más profundo de las controversias ambientales.

Descriptor:

LDA, isotopías, metodología cuantitativa-cualitativa, discurso ambiental, controversias ambientales.

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina
Topic modeling (LDA) applied to an environmental controversy
Isotopies of environmental discourse about fire in the Paraná River islands, Argentina.
Páginas 080 a 117 en **La Trama de la Comunicación**, Volumen 28 Número 1, enero a junio de 2024 - ISSN 2314-2634 (en línea)

Topic modeling (LDA) applied to an environmental controversy

Isotopies of environmental discourse about fire in the Paraná River islands, Argentina.

Summary:

This article explores the application of Latent Dirichlet Allocation (LDA), a quantitative topic modeling technique, to the environmental controversy surrounding the fires on the Paraná River islands from 2020 to 2022. It examines how integrating qualitative and quantitative methods can enhance research in anthropology, political ecology, and discourse analysis. The study utilizes LDA to investigate the formation of isotopies—emergent semantic clusters from interviews and an assembly held during environmental protests in the City of Rosario, Argentina. Findings highlight key metrics, topics, and insights crucial for defining isotopies, which are connected to the organizational and political identities of the groups involved, the dichotomy between conservation and development, and the interplay between global and local discourses. This case study thereby characterizes the environmental discourse shaped by participants in the conflict dynamics. The paper advocates for this methodological integration to deepen the analysis of environmental controversies

Describers:

LDA, isotopies, mixed-methods approach, environmental discourse, environmental controversies.

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina
Topic modeling (LDA) applied to an environmental controversy
Isotopies of environmental discourse about fire in the Paraná River islands, Argentina.
Páginas 080 a 117 en **La Trama de la Comunicación**, Volumen 28 Número 1, enero a junio de 2024 - ISSN 2314-2634 (en línea)

Introducción:

En la antropología como en el análisis del discurso estamos familiarizados con interpretaciones cualitativas de los corpus, más que cuantitativas. Sabemos que cada enfoque tiene potencialidades y deficiencias, pero no es nuestra intención indagar sobre la comparativa. De modo general, nos interesa dar cuenta de su complementariedad. Pues, ¿de qué modo podríamos generar un aporte en este sentido? Una forma posible es la que proponemos desarrollar, indagar sobre lo que estamos menos familiarizados, lo cuantitativo.

En el presente artículo aplicamos una metodología de procesamiento del lenguaje natural (PLN), minería de datos y aprendizaje no supervisado, que requiere capacidad computacional para su implementación. El objetivo es mostrar el potencial de la metodología aplicada en la investigación doctoral a cargo del Lic. Francisco J. Preiti. La técnica metodológico-cuantitativa es conocida como modelado de tópicos, *Latent Dirichlet Allocation* (LDA), propuesta por Blei (2003). Por su lado, la investigación doctoral versa sobre el análisis de cuestionamientos ambientalistas emergentes por la presencia del fuego en islas del río Paraná entre los años 2020 y 2022. Tal investigación se encuentra en curso (2024) y comparte los ámbitos disciplinares de la antropología social, la ecología política y el análisis del discurso.

Los incendios en las islas del río Paraná han generado una serie de cuestionamientos ambientales que reflejan la preocupación por el impacto ecológico y social de los eventos. El estudio aborda su emergencia y evolución durante el período mencionado, al proporcionar una perspectiva crítica sobre las dinámicas socio-ecológicas en la región. Las movilizaciones sociales, mediante repertorios de protesta y matrices narrativas, desplegaron valoraciones sobre el territorio, el rol del Estado y las políticas públicas. Se destaca como problemática el impacto de la expansión de explotaciones y lógicas del sector privado en el ecosistema de

islas, cuya cara visible para el ambientalismo es la producción ganadera y el fuego como práctica asociada. De esta manera, el ambientalismo cuestiona el modelo de desarrollo productivo vigente en la región a través de una discusión sobre los instrumentos normativos que se orientan a regular los usos del territorio (fue slogan y bandera de lucha la búsqueda de una “Ley de Humedales Ya!”). En suma, los incendios escenifican para los grupos movilizados la expresión cruenta de un vínculo extractivista con el ambiente y sus componentes. A través de la recuperación del punto de vista de los actores, la investigación en curso busca indagar en tales visiones sociales que narran el territorio. El principal insumo se constituye a partir de entrevistas de investigación a referentes y actores de las movilizaciones ambientalistas cuyo registro textual, a su vez, es también el insumo básico de la metodología empleada para el presente artículo. Se espera que este enfoque metodológico que integra el PLN, la minería de datos y el aprendizaje no supervisado, contribuya a una comprensión más profunda y matizada de las interacciones entre la sociedad y el ambiente, promoviendo nuevas vías de investigación en la antropología, la ecología política y el análisis del discurso.

El desarrollo del escrito está organizado del siguiente modo. Partimos del encuadre de la investigación para elaborar una hipótesis particular del artículo. Luego, describimos la obtención y definimos el corpus de análisis, como así también la metodología cualitativa previamente empleada. En tercer lugar, definimos PLN, minería de datos y aprendizaje automático no supervisado para una mayor comprensión de la metodología cuantitativa que aplicamos. Además, explicamos los fundamentos de LDA, los supuestos generales que lo sustentan y la relación con la identificación de isotopías, herramienta del análisis del discurso. Posteriormente, describimos el acondicionamiento del corpus para aplicar LDA. La última sección es de resultados e interpretaciones. Hacia el final, a modo de conclusión, establecemos proyecciones de investigación a partir de reflexiones sobre los resultados.

1. De la investigación general a una hipótesis particular.

1.1 Marco de investigación y focalización del caso de estudio.

La investigación propone contribuir a los debates de la antropología y la ecología política en torno de controversias ambientales mediante el estudio de un caso: la presencia del fuego en las islas del Delta del Paraná entre los años 2020 y 2022, donde multitudes¹ cuestionaron un campo de la experiencia social e instalaron en la esfera pública una preocupación por la distribución y pagos de externalidades² del capitalismo global. Consideramos que el abordaje de controversias contemporáneas participa de la cooperación social que se configura alrededor de cada desafío que emerge del devenir de las sociedades. Es por ello que la producción del trabajo de investigación doctoral estriba en una doble destinación: se orienta a constituir una contribución del saber científico-social y se propone nadar en las turbulentas corrientes de la producción del conflicto ambiental.

A los fines del escrito y desde un punto de vista teórico referimos a conflictos ambientales (Merlinsky, 2022; 2013; Telias, 2014) y a discusiones de la ecología política reconociendo la emergencia de una ambientalización de las luchas sociales (Leite López, 2006) y ecoterritoriales (Svampa, 2013). Constituyen un marco

1 Según distintas fuentes en los cortes de ruta del año 2020 se movilizaron entre 2000 y 15.000 personas, considerándose como unas de las más grandes movilizaciones urbanas por motivos ambientales del presente siglo en el país (Argento, 2023). Por ejemplo, en Wikipedia se señalan 10.000 personas en la movilización del 1/08/2020 (https://es.wikipedia.org/wiki/Incendios_en_el_delta_del_r%C3%ADo_Paran%C3%A1_de_2020) mientras que otras fuentes reconocen hasta 15.000 (Aranda, 2022).

2 Con este concepto referimos a los costes de los procesos de producción y consumo capitalista, a los impactos ecosistémicos que muchas veces recaen en las poblaciones locales. Para más información véase Martínez Alier (2006).

interpretativo también los discursos del desarrollo-biodiversidad (Escobar, 1998; 1999) y la antropología de la conservación (Beltrán Costa y Santamarina, 2016, Ferrero Et. Al, 2019). Dentro del amplio universo de controversias ambientales, nos hemos enfocado en los procesos de construcción de identidades colectivas y *querella* ambiental en el caso de estudio (Preiti, 2021a; 2021b; 2022; Preiti & Youkhana, 2023).

Una de las características de los discursos ambientales es la elaboración de lenguajes de valoración (Martinez Alier, 2006) si bien particulares, atravesados por expresiones globales (incluso estimulados por movimientos globales como Greenpeace, Fridays For Future o Extinction Rebellion) que abrevan en la producción discursiva de la biodiversidad a través de reivindicaciones territoriales (Escobar, 1999). En relación a ello, vamos a considerar el neologismo “glocal”. Es un término acuñado por Escobar (1999) quien lo retoma de Dirlik (1997): se busca designar la interrelación y mutua significatividad entre prácticas y sentidos propios de un lugar y aquellos que circulan más allá de lo regional. Los discursos de la biodiversidad son un ejemplo significativo de esta interdiscursividad. El autor llama la atención respecto de que lo “glocal” es una primera aproximación que sugiere “...prestar la misma atención a la localización de lo global y a la globalización de lo local.” (Escobar, 1999, p. 371). En la misma dirección reflexiona De Sousa Santos (2006) al interrogarse si no son, justamente, las narrativas y prácticas globales las que configuran rasgos culturales percibidos como locales:

Por su parte la globalización es una identidad que se expande en el mundo y, al expandirse, adquiere la prerrogativa de nombrar como locales a las entidades o realidades rivales. Es decir, no hay globalización sin localización. Cuando globalizas el McDonald's, localizas tus comidas: las tornas étnicas, locales. No hay universalismo sin particularismo (De Sousa Santos, 2006, p. 25).

