

# Parlagraph: Intervenciones del parlamento como una base de datos de grafos

Belén Brandino *Instituto de Computación*  
*Facultad de Ingeniería, Universidad de la República*  
 Montevideo, Uruguay  
 bbrandino@fing.edu.uy

Santiago Góngora *Instituto de Computación*  
*Facultad de Ingeniería, Universidad de la República*  
 Montevideo, Uruguay  
 sgongora@fing.edu.uy

## Resumen

En este trabajo se presenta un esfuerzo inicial para representar la actividad parlamentaria uruguaya, en particular los discursos dados por los legisladores en distintas cámaras, como una base de datos de grafos. Se explora también una posible forma de enriquecer la representación mediante el análisis de emociones y el análisis de tópicos. Sobre esta base de datos ampliada se realizan algunas consultas a modo de prueba de concepto sobre el potencial de contar con una representación así para los discursos parlamentarios.

## I. INTRODUCCIÓN

Los discursos dados por los legisladores en las cámaras gubernamentales reflejan una serie de aspectos sociales, culturales, económicos y políticos que son de relevancia en el día a día de la nación y forman parte de su legajo histórico. Tener estos datos abiertos permite que toda la población, independientemente de su ocupación, pueda libremente acceder a una fuente confiable y libre de ruido de las posiciones reales de algunos de sus representantes. Estos discursos originales también pueden ser profundamente analizados en pos de hallar patrones que luego se reflejan en posiciones concretas dentro del espectro político.

Sin embargo no basta con tener estos datos en crudo, puesto que la información desprendida de ellos resulta difícil de inferir o extraer. En este trabajo se presenta un pequeño esfuerzo en representar los discursos de las cámaras de representantes usando una base de datos de grafos, con el objetivo de que los datos así presentados permitan a expertos hacer consultas para realizar análisis discursivos complejos. Además de representar los datos básicos encontrados en el *scrapping* original se incluyen experimentos preliminares con análisis de tópicos y análisis de emociones. No es el objetivo de este trabajo una solución *performante* y definitiva para este problema, sino hacer una pequeña prueba de concepto para futuros proyectos.

El documento se estructura de la siguiente manera: en la sección II se lista el trabajo previo relevante para el presentado; en la sección III se muestra y discute el modelo de datos elegido para la base de datos de grafos; en la sección IV se describen los experimentos de Procesamiento de Lenguaje Natural (PLN) realizados; en la sección V se presenta como ejemplo una serie de consultas realizadas a la base de datos y por último la sección VI muestra las conclusiones del trabajo y presenta el trabajo a futuro.

## II. TRABAJOS RELACIONADOS

El enfoque de representar la actividad parlamentaria mediante una base de datos de grafos no es nuevo. [Van Aggelen et al., 2017] representan los debates del Parlamento Europeo mediante Resource Description Framework (RDF), enriqueciéndola con datos de otros conjuntos, como una representación jerárquica de tópicos. En [Dargis et al., 2018] se describe la creación de un corpus multilingüe usando los discursos del parlamento de Latvia. Parte del corpus está publicado utilizando RDF. [Perak and Rodik, 2018], [Perak, 2020] presenta un esfuerzo similar para el parlamento de Croacia, también utilizando RDF. En ambos casos se expresa parte del corpus con información morfosintáctica siguiendo el formato de Universal Dependencies <sup>1</sup>, pero en el último se hace una representación adicional de cada palabra de cada discurso como un nodo en una base de datos de grafos con Neo4J<sup>2</sup>.

El uso de bases de datos de grafos para asistir el análisis discursivo no se limita solamente al dominio parlamentario. En [Botzer and Weninger, 2023] se utiliza una representación de grafos para el modelado de discursos en redes sociales. Particularmente se opta por representar las entidades y conceptos como los nodos. [Gourgaris et al., 2018] utiliza un modelo

<sup>1</sup><https://universalddependencies.org/format.html>

<sup>2</sup><https://neo4j.com/>

de grafos para representar preguntas y respuestas de una manera tal que facilite la exploración de cadenas de argumentación complejas.

Finalmente uno de los trabajos que más relación tiene con el presentado en este documento es [Venkatakrisnan and Schmidt, 2018], puesto que realizan un modelado de grafos para el Parlamento Europeo, a la vez que se centran en la exploración de tópicos.

### III. MODELO DE DATOS

Los datos a representar en la base de datos de grafos son los disponibles a raíz de la línea de investigación en la que, entre otros resultados, se publicó el paquete *Speech* para el lenguaje R<sup>3</sup>. Particularmente este juego de datos contiene intervenciones de 1985 a 2015 y ya están organizadas en un *dataset* llamado `speech_intervencion`. Este archivo tiene 255.744 entradas y sus campos son [Luján et al., 2023]:

- **legislator**: el nombre completo del legislador.
- **legislature**: el período legislativo correspondiente, considerando las diferentes conformaciones desde 1985 a 2015.
- **chamber**: en qué cámara se enunció la intervención: cámara de representantes, cámara de senadores, comisión permanente o asamblea general.
- **date**: fecha de la sesión en el formato AAAA-MM-DD
- **id**: identificación de la sesión
- **speech**: apellido en mayúscula del senador seguido de su intervención (e.g. “ABISAB Voto por la afirmativa”).
- **party**: partido político del legislador.
- **doc\_type**: formato del documento original: html, pdf o scan\_pdf.
- **sex**: sexo del legislador; 0 si es mujer, 1 si es hombre.

