

**CRISTIAN GONZÁLEZ ARIAS**  
Pontificia Universidad Católica de Valparaíso  
Viña del Mar, Chile  
cristian.gonzalez@pucv.cl

## ***Recepción en twitter de un debate presidencial televisivo a través de la herramienta topic modelling<sup>1</sup>***

### **Introducción**

Podemos considerar las redes sociales como un lugar de ampliación del espacio público tradicional. Twitter en particular, se percibe como un espacio relevante en el ámbito de la discusión política, especialmente en periodos de campaña electoral. Por el volumen de mensajes que se intercambia en determinados periodos de tiempo, sabemos que la actividad de twitter está muy relacionada con lo que ocurre en los medios de comunicación tradicionales. No obstante, desconocemos qué motiva a los usuarios de twitter a comentar. En esta investigación exploratoria, analizamos el comportamiento de los usuarios de twitter ante un debate presidencial transmitido por televisión. El objetivo de esta indagación fue caracterizar la recepción del debate televisivo por parte de los usuarios de twitter, a través de la identificación de los tópicos de las interacciones entre los usuarios de twitter, durante la realización de dicho evento televisivo. Para llevar a cabo esta investigación, recolectamos un corpus de tuits asociados al *hashtag* del debate o a los nombres de los candidatos a la presidencia en los comicios de noviembre de 2017 en Chile. Sobre este corpus, utilizando la herramienta computacional *topic modelling*, identificamos los tres primeros tópicos de conversación en twitter y realizamos una relación entre los tópicos del debate y los tópicos de discusión de los usuarios de twitter.

Este trabajo se enmarca en el ámbito de los estudios del discurso público en la medida de que participa de los esfuerzos por comprender la relación que se establece entre el espacio mediático tradicional y el espacio mediático virtual, construido a través de las redes sociales. El debate, en el que participaron 8 candidatos, se realizó el 06 de noviembre de 2017 organizado por la Asociación Nacional de Televisión (ANATEL) que agrupa 6 canales de televisión chilenos.

En el contexto de esta investigación, entendemos recepción como la reacción que se manifiesta en los mensajes de twitter asociados al evento mediático “debate presidencial”. La noción de tópico la hemos conceptualizado de manera amplia como la respuesta a la pregunta ¿de qué se habla? En el caso del debate televisivo, en el que participaron 8 candidatos, identificamos el tema de cada una de las intervenciones de los candidatos y de las preguntas realizadas por los periodistas a los candidatos. En el caso de los mensajes de twitter, empleamos la herramienta *topic modelling*, que busca las relaciones entre palabras de mayor frecuencia por mensaje, por medio de la técnica *biterm Topic model* (BTM) usando relaciones de trigramas. Recientes trabajos recomiendan la técnica BTM para textos cortos como los de twitter (Jonsson&Stoole, 2016).

### **El estudio del espacio público en el auge de las redes sociales**

Con el auge de las redes sociales, el estudio de lo digital es ineludible y cada vez más importante para la investigación social. La preocupación por el estudio del comportamiento humano en el ambiente virtual se torna fundamental con el aumento explosivo de las redes sociales para comunicarse. De acuerdo con Rill, Reinel, Scheidt y Zicari (2014) la relevancia de twitter en el ámbito político está asociada al dos de sus grandes virtudes: la velocidad y la eficacia en la generación y trasmisión de contenidos. Esta verdadera revolución en las comunicaciones se presenta como un desafío para los científicos sociales tanto en el plano teórico como en el metodológico. Es por tal motivo que desde el análisis del discurso estamos buscando las herramientas para su comprensión y estudio, especialmente con el fin de explorar las características del discurso político y su organización estratégica (Charaudeau, 2005) asociado al desarrollo del espacio público virtual.

### **El debate televisivo y los resultados electorales**

El debate televisivo se realizó 13 días antes de las elecciones presidenciales que fueron el 19 de noviembre de 2017 y contó con la participación de los 8 candidatos. El debate fue organizado por la Asociación Nacional de Televisión, ANATEL, que reúne a 6 de los principales canales de televisión chilenos. El debate duró 2 horas 40 minutos, en el que todos los candidatos contaron la misma cantidad de tiempo para hablar.