Queda explícito que una teoría debiera anclar en algo para su comprensión. Es por ello que proponemos focalizar en un caso de estudio. Una caracterización regional señala al ecosistemas de islas del Delta del río Paraná como una ecorregión de suma importancia por su biodiversidad (su diversidad cultural no suele ser mencionada). En ella se superponen figuras de conservación y protección ambiental como la Reserva de Biósfera del delta del Paraná o la Reserva de Usos Múltiples Islas de Victoria. También están bajo conservación estricta por parte del Estado tres parques nacionales que involucran a tres provincias: el Parque Nacional (PN) PreDelta (Prov. De Entre Ríos); PN Islas de Santa Fe (Prov. De Santa Fe); el PN Ciervo de los Pantanos (correspondiente a la Provincia de Buenos Aires). La relación entre la ampliación de áreas protegidas y los cuestionamientos sociales al desarrollo son significativos y pueden analizarse en clave de un tensionamiento conflictivo y complementario entre la conservación del ambiente y el desarrollo socioproductivo (Ferrero, 2019; Preiti y Ferrero, 2023). A su vez, está declarada una porción de la región al norte como Sitio Ramsar “Delta del Paraná” en un acuerdo entre los gobiernos provinciales y la Administración de Parques Nacionales, que involucra los departamentos de Diamante (Entre Ríos) y San Gerónimo (Santa Fe). Junto con los parques nacionales, el Sitio Ramsar y otras figuras de conservación, el río Paraná representa un mosaico complejo de formas de gubernamentalidad territorial y producción social del espacio (Lefebvre, 2013) en donde se despliegan tensiones entre la conservación del ambiente y el desarrollo económico y productivo de la región.

Se configura así una delimitación geográfica donde se han desarrollado incendios que en buena medida sociedades extralocales, significaron y cuestionaron (como es el caso de la ciudad de Rosario, Argentina). En ellas se nuclea actores con distintos niveles de participación, formación experta, militancia social y preocupaciones por lo ambiental que abordan, visibilizan, reconstruyen y problematizan el territorio. Hemos tomado como

La Trama de la Comunicación - Volumen 28 Número 1 - enero a junio de 2024

ISSN 2314-2634 (en línea) / P. 080-117

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina

Francisco José Preiti y Francisco Javier Alomar

manifestaciones sus discursos y prácticas, cuya recolección ha constituido el corpus de la investigación. En el caso de estudio trabajamos sobre enunciaciones locales de la controversia ambiental a la luz del despliegue de discursos polémicos, ideológicos y construcción del adversario (Amossy, 2016, 2017; Montero, 2016; Van Dijk, 1996). Identificamos la impugnación de actores y prácticas productivas locales (ganadería extensiva, caza, pesca y modificaciones antrópicas tal como endicamientos, terraplenes, caminos, etc.) por parte de los movimientos ambientales de la ciudad (Preiti, 2021a; 2021b; 2022).

1.2 Obtención del corpus y metodología cualitativa utilizada.

En el presente artículo trabajamos sobre una parte del corpus total de la investigación doctoral³, compuesto por entrevistas estructuradas, semiestructuradas y abiertas a referentes de organizaciones socioambientales y activistas autoconvocados. Además se encuentra la desgrabación de una asamblea abierta que se llevó a cabo sobre el puente “Nuestra Señora del Rosario”, interrumpido su tránsito por una movilización de miles de personas el 1/08/2020, la mayor conflictividad socioambiental de la historia de la ciudad de Rosario. En tal asamblea participaron organizaciones sociales y no gubernamentales, referentes políticos, agrupaciones estudiantiles, universitarias, sindicatos y autoconvocados. Por su parte, las entrevistas se basaron en la reconstrucción, junto a referentes y participantes, de los diversos eventos e intervenciones urbanas que se desarrollaron como cuestionamientos a la presencia del fuego en las islas. El trabajo

3 El corpus total de la investigación doctoral comprende cuatro conjuntos de textualidades: 1) entrevistas y notas de campo de las movilizaciones ambientales; 2) documentos de políticas públicas; 3) recuperaciones de discursividades mediáticas digitales y 4) abordajes técnicos científicos sobre el delta paranaense.

de campo, metodología principalmente cualitativa, característica de la antropología para el estudio de grupos sociales, resultó fundamental en la recolección del corpus de la tesis, como así también para la identificación de informantes claves a quienes realizamos las entrevistas, texto que analizamos en el presente artículo (en conjunto con la desgrabación de la asamblea). El trabajo de campo fue realizado en la ciudad de Rosario y consistió en la observación participante de asambleas, actividades artísticas, movilizaciones, acampes, cortes de rutas y del puente.

Una de las actividades que orientó las entrevistas fue la propuesta de “revisitar” los eventos y movilizaciones ambientalistas en función de los acontecimientos, narrando sus percepciones y vivencias. Nuestra motivación fue reconstruir los hechos pasados, prestando atención a la distribución de roles sociales expuestos por los actores, las preocupaciones, reivindicaciones y demandas socioambientales, su relación con el Estado y el territorio, las características y adversidades de la lucha socioambiental, entre otros aspectos.

1.3 Hipótesis específica del artículo.

Durante el trabajo de campo y luego de realizar las entrevistas, generamos interpretaciones cualitativas que surgieron de la experiencia inmersiva y el retorno al producto de ello, como una vuelta hacia lo vivenciado, recolectado, registrado, incluso, mediatizado por el tiempo, habilitando ciertas preguntas, (sólo por mencionar algunas): ¿cómo es conceptualizada la relación ambiente - sociedad en los discursos emergentes de las movilizaciones socioambientales y en las entrevistas? ¿Qué visiones de mundo encontramos en las enunciaciones de los grupos movilizados que se han pronunciado en torno al conflicto en foco? ¿Cuáles son las preocupaciones del ambientalismo local en relación al territorio de islas? ¿Existe un discurso ambiental discernible de otros tras el conflicto sobre la presencia del fuego? ¿Qué regula-

ridades tiene? ¿Qué rol le es asignado al Estado? ¿Y a la naturaleza? ¿Cuáles son las formas organizacionales de los grupos?

Nos interesa destacar una singularidad que emergió del análisis cualitativo. En los resultados parciales de investigación quedaron expuestas recurrencias y lugares comunes al desgrabar las entrevistas. En tales regularidades se expresan como temáticas predominantes la organización política, la delimitación de identidades colectivas y la biodiversidad en tanto red de producción sociodiscursiva (Escobar, 1999). Por ejemplo, los nombres que las propias organizaciones dieron a los eventos e intervenciones desarrolladas durante esos años (2020-2021) fueron: “Marcha Viboreante”, “Abrazo al Río”, “Cruce de las orillas”, “Rosarizao Socioambiental”, “Juicio ético y popular a los ecocidas”. Tales designaciones parecieran aglutinar nominaciones provenientes de distintos ámbitos: el ambientalismo, la protesta social, la ecología, la biodiversidad, la militancia y el activismo, las identidades colectivas y la conservación de la naturaleza; pero a su vez, nos recuerdan a las isotopías⁴ del análisis del discurso (Greimas en Maingueneau y Charaudeau, 2005), dado que, al aglutinar distintos ámbitos, se configuran nuevas relaciones y categorías de sentido expresadas con ciertas palabras y no otras. Pues, si aceptamos la existencia de tales isotopías, ¿cuáles subyacen en las entrevistas realizadas? En caso que las hubiera, ¿otorgan una particularidad en el despliegue textual del caso que lo diferencia de otras controversias ambientales ya estudiadas por la antropología y la ecología política? ¿Acaso la preocupación por la organización política, las identidades colectivas, la biodiversidad y la conservación del ambiente son constitutivos de las controversias ambientales en general o estamos frente a las particularidades de un caso? Y por último, ¿de qué modo podríamos evidenciar cuantitativamente la presencia de isotopías en las entrevistas y

4 La categoría isotopía es conceptualizada en la sección 2.2. Además, planteamos su relación con el modelado de tópicos (LDA).

asamblea? Consideramos de manera preliminar que la organización política, las formas organizacionales del ambientalismo social (Gutiérrez e Isuani, 2014), definiciones de la identidad colectiva y los discursos de la biodiversidad (Escobar, 1999), se entremezclan, discursivamente, y conforman isotopías (Greimas en Maingueneau y Charaudeau, 2005) singulares al proceso que hemos delimitado estudiar. Tales campos semánticos pueden evidenciarse cuantitativamente aplicando el modelado de tópicos (LDA).