Luego de explorar diferentes posibilidades y analizar los trabajos anteriores detallados en la sección II, se decidió:

- Representar a los legisladores como nodos *Legislator*, usando como etiqueta sus nombres y agregando una propiedad que represente su sexo. Resulta natural que los legisladores sean nodos, ya que son una entidad de la realidad. Además, interesa obtener datos centrados en los legisladores.
- Representar a los partidos políticos como nodos *Party*, con etiqueta el nombre del partido. Nuevamente, interesa obtener datos centrados en los partidos políticos también.
- Representar la asociación de un legislador a un partido político como una relación, llamada *represents\_party* (*representa\_partido*), entre nodos del tipo *Legislator* y *Party*.
- Representar las cámaras como nodos *Chamber*, con la propiedad del año. De esta manera se generará un nodo por cámara por año. Un ejemplo de esto se puede ver en la figura 1 donde se tiene un nodo que representa la Comisión Permanente de 1998. Esta decisión se desprende del hecho de que en el correr de los años los legisladores que pertenecen a las cámaras van cambiando, haciendo que sea importante conocer no solo en las varias cámaras que pudo haber participado un legislador sino también en qué años.
- Representar la participación de un legislador en una cámara con la relación llamada *participated\_in* (*participó\_en*), entre nodos del tipo *Legislator* y *Chamber*.
- Representar las distintas sesiones de una cámara como nodos *Session*, con el identificador de la sesión como etiqueta y su fecha como propiedad. Resulta sensato que las sesiones sean un nodo para poder asociar las intervenciones a éstas. Como se comentará más en detalle en la sección VI, este aspecto del modelo de datos podría cambiar si se representara también el orden de las intervenciones.
- Representar las sesiones pertenecientes a una cámara con la relación *session\_number* (*numero\_de\_sesión*), entre nodos del tipo *Chamber* y *Session*.
- Representar las intervenciones de los legisladores como nodos *Intervention*, y el texto de la intervención como etiqueta. Esta idea fue obtenida del trabajo de [Perak and Rodik, 2018], que representaban de esta forma los textos de las intervenciones.
- Representar la ocurrencia de una intervención dentro de una sesión con la relación *session\_contains\_intervention* (*sesión\_contiene\_intervención*), entre nodos del tipo *Session* e *Intervention*.
- Representar qué legislador hizo qué intervención con la relación *made\_intervention* (*realizó\_intervención*), entre nodos del tipo *Legislator* e *Intervention*.

La carga del grafo se realizó en AuraDB<sup>4</sup> con el fin de tener la base en la nube. Dadas las limitaciones de memoria (por temas de seguridad) que posee AuraDB, solo fue posible cargar veinte mil entradas del dataset. En la imagen 2 se muestra cómo se ve una parte del grafo cargado en Neo4J con estos datos. Es posible ver dos nodos celestes representando legisladores, ambos pertenecientes al mismo partido (FA) representado por el nodo en violeta. A su vez se pueden ver las intervenciones realizadas por ambos legisladores en color verde lima, en verde medio las cámaras donde participaron y en verde oscuro las sesiones. Se omite mostrar todas las conexiones ya que deja muy cargado el grafo para el ejemplo.

<sup>3</sup><https://nicolas-schmidt.github.io/speech/>

<sup>4</sup><https://neo4j.com/cloud/platform/aura-graph-database/>

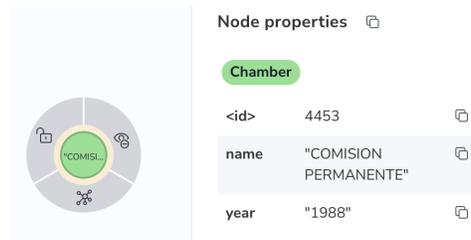


Figura 1: Esta imagen muestra un nodo del tipo *Chamber*, donde se puede ver que tiene la propiedad año.

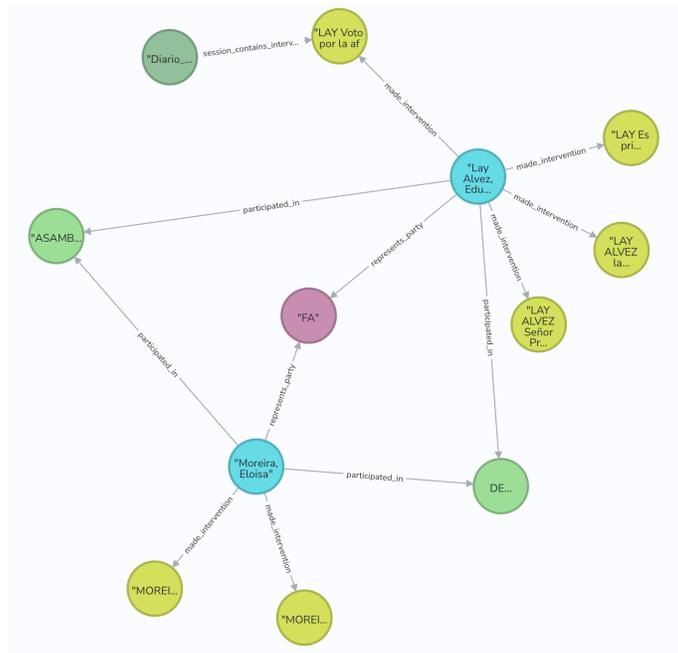


Figura 2: Esta imagen muestra la estructura del grafo que representa los datos de *speech\_intervencion*.

Para la carga del grafo se utilizaron los comandos `LOAD CSV` (para importar datos de un csv), `MERGE` (para crear nuevos nodos o relaciones, o utilizar los ya creados) y `CREATE` (para crear nuevos nodos) de Cypher [Neo4j, 2023], con el código desplegado en 1. En este código primero se obtienen los datos del archivo CSV (con los cabezales que indican qué representa cada valor), identificando cada línea como *row*. Luego se realiza un `MERGE` con cada nodo mencionado anteriormente y sus relaciones.

#### Listing 1 Código para la carga del grafo, sin sentimientos ni tópicos asociados a las intervenciones

```

1      LOAD CSV WITH HEADERS FROM '<link_descarga_csv>' AS row
2      MERGE (l:Legislator {name: row.legislator, sex: row.sex})
3      WITH l, row
4      MERGE (p:Party {name: row.party})
5      MERGE (l)-[r_p:represents_party]->(p)
6      MERGE (c:Chamber {name: row.chamber, year:left(row.date, 4)})
7      MERGE (l)-[r_c:participated_in]->(c)
8      MERGE (s:Session {name: row.id, date: row.date})
9      MERGE (c)-[c_s:session_number]->(s)
10     CREATE (i:Intervention {speech: row.speech})
11     MERGE (s)-[s_i:session_contains_intervention]->(i)
12     MERGE (l)-[l_i:made_intervention]->(i)

```

Sin embargo el uso de `MERGE` es a excepción de las intervenciones para las que se utiliza `CREATE`, ya que sino se genera un nodo para intervenciones iguales de un mismo legislador sin importar la sesión. En la figura 3 se puede ver un ejemplo para intervenciones similares, como es frecuente en el caso de las peticiones de palabra, donde se unifican las intervenciones, lo cual no parece deseable. Además en esta figura se puede ver el primer error de calidad de datos, donde existen dos nodos para el mismo legislador ya que aparece en dos intervenciones con la propiedad de sexo distinta.

