

Condiciones para el estudio legislativo: un análisis de la disponibilidad documental en el Congreso del Estado de México

Artículos académicos

Itzel Cruz Pérez¹
Comecyt

 <https://orcid.org/0000-0003-4931-8178>

Resumen: El análisis computacional de textos se ha consolidado en la ciencia política como una herramienta fundamental; sin embargo, su aprovechamiento puede enfrentar barreras significativas, como la escasez de archivos históricos o la falta de estandarización documental, entre otras. Este trabajo aborda esta brecha, examinando la infraestructura de datos legislativos del Congreso del Estado de México. Mediante un análisis documental exploratorio, se identifica la existencia de diversos textos, pero se revelan tres obstáculos cruciales para la investigación empírica y la transparencia efectiva: 1) la restricción de la información a solo la del periodo de la legislatura actual, sin que esté disponible la de legislaturas anteriores; 2) la inconsistencia y falta de homologación de los formatos, y 3) la prevalencia de documentos no legibles por máquina (que impiden el análisis automatizado).

Palabras clave: Poder Legislativo; Datos abiertos vinculados; Gobierno abierto.

Recepción: 17/10/2025

Aprobación: 09/02/2026

<https://doi.org/10.67555/tlv12.4>



Esta contribución se encuentra publicada bajo una licencia Creative Commons Atribución/Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional

SOBRE LA IMPORTANCIA DE LOS TEXTOS LEGISLATIVOS

Las legislaturas, también llamadas, de acuerdo con el país, como congresos, parlamentos, asambleas o cortes son espacios vitales de nuestras democracias. En los ámbitos subnacional, nacional e, incluso, transnacional (como el Parlamento Europeo), se encuentran presentes como elemento constitutivo de regímenes democráticos. En México, podemos contabilizar 32 congresos locales y un Congreso de la Unión (conformado por las cámaras de diputados y de senadores), cada uno con sus propios elementos característicos, desde su conformación, hasta su manera de legislar.

Entre sus funciones puede mencionarse que son responsables de proponer o enmendar proyectos de ley; debatir y, potencialmente, moldear una diversidad de políticas; aprobar los presupuestos, y aprobar o ratificar el nombramiento de ciertos funcionarios públicos, entre otras. Para realizar estas funciones se comunican públicamente entre sí y con la ciudadanía; sus integrantes pronuncian discursos con el

1. Doctora en Marketing Político, Actores e Instituciones de la Sociedades Contemporáneas y máster en Estudios Internacionales por la Universidad de Santiago de Compostela, España; licenciada en Ciencias Políticas y Administración Pública por la Universidad Autónoma del Estado de México. Investigadora del Consejo Mexiquense de Ciencia y Tecnología (Comecyt) adscrita al Instituto de Estudios Legislativos del Congreso del Estado de México (INESLE), y profesora de asignatura en la Universidad Autónoma del Estado de México.
Correo electrónico: itzelcruzp@outlook.com

objetivo de posicionarse sobre líneas divisorias amplias o iniciativas específicas (Borghetto y Chaqués-Bonafont, 2019), redactan documentos y publican informes de las comisiones, así como informes especiales, comunicados de prensa, etcétera.

Por ello es posible suponer que en México las legislaturas locales y la federal producen millones de documentos año con año. En otros países, como el Reino Unido, se ha identificado que, por ejemplo, entre 1988 y 2019 los parlamentarios de la Cámara de los Comunes pronunciaron casi dos millones de discursos con casi 370 millones de palabras (Rauh et al., 2020, como se citó en Sebök et al., 2023). Sin lugar a dudas, de manera amplia la labor legislativa genera un vasto corpus de textos que no solo expresa la cantidad de trabajo diario, sino que “encapsula las prioridades, posturas y decisiones políticas que mueven y estructuran las sociedades democráticas” (Slapin y Proksch, 2014), por lo que, cada vez más, los documentos legislativos han atraído la atención de la comunidad académica.