Hemos dado un marco general de la tesis doctoral y presentamos una hipótesis particular para el artículo. Ahora, debemos ocuparnos de la metodología y definir ciertos conceptos, tal vez, un tanto ajenos a los enfoques que estamos acostumbrados, más cualitativos. Porque de otro modo, no se comprendería la herramienta metodológica que utilizamos (LDA). Para ello, pasamos al apartado siguiente.

2. Metodología empleada: LDA.

2.1 Algunas definiciones preliminares: PLN, minería de datos y aprendizaje automático no supervisado.

Natural language processing conlleva en sí un problema de traducción: *language* es indistintamente lenguaje o lengua según el contexto. Desde un punto de vista lingüístico y antropológico sabemos que no es lo mismo. Excede al artículo una exhaustiva disquisición sobre el tema. En el campo de las ciencias de la computación, PLN se refiere al habla o la escritura de lenguas: quechua, español, flamenco, etc. El atributo de natural plantea una diferencia de aquellos sistemas semióticos artificiales porque no han sido inventados como sí lo fueron, por ejemplo, los lenguajes de programación. El término procesamiento indica obtener información de ellos a través de algoritmos. El PLN abarca diversas

tareas computacionales, desde básicas, como el reconocimiento de palabras, sintaxis, estructuras gramaticales, hasta cuestiones de mayor complejidad como traducción automática, el análisis de sentimientos, generación de texto y procesamiento del habla, entre otras (Jurafsky y Martin, 2024).

LDA es un algoritmo de PLN que genera un modelo estadístico, de minería de datos y aprendizaje no supervisado. Cuando decimos minería de datos referimos a “...the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable and predictive models from large-scale data.” (Zaki y Meir, 2020, p. 4). Los algoritmos de minería son de gran utilidad dado que permiten establecer relaciones subyacentes a los datos, muchas veces inasequibles al investigador, ya sea por la gran escala del corpus o porque son “anti-intuitivas”.

Una técnica de aprendizaje automático no supervisado es aquella que busca patrones y estructuras sin disponer de una variable objetivo y carece de “vigilancia” humana. Los algoritmos no supervisados buscan en los datos identificar diferentes características o instancias que los agrupen, generando *clusters*. En el caso del algoritmo LDA los *clusters* son agrupaciones de palabras⁵, siendo una entre varias técnicas de modelado de tópicos. Estas agrupaciones, por sí mismas, carecen de sentido desde el punto de vista de la investigación si no son sometidas a un proceso interpretativo que permita nominalizarlas e identificar qué rol juegan en la escena textual. Pero esto ya nos lleva a los resultados. Antes debemos reparar en la siguiente sección.

2.2 Modelado de tópicos y su relación con las isotopías del análisis del discurso.

El *output* del algoritmo LDA, a partir del corpus de análisis

5 Como explicitamos más adelante, no son palabras, sino raíces nominales.

como *input*, es un modelado de tópicos. Ahora bien, ¿qué significa modelado de tópicos (*output*)? En nuestro caso, los *clusters* son agrupaciones de palabras que llamamos tópicos. En términos del algoritmo, consisten en una distribución de probabilidades de aparición de componentes del vocabulario que, al ser interpretadas y luego de un proceso de inferencia por parte del investigador, se traducen en relaciones semánticas, latentes, tal vez no detectables tras una lectura tradicional. Es decir, que estos conjuntos de palabras que tienen una distribución probabilística y *aparecen* juntas, si logramos darle sentido, son análogos a las isotopías del análisis del discurso: palabras que en la textualidad *funcionan* juntas.

El concepto de isotopía fue acuñado por Greimas (1966). En principio, entendemos que la isotopía “...designa globalmente los procedimientos que contribuyen a la coherencia de una secuencia discursiva o de un mensaje. Basada en la redundancia de un mismo rasgo en el suceder de los enunciados, esa coherencia concierne principalmente a la organización semántica del discurso” (Maingueneau y Charaudeau, 2005, p. 342). En tal sentido, existe cierto consenso teórico acerca de que la isotopía “...define los mecanismos reguladores que contribuyen a hacer de un enunciado o de un texto ‘una totalidad de significación’” (Greimas En Maingueneau y Charaudeau, 2005, p. 342). Finalmente es de interés destacar según estas vertientes teóricas (semiótica, semiolingüística, análisis del discurso) que, “...en el plano funcional, la coherencia discursiva producida por la isotopía condiciona la legibilidad de los textos (Maingueneau y Charaudeau, 2005, p. 342), ‘desde el punto de vista del enunciatario, la isotopía constituye una grilla de lectura que vuelve homogénea la superficie del texto, por cuanto permite suprimir las ambigüedades’ (Greimas y Courtés En Maingueneau y Charaudeau, 2005, p. 342). Considerando que en el devenir teórico de la categoría se establecieron tipologías, para nuestro caso, es de interés retener aquellas identificadas por los autores que han acuñado el concepto, llamadas

isotopías semánticas estrictas (definidas por la recurrencia de una misma categoría de sentido) (A.-J Greimas, J. Courtés En Maingueneau y Charaudeau, 2005, p. 342).

El algoritmo asignará una distribución de probabilidad a cada palabra para un tópico, y la contribución del tópico, también en términos probabilísticos, a un conjunto textual⁶. Luego, al obtener el *output*, debiéramos interpretar y nominar al tópico, justamente, transformarlo en isotopía: grupo de palabras que conforman un campo semántico, de coherencia, recurrencia y organización en la escena textual.

2.3 Terminología de LDA y supuestos generales del algoritmo.

LDA parte de un corpus total que, según la terminología de Blei (2003) llamamos D , el cual es una colección finita de documentos w (que en nuestro caso son las entrevistas y la asamblea), y, a su vez, cada w está compuesto por un conjunto de palabras que pertenecen a un vocabulario, también finito, denominado V . Uno de los supuestos principales de LDA es que cada w del corpus D es susceptible de contener varios tópicos (Rosati, 2022). En nuestro caso, quiere decir que en cada entrevista se habla de varios temas a la vez (política, identidad, etc.), pero estos temas (representados por los tópicos) se distribuyen en el conjunto total de entrevistas y asamblea, aunque no necesariamente en todas.

Siguiendo a Rosati (2022), pasamos a explicitar los supuestos generales del algoritmo. (1) Una palabra del vocabulario V tiene relevancia si es frecuente en un documento w (entrevista / asamblea). (2) Por otra parte, esa palabra, otorga más información sobre ese w si está presente en pocos w y no en todos. (3) Cada w puede presentar varios tópicos y cada tópico es una distribución

⁶ . Esta cuestión se comprenderá mejor cuando describamos la preparación del corpus (sección 2.4.) y en la interpretación de los resultados (sección 3.).

de probabilidad de aparición sobre el vocabulario V . (4) Los tópicos existen con anterioridad a cada w y la distribución de probabilidad sobre el V es constante. Para el algoritmo, (5) el orden de documentos (w) y (6) palabras del vocabulario (V) es indiferente. Por último, (7) se admite que existe una cantidad finita y no variable de tópicos.

Los supuestos que mencionamos parecieran arbitrarios, pero son totalmente imprescindibles para comprender el funcionamiento de la metodología, su potencialidad y limitaciones. Consideramos que la razón se esclarecerá en la siguiente sección, donde explicamos la preparación del corpus. Es que hasta aquí, sólo hemos discurrido sobre el *output* de LDA, su relación con las isotopías del análisis del discurso y los supuestos, pero nada hemos dicho aún de la preparación del corpus para llegar a ser el *input* del algoritmo.

2.3 Preparación del corpus de análisis.

El punto de partida metodológico fue el trabajo de campo y la observación participante. De ello obtuvimos los informantes claves a quienes hicimos las entrevistas, y, además, generamos un registro de la asamblea abierta. Los registros fueron desgrabaciones que organizamos metodológicamente en: 1) entrevistas, universo de estudio (UE) y 2) asamblea, universo de contraste (UC), más allá de que estudiamos todo el conjunto. La división cualitativa del corpus se fundamenta en que UE se compone de entrevistas abiertas, guionadas y semiestructuradas mientras que la asamblea, UC, tuvo una directriz no organizada por el investigador: se desarrolló en correspondencia con dinámicas grupales de la actividad contenciosa. Pero además, esta división surgió de considerar como una diferencia la dualidad público / privado en la recolección del corpus de análisis. La asamblea se desplegó en un ámbito público atípico (sobre una ruta interprovincial) y con una dinámica colectiva reglada

La Trama de la Comunicación - Volumen 28 Número 1 - enero a junio de 2024

ISSN 2314-2634 (en línea) / P. 080-117

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina

Francisco José Preiti y Francisco Javier Alomar

(con turnos y lista de oradores). Mientras que las entrevistas se desarrollaron en situaciones diádicas, con una menor restricción de la performance discursiva y en lugares del discurrir cotidiano (casas particulares, bares, plazas, espacios de coworking, ámbitos laborales). La situación de enunciación (Maingueneau, 2004) que se configura en cada escena enunciativa (movilización social / entrevista) imprime sobre el corpus las especificidades del género.