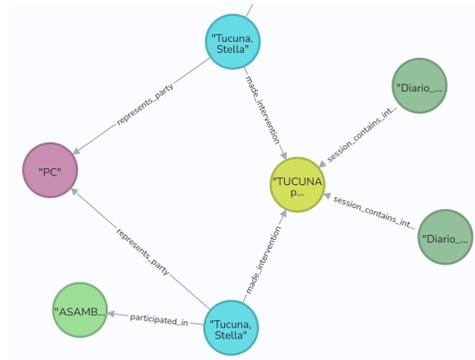


Figura 3: Esta imagen muestra como se unifican dos intervenciones distintas (con el mismo diálogo pero distinta sesión) por el uso del comando MERGE.

El grafo fue posteriormente enriquecido mediante un análisis automático de los textos “speech”, donde a cada intervención se le asigna una emoción denotada por el orador y un tópico abordado en ella. Los detalles de este análisis se profundizan en la sección IV. Para la representación en el grafo se decidió:

- Representar las emociones como nodos *Emotion*, usando como etiqueta el nombre de la emoción.
- Representar la asociación de una intervención a una emoción como una relación, llamada *main\_emotion* (principal\_emoción), entre nodos del tipo *Legislator* y *Emotion*.
- Representar los tópicos como nodos *Topic*, usando como etiqueta el nombre del tópico.
- Representar la asociación de una intervención a una emoción como una relación, llamada *main\_topic* (principal\_tópico), entre nodos del tipo *Legislator* y *Topic*.

Cada emoción y tópico se representan con un nodo con el objetivo de poder realizar consultas encadenadas en función de estos aspectos, como se mostrará en la sección V. Esta evolución del grafo se muestra en la figura 4, donde en rosa claro se puede ver el nodo que representa el tópico, y en lila el nodo que representa la emoción, ambos relacionados con una intervención.

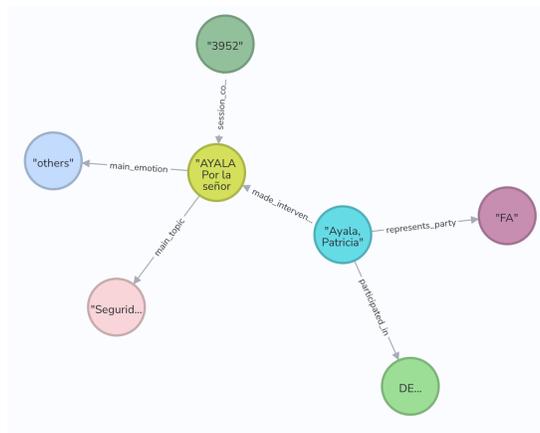


Figura 4: Esta imagen muestra la evolución del grafo una vez procesadas las intervenciones con los enfoques que se detallan en la sección IV.

En cuanto al código de Cypher utilizado para agregar estos nodos, se agrega a continuación del código presentado en 1 el código análogo al utilizado para la creación del resto de los nodos, creando los nodos *Emotion* y *Topic* y las relaciones correspondientes con el comando MERGE, como se puede ver en 2.

---

**Listing 2** Código para anexar al código presentado en 1, para agregar las emociones y tópicos al grafo

---

```

1     MERGE (e:Emotion {emotion: row.emotions})
2     MERGE (i)-[i_e:main_emotion]->(e)
3     MERGE (t:Topic {topic: row.topics})
4     MERGE (i)-[i_t:main_topic]->(t)

```

---

#### IV. EXPERIMENTACIÓN CON PROCESAMIENTO DE LENGUAJE NATURAL

Como se mencionó en la sección anterior, luego de tener los datos del *dataset* original representados como un grafo surge la idea procesar los discursos automáticamente, con el objetivo de enriquecer la base de datos y hacer una prueba de concepto de futuros caminos a seguir. En un contacto con Nicolás Schmidt y Elina Gómez, investigadores trabajando en estos temas, se determinó que una combinación interesante podría ser el análisis de emociones y el análisis de tópicos de cada intervención. De esta forma se podría realizar consultas como por ejemplo qué emoción denotan legisladores de cierto partido político al hablar de un tema en concreto.

En esta sección se detallan ambos experimentos. Todos ellos fueron corridos sobre todo el conjunto de datos `speech_intervencion` usando Google Colab<sup>5</sup>, llevando aproximadamente un total de 3 horas.

##### IV-A. Análisis de emociones

Este experimento consiste en hallar para cada una de las intervenciones cuál es la emoción predominante. Dependiendo del marco teórico se puede tener diferentes conjuntos de emociones a analizar. En este trabajo se utilizó `pysentimiento` [Pérez et al., 2021]<sup>6</sup>, que provee un modelo que clasifica un texto entre las categorías: *joy*, *surprise*, *fear*, *sadness*, *disgust*, *anger* y *others*. Hay que tener en cuenta que dicho modelo está entrenado predominantemente sobre *tweets*, por lo que a priori el largo de los textos con los que fue entrenado puede distar en gran medida del largo de las intervenciones del parlamento, lo que podría impactar en su *performance*. En la figura 5 se muestra la distribución de cantidad de caracteres de las intervenciones presentes en el *dataset* según ciertos rangos. Estos rangos están determinados por el límite histórico que han tenido los *tweets*: primero hasta 140 caracteres, luego hasta 280 caracteres y a partir de la aparición de *Twitter Blue*<sup>7</sup> se permite más de 280 caracteres.