Por supuesto, no se trata de un fenómeno nuevo, los estudiosos han reconocido desde hace mucho tiempo que la política y el conflicto político a menudo ocurren en la palabra escrita y hablada. Como señalan Grimmer y Stewart (2013), los candidatos individuales y los partidos políticos articulan sus puntos de vista mediante plataformas y programas políticos; una vez elegidos, los y las representantes escriben y debaten la legislación; después de que se aprueban las leyes, la burocracia solicita comentarios antes de emitir regulaciones. Estos ejemplos y muchos otros muestran que para entender de qué se trata la política necesitamos saber qué dicen y escriben los actores políticos.

En la ciencia política, el análisis de contenido se ha posicionado como una herra-

mienta poderosa que se utiliza fundamentalmente para analizar textos políticos y posicionar a la élite política en el debate político. El *texto político* es una categoría global que puede incluir textos generados por políticos individuales y por partidos. Los primeros incluyen, por ejemplo, entrevistas, debates parlamentarios y discursos del debate electoral, mientras que los segundos son documentos que representan al partido en su conjunto e incluyen comunicados de prensa y panfletos. Pero los documentos que han recibido una amplia atención por parte de los investigadores han sido los correspondientes a los procesos electorales, conocidos como programas electorales, plataformas o manifiestos.

Hasta hace no mucho tiempo la extracción de información de voluminosos corpus textuales se limitaba a análisis manuales costosos y de alcance limitado. Uno de los mayores retos para los investigadores era el volumen y el costo masivo que implicaba el análisis de textos; incluso, colecciones de textos de tamaño moderado. Raramente los estudiosos pueden leer en formato impreso todos los textos de un corpus (Grimmer y Stewart, 2013). Ante esta situación, emergieron propuestas como el *Manifesto Project* (Lehmann et al., 2025) que, desde su creación en 1979, se embarcó en la ambiciosa tarea de recopilar y codificar manualmente los programas electorales de los principales partidos de diferentes regiones del mundo. Con los programas electorales convertidos en datos se han abordado más de 85 temas de investigación en relación con los partidos políticos, como la competición partidista, representación, posicionamiento ideológico, polarización, relevancia de temas, etcétera (ver más en Volkens et al., 2015).

En los últimos años, el análisis de textos políticos ha experimentado un crecimiento exponencial en el campo de la ciencia

política, impulsado por los avances en el poder computacional, las técnicas de aprendizaje automatizado y los métodos estadísticos (Slapin y Proksch, 2014). Con estos avances se han desarrollado herramientas de análisis y clasificación de contenido automatizado (Jurafsky y Martin, 2000), abriendo fronteras para la investigación empírica, pero también hacia nuevos enfoques teóricos y respecto a los textos legislativos. En este sentido, un ámbito en donde se ha ampliado el interés académico es precisamente en los documentos que se generan desde las legislaturas (p.ej. Bräuninger, Debus y Wüst, 2015; Kiss y Sebők, 2022; Slapin y Proksch, 2014).

En parte, el gran corpus de textos que se desarrollaba en las legislaturas quedaba sin explorar debido a la abrumadora tarea de compilarlo en un formato utilizable; la falta de técnicas de análisis de contenido, de personal calificado y de financiación para el procesamiento de archivos más antiguos, así como la poca disponibilidad de documentos legibles por máquina, etcétera.

No obstante, ahora las legislaturas también han favorecido que los registros parlamentarios se hagan electrónicos, almacenando bases de datos en línea de fácil acceso, volviendo así más accesible la disponibilidad para los parlamentarios, investigadores y sociedad en general. Un ejemplo es la División de Documentación Parlamentaria de Alemania, que cataloga e indiza los materiales de acceso público del Bundestag y el Bundesrat alemanes según aspectos formales y temáticos (incluidas las actas plenarias, los proyectos de ley, las mociones, las interpelaciones y varios informes) (Kiss y Sebők, 2022).

Además de la disponibilidad desde las legislaturas para el acceso a los datos, se han hecho importantes esfuerzos desde las universidades y centros de investigación, especialmente en países con democracias consideradas avanzadas que han demostrado el potencial que ofrecen los métodos de texto como datos para comprender el funcionamiento interno de las democracias. Algunos proyectos de relevancia son el Observatorio de Textos Políticos en las Democracias Europeas (OPTED)², financiado por la Unión Europea y que reúne a 17 instituciones de investigación de Europa, buscando sentar las bases para una infraestructura de análisis de textos políticos (múltiples y de diferentes actores) en Europa; y ParlLawSpeech³, una base de datos integrada y multilingüe de discursos y legislación parlamentaria, que se ha publicado como una base de datos de código abierto.