La desgrabación fue el primer paso de preparación del corpus para LDA, porque sobre cada w (entrevistas / asamblea) fue necesario aplicar una serie de técnicas para que, básicamente, cada w del corpus D , quede reducido a un conjunto que pueda ser procesado por LDA. Es importante retener que el algoritmo, luego de acondicionar los datos, sólo tiene información de cuántas veces aparece una palabra en cada w del corpus D , y que esa palabra es igual o distinta a otra sin importar su contenido semántico: “fuego” será igual a “fuego”, pero tan distinto de “agua” como de “incendio”. La secuencia de preparación del corpus e implementación del algoritmo, una vez que cada w es texto escrito (en nuestro caso, formato *txt), es la siguiente:

1. Se hace una extracción del texto de cada w y se lo nombra como entrevista 0, 1, 2,..., 20. La entrevista 0 es la asamblea. Hemos decidido poner números a las entrevistas para resguardar la identidad de los entrevistados.
2. Luego, a cada uno de los textos, se aplica un proceso de normalización. Por ejemplo, si el texto original es: “Esto se va a apagar, pero no pasa nada. [...]”, se reduce a “esto se va a apagar pero no pasa nada [...]”. Es decir, se sacan mayúsculas y transforman a minúsculas, se retiran los signos de puntuación, caracteres especiales, números y palabras que sean de un solo carácter.
3. El tercer paso es la tokenización. Esta técnica consiste en dividir un texto en unidades significativas, discretas. A par-

- tir de aplicar (b), se transforma el texto en una lista de palabras: [esto, se, va, apagar, pero, no, pasa, nada, ...].
4. Para muchos algoritmos de PLN, como es el caso de LDA, es necesario aplicar el proceso a las *stopwords*, que son palabras que desde el punto de vista del algoritmo no aportan sentido por su alta frecuencia en todos los w^7 , tales como, conectores, artículos, pronombres relativos, preposiciones, entre otras. Pues, la lista anterior, luego de aplicar *stopwords*, queda: [va, apagar, pasa,...]
 5. Ocurre que a los fines del algoritmo, palabras como “apagar”, “apagó” y “apagará” quisiéramos que sean tomadas todas bajo una misma unidad, porque de lo contrario, al ser tomadas como distintas, generarían una distorsión en la distribución estadística del modelado de tópicos. Es por ello que se aplica la técnica de *stemming*, que reduce todas ellas a su raíz nominal: “apag”. Siguiendo el ejemplo anterior, la lista ahora es: [va, apag, pas, ...]
 6. El siguiente paso, es generar un diccionario, objeto particular de la clase *dictionary* de la librería Gensim de Python⁸. El diccionario asocia cada raíz nominal a un número natural (1, 2, 3,..., n). Esto se hace porque en términos de rendimiento computacional es menos costoso trabajar con números que con agrupaciones de letras. La ordinalidad de los números naturales no es tenida en cuenta, es sólo para

7 Recordar los supuestos (1) y (2) de LDA. De ahí la importancia de retirar las stopwords del conjunto de datos, porque de otro modo, predominarían por su alta frecuencia en todos los tópicos, desvirtuando las posibles identificación de tópicos latentes

8 Python es el lenguaje de programación que utilizamos para implementar la metodología. Una librería o biblioteca es un conjunto predefinido de herramientas y funciones con fines particulares. En nuestro caso, es mediante Gensim que implementamos LDA. Nosotros no diseñamos el algoritmo LDA que, de por sí, se basa en matemáticas probabilísticas complejas (Blei, 2003). Es decir, utilizamos el algoritmo que la librería ofrece y le pasamos una estructura de datos.

diferenciarlos.

7. Ahora necesitamos asociar los números del diccionario que representan raíces nominales a su frecuencia en cada w , es decir, a cuántas veces aparece cada raíz nominal en una entrevista o en la asamblea. Un modelo utilizado para esto es *bags of words* (BoW): pensar cada documento como una bolsa de palabras donde el orden no importa (supuesto 5 y 6 del punto anterior). Para el concepto de BoW sólo interesa identificar y diferenciar las raíces nominales e indicar su frecuencia en los documentos w . De este modo y en el contexto de LDA, cada raíz se representa como una tupla⁹. Por ejemplo, supongamos que la raíz “apag” es igual a 1 y aparece 10 veces en un documento: esto se representaría (1, 10). Un documento, como BoW, se figurara como $w = [(1, 10) , (4, 8) , (11, 5) , \dots , (20, 3)]$, donde lo que está entre corchetes (lista), representa a un documento (entrevista / asamblea), y, los pares entre paréntesis (tuplas) son las raíces de ese documento. El primer número de la tupla es la raíz nominal, y el segundo, su frecuencia en el documento w .
8. Aquel corpus desgravado a formato *txt, ahora se modeliza como corpus (D) de documentos (w). Recordemos que llamamos D al corpus total y w a cada uno de los documentos. En nuestro caso, tenemos 21 documentos -la entrevista o es la asamblea-, y de modo abstracto el corpus queda $D = [w_0, w_1, w_2, \dots, w_{19}, w_{20}]$. El *input* del algoritmo LDA recibe un objeto concreto que es una lista (entre corchetes, que representa al corpus) de listas (también entre corchetes, que representan a la asamblea y entrevistas); estas últimas, a su vez son conformadas por las tuplas (entre paréntesis, que representan a cada raíz nominal con la frecuencia de

9 Tanto las listas, las tuplas como los diccionarios, son objetos del lenguaje de programación Python, que es uno, justamente, orientado a objetos. Cada objeto tiene sus atributos y métodos que permiten atender a diversos fines computacionales.

aparición en ese documento). Pues, a modo de ejemplo, el corpus como *input* del algoritmo LDA queda: $[[(0,9), \dots, (500, 8)], \dots, [(2006,4), \dots, (500, 2)]]$. Se aprecia que una raíz nominal puede estar en varios documentos, y, de los supuestos (5) y (6) el orden de las tuplas (raíces nominales con frecuencia) y listas (documentos, w) es indistinto para LDA. De ahí la idea de bolsa de palabras (BoW).

De este modo construimos un corpus D que consta de un vocabulario V conformado por 2006 números que representan raíces nominales con su respectiva frecuencia de aparición en 21 documentos w . En total, la cantidad de raíces nominales que tiene el corpus (contemplando las repeticiones en todos los documentos) es 31547.

Hasta aquí es la preparación del corpus para ser *input* de LDA.

3.Resultados

3.1 Interpretaciones preliminares.

Antes de obtener los resultados fue necesario fijar hiperparámetros, que son parámetros establecidos con anterioridad de ejecutar el algoritmo. La versión LDA de la librería Gensim admite varios hiperparámetros, de los cuales sólo explicaremos uno: el número de tópicos. El algoritmo devolverá resultados diferentes si uno varía esta cantidad. En nuestro caso determinamos que el número de tópicos sea 3. Hicimos esto luego de una exploración algorítmica: variamos arbitrariamente el hiperparámetro hasta que, luego de sucesivas evaluaciones, optamos con el siguiente *output*:

(0, '0.010*"acamp" + 0.008*"comision" + 0.007*"asamble" + ... + 0.003*"arte"')

(1, '0.009*"xr" + 0.006*"are" + 0.006*"contact" + ... + 0.003*"nombr"')

La Trama de la Comunicación - Volumen 28 Número 1 - enero a junio de 2024

ISSN 2314-2634 (en línea) / P. 080-117

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina

Francisco José Preiti y Francisco Javier Alomar

(2, '0.008*"naturalez" + 0.008*"vid" + 0.006*"exist" + ... + 0.003*"soci"')

Los tópicos son 3. El output contiene las primeras 20 raíces nominales de cada tópico que en su mayoría omitimos dado que las presentaremos en un gráfico. Solamente trabajar con las primeras 20 es una decisión metodológica, porque cada tópico se constituye de la totalidad del vocabulario, es decir, en nuestro caso, 2006 raíces. Para comprender esto es necesario destacar que el output de tópicos es devuelto como una suma en orden decreciente donde cada raíz hace una contribución de probabilidad al tópico: por ejemplo, para el tópico 2, la raíz “naturalez” es la primera y más importante, siendo poco más del doble de “asamble”, que es la tercera. Ahora bien, ¿qué significa esto? Básicamente, en relación a un w particular, la sumatoria establece la probabilidad de obtener el w (entrevistas / asamblea) a partir del conjunto de palabras a la que aluden las raíces y conforman un tópico. Como se aprecia, las probabilidades son pequeñas. La intuición que debemos retener del output es que, si bien las chances en general resultan bajas, otorgan una probabilidad inmensamente mayor a la que obtendríamos si eligiéramos al azar un conjunto de raíces de V para obtener un w original (Blei, 2003)¹⁰.