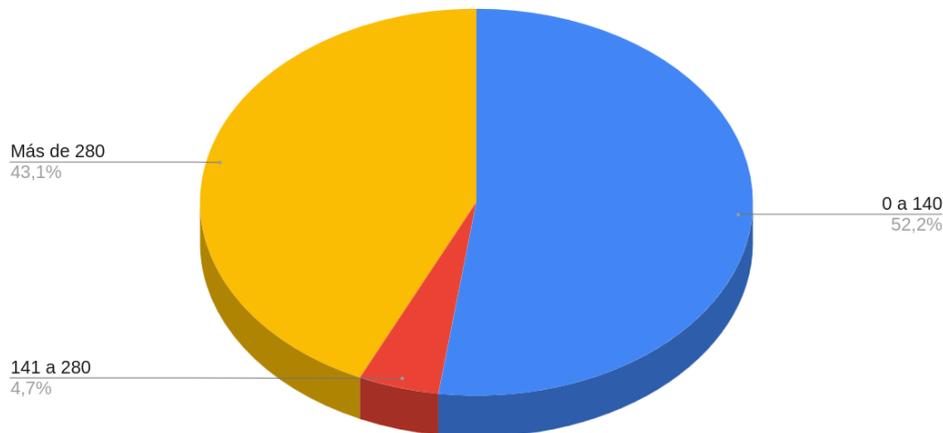


Figura 5: Esta figura muestra la proporción de intervenciones en el *dataset* que tienen de 0 a 140 caracteres, de 141 a 280 caracteres o más de 280 caracteres.

Una vez usado el modelo para hallar la emoción predominante para cada intervención se puede apreciar un gran desbalance en la cantidad por categoría. De las 255.744 intervenciones en el *dataset*, 240.063 (el 93,86%) fueron clasificadas como *others*. En la figura 6 se observa la distribución para el resto de categorías<sup>8</sup>. En ella se visualiza claramente que la gran mayoría de intervenciones fueron clasificadas como *joy* (alegría), *sadness* (tristeza) y *anger* (enojo). Este resultado cuantitativo parece ser razonable con el tipo de discursos que comúnmente se realizan, donde los parlamentarios suelen celebrar positivamente una idea o citar una anécdota para apoyar su discurso (*joy*), o bien discrepar con una idea o abordar un tema sensible o controversial (*anger* y *sadness*). Está claro que un grupo de expertos podría hacer mejores y más certeras valoraciones, por ejemplo realizando una curación manual de un muestreo de las intervenciones clasificadas. En la siguiente sección, en la tabla I, se muestran algunos ejemplos de intervenciones clasificadas según emociones.

<sup>5</sup><https://colab.research.google.com>

<sup>6</sup><https://huggingface.co/finiteautomata/beto-emotion-analysis>

<sup>7</sup><https://help.twitter.com/en/using-twitter/twitter-blue>

<sup>8</sup>No se incluye en el gráfico el 93,86% de las intervenciones etiquetadas como *others* para hacer más legible el resto de proporciones.

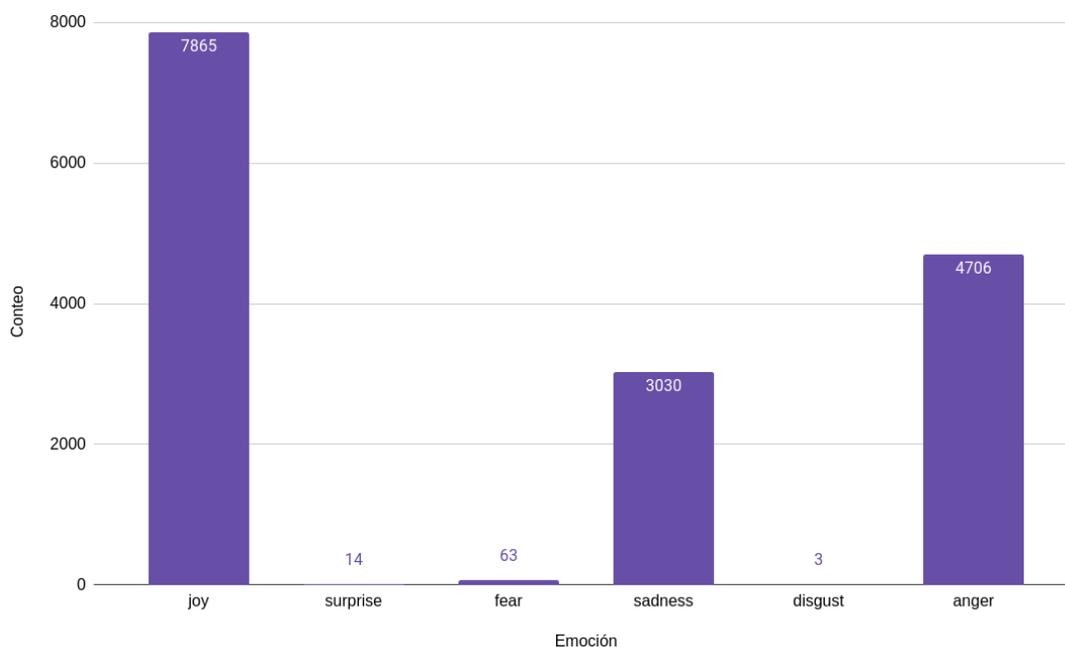


Figura 6: Esta figura muestra el conteo sobre el *dataset* según la emoción calculada con *pySentimiento* sobre cada discurso.

#### IV-B. Análisis de tópicos

El análisis de tópicos es uno de los problemas más antiguos del PLN y también de la clasificación de textos [Jurafsky and Martin, 2023]. Existen diferentes formas de afrontar este problema. En el caso de que se tuviera un *dataset* de textos anotados con su correspondiente tópico, de la forma  $\langle \text{texto}, \text{tópico} \rangle$ , se podría entrenar un algoritmo que dado un texto nuevo permita hallar el tópico correspondiente. Este sería un enfoque **supervisado**.

Otro camino para esto puede ser mediante un enfoque **no supervisado**, donde no se necesita texto anotado con la categoría que le corresponde. Hay muchos enfoques no supervisados para seguir, como puede ser el *clustering* [Mitchell et al., 2007], métodos que usen formalismos sintácticos y semánticos (e.g. si el sujeto de la oración es “helados” entonces el texto trata sobre *alimentos*), utilizando algún otro tipo de representación de textos como los *embeddings* [Jurafsky and Martin, 2023], entre otros.