Gracias a estos esfuerzos, principalmente en Europa, de acuerdo con Slapin y Proksch (2014) los académicos han utilizado el análisis de contenido para examinar al menos siete características de las legislaturas: ideología partidaria y polarización, posicionamiento del gobierno, escrutinio parlamentario, representación basada en circunscripciones, agenda política, calidad del debate y el papel de los medios. Asimismo, ha permitido plantear nuevas preguntas y responder a otras cuya solución antes no era posible e, incluso, cubrir periodos más largos y un mayor número de casos para la investigación comparada. Como sugieren Sebők *et al.* (2023), entre las preguntas de investigación que este enfoque puede responder se encuentran: ¿Qué temas políticos se abordan con mayor frecuencia en los discursos y leyes? ¿Cómo evoluciona la agenda política del Congreso a lo largo del tiempo? ¿Cómo se posicionan ideológicamente los legisladores en el espacio político? ¿Cuál es la tonalidad de la comunicación de los le-

²Consultar más en: <https://opted.eu/about/>

³Consultar más en: <https://parllawspeech.org/research/>

gisladores? ¿Cómo se manifiesta el conflicto entre gobierno y oposición en los debates legislativos? ¿Qué estrategias retóricas utilizan los legisladores para la movilización o persuasión? ¿Cuánto influye el poder legislativo en la elaboración de leyes?, entre muchas otras y que los investigadores ya han empezado a abordar con base en la evidencia empírica (p.ej. Greene y Cross, 2017; Quinn *et al.*, sin fecha).

No obstante, esta relevancia del análisis de contenido de los documentos legislativos no ha recibido la misma atención en otras regiones del mundo, particularmente en América Latina. Si como identifican Crisp y Schibber (2014) en la región latinoamericana se ha dado un aumento acelerado respecto a la política legislativa, acumulando grandes cantidades de investigación sistemática y empírica, esta se ha centrado sobre todo en la relación entre personas legisladoras y sus electores y las relaciones Legislativo-Ejecutivo, entre otros.

En parte, esto se debe a los problemas que se identificaron también en otras partes del mundo en otros momentos. Por ejemplo, la implementación de sistemas de información para la creación, almacenamiento, recuperación y difusión fiables de documentos legislativos. En los casos en que los datos son accesibles, sus formatos pueden no ser necesariamente útiles para los análisis computacionales; incluso las legislaturas pueden modificar la estructura de sus bases de datos dependiendo de los intereses de los gobiernos. Por lo tanto, las limitaciones de los datos obligan regularmente a los académicos a acotar el alcance de sus trabajos.

Por lo tanto, dada la relevancia del tema de los documentos legislativos en el campo de la investigación en ciencia política, este trabajo se plantea como objetivo identificar elementos relacionados con la actividad del

Congreso del Estado de México. Concretamente, se busca explorar cuáles son los documentos que se encuentran disponibles en la página *web* del Congreso, cuáles son las características de los documentos disponibles y cómo está organizada la información documental en el sitio *web*.

DATOS

Se recurrió al análisis documental, un proceso de construcción del conocimiento en el que los insumos de trabajo son principalmente documentos escritos (Marcelino, Martínez y Camacho, 2024). Esencialmente, el análisis documental busca generar nueva información o encontrar la respuesta a una interrogante de forma coherente y argumentada. Particularmente, se ha señalado que el análisis documental está constituido por varias operaciones fundamentales y diferentes (Castillo, 2004). Por un lado, puede ser una descripción física (bibliográfica/documental, análisis formal/externo); por otro, se realiza un análisis de contenidos (indización, resumen, clasificación). Este trabajo se enfoca en el primer elemento.