Antes de mostrar una representación gráfica de los tópicos, mencionamos una dificultad al implementar LDA. Si bien tras la exploración de los sucesivos output optamos por el presentado, encontramos raíces de palabras tales como “che”, “ponele”, “obvio”, provenientes de la oralidad que no fueron identificadas como stopwords, pero dada su alta frecuencia aparecían en los tópicos, desvirtuando los posible sentidos que pudiéramos otor-

10 A los fines del algoritmo cada w es un BoW donde la información con la que cuenta es que cada raíz nominal (representada por un número) es distinta a otra y su frecuencia en cada w .

garles al tomarlos como cadenas isotópicas. Esto fue solucionado haciendo una “limpieza” ad hoc, retirándolas del corpus.

A continuación, presentamos a cada tópico como una nube de palabras. Este tipo de gráficos resulta muy intuitivo para establecer, por una parte, las isotopías, y por otra, la jerarquía de las palabras a la que aluden las raíces, dada por su contribución a cada tópico

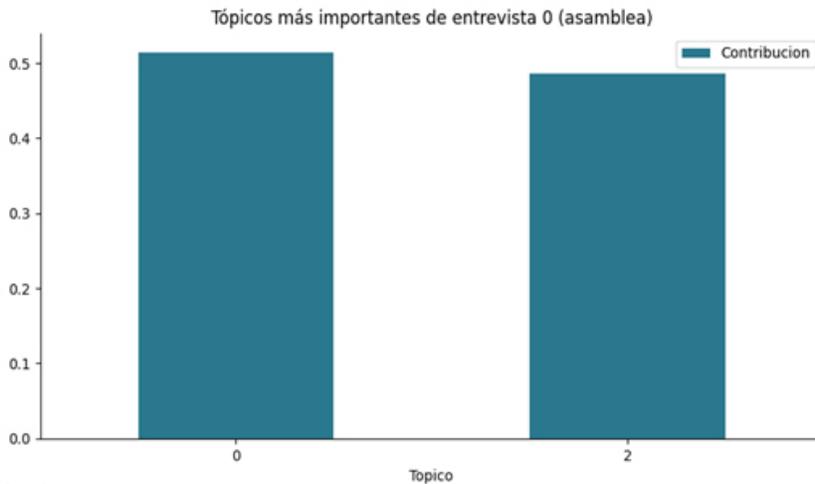
Lejos de ser arbitrarios, los tópicos tienen su fundamento en la frecuencia de las raíces nominales, y comienzan a cobrar sentido como cadenas isotópicas. Por ejemplo, para el tópico 0, dadas las probabilidades, contiene a las raíces “acamp”, “comision” y



Gráf.3.1.a

“asamble” en las siguientes proporciones: 1%, 0.08%, y 0.07%, respectivamente. Entonces, podemos decir que para el tópico 0, las palabras como acampar, acampe, comisión y asamblea, etc., en conjunto con las demás que alude el tópico, permite pensar una isotopía. Este no es el único resultado directo que devuelve el algoritmo. Además, de cada w nos dice en qué proporción está compuesto por cada tópico. Recordemos que la entrevista 0 es la asamblea. Pues para ella, tenemos el siguiente gráfico:

Se aprecia que la probabilidad de obtener el w original a partir del tópico 0 es de 0.514, y del tópico 2, 0.486. Esto quiere decir que el w que representa a la asamblea está compuesto, proporcionalmente, en un 51,4% por el tópico 0, y, en 48,6% por el tó-



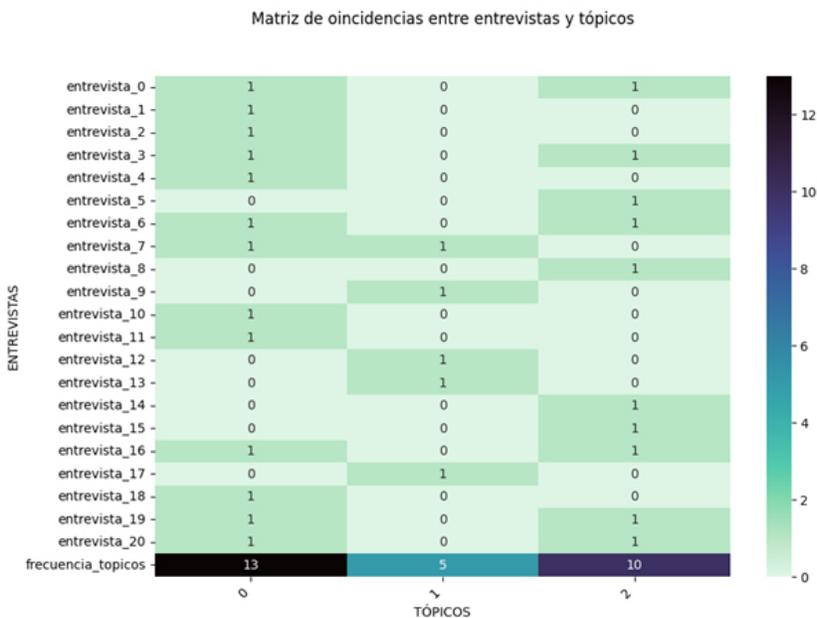
Gráf.3.1.b

pico 2. Pues, las isotopías que asociemos a estos tópicos, dado el alto porcentaje que el w contiene de ellos, las vinculamos a la asamblea. Para cada w del corpus podríamos mostrar un gráfico similar a Graf.3.1.b, pero hemos optado por el anterior dado que deviene en un refuerzo argumentativo de la siguiente sección.

3.2 Resultados generales: inferencias sobre la totalidad del corpus.

A partir de los resultados preliminares, diseñamos una matriz que relaciona la presencia (1) o ausencia (0) de un tópico con cada w para establecer su frecuencia en el corpus D :

Se observa que en la asamblea (entrevista 0) abrevan los dos tópicos principales (0 y 2) que predominan en la mayoría de las entrevistas. Además, si retomamos el Gráf.3.1.b, los dos tópicos contribuyen en considerable proporción al w asamblea. Pues,

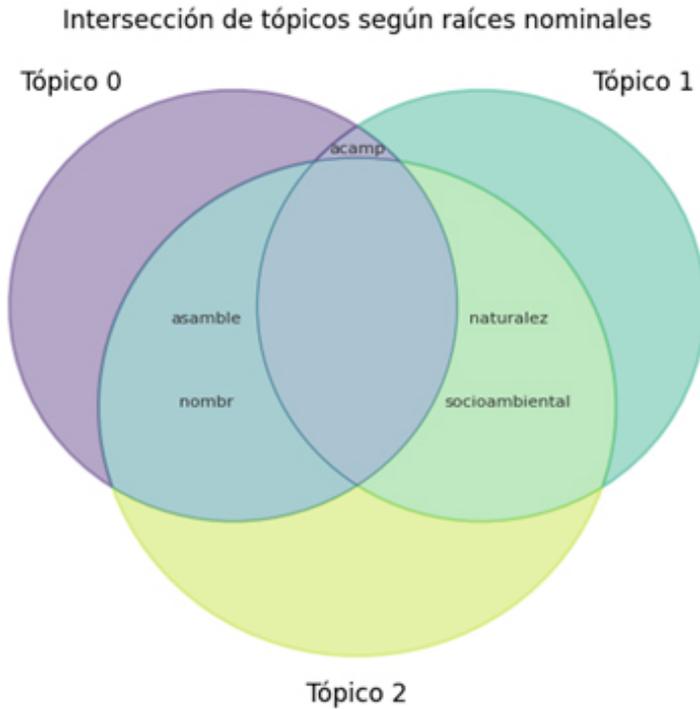


Gráf.3.2 a

queda denotado que los tópicos de la asamblea (UC), a su vez, emergen mayoritariamente en las entrevistas (UE), lo cual permite inferir que comparten un horizonte de sentido.

Determinamos 20 raíces para cada tópico, por lo que, de un total de 60 de ellas, las intersecciones entre los conjuntos se dan sólo sobre 5. Nos pareció importante indagar sobre la relación hacia el corpus *D*, particularmente en la frecuencia de aparición

Encontramos que, por ejemplo, “acamp” no aparece en 7 documentos, de los cuales uno es la asamblea. Hay un comportamiento similar con el resto de las raíces, es decir, no están contenidas en ciertos documentos. Por otra parte, la frecuencia de aparición



Gráf.3.2 b

en el corpus *D* es: “acamp”, 201; “asamble”, 180; “nombr”, 117; “socioambiental”, 103; “naturalez”, 125. Suman 726 sobre un total de 31547 del corpus . La proporción es el 2.03% sobre el corpus total, lo cual, en primera instancia, pareciera ser despreciable.