Dada la falta de datos anotados y el tiempo disponible para abordar el proyecto, se optó por utilizar representaciones vectoriales de texto (*sentence embeddings*) junto con una medida de similitud para compararlos (i.e. cuál vector es más similar a cuál otro). Para ello la lista de tópicos fue diseñada en base a una lista de interés dada por Nicolás Schmidt y Elina Gómez para cubrir en un análisis primario:

- Aborto
- Reforma electoral
- Ley de puertos
- Leyes de presupuesto
- Eutanasia
- Seguridad social
- Regulación del mercado de marihuana

Como el algoritmo que se presentará en esta sección funciona mediante la codificación de textos en vectores, cada uno de los tópicos a tener en cuenta tiene que ser representados mediante un texto. Qué tipo de texto es el que funciona mejor está sujeto a la experimentación, pero a priori se podría intentar solamente con los nombres de los tópicos, los nombres y una lista de palabras clave o con una descripción rica de cada tópico. Es importante destacar que en un proyecto de mayor duración sería esperable poder experimentar con diferentes conjuntos de tópicos y diferentes formas de describirlos; se comenta más en profundidad algunas ideas de trabajo a futuro para esto en la sección VI. Como el objetivo de este trabajo es realizar una prueba de concepto, se experimentó solamente con una única descripción para los siguientes tópicos<sup>9</sup>:

- **Aborto:** “Estoy hablando del aborto. Interrupción del embarazo, ley de salud sexual y reproductiva, feto.”

<sup>9</sup>Notar que se quitó el tópico *Reforma electoral* (dentro de los tópicos sugeridos) y se agregó *Seguridad ciudadana*, *Tragedias* y *Petición de palabra*.

- **Ley de puertos:** “Estoy hablando de la ley de puertos. Servicios portuarios, puertos comerciales, circulación de mercancías.”
- **Leyes de presupuesto:** “Estoy hablando de la ley de presupuesto. PBI (Producto Bruto Interno), asignación de recursos, inversiones, aumentos salariales, inflación.”
- **Eutanasia:** “Estoy hablando de la eutanasia. Muerte voluntaria para acabar el sufrimiento, enfermedades terminales.”
- **Seguridad social:** “Estoy hablando de la seguridad social. Jubilación, pensiones, invalidez, vejez, cobertura.”
- **Seguridad ciudadana:** “Estoy hablando de la seguridad ciudadana. Robo, rapiña, copamiento, arma, policía, delincuente.”
- **Tragedias:** “Estoy recordando un hecho o contando una anécdota trágica para la sociedad. Accidentes de tránsito, asesinatos, muertes, atentados, nazismo, holocausto, terrorismo, dictadura”
- **Regulación del mercado de marihuana:** “Estoy hablando de la regulación del mercado de marihuana. Cultivo, autocultivo, cannabis, droga psicoactiva”
- **Petición de palabra:** “Solamente estoy pidiendo la palabra, haciendo una interrupción o finalicé mi intervención”

Para hallar cuál de las anteriores oraciones es la que mejor describe a la intervención parlamentaria se utiliza la técnica de `sentence-transformers`<sup>10</sup>, que codifica a los textos como un *vector denso* (embedding). Luego de codificar la intervención y a cada una de las descripciones de los tópicos como vectores se intenta hallar el vector del tópico más cercano al de la intervención.

En el siguiente pseudocódigo 3 se describe cómo se obtiene el valor de similitud máximo entre un discurso y los tópicos disponibles. Luego el tópico elegido es el que se corresponde a ese valor máximo:  $\arg \max_{\text{tópico}} \text{cosine\_similarity}(\text{discurso}, \text{tópico})$ .

**Listing 3** Pseudocódigo: Hallar el tópico para una intervención “speech”

```

1     input speech: str
2     input topic_descriptions: list of str
3
4     text_embedding = encode(speech)
5     topics_embeddings = encode(topic_descriptions)
6
7     max_value = -10
8     chosen_topic = "placeholder"
9
10    for topic_embedding in topics_embeddings:
11        similarity = cosine_similarity(text_embedding, topic_embedding)
12        if similarity > max_value:
13            max_value = similarity
14            chosen_topic = get_topic_id(topic_embedding)
15
16    return chosen_topic

```

En la tabla I se muestra una serie de ejemplos de intervenciones del *dataset* clasificadas según la previa lista de tópicos utilizando el algoritmo 3 así como también la emoción detectada según *pysentimiento* [Pérez et al., 2021].

Es notorio que el algoritmo de tópicos está lejos de funcionar de la mejor manera posible y hay dos motivos claros que seguramente están deteriorando su *performance*. En primer lugar cada intervención está precedida del nombre del legislador orador, lo que suele introducir palabras extrañas (como apellidos) con una semántica poco clara. En segundo lugar cada una de estas intervenciones son clasificadas de manera aislada, sin tener en cuenta su contexto. Esto último es vital para algunas intervenciones que surgen de un diálogo, como el cuarto ejemplo en la intervención de Vener Carboni.

## V. CONSULTAS REALIZADAS A LA BASE

Una vez cargada la base de datos se realizaron algunas consultas con el lenguaje Cypher para explorar las diferentes posibilidades que ofrece una representación como esta para las intervenciones parlamentarias. En esta sección se presentan diferentes consultas y sus resultados.

### V-A. Hallar todos los oradores que opinaron con una emoción particular

En la imagen 7 se muestra el resultado para esta consulta, con la emoción *fear* (miedo), retornando los legisladores y la intervención en cuestión.

El código de la consulta se presenta en 4. Para esto, se parte de un nodo *Legislator* con el comando `MATCH`, recorriendo la relación *made\_intervention*, y con la intervención obtenida recorriendo la relación *main\_emotion*, limitando la emoción obtenida al tipo *fear* con el comando `WHERE`.