El análisis documental se realizó en varias fases. La primera, esencialmente exploratoria, consistió en identificar dónde se registran los documentos que emanan del trabajo legislativo del Congreso del Estado de México. Se identificó que existe un área particular dentro del Congreso que organiza y concentra la gran mayoría de textos legislativos, llamada Secretaría de Asuntos Parlamentarios⁴, la cual tiene como misión “Proveer en lo que a ella compete, lo necesario para un trabajo legislativo eficiente; y con ello permitir que las Asambleas cumplan con su cometido de hacer leyes de calidad”.

⁴Para la identificación del área, se puede consultar: <https://www.secretariadeasuntosparlamentarios.gob.mx/secretaria>

En la página *web* de la Secretaría de Asuntos Parlamentarios se identifican dos apartados que contienen varios tipos de documentos. Un primer apartado, denominado “Trabajo Legislativo”, y un segundo apartado denominado “Órganos de Difusión”. En el apartado de Trabajo Legislativo se encuentran las sesiones, las iniciativas de ley o decreto, los puntos de acuerdo, minutas, decretos y acuerdos, mientras que en el apartado de Órganos de Difusión se encuentran el Diario de Debates y las gacetas parlamentarias.

Una vez que se identificaron los apartados y los tipos de documentos, se analizaron las características de estos, identificando el periodo que abarca la información disponible de documentos legislativos dentro de la página *web*, la cantidad de páginas, el tipo de texto (formato de documento portátil [PDF] o documentos de procesamiento de textos) y si es legible a máquina. En esta fase de análisis, el primer aspecto que llama la atención es que los documentos disponibles solamente cubren la legislatura actual, es decir, no es posible acceder a documentos de otros periodos. La LXII Legislatura inició el 5 de septiembre del 2024 y concluirá el 4 de septiembre del 2027. Esta fase de análisis permitió identificar concretamente los datos que se presentan en la Tabla 1.

Tabla 1. Textos legislativos identificados en el Congreso del Estado de México

Tipo de documento	Periodo disponible	Total de documentos	Número total de páginas	Formato disponible
Gacetas	12/09/2024 al 5/10/2025	78	11 164	PDF
Iniciativas	5/09/2024 al 1/10/2025	806	8 024	Mixto
Puntos de acuerdo	4/09/2024 al 1/10/2025	119	759	Mixto
Acuerdos	4/09/2024 al 1/10/2025	78	684	Mixto
Decretos	5/09/2024 al 24/09/2025	175	2 301	Mixto
Diarios de debates	4/09/2024 al 10/12/2024	23 (7 tomos)	3 148	Mixto

Fuente: elaboración propia.
Nota: el análisis se realizó entre septiembre y octubre de 2025, por lo que los datos reflejan la disponibilidad hasta ese momento.

RESULTADOS

Después de haber identificado los elementos del análisis documental respecto a los tipos de documentos, el periodo que abarcan y sus características, se observan varios elementos importantes. Comenzando con los tipos de documentos, encontramos hasta seis tipos. Las

gacetas parlamentarias se encuentran entre los documentos más importantes del Congreso, pues son el medio por el cual las leyes, decretos y demás disposiciones entran en vigor y se dan a conocer a la ciudadanía. Se identificaron en total 78 gacetas, las cuales varían en extensión (hay gacetas de cinco páginas, hasta gacetas de 700 páginas), contabilizando en conjunto 11 164 páginas. Son documentos que cuentan con cierta estructura, donde se establece un índice que especifica los asuntos tratados por la legislatura y el periodo que abarca; concentran diferentes tipos de textos, desde los órdenes del día, acuerdos, proyectos de decreto, etcétera, por lo que su extensión es muy variada. En su mayoría son formatos PDF que además de ser un formato de archivo propietario, no son legibles a máquina.

Respecto a las iniciativas, en el periodo disponible, que abarca un año, se encontraron 806. La extensión también es variada; se identificaron documentos con una página y otros hasta con 211 páginas. La cuestión es que muchas veces solo se registran documentos de una página donde se identifican las firmas, pero no el contenido completo. Además, no existe congruencia en cuanto al formato de los documentos registrados en el sistema, algunos son documentos de procesamiento de textos (derivados del *software* propietario Microsoft Word), que se descargan automáticamente, y otros en PDF.

