

Semantic change in health-related legislative discourse in Argentina: a case study on sudden death

Amalia Guaymás Canavire¹, Valeria Laura Burgos², and María Elena Buemi^{1,3}

¹ Maestría en Explotación de Datos y Descubrimiento del Conocimiento, UBA,
Buenos Aires, Argentina

² Universidad Hospital Italiano, Buenos Aires, Argentina

³ DC-ICC, UBA-CONICET

acarolinagc@gmail.com, valeria.burgos@hospitalitaliano.org.ar,
mebuemi@gmail.com

Abstract. Legislative language dynamically reflects political priorities, social values, and conceptual transformations within a society. Analyzing its semantic evolution can provide insights into emerging normative and conceptual changes. This study examines semantic change in health-related legislative discourse in Argentina using small corpora derived from open data on legislative initiatives from the Honorable Chamber of Deputies, covering parliamentary periods 127 to 141 (2009–2024). The data were organized into three comparative time intervals to enable a diachronic analysis using Natural Language Processing (NLP) techniques. Word embeddings were employed to capture semantic representations of terms and detect sociocultural shifts over time. Semantic spaces were aligned across periods using orthogonal Procrustes transformation, and changes were assessed through cosine similarity and nearest-neighbor analysis on static embeddings. Additionally, contextualized embeddings in Spanish (BETO) were combined with k-means clustering to identify usage patterns. The approach was further applied to a case study on terms related to sudden death.

Results indicate the emergence of cross-cutting concepts such as ‘vida’ and ‘deportivo’, while technical terms such as ‘desfibrilador’ remained relatively stable. Moreover, the analysis highlights the role of legislative committees and their connection to regional issues, providing contextual depth to the observed semantic patterns. These findings demonstrate the potential of NLP methods for analyzing semantic change in specialized, small-scale legislative corpora, offering a novel perspective for understanding the evolution of public health discourse and policy priorities.

Keywords: Semantic change, open data, Natural Language Processing, public health, legislative speech, sudden death.

Cambio semántico en el discurso legislativo de salud en Argentina: un caso de estudio en muerte súbita

Abstract. El lenguaje legislativo refleja las dinámicas de las prioridades políticas, los valores sociales y las transformaciones conceptuales en una sociedad. El análisis de su evolución semántica puede contribuir a anticipar cambios normativos y conceptuales. Nuestro país dispone de numerosos datos abiertos que son insumos fundamentales para la transparencia y la rendición de cuentas públicas. Este trabajo se basó en el uso de *corpus* pequeños provenientes de Datos Abiertos de iniciativas de ley sobre salud de la Honorable Cámara de Diputados de la Nación Argentina, correspondientes a los períodos parlamentarios 127 a 141 (2009-2024). Los mismos fueron organizados en tres tramos temporales comparativos con el objetivo de evaluar las características de los cambios semánticos mediante métodos de Procesamiento del Lenguaje Natural (PNL).

Se utilizaron representaciones vectoriales de palabras (*word embeddings*) para detectar cambios semánticos de origen sociocultural. Para comparar espacios semánticos entre períodos, se aplicaron métodos de alineación ortogonal (*Procrustes*) y vecinos más cercanos sobre *embeddings* estáticos. Complementariamente, se usaron *embeddings* contextualizados en español (BETO) combinados con un algoritmo de agrupamiento (*k-means*) para identificar usos. Los modelos fueron luego aplicados sobre un conjunto de términos asociados a muerte súbita. Los resultados mostraron el surgimiento de conceptos transversales como ‘vida’ y ‘deportivo’, mientras que ‘desfibrilador’ permaneció relativamente estable. De manera interesante, el análisis de los proyectos asociados a los términos técnicos destaca el rol de las comisiones legislativas participantes y su vínculo con problemáticas regionales, aportando una dimensión contextual para la interpretación de los resultados.

Palabras clave: Cambio semántico, datos abiertos, Procesamiento de Lenguaje Natural, salud pública, muerte súbita.