El resultado de la contienda electoral estableció las dos primeras mayorías: los candidatos Sebastián Piñera (Derecha) y Alejandro Guillier (Centro-izquierda). Dado que ninguno obtuvo más del 50% de votos, estos dos candidatos pasaron a la segunda vuelta para definir las elecciones entre ellos. El detalle de los resultados electorales se puede apreciar en la Tabla 1.

---

<sup>1</sup> Investigación realizada en el marco del proyecto DI *Modelo predictivo de resultados electorales basado en la interacción comunicativa de usuarios en redes sociales aplicando técnicas de inteligencia computacional y Big Data*. Etapa 2018. Pontificia universidad Católica de Valparaíso.

Candidato	Votos	%
Carolina Goic	387 780	5,9
José Antonio Kast	523 213	7,9
<b>Sebastián Piñera</b>	<b>2 417 216</b>	<b>36,6</b>
<b>Alejandro Guillier</b>	<b>1 497 116</b>	<b>22,7</b>
Beatriz Sánchez	1 336 824	20,2
Marco Enríquez-Ominami	376 471	5,7
Eduardo Artés	33 690	0,5
Alejandro Navarro	24 019	0,4

**Tabla 1: resultados de la primera vuelta electoral de la elección presidencial Chile 2017**

### Procedimientos metodológicos

En primer lugar, se extrajo manualmente los tópicos del debate televisivo. En segundo lugar, se trabajó sobre un corpus de tuits asociados al debate, al que se le aplicó la herramienta *topic modelling* para la extracción automática de los tópicos. Finalmente, se compararon los tópicos abordados en el debate televisivo y los generados en la conversación en twitter.

### Los temas del debate televisivo

En el caso del debate televisivo presidencial, identificamos tópicos utilizando la información que fue proporcionada por los propios organizadores del debate y del análisis de los videos del debate. El debate se dividió en siete partes: la primera de tema libre para cada uno de los periodistas que preguntaron a los 8 candidatos; la segunda parte, estuvo dedicado a la Seguridad; la tercera parte a Economía y la cuarta a Salud. La sexta parte fue de tema libre para los periodistas y la séptima y última parte fue de tema libre para cada uno de los candidatos. El detalle de los temas abordados por cada uno de los candidatos en cada una de las secciones se puede consultar en el Anexo 1.

### El Topic Modelling

Para analizar el corpus de mensajes de twitter utilizamos la técnica *Topic Modelling*, particularmente, la técnica BTM con trigramas. En el campo del aprendizaje automático y el procesamiento del lenguaje natural, el *Topic Modelling* (modelado de tópicos) es un tipo de modelo estadístico construido para identificar "temas" abstractos que están presentes en un conjunto de documentos o mensajes. Esta herramienta permite identificar estructuras semánticas ocultas en un texto. Esta técnica se basa en el supuesto de que si un documento trata sobre un tópico en particular, se espera que determinadas palabras particulares aparezcan en el documento más o menos frecuentemente. Si bien, regularmente un documento se refiere a tópicos múltiples en diferentes proporciones, los "tópicos" se identifican a través de grupos de palabras que coocurren con mayor frecuencia en un documento determinado. Dichos modelos de tópicos también se conocen como modelos de tópicos probabilísticos.

Actualmente, debido a la cantidad de material escrito que se producen en los ambientes virtuales, los modelos de tópicos pueden ayudar a organizar y ofrecer ideas para que comprendamos grandes volúmenes de textos.

### Descripción datos recolectados de twitter

Para el seguimiento de los mensajes emitidos en la red social Twitter, se realizó una búsqueda de todas las menciones a los candidatos, ya sea por su nombre o mención a la cuenta del candidato (@candidato). Posteriormente, se procedió a limpiar los datos de elementos que podían producir ruidos o sesgar los resultados. Estos procedimientos se detallan en el Anexo 2. El corpus quedó constituido por tuits y retuits producidos hasta una hora antes y una hora después de realizado el debate. A continuación, en la tabla 2 se pueden apreciar el volumen de tuits que motivó cada candidato.