Por otro lado, los puntos de acuerdo y acuerdos se encuentran en un mismo apartado. En la mayoría de los casos se localizan los dos tipos de documentos respecto al número de acuerdo, pero en ocasiones solo se identifica uno de ellos, ya sea el punto de acuerdo o el acuerdo, por lo que existen variaciones en el total de documentos. De los puntos de acuerdo se identificaron 119 documentos; su exten-

sión es más reducida (el documento más amplio tiene 41 páginas, mientras que la mayoría es de menos de 10 páginas).

En cuanto a los acuerdos, se identificaron 78. Las características de estos documentos también son confusas, ya que hay documentos de una página (la mayoría), hasta acuerdos de 118 páginas. El contenido también varía y no existe un formato homologado.

En lo relativo a los decretos, se identificaron 175 documentos, los cuales sí tienen un formato más estandarizado y se presentan en PDF; la mayoría son legibles a máquina. Además, en la página web se identifica una organización más clara de la información, al especificarse a los autores y las iniciativas en las que se base el decreto; incluso se agrega la información de la votación.

Por último, respecto al Diario de Debates se identificaron 23 documentos que se concentran en siete tomos. No obstante, la información disponible apenas cubre de septiembre a diciembre del 2024. Se observa que son documentos extensos, ya que, a pesar de la poca disponibilidad, concentran 3 148 páginas.

El panorama general muestra que se generan diferentes tipos de documentos en el Congreso del Estado de México; que cuenta con un área específica para concentrarlos; que se establece dónde se puede encontrar la información, y que las gacetas parlamentarias están disponibles para el público en general de manera constante y actualizada. No obstante, respecto a los otros tipos de documentos, como las iniciativas, los puntos de acuerdo y acuerdos, los disponibles son más diversos y con formatos distintos, mientras que los documentos del Diario de Debates casi no se encuentran disponibles. Asimismo, hay

que enfatizar que la disposición limitada de los documentos no permite la investigación y comparación a lo largo del tiempo.

El análisis sobre la documentación legislativa del Congreso del Estado de México, si bien es un estudio de caso concreto, nos obliga a reflexionar sobre la importancia de los textos legislativos como pilar de la transparencia y como materia prima esencial para el avance de la ciencia política. Como ha quedado claro en este trabajo, la labor legislativa genera un vasto corpus de textos que va mucho más allá del simple registro burocrático. Estos documentos, como ya se ha enfatizado, encapsulan las prioridades, posturas y decisiones políticas que definen y estructuran nuestra sociedad democrática. Son, en esencia, la evidencia empírica de la gobernanza.

La disponibilidad y accesibilidad de este material influye o tiene impacto en varias aristas, desde el estudio de la representación, la comprensión del conflicto político y como elemento de rendición de cuentas. En este último aspecto debe destacarse que la disponibilidad del material legislativo se conecta directamente con el principio de transparencia. Una democracia exige que el trabajo de sus legisladores no solo sea público, sino que esté disponible en un formato accesible y organizado. En el caso del Congreso del Estado de México, la existencia de la Secretaría de Asuntos Parlamentarios y la publicación de la información son elementos esenciales y positivos.

No obstante, el análisis documental revela que la transparencia formal no siempre se traduce en una disponibilidad efectiva y útil para el escrutinio. Este trabajo pone el foco de atención en tres puntos clave que ameritan una profunda reflexión sobre el trabajo aún pendiente:

1. Formatos legibles por máquina. La prevalencia de documentos PDF no legibles por máquina (no analizables automáticamente) para documentos cruciales, como las gacetas parlamentarias y el Diario de Debates que concentran el mayor volumen de texto impone una barrera técnica que limita la investigación de gran escala y el análisis computacional moderno (Text-as-data).

2. Inconsistencia y homologación. La falta de congruencia en el registro de documentos como iniciativas y puntos de acuerdo (variación en formatos, textos incompletos, extensión confusa) obstaculiza la recolección sistemática de datos y añade un costo masivo de limpieza de datos para cualquier investigador.

3. Memoria histórica limitada. La restricción de información documental únicamente para el periodo de la legislatura actual limita drásticamente la posibilidad de realizar investigación comparada a largo plazo. Sin registros históricos accesibles, es imposible analizar la evolución de la agenda política, la polarización o el impacto de las políticas a lo largo de décadas. Esto perpetúa un ciclo de investigación limitada al aquí y ahora.