1 Introducción

La convergencia entre Datos Abiertos y el Procesamiento del Lenguaje Natural (PLN) abre oportunidades en el ámbito legislativo (Bertoni et al., 2023). En este marco, la Detección de Cambios Semánticos (DCS) por cambios culturales emerge como una herramienta para vincular los discursos legislativos con actores políticos, grupos de interés y transformaciones históricas (Dritsa et al., 2022; Kutuzov et al., 2018; Rodríguez et al., 2023). Los Datos Abiertos, definidos por la *Open Knowledge Foundation* (Open Knowledge Foundation, s.f.) como datos de libre acceso, uso y redistribución, son insumos fundamentales para la transparencia pública (Xhardez, 2020). En nuestro país, la HCDN ofrece un portal con nóminas de legisladores, comisiones, proyectos, sesiones y votaciones (Honorable Cámara de Diputados de la Nación Argentina, 2025). La aplicación de la ciencias de datos a estos datos ha impulsado iniciativas de análisis basado en grafos (Alemán & Clerici, 2022; Álvarez-Ugarte, 2020), aunque los estudios sistemáticos con PLN sobre datos abiertos legislativos continúan siendo incipientes. La DCS consiste en identificar variaciones en el significado de una palabra a lo largo del tiempo, manifestadas como derivas graduales o cambios abruptos determinados por contextos socioculturales específicos (Kutuzov et al., 2018). Los enfoques computacionales basados en *embeddings* han demostrado ser efectivos para capturar estas evoluciones de forma cuantificable (Betancourt & Murillo, 2023; Dritsa et al., 2022; Manrique-Gómez et al., 2025). Entre los métodos más empleados se encuentran: (i) la **alineación ortogonal de Procrustes**, que proyecta modelos de distintos períodos a un espacio común y utiliza la similitud coseno (SC) como métrica de detección de cambio semántico; (ii) la **vecindad semántica**, que considera los vecinos más cercanos y usa el puntaje de cambio (PC) como métrica para medir los cambios; y (iii) ***embeddings* contextualizados con clustering**, BETO más *k-means*, evaluados a través de la divergencia de *Jensen-Shannon* (DJS) y el índice de silueta (IS).

La prevención de la Muerte Súbita (MS) constituye un caso de estudio para DCS relevante porque evidencia una brecha entre el marco normativo y su implementación efectiva (Ministerio de Salud de la Nación Argentina, s.f.). La Ley 27.159 (promulgada en 2015 y reglamentada en 2022) establece un sistema integral de prevención que enfatiza la capacitación comunitaria y la coordinación entre jurisdicciones (“Decreto 402/2022 - Reglamentación de la Ley 27.159”, 2022). Así, aplicar DCS a términos de salud asociados con MS permite explorar cómo evoluciona el lenguaje legislativo e identificar cambios en las prioridades institucionales. Este trabajo propone una metodología y un modelo inicial para DCS aplicados a los títulos de proyectos de ley en salud iniciados en la Honorable Cámara de Diputados de la Nación Argentina (HCDN).

2 Metodología

Este trabajo utiliza proyectos legislativos anuales presentados durante el período comprendido entre los años 2009 al 2024, que se corresponden con los perío-

dos legislativos 127 al 141. De un total de 92535 proyectos legislativos se seleccionaron los 2848 que fueron girados primariamente a comisiones de salud. Con la selección mencionada, se construyeron tres *corpus* diacrónicos que denominaremos C_{t_1} , C_{t_2} y C_{t_3} , de modo que se corresponden con C_{t_1} (127–131, 2009–2014), C_{t_2} (132–136, 2014–2019) y C_{t_3} (137–141, 2019–2024). A su vez, cada par de *corpus* $C_{t_1}-C_{t_2}$ y $C_{t_2}-C_{t_3}$ fue tratado con tres enfoques basados en *embeddings*: *Procrustes*, Vecindad semántica y *Embeddings* contextualizados con *clustering*, comentados en la sección anterior. Se consideró cambio semántico cuando se cumplió al menos uno de los siguientes criterios: SC < 0.5, PC superior al promedio del *corpus*, o DJS > 0.5 acompañado de IS superior a la media. Se definieron 16 términos de interés, provistos por 9AMAS⁴ ‘muerte súbita’, ‘desfibrilador’, ‘desfibrilación’, ‘cardiovascular’, ‘cardioasistido’, ‘DEA’, ‘dispositivo’, ‘muerte’, ‘vida’, ‘capacitación’, ‘emergencia’, ‘supervivencia’, ‘reanimación’, ‘prevención’, ‘incidencia’ y ‘deportivo’. Dado que ‘muerte súbita’ constituye un concepto central para el estudio, ambas palabras se consideran como un único *token* en el análisis.