Candidato/a	Personas	tuit	retuit
Carolina Goic	2715	3609	12909
José Antonio Kast	<b>4716</b>	4617	11991
Sebastián Piñera	<b>9496</b>	6639	15228
Alejandro Guillier	3734	2316	4476
Beatriz Sánchez	<b>3940</b>	2633	6253
Marco Enríquez-Ominami	<b>3828</b>	4286	11742
Eduardo Artés	480	373	307
Alejandro Navarro	965	734	1766

**Tabla 2: mensajes producidos por candidatos Debate televisivo**

### Resultados

Los resultados incluyen los tópicos que fueron identificados en el debate televisivo y los tópicos identificados en las interacciones de los usuarios de twitter. En el debate televisivo se identificó cuatro tópicos principales propuestos por los organizadores: Seguridad, Economía, Educación y Salud y cada uno de los tópicos derivados

de las intervenciones de los candidatos y a las preguntas de los periodistas. En el Anexo 1 se puede apreciar el detalle de los temas abordados por candidato en el orden en que transcurrió el debate.

En el caso del análisis del corpus de tuits, los 3 tópicos principales identificados son:

Tópico 1: Comentarios a un gráfico distorsionado presentado por Piñera en pantalla

Tópico 2: El indulto de narcotraficantes en el periodo anterior en que gobernó Piñera

Tópico 3: Evaluación de cada uno de los candidatos realizado por una actriz famosa

A modo de ejemplo de la manera como se interpretaron y elaboraron los tópicos, en la tabla 3 se aprecia la forma como la herramienta *topic modelling* presenta la información para el inferencia y elaboración del tópico.

Tópico 1	Probabilidad
27_doble_22	0.011479
impresionante_128584_miente	0.011471
128584_miente_datos	0.011471
miente_datos_distorciona	0.011471
datos_distorciona_proporción	0.011456
distorciona_proporción_cifras	0.011456
proporción_cifras_27	0.011456
cifras_27_doble	0.011456
doble_22_debatea	0.011439
128514_128514_128514	0.010751
<b>Coherence</b>	<b>0.212655</b>

**Tabla 3: datos proporcionados por *topic modelling* para la identificación de un tópico**

De esta información proporcionada por el algoritmo y, por nuestro conocimiento del contexto político y cultural, es que podemos inferir el tópico correspondiente. Cabe mencionar que la serie de números corresponden a emoticones que acompañan el texto, los que también fueron incorporados al análisis. A continuación, describiremos los tres tópicos principales detectados. Dado que se pueden rastrear los mensajes de twitter, ilustraremos los tópicos, con los que son, probablemente, las situaciones o tuits que gatillaron la generación de los tópicos.

#### Caracterización por tópico en twitter

El tópico 1 hace referencia a un gráfico presentado por el candidato Piñera, donde se aprecia una desproporción de las barras que lo conforman. La Imagen 1 muestra el momento en el que el candidato exhibe el gráfico. La reacción de los tuiteros fue de denuncia y burla ante el intento de manipulación de la opinión pública. De esta manera, el tópico aborda el absurdo del gráfico y se acusa al candidato de mentiroso.

Podemos ver que este tópico de la conversación en twitter se relaciona con la intervención del candidato en la segunda parte del debate. En ese momento, el candidato responde a una pregunta sobre el tópico general de Seguridad. En este caso, es notoria la relación del tópico generado en twitter con un momento específico del debate, no obstante, el tópico 1 se construye sobre el cuestionamiento de la honestidad del candidato.



**Imagen 1: Polémico gráfico de Piñera en el debate presidencial ANATEL 2017**

El tópico 2, hace referencia a un indulto que realizó el candidato Piñera mientras era presidente de la república un primer período. Este indulto favoreció ampliamente a narcotraficantes, como lo recuerda el tuit presentado en la Imagen 2. Es fácilmente identificable que el tópico se genera sobre una información que se agrega al debate. Es decir, se trata de un tópico que no fue tratado directamente en el debate, pero que es aportado por un usuario de twitter. Al igual que el tópico anterior, se construye sobre el cuestionamiento de la honestidad del candidato.



**Imagen 2: Tuit representativo del tópico 2**

El tópico 3 se podría calificar como una evaluación de los candidatos. De manera lúdica e irónica, una tuitera califica a cada uno de los candidatos de acuerdo a su desempeño en el debate. Este tuit es el que genera otros tuits y retuits que llega a configurarse como uno de los tres tópicos principales. Claramente, en el tuit se aprecia la preferencia por una de las candidatas y expresa una visión negativa de la mayoría de los otros candidatos. Este tópico no se relaciona con ninguno de los tópicos tratados en el debate, sino que se desprende de un posicionamiento general ante el evento televisivo.