Ahora bien, a falta de evidencia cuantitativa otorgamos una interpretación cualitativa. No caben dudas que “socioambiental” y “naturalez” se vinculan al ambientalismo. La relación no es tan directa para el caso de “acamp” y “asamble”, que sugieren una instrumentación política de las anteriores: formas de organización particulares, verbalizadas, que se inscriben en el discurso como acción política. La identidad se manifiesta de manera directa a través del conjunto bajo los significantes de raíz nominal

“nombr”, intersección de los tópicos 0 y 2, pero su relevancia posee otras implicancias: ¿cómo nombrar la acción política que supone discusiones acerca de la conservación, desarrollo, sociedad y naturaleza? Es a través de la interacción social, en este caso vehiculizada por la controversia ambiental, que se configura un problema de presentación e identidad del sujeto. A partir del rastreo de la funcionalidad semántica de “nombr” encontramos que se vincula a dos cuestiones: por un lado, refiere a la identidad en los avatares organizacionales del ambientalismo, mientras que, en otras ocasiones, se vincula a discursos polémicos e ideológicos (Amossy, 2016, 2017; Van Dijk, 1996). Una forma recurrente mediante la cual opera la raíz es en los procesos de nominalización (de las comisiones, de las organizaciones locales, de referentes de ONGs). Mientras que otra forma de uso aparece en función de una designación polémica del otro (“se conocen los nombres de los denunciados”; “ocultar los nombres de los grandes intereses”), sean sujetos discursivos adversariales o bien objetos discursivos con los cuales se polemiza (en el caso de la asamblea el término aparece en una ocasión de la siguiente manera: “sistema llamado capitalismo pongamos los nombres como son, se llama capitalismo”)¹¹.

3.3 Del modelado de tópicos a la construcción de isotopías

A partir de la visualización de los gráficos encontramos tres tópicos que arroja el modelo ¿qué significan estos agrupamientos? ¿de qué modo son interpretados como isotopías? La definición de isotopía semántica estricta habilita la relación en base a la recurrencia de una misma categoría. Para la construcción de una cadena isotópica es preciso reconstruir campos semánticos que

¹¹ Estas son citas textuales del corpus de entrevistas realizadas a los activistas ambientales.

corresponden a esferas de la praxis lingüística, dado que las palabras funcionan juntas, aportan sentido y amalgaman el texto. Una forma de construir un campo semántico es a través de nombrar al conjunto. De tal modo, los tópicos a los que hemos arribado a través de implementar el algoritmo LDA son susceptibles de ser nominados de la siguiente manera:

| CATEGORÍAS | | | | | |
|--------------------|-----------------------|-----------|------------|----------|-------|
| TÓPICO o | Organización Política | Identidad | Naturaleza | Economía | Otros |
| 0.010 “acamp” | X | | | | |
| 0.008 “comisio” | X | | | | |
| 0.007 “asambl” | X | | | | |
| 0.007 “multi” | X | | | | |
| 0.005 “asambl | X | | | | |
| 0.004 “animal” | | | X | | |
| 0.004 “vot” | X | | | | |
| 0.004 “autoconvoc” | X | X | | | |
| 0.004 “bander” | X | X | | | |
| 0.004 “labur” | X | | | X | |
| 0.003 “ong” | X | | | | |
| 0.003 “nombr” | X | X | | | |
| 0.003 “izquierd” | X | X | | | |
| 0.003 “apag” | | | | | X |
| 0.003 “ibam” | X | X | | | |
| 0.003 “tierr” | | | X | | |
| 0.003 “puebl” | X | X | | | |
| 0.003 “cuerp” | | X | | | |
| 0.003 “suced” | | | | | X |
| 0.003 “arte” | | | | | X |
| TOTALES | 14 | 7 | 2 | 1 | 3 |

Cuadro.3.3.a

Es evidente que las categorías podrían ser más, u otras, y responden a supuestos *a priori* del tema. Dada la predominancia de términos asociados a un campo semántico orientado a la organización política (14 de 20), y otro a identidad (7 de 20), denominamos al tópico o: organización e identidad política. El ejercicio de clasificación semántica, si bien se basa en la pertenencia de raíces nominales a categorías, conlleva un alto grado de interpretación cualitativa a partir del trabajo de campo y la realización de

entrevistas. Es por ello, que el ejercicio constituye una intersección entre interpretaciones cuantitativas y cualitativas.

Generar un cuadro como el Cuadro.3.3.a, es un modo de construir isotopías en función de los tópicos. También es válido recuperar raíces nominales de acuerdo a otros criterios semánticos de agrupación. Si observamos el Gráf.3.1.a, para el tópico 2, se evidencia la antinomia conservación / desarrollo, discusión relevante en el abordaje de controversias ambientales. En el gráfico encontramos: “capital”, “model”, “gestion”, “recurs”, “desarroll”, compartiendo el mismo tópico con otro subconjunto: “naturaleza”; “vida”; “exist”; “univers”. La identificación de ambos subconjuntos señala la presencia de una tensión entre los vocabularios de la conservación y el desarrollo. Por otro lado, varios términos remiten al ámbito de lo social o refieren a categorías de agrupamiento social: “soci”, “socio”, “socioambiental”; “puebl”. Pues, por derivación, nombramos al tópico 2 como: sacionaturaleza, conservación y desarrollo.

Admitimos que ha costado dar nombre al tópico 1. Los 20 términos que reunidos por el algoritmo son: ‘0.009*”xr” + 0.006*”are” + 0.006*”contact” + 0.005*”sts” + 0.004*”socioambiental” + 0.004*”naturalez” + 0.004*”plan” + 0.004*”piec” + 0.004*”practic” + 0.004*”territorial” + 0.003*”global” + 0.003*”incident” + 0.003*”acamp” + 0.003*”futur” + 0.003*”ciudadan” + 0.003*”cumpl” + 0.003*”exact” + 0.003*”present” + 0.003*”basur” + 0.003*”nombr”.

Las cuatro raíces que aportan mayoritariamente al tópico son: ‘0.009*”xr” + 0.006*”are” + 0.006*”contact” + 0.005*”sts”. Nuestra inferencia al observar el tópico es la reunión de lexemas y raíces nominales que señalan singularidades del caso en contraposición a referencias globales. Esto se aprecia en la oposición: (1) “xr”, abreviatura del movimiento ambientalista mundial Extinction Rebellion, que señala la importancia de los sujetos colectivos del ambientalismo junto a la intromisión de un sujeto global en el caso particular; y (2) “sts”, que son las siglas de la

organización local Soluciones Tecnológicas Sustentables, la cual tuvo relevancia en las intervenciones ambientales de la ciudad. También se inscribe como huella de la singularidad del caso los elementos “plan” y “piec”¹².

Hallamos otra dimensión semiótica de globalidad en el despliegue de un vocabulario del ambientalismo a escala mundial expresado en palabras como: “socioambiental”; “naturalez”; “global”; “ciudadan”; “basur”; “futur”. De modo general, los movimientos ambientalistas internacionales utilizan estas nominaciones. Respecto de “futur” es interesante no sólo vincular el término con Fridays For Future, el conocido movimiento ambiental internacionalista fundado por la activista sueca Greta Thunberg, sino que también el futuro es un argumento y lugar narrativo recurrente de las propuestas ambientalistas, cuya senda la marca, por ejemplo, el conocido informe titulado Nuestro Futuro Común (Naciones Unidas, 1987)¹³. Nuevamente, de modo antagónico, la raíz “are” corresponde a la recurrencia de “área”, bajo dos acepciones: o bien se refiere a “áreas quemadas” por el fuego, o bien a “áreas de trabajo” dentro de las organizaciones ambientalistas. Es decir, se alude o bien a una espacialidad (“territorial”) o a una preocupación de la organización, ambas cuestiones señalan particularidades en un nivel local. En síntesis, nuevamente, la tensión está dada por una dualidad: mientras que algunos lexemas refieren a una localidad, encontramos términos pertenecientes a un interdiscurso global.

Por último, emerge un campo semántico vinculado a la urbanidad: “ciudadan”; “basur”; “global”. Si bien resulta evidente asociar palabras como basura y ciudadano a lo urbano, no pareciera ser tan directa la relación con “global”. Sin embargo, según

12 El Plan Integral Estratégico para la Conservación y Aprovechamiento Sustentable del Delta del Paraná (PIECAS - DP) es un documento de política pública local creado como corolario de los grandes incendios del 2008. El mismo se volvió a poner en discusión en los incendios del 2020 junto con una multiplicidad de proyectos de Ley de protección de humedales.