El camino por seguir requiere un esfuerzo consciente por parte del Congreso del Estado de México para transformar su documentación en una infraestructura de datos de alta calidad. El avance de la ciencia política en países con democracias avanzadas demuestra el enorme potencial de estos métodos para responder a preguntas fundamentales sobre la política, por lo que, para alcanzar un buen nivel de trans-

parencia y utilidad documental, se sugiere concretamente la estandarización de formatos, cobertura temporal amplia y metadatos enriquecidos. Mejorar la disponibilidad y organización de los documentos legislativos no es solo una tarea técnica; es un imperativo democrático que fortalece la rendición de cuentas, estimula la investigación empírica rigurosa y, en última instancia, mejora la calidad de la deliberación pública en el Estado de México.

Por otro lado, se pone de manifiesto la necesidad de ampliar la investigación de los textos legislativos en nuestro país y que los investigadores puedan capturar la información proporcionada por los congresos y la almacenen en formatos más permanentes y estables, mediante la creación de bases de datos para garantizar la replicabilidad de los estudios que utilizan datos de los archivos parlamentarios (Slapin y Proksch, 2014).

Es importante señalar que las limitaciones identificadas en el Estado de México no son casos aislados, sino que forman parte de un desafío también para otras entidades y otros países de América Latina. Asimismo, resaltan algunos ejemplos donde se ha puesto mucho interés en el acceso a los datos en formatos legibles por máquina, como el de Brasil, que cuenta con un portal de Publicación de Datos Abiertos⁵ impulsado desde el congreso, donde se pueden descargar datos en lenguaje XML y formato JSON, entre otras posibilidades.

Para superar estas barreras en el contexto local, resulta imperativo transitar hacia estándares internacionales y lenguajes de procesamiento documental, como

Akoma Ntoso, que facilita la estructuración semántica de los textos. Asimismo, la adopción de arquitecturas de Datos Abiertos Enlazados (Linked Open Data) permitiría no solo la lectura automatizada, sino la conexión del corpus legislativo con otras bases de datos públicas, elevando el análisis de la transparencia de una simple consulta documental a un ecosistema de rendición de cuentas inteligente y multidimensional.

Este foco de atención puede tener impacto en varios sentidos. Por un lado, en el campo de la investigación, como afirman Sebők et al. (2023), “la extracción de información sistemática de corpus textuales ofrece el mayor potencial para el futuro de los estudios legislativos comparativos empíricos”. Por otro lado, también tiene una importante implicación al favorecer directamente la transparencia, la rendición de cuentas y la confianza en las instituciones gubernamentales. Como señala Loewenberg (2011), la evidencia disponible sugiere que, a pesar de la relativa transparencia de las cámaras legislativas, son poco entendidas y apreciadas por los públicos democráticos de todo el mundo.

CONCLUSIONES

Este trabajo puso énfasis en la literatura sobre el análisis de textos legislativos con herramientas automatizadas que está creciendo en el campo de la ciencia política alrededor del mundo. Asimismo, se puso énfasis en los diferentes temas de investigación que se pueden abordar retomando los textos legislativos. Este análisis exploratorio ha cumplido su objetivo al identificar las condiciones actuales de disponibilidad documental en el Congreso del Estado de México y sus características, revelando barreras concretas para el escrutinio empírico. Se identificaron hasta seis tipos de documen-

⁵Ver en: <https://datosabertos.camara.leg.br/swagger/api.html>

tos disponibles: gacetas parlamentarias, iniciativas, puntos de acuerdo, acuerdos, decretos y el Diario de Debates. Sin embargo, se señalaron tres obstáculos críticos: la limitación de la información al periodo de la legislatura actual, la falta de homologación de los formatos y la alta prevalencia de documentos no legibles por máquina.

Estos hallazgos demuestran que, aunque existe una voluntad formal de transparencia manifestada por la Secretaría de Asuntos Parlamentarios, la disponibilidad efectiva y utilizable de la información aún enfrenta desafíos técnicos. Abordar estos retos no es solo una tarea administrativa, sino un imperativo democrático que fortalece la rendición de cuentas.