<sup>10</sup>[https://huggingface.co/hiiamsid/sentence\\_similarity\\_spanish\\_es](https://huggingface.co/hiiamsid/sentence_similarity_spanish_es)

Speech	Tópico	Emoción
ABDALA Pido la palabra	Petición de palabra	others
ASTI Me permite una interrupción	Petición de palabra	others
LEGLISE Señor Presidente quedamos muy preocupados con la cifra que mencionó el señor Ministro en cuanto al costo que tuvieron los fondos de liquidación ya que habló de más de US 10 000 000 No sé si el Ministro tiene en su poder cómo se desagrega ese monto o si puede aportar la información el señor Presidente del Banco Central pero nosotros quisiéramos saber en su debido momento de dónde surge esa cifra puesto que realmente nos preocupa Muchas gracias	Ley de presupuesto	others
VENER CARBONI Seguimiento por parte de quién	Marihuana	others
ROSSI Muchas gracias El Poder Ejecutivo entiende que sería posible establecer un camino que de alguna manera nos permitiera profundizar en las preocupaciones las interrogantes y los intereses que el país el departamento los ciudadanos y este propio Poder Legislativo tienen en cuanto a este proyecto para poder crear condiciones que permitan sacar adelante lo que resulte mejor para la economía del país y por cierto para los vecinos del departamento de Rocha en particular los de La Paloma Termino con estas preguntas y quedo a la orden del señor Ministro y de sus asesores para seguir profundizando en este tema	Ley de puertos	others
GARCIA PINTOS Es inimputable parecería señor Presidente Es lamentable Lamentable que tengamos que estar en esto Lamentable	Hecho trágico	sadness
MUJICA Debo decir al señor Ministro que en realidad ha habido una reconversión exitosa para una empresa Hay una empresa a la que le ha ido muy bien Los remolacheros quedaron por el camino ahora ya no hay que reconvertirlos Se reconvirtieron en una ruina por lo que reitero ya no hay que reconvertirlos Apenas se gastaron US 100 000 con un experimento de algodonereros parece una cifra ridícula en cuanto a lo que se distribuyó en otra parte Yo tácitamente tengo que ver en esta nueva propuesta de reconversión y de la misma se desprende una crítica dura en el sentido de que lo que se estuvo gastando se gastó y se distribuyó mal aunque las palabras del señor Ministro y del Equipo de Gobierno no lo han dicho De todas formas les corresponde a ellos interpretarlo Acá se da vuelta la pisada en lo que hace a la manera de distribuir los fondos y los recursos que se dan escasos mitigados como siempre y por eso mismo se debe adoptar otro criterio Pero el Uruguay ha estado hasta ahora gastando plata en la reconversión con muy malos resultados desde el punto de vista social aunque no tanto desde el de las empresas Por mi parte entiendo que AZUCITRUS es una empresa bastante próspera Muchas gracias	Ley de presupuesto	anger
TROBO Por el señor senador Zumarán y voy a fundar el voto Para mí es una satisfacción poder votar al señor senador Zumarán como presidente de la Comisión Permanente que integramos por primera vez en nuestra actividad legislativa en su condición de legislador y hombre del Partido Nacional que integramos con mucho orgullo Señalamos que vamos a aportar a la labor de este Comisión Permanente el mayor de nuestros esfuerzos todo nuestro esmero y la mejor disposición	Seguridad social	joy
FALCO Señor Presidente como miembro de la Comisión no fui consultado sobre ese agregado De cualquier manera ya se conoce mi opinión sobre el proyecto original Pero que todavía la Cámara declare la vigencia del Decreto N° 102 de la Junta me deja sin palabras señor Presidente	Petición de palabra	surprise
KORZENIAK Voy a plantear el tema con mucha franqueza Apoyé esta alteración de orden del día concretamente porque tengo el temor fundado en la historia de las últimas sesiones de la Asamblea General de que una vez que se trate la fijación de los sueldos de los legisladores no haya quórum para considerar los otros temas Murmullos	Seguridad social	fear
SINGER Eso es horroroso	Hecho trágico	disgust

Cuadro I: Ejemplos de intervenciones clasificadas en tópicos según el algoritmo 3 y en emociones según *pysentimiento* [Pérez et al., 2021].

#### Listing 4 Código de la consulta 1

```

1 MATCH (l:Legislator)-[:made_intervention]->(i:Intervention), (i)-[:main_emotion]->(e:Emotion)
2 WHERE e.emotion="fear"
3 RETURN l as Legislador, i as Intervention

```

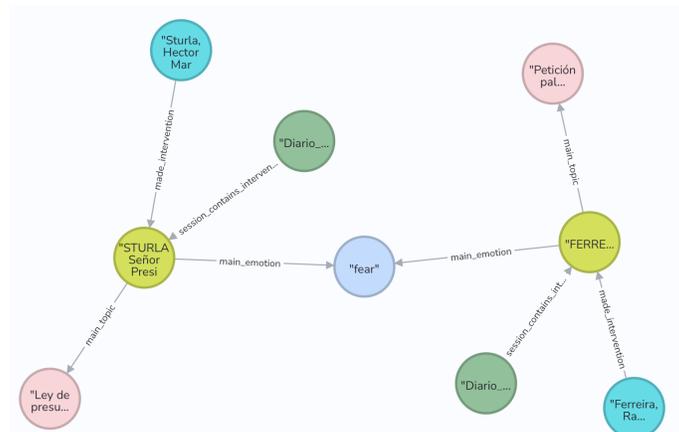


Figura 7: Consulta 1 en Cypher. En este caso se retornan los nodos *Legislador* e *Intervention* y se despliegan sus relaciones.

### V-B. Hallar todos los partidos políticos que tienen algún orador que opinaron sobre “Ley de puertos”

En la imagen 8 se muestra el resultado para esta consulta, retornando los partidos políticos en cuestión.

El código de la consulta se presenta en 5. Para esto, se parte de un nodo *Legislator* con el comando `MATCH`, recorriendo la relación *represents\_party*. Luego, a partir del legislador se recorre la relación *made\_intervention*, y con la intervención obtenida recorriendo la relación *main\_topic*, y por último limitando el tópico obtenido al tipo “Ley de puertos” con el comando `WHERE`. Para retornar, se utiliza el comando `DISTINCT` para retornar solo una vez cada partido político.