3 Resultados

A partir de los modelos construidos sobre el *corpus* de salud, se identificaron términos con posible cambio semántico entre pares de *corpus*. De los 16 términos asociados a MS, únicamente 9 de ellos (56,25%) son identificados, lo cual sugiere una cobertura parcial de los enfoques aplicados. La Tabla 1 presenta los términos seleccionados y evaluados por cada enfoque junto a sus respectivas métricas de cambio semántico. Para la transición $C_{t_1}-C_{t_2}$ (2009-2014) el enfoque de vecindad

Tabla 1. Cambio semántico para términos asociados a MS evaluado mediante enfoques basados en *embeddings* para cada par de *corpus*. SC = similitud coseno, PC = puntaje de cambio, DJS = divergencia de *Jensen-Shannon*, IS = índice de silueta.

Par de corpus	Procrustes		Vecindad semántica		BETO + k-means		
	SC	Palabra	PC	Palabra	DJS	IS	Palabra
$C_{t_1}-C_{t_2}$	0.53	muerte	-2.32	muerte	0.54	0.39	vida
	0.55	emergencia	-5.08	prevención	0.51	0.57	deportivo
	0.58	prevención	-5.25	vida	0.42	0.55	muerte
	0.65	vida	-6.17	capacitación	0.21	0.39	emergencia
	0.67	capacitación	-6.33	desfibrilador	–	–	–
	0.68	deportivo	–	–	–	–	–
$C_{t_2}-C_{t_3}$	0.22	dispositivo	-0.14	dispositivo	0.62	0.42	vida
	0.53	vida	-1.04	muerte	0.52	0.38	muerte
	0.54	muerte	-2.50	vida	0.50	0.50	deportivo
	0.60	prevención	-3.73	prevención	0.38	0.23	dispositivo
	0.62	emergencia	-4.42	deportivo	0.35	0.27	emergencia
	0.67	deportivo	-4.92	emergencia	–	–	–
	–	–	-5.00	dea	–	–	–

semántica identifica al término ‘muerte’ como candidato a cambio semántico dado que presenta un valor superior a -4.5. Por su parte, el método de *embedding* dinámico detecta los términos ‘vida’ (0.54) y ‘deportivo’ (0.51) como posibles

⁴ Ver Agradecimientos

casos de cambio. Para la transición $C_{t_2}-C_{t_3}$ (2014-2019) los términos ‘dispositivo’ (0.22 y -0.14), ‘muerte’ (-1.04 y 0.52) , ‘vida’ (-2.50 y 0.62) y ‘deportivo’ (0.5) se identifican como candidatos a cambio semántico.

El análisis de calidad de los agrupamientos para el método de *embedding* dinámico muestra que el término ‘deportivo’ presenta un valor de índice de silueta mayor a las medias de cada par de períodos (0.57 y 0.50), lo que sugiere una mejor definición de sus agrupamientos semánticos.

El análisis de vecindades (Tabla 2) ilustra los desplazamientos semánticos más relevantes. El término ‘muerte’ pasa de una dimensión biomédica y forense en C_{t_1} hacia usos normativos y de salud pública en C_{t_2} . El término ‘dispositivo’ exhibe una especialización semántica desde significados tecnológicos amplios en C_{t_2} hacia asociaciones biomédicas específicas, incluyendo ‘muerte súbita’ en C_{t_3} . El término ‘vida’ evoluciona de un objeto de protección sanitaria general hacia etapas específicas del ciclo vital.