Finalmente, este texto plantea la necesidad de una investigación más profunda y sistemática. Por esta razón, es crucial trascender la descripción de la disponibilidad y sumergirse en la codificación y el análisis computacional a gran escala de este corpus. Esto permitirá avanzar en el uso de métodos de texto-como-datos para responder a las preguntas de fondo de la ciencia política en el contexto subnacional mexicano, tales como la evolución de la agenda política del Congreso, el posicionamiento ideológico de los actores o el impacto de las políticas a lo largo de un periodo más amplio.

TL

Referencias

- Borghetto, E. y Chaqués-Bonafont, L. (2019) *Comparative Policy Agendas: Theory, Tools, Data*. *Comparative Policy Agendas: Theory, Tools, Data*. 1a ed. Editado por F.R. Baumgartner, C. Breunig, y E. Grossman. Estados Unidos: Oxford University Press. Disponible en: <https://doi.org/10.1093/oso/9780198835332.003.0029>.
- Bräuninger, T., Debus, M. y Wüst, F. (2015) «Governments, Parliaments and Legislative Activity», *Political Science Research and Methods*, 5(3), pp. 529-554. Disponible en: <https://doi.org/https://doi.org/10.1017/psrm.2015.29>.
- Castillo, L. (2004) «Análisis documental». Disponible en: <https://www.uves/macass/T5.pdf>.
- Crisp, B.F. y Schibber, C.F. (2014) «The Study of Legislatures in Latin America», *The Oxford Handbook of Legislative Studies*. Oxford University Press, pp. 628-646. Disponible en: <https://doi.org/https://doi.org/10.1093/oxfordhb/9780199653010.013.0031>.
- Greene, D. y Cross, J.P. (2017) «Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach», *Political Analysis*, 25(1), pp. 77-94. Disponible en: <https://doi.org/https://doi.org/10.1017/pan.2016.7>.
- Grimmer, J. y Stewart, B.M. (2013) «Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts», *Political Analysis*, pp. 1-31. Disponible en: <https://doi.org/10.1093/pan/mps028>.
- Jurafsky, D. y Martin, J.H. (2000) *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 1st edition. Pearson Prentice Hall.
- Kiss, R. y Sebök, M. (2022) «Creating an Enhanced Infrastructure of Parliamentary Archives for Better Democratic Transparency and Legislative Research», *International Journal of Parliamentary Studies*, 2(2), pp. 278-284. Disponible en: <https://doi.org/https://doi.org/10.1163/26668912-bja10053>.
- Lehmann, P. et al. (2025) «Manifiesto Corpus». Berlin. Disponible en: <https://manifiesto-project.wzb.eu/information/documents/corpus>.
- Loewenberg, G. (2011) *On Legislatures The Puzzle of Representation*. 1a ed. New York. Disponible en: <https://doi.org/https://doi.org/10.4324/9781315633145>.
- Marcelino, M., Martínez, M. del C. y Camacho, A.D. (2024) «Análisis documental, un proceso de apropiación del conocimiento», *Revista Digital Universitaria*, 25(6). Disponible en: <https://doi.org/http://doi.org/10.22201/ceide.16076079e.2024.25.6.1>.
- Quinn, K.M. et al. (sin fecha) «How to Analyze Political Attention with Minimal Assumptions and Costs», *American Journal of Political Science*, pp. 209-228.
- Sebök, M. et al. (2023) «Comparative European legislative research in the age of large-scale computational text analysis: A review article», *International Political Science Review*. Disponible en: <https://doi.org/https://doi.org/10.1177/01925121231199904>.
- Slapin, J.B. y Proksch, S.-O. (2014) «Words as Data: Content Analysis in Legislative Studies», *The Oxford Handbook of Legislative Studies*, pp. 126-144. Disponible en: <https://doi.org/https://doi.org/10.1093/oxfordhb/9780199653010.013.0033>.
- Volgens, A. et al. (2015) *Scope, Range, and Extent of Manifesto Project Data Usage: A Survey of Publications in Eight High-Impact Journals. Handbook for Data Users and Coders. Version I*. Berlin: WZB Berlin Social Science Center. Disponible en: <https://manifestoproject.wzb.eu/datasets/mpduds>.