13 El documento es conocido también como informe Brundtland.

el Banco Mundial se estima que para el año 2050 el 70% de la población del planeta vivirá en ciudades, por lo que, mayoritariamente, lo global será urbano¹⁴. Finalmente, la emergencia de “nombr” en este tópico no cesa de llamarnos la atención respecto de la importancia del nombre en la manifestación política y la organización colectiva, ya que es sabido que quien no tiene un nombre, no tiene voz ni logos en la *querella* pública (Rancière, 2000; 2007).

Por todo ello, llamamos al tópico: ambientalismo glocal¹⁵ urbano.

Conclusión:

Aplicamos la metodología cuantitativa LDA desarrollada por Blei (2003) en función de examinar singularidades de la controversia ambiental a partir del fuego en las islas del Río Paraná desde el año 2020. La posibilidad de tensionar las prenociones con el modelo devino en una validación de las reflexiones respecto del problema, pero además, produjo algo nuevo. Sabíamos de la relevancia de la organización e identidad política para los grupos movilizados como así también que los discursos de la biodiversidad atraviesan el conflicto en cuestión. Los hallazgos fueron metodológicos y de interpretación de resultados que trascendieron a la hipótesis y objetivo planteados, lo cual configura un campo de prospecciones que hace avanzar la investigación.

En primer lugar, interesa destacar el movimiento metodológico. Recordemos que el objetivo principal del artículo era indagar

14 Estos datos son recuperados del sitio web oficial del Banco Mundial (<https://www.bancomundial.org/es/topic/urbandevelopment/overview#:~:text=En%20la%20actualidad%2C%20alrededor%20del,de%20habitantes%20%80%94%20vive%20en%20ciudades.>).

15 Recordemos que glocal es un neologismo acuñado por Escobar (1999) que condensa la referencia a lo local, a lo global y la mutua dependencia de tales designaciones.

sobre LDA como metodología cuantitativa y establecer su complementariedad con las cualitativas. El contrapunto se aprecia tras la lectura del artículo: comenzamos por el trabajo de campo, las entrevistas y la observación participante (metodología cualitativa). Luego aplicamos el modelado de tópicos (metodología cuantitativa). Por último, establecimos las isotopías como giro de interpretación. Por otra parte, los resultados obtenidos se conciben con la hipótesis, dado que habilitan isotopías en alusión a la organización e identidad política, las formas organizacionales del ambientalismo, definiciones de la identidad colectiva y a los discursos de la biodiversidad. Pues, como primera conclusión afirmamos que el ejercicio realizado permitió delinear un tipo de producto social que podemos denominar discurso ambiental. Hablar de discurso ambiental, implica suponer que posee especificidades dentro de un espacio de interdiscurso a través del cual activa (y antagoniza) con otros (el político, socioambiental, conservacionista, desarrollista); pero además, mediante la construcción de isotopías se establece de qué modo se escenifica en el caso estudiado.

Por otra parte, obtuvimos resultados generales. Uno de ellos fue aislar raíces nominales comunes a los tópicos que han devenido en isotopías. No hallamos evidencia cuantitativa para establecer la relación dado que su proporción es pequeña respecto de la totalidad del corpus. Sin embargo, la intersección lógica de conjuntos habilitó una interpretación: las palabras asociadas a las raíces nominales: “acamp”, “asamble”, “nombr”, “socioambiental” y “naturalez” (Gráfico 3.1.d), son vasos comunicantes de sentido, una suerte de vascularidad semántica que atraviesa las diferentes isotopías (y tópicos). Tales términos caracterizan al caso particular como un emergente situado del discurso ambiental. Estas palabras conforman una macro isotopía y constituyen una mayor condensación, un acervo que podríamos extraer del caso dentro del discurso ambiental.

Otro de los resultados generales surgió de la distinción me-

todológica asamblea / entrevista, en la cual identificamos una distinción *a priori* entre lo público y lo privado. Encontramos una serie de oposiciones y restricciones diferenciales por la escena de discurso entre las entrevistas (dialogales / privadas) y la asamblea (grupal / pública) que nos condujo a considerarlas UE y UC, respectivamente. La división entre UC y UE suponía una separación. Por el contrario, encontramos una continuidad: tanto las entrevistas como la asamblea comparten un horizonte de sentido: ¿qué se deduce de esto? En primera instancia, la inscripción de enunciaciones públicas en entrevistas desplegadas en un ámbito privado. Es decir, que los actores recuperan problemáticas colectivas en la definición de sus percepciones acerca de los sucesos y ello supone una investidura pública en lo privado.

El ejercicio de asignación de nombre a los tópicos mediante la construcción de isotopías es el resultado de un esfuerzo cualitativo y cuantitativo que proyecta líneas de investigación vinculadas a diferentes problemáticas de la antropología y la ecología política. Tales isotopías son manifestaciones particulares del discurso ambiental que pueden ser interpretadas a través de la teoría social, dimensión relevante que será retomada en el trabajo de investigación general.

A modo de corolario, los nombres de las isotopías reconstruidas son:

- A) Organización e identidad política (tópico 0)
- B) Ambientalismo glocal urbano (tópico 1).
- A) Socionaturaleza, conservación y desarrollo (tópico 2).

Finalmente, mencionamos una dificultad que es imprescindible precisar si quisiéramos aplicar nuevamente la metodología empleada. Al implementar LDA con algoritmos asociados a la pre-

paración del corpus, como el proceso de *stemming* y *stopwords*, notamos que la herramienta presenta dificultades al procesar texto hablado. LDA funciona mejor en base a texto escrito que no fue originado en la oralidad, lo cual resulta lógico: la reiteración y frecuencia de palabras disminuyen considerablemente respecto del habla. También varía el tamaño del vocabulario, dado que las lenguas que han desarrollado escritura presentan la posibilidad de una mayor variedad de palabras. Por otra parte, si bien el corpus estudiado es extenso, resulta abordable para una sinopsis de lectura tradicional. La potencialidad de LDA está dada justamente para investigar corpus que contemplen miles de documentos textuales escritos. Pues, figuramos como proyecciones de trabajo de la investigación aplicar LDA a partes del corpus proveniente de fuentes escritas, documentos orientados a la formulación de políticas públicas y rastreo de noticias digitales.

Bibliografía

- Alimonda, H. (2011). *La Naturaleza colonizada. Ecología política y minería en América Latina*. Buenos Aires, Argentina: CLACSO.
- Amossy, R. (2017). *Apología de la polémica*, Ciudad Autónoma de Bs. As, Argentina: 1° ed. Editorial Prometeo.
- (2016). Por una retórica del dissensus. Las funciones de la polémica. En A. Montero (Ed): *El análisis del discurso polémico: disputas, querellas y controversias*. Buenos Aires, Argentina: Editorial Prometeo.
- Aranda, L. (2022). Incendios en el Delta del Paraná: hubo más de 15.000 personas en el puente Rosario-Victoria. *Clarín*. Recuperado de: https://www.clarin.com/sociedad/incendios-delta-parana-15-000-personas-puente-rosario-victoria_0_7956zQnBly.html
- Astelarra, S. [et al.] (Ed.) (2022). *Problemáticas socioculturales del delta del*

río Paraná: enfoques desde las ciencias sociales. Buenos Aires, Argentina: Editorial Teseo Press.

- Beltrán Costa y Santamarina, (2016). Antropología de la conservación. Balance y Perspectivas. En *Revista de Antropología Social*, 25(1), Ediciones Complutense. Pp. 85-109. Recuperado de: https://www.academia.edu/25838575/Antropolog%C3%ADa_de_la_Conservaci%C3%B3n_en_Espa%C3%B1a._Balance_y_perspectivas
- Blei; Andrew Y. NG & Michael I. Jordan (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3.
- Castro, H. (2011) Naturaleza y ambiente. Significados en contexto. En Raquel Gurevich (Ed.) *Ambiente y educacion: una apuesta al futuro* (pp. 43-71). - Buenos Aires: 1ª ed. Paidós.
- Charaudeau, P. Y Maingueneau, D. (2005) *Diccionario de Análisis del Discurso.* Buenos Aires, Argentina: Amorrortu Editores, 1 ed.
- De Sousa Santos, B (2006).