#### Listing 5 Código de la consulta 2

```

1 MATCH (l:Legislator)-[:represents_party]->(p:Party), (l)-[:made_intervention]->(i:Intervention),
2 (i)-[:main_topic]->(t:Topic)
3 WHERE t.topic="Ley de puertos"
4 RETURN DISTINCT p as Party

```

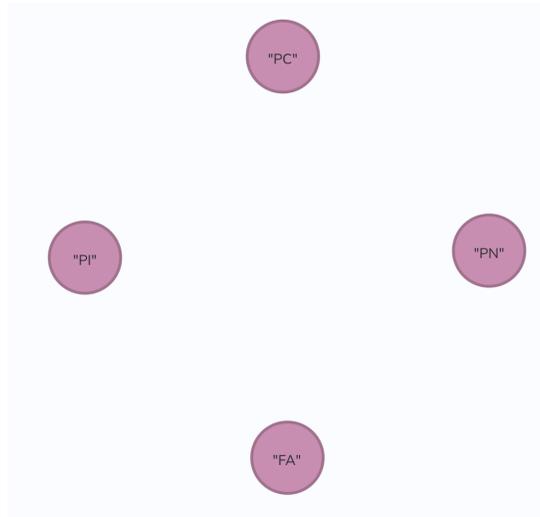


Figura 8: Consulta 2 en Cypher

### V-C. Qué cámara tiene más intervenciones con una emoción particular

En la imagen 9 se muestra el resultado para esta consulta, retornando las distintas cámaras y la cantidad de intervenciones asociadas a la emoción, en este caso *sadness* (tristeza)

El código de la consulta se presenta en 6. Para esto, se parte de un nodo *Chamber* con el comando `MATCH`, recorriendo la relación *session\_number*. Luego, a partir de la sesión se recorre la relación *session\_contains\_intervention*, y con la intervención obtenida recorriendo la relación *main\_emotion*, y por último limitando la emoción obtenida al tipo *sadness* con el comando `WHERE`. Se retorna la cámara y la cantidad de intervenciones utilizando el comando `count`, por último ordenándolas de forma descendiente por la cantidad de intervenciones con el comando `ORDER BY`.

#### Listing 6 Código de la consulta 3

```

1 MATCH (c:Chamber)-[:session_number]->(s:Session), (s)-[:session_contains_intervention]->
2 (i:Intervention), (i)-[:main_emotion]->(e:Emotion WHERE e.emotion="sadness")
3 RETURN c.name as Chamber, count(i) as CountInterventions
4 ORDER BY CountInterventions DESC

```

Chamber	CountInterventions
"CAMARA DE REPRESENTANTES"	192
"ASAMBLEA GENERAL"	51
"COMISION PERMANENTE"	11

Figura 9: Consulta 3 en Cypher

#### V-D. Legisladores de partidos distintos que opinan con una misma emoción sobre un mismo tópico

En la imagen 10 se muestra el resultado para esta consulta, retornando los dos legisladores, sus partidos políticos, el tópico y la emoción.

El código de la consulta se presenta en 7. Para esto, se parte de dos nodos *Legislador* con el comando `MATCH`, recorriendo la relación *represents\_party*. Luego, a partir de los legisladores se recorre la relación *made\_intervention*, y con la intervención obtenida recorriendo la relación *main\_emotion* y *main\_intervention*, se pide que ambos nodos obtenidos sean el mismo, para que las intervenciones tengan el mismo tópico y misma emoción. Luego, con el comando `WHERE` se limitan varias cosas:

- Que los partidos políticos de los legisladores sean distintos, ya que era la esencia de la consulta.
- Que el identificador de un legislador sea mayor que el del otro. De esta manera se evita contener pares repetidos de legisladores.
- Que el tópico no sea “Petición de palabra”, ya que se quiere hallar un tema abordado concreto.
- Que la emoción no sea *others* (otros) ya que no tiene valor para el análisis.

Por último se retornan los valores de interés.

#### Listing 7 Código de la consulta 4

```

1 MATCH (l1:Legislador)-[:represents_party]->(p1:Party), (l2:Legislador)-[:represents_party]->(p2:Party),
2 (l1)-[:made_intervention]->(i1:Intervention), (i1)-[:main_topic]->(t:Topic),
3 (i1)-[:main_emotion]->(e:Emotion), (l2)-[:made_intervention]->(i2:Intervention),
4 (i2)-[:main_topic]->(t), (i2)-[:main_emotion]->(e)
5 WHERE p1 <> p2 AND elementId(l1)>elementId(l2) AND t.topic <> "Petición de palabra" AND
6 e.emotion <> "others"
7 RETURN l1.name as Legislador1, p1.name as Party_Legislator1, l2.name as Legislador2,
8 p2.name as Party_Legislator2, t.topic as Topic, e.emotion as Emotion

```

Legislador1	Party_Legis...	Legislador2	Party_Legis...	Topic	Emotion
"Araujo, Jose German"	"FA"	"Espinosa, Gustavo A."	"PC"	"Seguridad social"	"anger"
"Araujo, Jose German"	"FA"	"Ziminov, Luis Alberto"	"PC"	"Seguridad social"	"anger"
"Araujo, Jose German"	"FA"	"Radio, Daniel"	"PI"	"Seguridad social"	"anger"
"Araujo, Jose German"	"FA"	"Amarilla, Gerardo"	"PN"	"Seguridad social"	"anger"
"Araujo, Jose German"	"FA"	"Arocena, Jose Andres"	"PC"	"Seguridad social"	"anger"

Figura 10: Consulta 4 en Cypher

## VI. CONCLUSIONES Y TRABAJO FUTURO

Se presentó una prueba de concepto para representar a las intervenciones de legisladores del parlamento uruguayo como una base de datos de grafos. Además esta base de datos está enriquecida con un análisis automático de los discursos, asignando a cada uno de ellos una categoría según la emoción que transmite el orador y el tópico del que está hablando.

Si bien los resultados alcanzados no son buenos (e.g. la gran mayoría de los discursos tienen asignado un tópico incorrecto), esta prueba de concepto permite plantear un posible camino a seguir para generar una plataforma que asista a expertos a realizar consultas para realizar análisis complejos de los discursos en el parlamento.