Tabla 2. Análisis de vecindad para términos clave. La vecindad se conforma por los vecinos más cercanos identificados en el modelo estable.

Par de corpus	Enfoque	Palabra	Vecindad C_{t_i}	Vecindad $C_{t_{i+1}}$
$C_{t_1}-C_{t_2}$	Vecindad semántica	muerte	{práctica, historia, cuidado, institución, embarazado, clínico, paciente, menor}	{efecto, cuidado, dengue, promover}
$C_{t_2}-C_{t_3}$	<i>Procrustes</i>	dispositivo	{dato, similar, electrónico, no-civo}	{oftalmológico, tecnológico, mínimo, muerte súbita}
$C_{t_2}-C_{t_3}$	Vecindad semántica	vida	{cuidado, calidad, trombofilia, embarazado, instituto, fibromi-algia}	{embarazado, encontrar}

Exploración de comisiones dominantes

La exploración de las comisiones legislativas a las cuales se giran proyectos permite identificar su recorrido institucional, por ello hemos examinado las comisiones por las que transitan los proyectos de ley vinculados a salud, en particular, las relacionadas a muerte súbita. Se identifican 52 proyectos de ley asociados a términos más específicos (‘muerte súbita’, ‘desfibrilador’, ‘desfibrilación’, ‘dea’, ‘reanimación’, ‘cardiovascular’ y ‘cardioasistido’), este subconjunto representa aproximadamente el 0.20% del total de iniciativas legislativas sobre salud en el *corpus* analizado. Con el objetivo de identificar las comisiones con mayor actividad legislativa, se identificaron las 10 comisiones con mayor frecuencia de giros para tres conjuntos de proyectos: (1) el total de proyectos de ley, (2) los proyectos vinculados a salud y (3) los proyectos de salud que contienen términos específicos asociados a MS (Tabla 3).

Se observa que, independientemente de los proyectos, la Comisión de Presupuesto y Hacienda cuenta con la mayor proporción de giros (46,96 % en proyectos de salud y 61,11 % en los relacionados a MS). En cuanto a los proyectos específicos sobre muerte súbita, la Comisión de Educación ocupa la segunda posición en frecuencia de giros (15.28%). Esta presencia puede estar vinculada con el enfoque preventivo que caracteriza a muchas de estas iniciativas. Asimismo,

Tabla 3. Identificación de las principales comisiones legislativas según tipo de proyecto en la Cámara de Diputados de Argentina entre los períodos parlamentarios 127–141 (2009–2024).

Top	Proyectos de ley				Proyectos de ley en salud				Proyectos específicos de MS			
	Comisión	Frec.	%		Comisión	Frec.	%		Comisión	Frec.	%	
1	Presupuesto y Hacienda	17764	25.44		Presupuesto y Hacienda	1882	46.96		Presupuesto y Hacienda	44	61.11	
2	Legislación General	6263	8.97		Legislación General	375	9.36		Educación	11	15.28	
3	Legislación del Trabajo	4063	5.82		Familia, Mujer, Niñez y Adolescencia	216	5.39		Deportes	3	4.17	
4	Legislación Penal	3685	5.28		Educación	165	4.12		Legislación General	3	4.17	
5	Asuntos Constitucionales	2983	4.27		Comercio	153	3.82		Discapacidad	2	2.78	
6	Justicia	2547	3.65		Legislación del Trabajo	142	3.54		Familia, Mujer, Niñez y Adolescencia	2	2.78	
7	Educación	2367	3.39		Industria	140	3.49		Familias, Niñez y Juventudes	2	2.78	
8	Familia, Mujer, Niñez y Adolescencia	2270	3.25		Discapacidad	122	3.04		Transportes	2	2.78	
9	Cultura	1685	2.41		Prev. Adicc. Control Narcoctrf.	105	2.62		Banca de la Mujer	1	1.39	
10	Transportes	1673	2.40		Comunicaciones e Informática	97	2.42		Comunicaciones e Informática	1	1.39	

algunas comisiones aparecen de manera exclusiva en este subconjunto, tal como la Comisión de Deportes que ocupa la tercera posición (4.17% de proyectos girados), y otras con menor frecuencia como Banca de la Mujer. Otro aspecto a destacar es la participación de la Comisión de Comunicaciones e Informática entre las de mayor frecuencia de giros de proyectos, tanto en los vinculados a salud en general (2.42%) como en los proyectos específicos sobre muerte súbita (1.39%).