Imagen 3: Tuit representativo del tópico 3

Podemos apreciar que los tres tópicos principales de la conversación de twitter generan contenido que no fue abordado directamente en el debate presidencial. Se trata de información que se desprende del debate, como en el primer tópico, donde hay burlas y repudio al intento del candidato de manipular a los electores, mediante la adulteración de un gráfico. En el tópico 2, se aporta información que relativizan o contradicen las palabras del candidato. En el mismo sentido que el anterior, mientras el candidato expone su plan antidelincuencia, los tuiteros comentan la indulgencia del candidato con narcotraficantes que se beneficiaron de un indulto que el realizó mientras era presidente de la república cuando fue presidente. En el tópico 3 se combina una apreciación general del evento con un acto de apoyo a una candidata en específico. Llama la atención que el tuit también se construye sobre el cuestionamiento de la honestidad del candidato Piñera.

### Conclusiones

Una primera reflexión que motivan los datos es respecto de la paradoja de la exposición negativa de los candidatos en las redes sociales y el éxito electoral. Tenemos que considerar que los tres tópicos identificados en twitter fueron fundamentalmente negativos para el candidato que ganó las elecciones.

También resulta interesante constatar que no podemos asociar el número de menciones con el éxito electoral. Si bien Piñera, ganador de las elecciones, fue el candidato que tuvo más menciones y que movilizó a más personas a hacer comentarios en twitter, el segundo lugar en las elecciones lo obtuvo Alejandro Guillier, quien se posicionó en quinto lugar en la actividad en twitter. Sin duda, esto, cuestiona la idea de que el volumen de votos se relaciona de manera directa con las preferencias electorales.

Podemos sostener además, que los tres tópicos principales de las conversaciones en twitter aportan nueva información al debate. Ya sea porque un tópico retematiza el debate o porque introduce ingenio y humor. Por ejemplo, en el primer tópico, un gráfico de información sobre delincuencia, se transforma en evidencia de una falta ética del candidato o en el segundo tópico, la entrega de información que no fue dicha en televisión, cambia el tenor de tópico que introduce una propuesta de política antidelincuencia al cuestionamiento de la honestidad del candidato. En el caso del tercer tópico, se posiciona claramente como espectador y califica el comportamiento de los candidatos, introduciendo un tono que no se puede dar en el medio tradicional, pero si en redes sociales. Asimismo, se puede constatar que los tres tópicos identificados en twitter intentan introducir una polémica y se configuran en el ánimo de enlodar a una persona, lo que podría identificarse preliminarmente, como una de las características más salientes del debate público virtual en contexto electoral.

Sin lugar a dudas que estas observaciones son parciales y deben ser estudiadas más ampliamente. En trabajos posteriores queremos abordar la actividad de twitter en el periodo general de campaña, analizando todos los mensajes que se produjeron en casi 60 días. Si avanzamos en esta línea de investigación, creemos que podremos conseguir técnicas y procedimientos para comprender mejor el espacio público virtual y su relación con los medios de comunicación tradicionales.

### Referencias bibliográficas

- Charaudeau, P. (2005). *Le Discours politique. Les masques du pouvoir*. Paris: Viubert.
- Jonssón, E. &Stolee, J. (2016). *An Evaluation of TopicModelling Techniques for Twitter*
- Rill, S., Reinel, D., Scheidt, J., &Zicari, R. (2014). PoliTwi: Early detection of emerging political topics on twitter and the impacton concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 24-33.
- Yaqub, U., Chun, S. A., Atluri, V., &Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, 34(4), 613-626.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In European conference on information retrieval (pp. 338-349). Springer, Berlin, Heidelberg.