La Sociología de las Ausencias y la Sociología de las Emergencias: para una ecología de saberes. Capítulo I. En *Renovar la teoría crítica y reinventar la emancipación social (encuentros en Buenos Aires)*. ISBN 987-1183-57-7. Recuperado de: <http://bibliotecavirtual.clacso.org.ar/ar/libros/edicion/santos/Capitulo%20I.pdf>

- Escobar, A. (1999) *El final del Salvaje: naturaleza, cultura y política en la Antropología Contemporánea*, Bogotá, Colombia: Editorial CEREC.
- (1998) *La invención del Tercer Mundo. Construcción y deconstrucción del desarrollo.* Caracas, Venezuela: 1ra.. edición Fundación Editorial El Perro y La Rana, Traducción de Diana Ochoa.
- Ferrero, B. (2019). *Islas de naturaleza. Perspectivas antropológicas sobre las políticas de conservación* (comp.) Rafaela: UNRaf Ediciones.
- Gutiérrez, R. A., & Isuani, F. (2014). La emergencia del ambientalismo estatal y social en Argentina. *Rev. Adm.*

- Pública* — Rio de Janeiro 48 (2), 295-322, Recuperado de: <http://dx.doi.org/10.1590/0034-761217005>
- Jurafsky D. y Martin J.H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. University of Colorado at Boulder, EEUU. Recuperado de: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
 - Latour, B. (2007). *Nunca fuimos modernos. Ensayo de Antropología Simétrica*. Buenos Aires, Argentina: 1ª Ed. S. XXI, Editores Argentina, 224 Pp.; 21 x 14 cm.
 - (2001) *La esperanza de Pandora. Ensayo sobre la Realidad en los Estudios de la Ciencia*. Barcelona, España: 1º Ed. Gedisa.
 - Lefebvre, H. (2013) *La producción social del espacio*. Madrid, España: Capitán Swing Libros, S. L.
 - Leff, E. (2003). La ecología política en América Latina. Un campo en construcción. *Rev. Sociedade e Estado*, Brasilia, v. 18, Pp. 17-40
 - (2004) *Ecología Política y Saber Ambiental*, en Leff “*Racionalidad Ambiental. La reapropiación social de la naturaleza*”, México: Editorial Siglo XXI.
 - Leite Lopes, J. S. (2006). Sobre procesos de ‘ambientalización’ de los conflictos y sobre los dilemas de la participación. *Horizontes Antropológicos*, 12(25), 31–64.
 - Martínez Alier, J. (2006). *El ecologismo de los pobres: Conflictos Ambientales y Lenguajes de Valoración*, Barcelona: Icaria/FLACSO, .
 - Merlinsky, G. (2013). *Cartografías del conflicto ambiental en Argentina*, (comp) - Ciudad Autónoma de Buenos Aires, Argentina: 1a ed. Fundación CICCUS. 320 p. ISBN 978-987-693-033-8.
 - (2022). *Toda ecología es política*. Buenos Aires, Argentina: Siglo XXI.
 - Maingueneau, D. (2004) *¿Situación de enunciación o situación de comunicación?* - Université Paris XII, Francia. Traducción realizada por Laura Miñones (UBA), 3(5).

- Montero, A. S. (Ed.) (2016). *El análisis del discurso polémico. Disputas, querellas y controversias*, (comp.). Buenos Aires, Argentina: Editorial Prometeo.
- Mohammed J. Zaki, Wagner Meira (2020). *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Jr. Cambridge University Press. 2do Edition. Recuperado de: https://pzs.dstu.dp.ua/DataMining/bibl/mohammed_j_zaki_wagner_meira_jr_data_mining_and_analysis_fun.pdf
- Naciones Unidas (1987). *Nuestro futuro común*. Recuperado de: https://www.ecominga.uqam.ca/PDF/BIBLIOGRAPHIE/GUIDE_LECTURE_1/CMMAD-Informe-Comision-Brundtland-sobre-Medio-Ambiente-Desarrollo.pdf
- Nogueira, (2016). ¿El ambiente como cuestión? Algunas reflexiones desde el análisis de políticas públicas para el caso argentino (2003-2015). *Gestión y Ambiente* 19(2), 273-288.
- Preiti, F. & Ferrero, B. (2023) Introducción al Dossier Ambiente, Territorio y Sociedad. En *Revista de la Escuela de Antropología*, 23(01), pp. 1-8 DOI 10.35305/rea.viXXXIII.283
- Preiti & Youkhana, E. (2023). Environmental Movements in Insurgent Territories. An anthropological framework to study the performance art in Rosario, Argentina. *Kult-ur. Revista Interdisciplinaria sobre la cultura de la ciudad*, 10 (20). Recuperado de: <https://www.e-revistas.uji.es/index.php/kult-ur/article/view/7538>
- Preiti, F. (2022) Multitudes sobre el puente. Reflexiones antropológicas acerca del ambientalismo social en torno del fuego en las islas del delta durante el año 2020. En Astelarra Et. al. (Ed.) *Problemáticas socioculturales del delta del río Paraná: enfoques desde las ciencias sociales*. Recuperado de: <https://www.teseopress.com/roblematicassocioculturalesdeldeltadelrioparana/>
- (2021a). Patria o fuego. Análisis discursivo de una Carta Abierta en torno de incendios en islas del Delta paranaense (año 2020). *Revista Uruguaya de Antropología y Etnografía*, vol. 6, n.º 1, 89-98. Recuperado

- de: <https://ri.conicet.gov.ar/handle/11336/167325>
- (2021b) Escenas del discurso mediático en torno de incendios en el Delta del Río Paraná (año 2020) en *Actas del XII Congreso Argentino de Antropología Social (CAAS)*. La Plata, junio, julio y septiembre de 2021. Recuperado de: <https://sedici.unlp.edu.ar/handle/10915/131505>
 - Rancière, J. (2000). Política, identificación y subjetivación. En B. Ardití (Ed.), *El reverso de la diferencia* (pp. 219-248). Caracas, Venezuela: Nueva Sociedad
 - (2007). *El odio a la democracia.*, Buenos Aires, Argentina: Amorrortu.
 - Rosati, G. (2022) Procesamiento de Lenguaje Natural aplicado a las ciencias sociales. Detección de tópicos en letras de tango. *Revista Latinoamericana de Metodología de la Investigación Social*, N° 23. Año 12, (pp. 38-60). Recuperado de: http://relmis.com.ar/ojs/index.php/relmis/article/view/procesamiento_lenguaje_natural/217
 - Svampa, M. (2013). Consenso de los Commodities y lenguajes de valoración en América Latina. Ciudad Autónoma de Buenos Aires: *Nueva Sociedad* N° 244.
 - Swyngedouw, E. (2011) La Naturaleza no existe. La sostenibilidad como síntoma. *Urban. Revista del Departamento de Urbanística y Ordenación del Territorio, Vol. 1*, Recuperado de: <https://polired.upm.es/index.php/urban/issue/view/106>
 - Telias, (2014) *La educación ambiental en la Argentina: actores, conflictos y políticas públicas* / Aldana Telias; María Laura Canciani; Pablo Sessano. Buenos Aires: - 1a ed. - San Fernando: La Bicicleta”
 - Thompson, J. (2011) Los límites cambiantes de la vida pública y la privada. *Nueva época*, núm. 15, enero-junio, pp. 11-42. ISSN 0188-252x.
 - Vaccaro, I., Paquet, P. A. y Beltrán Costa, O. (2012) *Antropología de la conservación. Naturaleza, Estado, mercado y cultura.* Santamarina, B. (coord.):

Geopolíticas patrimoniales. De culturas, naturalezas e inmaterialidades. Una mirada etnográfica, Publisher:

Germanies, pp. 29-53.

- Van Dijk, A T. (1996). *Análisis del discurso ideológico*. México: UAM-X, Versión 6; Pp.15-43.

Datos de los autores:

Lic. Francisco José Preiti

Antropología, discurso y ambiente

Escuela de Antropología (UNR) / CONICET

Santiago 557 (2000 Rosario)

franciskojo03@gmail.com

Tel. +54 9 3415 94-9238

ORCID: 0000-0003-1835-2533

15/04/2024

Lic. Francisco Javier Alomar

Antropología, discurso e inteligencia artificial

Escuela de Antropología (UNR)

Tecnicatura Universitaria en

Inteligencia Artificial (UNR)

9 de julio 3261 (2000 Rosario)

franciscojalomar@gmail.com

Tel. +54 9 3413 54-4281

ORCID: 0009-0001-3010-7533

15/04/2024

Registro Bibliografico:

Lic. Francisco José Preiti y

Lic. Francisco Javier Alomar.

“Modelado de tópicos (LDA) aplicado a una controversia ambiental: Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina.” en *La Trama de la Comunicación*, Vol. 28 Número 1, Anuario del Departamento de Ciencias de la Comunicación. Facultad de Ciencia Política y Relaciones Internacionales, Universidad Nacional de Rosario. Rosario, Argentina. UNR Editora, julio a diciembre de 2023 p. 080-117.

ISSN 2314-2634 (en línea).

Recibido: 15/04/2024

Aceptado: 24/7/2024

La Trama de la Comunicación - Volumen 28 Número 1 - enero a junio de 2024

ISSN 2314-2634 (en línea) / P. 080-117

Modelado de tópicos (LDA) aplicado a una controversia ambiental. Isotopías del discurso ambiental sobre el fuego en islas del Paraná, Argentina

Francisco José Preiti y Francisco Javier Alomar