4 Conclusiones

El análisis de DCS mediante *embeddings* evidencia una baja presencia de términos vinculados a la prevención de muerte súbita en el *corpus* legislativo. Palabras como “vida”, “muerte”, “dispositivo” y “deportivo” muestran desplazamientos semánticos relevantes entre períodos. En particular, “dispositivo” incorpora “muerte súbita” en su vecindad en C_{t_3} , mientras que “desfibrilador” se mantiene estable, sugiriendo una terminología técnica consolidada. El análisis de comisiones revela un enfoque intersectorial, con participación de Educación y Deportes, y ausencia de sectores como Turismo o Minería. Entre las limitaciones, se destaca el tamaño reducido del *corpus* y el uso exclusivo de títulos. No obstante, los resultados confirman el potencial del enfoque para analizar la evolución del discurso en salud pública.

Agradecimientos. A la empresa 9AMAS (Salta, Argentina) por su contribución en la selección de las palabras claves que se utilizan en este artículo. Certificación IRAM ISO 9001- 2015 de Gestión de Calidad (<https://9amas.com/>).

References

- Alemán, E., & Clerici, P. (2022). Centralidad y productividad legislativa en la cámara de diputados argentina, 1983–2019. *Revista de Ciencia Política (Santiago)*, 42.
- Álvarez-Ugarte, R. (2020). Social network analysis and the law: A proof of concept for argentina [SSRN Electronic Journal]. <https://ssrn.com/abstract=3711556>
- Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (2023). *Handbook of computational social science for policy* [Edited volume]. Springer International Publishing.
- Betancourt, E., & Murillo, E. (2023). Analysis of the semantic shift in diachronic word embeddings for spanish before and after covid-19. *CLEI Electronic Journal*, 26.
- Decreto 402/2022 - reglamentación de la Ley 27.159 [Consultado el 17 de Abril de 2026]. (2022). <https://www.boletinoficial.gob.ar/detalleAviso/primera/266562/20220715>
- Dritsa, K., Thoma, A., Pavlopoulos, J., & Louridas, P. (2022). A greek parliament proceedings dataset for computational linguistics and political analysis. *Advances in Neural Information Processing Systems*, 28874–28888.
- Honorable Cámara de Diputados de la Nación Argentina. (2025). Portal de datos públicos en formato abierto de la honorable cámara de diputados de la nación [Consultado el 1 de junio de 2025]. <https://datos.hcdn.gob.ar/>
- Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic word embeddings and semantic shifts: A survey. *arXiv preprint arXiv:1806.03537*. <https://arxiv.org/abs/1806.03537>
- Manrique-Gómez, L., Montes, T., & Manrique, R. (2025). Gender semantics and historical feminisms: An interdisciplinary approach through natural language processing. *Revista de Estudios Sociales*, 93, 19–38.
- Ministerio de Salud de la Nación Argentina. (s.f.). Atlas: Muerte súbita [Consultado el 15 de abril de 2026]. http://www.legisalud.gov.ar/atlas/muerte_subita.html#6
- Open Knowledge Foundation. (s.f.). Open definition [Consultado el 1 de junio de 2025]. <https://opendefinition.org/>
- Rodriguez, P., Spirling, A., & Stewart, B. M. (2023). Embedding regression: Models for context-specific description and inference. *American Political Science Review*, 117(4), 1255–1274.
- Xhardez, V. (2020). *Datos abiertos en la argentina: Desafíos para la apertura y reutilización de datos públicos de gobierno*. CIECTI.