## ANEXO 1: Estructura de tópicos del Debate Televisivo ANATEL

### Parte 1: Tema libre para los periodistas

C. Goic	Eliminar el PC de la coalición; Salud pública y privada
J. A. Kast	Religión (educación y aborto)
S. Piñera	Austeridad y gasto público (fideicomiso ciego)
A. Guillier	Recolección de firmas y apoyo político
B. Sánchez	Frente Amplio (conflictos y Sharp, alcalde de Valparaíso)
MEO	Alianza con Guillier (narcotráfico)
E. Artés	Recuperación de recursos naturales y apoyo popular
A. Navarro	Vacunas y problemas de salud

### Parte 2: Seguridad

C. Goic	La Legua (blindar casas, intervenciones); sí hay terrorismo
J. A. Kast	Porte de armas
S. Piñera	Narcotráfico (La Legua y Fiscal Peña)
A. Guillier	Politización de las instituciones y fiscalía nacional como herramienta política
B. Sánchez	Luchsinger-McKay; drones
MEO	SENAME controlado por FF.AA. y oposición a las armas
E. Artés	Intervención social a La Legua; Pena de muerte a los narcos
A. Navarro	Registro de las comunicaciones privadas

### Parte 3: Economía

C. Goic	Reforma laboral; Fondo E
J.A. Kast	Achicar el estado (recortes en viáticos y ministerios)
S. Piñera	Reducir viáticos; patrimonio personal
A. Guillier	Inversión extranjera; contaminación norte; tributar en regiones
B. Sánchez	No+AFP
MEO	Banco Central y ahorros
E. Artés	Nacionalización y expropiación de recursos naturales
A. Navarro	Estatización de servicios básicos, terminar Transantiago

### Parte 4: Educación

C. Goic	Gratuidad total a técnicos
J.A. Kast	Ausencia a clases; Educación sexual integra
S. Piñera	Preferencia pensiones sobre gratuidad; Educación parvularia
A. Guillier	Bajar tasa CAE; 70% de gratuidad
B. Sánchez	Condonar CAE; Gratuidad
MEO	Gratuidad; Estatizar deudas
E. Artés	Reformular el carácter de la educación; Nuevo estado patriótico y popular
A. Navarro	Deuda histórica profesores

### Parte 5: Salud

C. Goic	Construir hospitales, oncológicos; Fortalecer sector público
J.A. Kast	Sexualidad (Método natural)
S. Piñera	Camas nuevas; construcción; listas de espera; cultura vivir sano
A. Guillier	Sedentarismo; estrés; ordenar recursos
B. Sánchez	Seguro único de salud;
MEO	Traer médicos extranjeros; CESFAM y EUNACOM
E. Artés	Derecho asegurado; expropiar Clínica Las Condes
A. Navarro	Médicos de planta sector público; Especialistas

### Parte 6: Tema libre para los periodistas

C. Goic	Guillier más de lo mismo
J.A. Kast	Reducir el estado; Indulto a condenados por DD.HH.
S. Piñera	Transparencia; Minas y LAN
A. Guillier	Araucanía (No terrorismo)
B. Sánchez	Impuesto súper ricos
MEO	Perseverancia
E. Artés	Representación real de izquierda
A. Navarro	Asamblea constituyente; AFP; Venezuela

### Parte 7: Mensaje final candidatos

C. Goic	Nueva constitución
J.A. Kast	Dios, patria y familia
S. Piñera	Tiempos mejores
A. Guillier	Inclusión
B. Sánchez	Mujer, regionalismo, medioambiente
MEO	Empleo
E. Artés	Izquierda verdadera
A. Navarro	Recursos naturales

### Anexo 2 Procedimientos de “limpieza” de los datos

<p>Los procedimientos de limpieza de datos, corresponden a la eliminación de:</p> <ul style="list-style-type: none"> <li>- Enlaces (<a href="http://direccion.co">http://direccion.co</a>)</li> <li>- Cuentas de usuarios (@Usuario)</li> <li>- El token “RT”, indicando que el mensaje corresponde a un re-tweet</li> <li>- Se remueven stopwords*</li> <li>- Para cada candidato se remueven los nombres de los mismos en distintas variaciones (nombres con y sin tildes, variaciones del nombre mal escrito, y nombre y apellido juntos)</li> <li>- Para cada candidato se remueven cada conjugado con la palabra vota/vota, por ejemplo votaKast, votaGoic, votopiñera, etc.</li> <li>- Símbolos (por ej: #, !, ?, etc.)</li> <li>- Para los eventos, se eliminan los hashtags relacionados al evento (#debatetelevisivo)</li> </ul>
---