Sobre la representación en la base de datos de grafos se observa que los discursos podrían ser separados en diferentes unidades, como se realiza en el trabajo de [Perak and Rodik, 2018]. Las intervenciones son en general complejas, abordando más de una arista del mismo problema a la vez, refiriéndose o respondiendo a más de un legislador, citando ciertas leyes y otras veces incluso contando anécdotas relacionadas al tema de debate. Cuál separación realizar dependerá de un análisis junto a expertos y de diferentes experimentos que permitan bajar a tierra las posibilidades que ello permita. Tener estas separaciones permitiría también un análisis más cuidadoso y certero, pensando por ejemplo en que actualmente el análisis de emociones se hace para la intervención completa y quizá el orador expresa diferentes emociones en diferentes partes de su discurso. Además esta separación podría también mejorar el desempeño de estos y otros procesamientos automáticos, puesto que el largo de los textos de entrada suele ser un desafío para los modelos de PLN [Jurafsky and Martin, 2023].

Otro aspecto fundamental es que las intervenciones fueron clasificadas (en tópicos y emociones) sin hacerle ningún tipo de preprocesamiento al texto original, en el que se observan nombres propios al inicio de cada texto, palabras que no son semánticamente relevantes (i.e. *stopwords*) y ausencia de símbolos de puntuación que permitan una clara separación en oraciones. Adicionalmente los textos fueron procesados de manera aislada sin tener en cuenta qué era lo que se venía discutiendo previamente. Esto no solamente dificulta el análisis automático, sino que incluso representa una dificultad significativa en caso de que un humano quiera interpretar qué es lo que se expresa (e.g. ¿qué es lo “horroroso” que menciona Singer en el último

ejemplo de la tabla I?). Esto está también relacionado al modelado de las intervenciones; es de conocimiento que los datos con los que se realizó este trabajo cuentan con una versión ordenada por turnos de intervenciones, lo que permitiría también plasmarlo en el modelo de datos. En ese caso se podrían generar relaciones entre intervenciones que den una noción de orden (i.e. qué intervención precede/sucedee a cuál otra), lo que podría dar paso a un análisis discursivo más rico, ya sea con el análisis automático de estas intervenciones encadenadas o con consultas más complejas en Cypher.

En cuanto al análisis de tópicos hay tres aspectos fundamentales a explorar. En primer lugar el método a seguir; el entrenamiento supervisado de un clasificador podría dar mejores resultados, aunque también valdría la pena experimentar con técnicas de *prompting* con modelos de lenguaje [Brown et al., 2020].

En segundo lugar es importante experimentar con diferentes conjuntos de tópicos. Esto es vital en el caso del entrenamiento de un clasificador, puesto que cuanto más tópicos mayor dispersión habrá en la distribución de probabilidad, lo que podría traducirse en que es mejor fusionar ciertos tópicos similares como por ejemplo *aborto* y *eutanasia*.

En tercer lugar está la forma de representar estos tópicos. En caso de que se siga optando por un enfoque basado en *embeddings* es importante experimentar con diferentes descripciones para cada uno de los tópicos. Una posibilidad es que expertos realicen una descripción lo más objetiva posible de cada tópico, intentando no replicar ciertos sesgos. Otra posibilidad es obtener una descripción para estos tópicos mediante un enfoque automatizado basado en *scrapping* (e.g. que se obtenga la definición de “Seguridad ciudadana” de Wikipedia<sup>11</sup> y se la combine con la descripción dada en la página del IMPO<sup>12</sup>).

<sup>11</sup>[https://es.wikipedia.org/wiki/Seguridad\\_ciudadana](https://es.wikipedia.org/wiki/Seguridad_ciudadana)

<sup>12</sup><https://www.impo.com.uy/bases/leyes/16707-1995>

## REFERENCIAS

- [Botzer and Weninger, 2023] Botzer, N. and Weninger, T. (2023). Entity graphs for exploring online discourse.
- [Brown et al., 2020] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners.
- [Dargis et al., 2018] Dargis, R., Auzina, I. A., Bojars, U., Paikens, P., and Znotins, A. (2018). Annotation of the corpus of the saeima with multilingual standards.
- [Gourgaris et al., 2018] Gourgaris, P., Kanavos, A., Karacapilidis, N., and Tampakas, V. (2018). Argumentative discourse concepts as revealed by traversing a graph. In Iliadis, L., Maglogiannis, I., and Plagianakos, V., editors, *14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, volume AICT-520 of *Artificial Intelligence Applications and Innovations*, pages 123–132, Rhodes, Greece. Springer International Publishing. Part 3: MHDW.
- [Jurafsky and Martin, 2023] Jurafsky, D. and Martin, J. H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd - draft edition.
- [Luján et al., 2023] Luján, D., Schmidt, N., and Moraes, J. A. (2023). Estimating parties' policy positions in uruguay: Comparing scaling methods based on legislative speeches and roll-call votes. *Latin American Politics and Society*, page 1–17.
- [Mitchell et al., 2007] Mitchell, T. M. et al. (2007). *Machine learning*, volume 1. McGraw-hill New York.
- [Neo4j, 2023] Neo4j, I. (2023). Cypher manual. <https://neo4j.com/docs/cypher-manual/current/introduction/>. [Online; Accessed: Julio de 2023].
- [Perak, 2020] Perak, B. (2020). Modeling semantic relations from a dependency-based graph: A corpus-based network analysis of croatian parliamentary debates.
- [Perak and Rodik, 2018] Perak, B. and Rodik, F. (2018). Building a corpus of the croatian parliamentary debates using udpipes open source nlp tools and neo4j graph database for creation of social ontology model, text classification and extraction of semantic information. In *Conference on Language Technologies & Digital Humanities*. <https://www.bib.irb.hr/960280>.
- [Pérez et al., 2021] Pérez, J. M., Giudici, J. C., and Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.
- [Van Aggelen et al., 2017] Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., and Beunders, H. (2017). The debates of the european parliament as linked open data. *Semantic Web*, 8(2):271–281.
- [Venkatakrisnan and Schmidt, 2018] Venkatakrisnan, A. R. and Schmidt, N. (2018). A network analytical take on the european parliament.