

# ¿Los movimientos sociales provocan cambios culturales?: una propuesta de metodología computacional desde el análisis del discurso en los medios de prensa aplicada al movimiento #MeToo

Fabian Ruiz

**Resumen**—The analysis of discourse in news media about social movements, has been present on different studies regarding social science. Taking in account the objective of social movement is to make cultural changes, and that news media reflects the society, it is in the interest of social investigations to understand how social movements impacts news media. Nevertheless, given the high volume of news produced by the news media everyday, it is complex to make a general analysis about the media. Even though, there have been approaches that analyze the news media trough data science techniques, is hard to replicate, since these studies usually have an qualitative approach. Regarding this, the objective of this paper, is to create a computational protocol, that is complementary with the analysis of social science investigators, that helps to analyze the changes on news media regarding social movements. To do this, we propose computational protocol that is based on entity recognition and topic modeling, to analyze representativity and thematics presented on news media, on different levels, regarding a social movement. To illustrate the replicability of the protocol, we analyze the case of study of the movement #MeToo, and we analyze if there is a change on discourse in Chilean news media. Finally, we make statistical analysis regarding correlation and distribution, to understand if there is a change on representativity or thematic, provoked by the #MeToo movement.



## INTRODUCCION

El análisis del discurso de los medios de comunicación en el contexto de movilizaciones sociales ha sido objeto de extensos estudios. Leopold & Bell (2017) destacan que es a través de los medios que los movimientos sociales pueden comunicar su mensaje y elaborar estrategias para atraer la atención. En sí, el movimiento social se define justamente como un desafío colectivo para generar un cambio en una estructura de poder establecida, representada por un élite, un adversario o una autoridad (Meyer & Tarraow, 1998). +

Por lo tanto el objetivo principal de los los movimientos sociales, es lograr un cambio en la cultura. Tomando esto en cuenta, los investigadores en ciencias sociales, han demostrado interés en analizar el impacto de los movimientos sociales en la sociedad, específicamente mediante los medios de prensa (Haller et al, 2006).

Si bien, existen diversos estudios enfocados en analizar el cambio en el discurso de los medios de prensa debido a movimientos sociales (Leopold & Bell, 2017), tiende a ser un análisis reducido debido al enfoque manual, en donde se analiza una cantidad pequeña de medios para crear conclusiones generales con respecto a los medios de prensa. Cabe destacar, que esto se produce dada la gran cantidad de medios existentes, al igual que la cantidad de noticias producidas por estos, por lo que hacer análisis a gran escala resulta complejo. Por otra parte, existe interés desde el área de la ciencia de

• *Fabian Ruiz is with Austral University of Chile*

•

los datos para lograr un análisis más profundo de los medios (Bahamonde et al, 2018) , sin embargo tienden a tener enfoque cualitativo, por lo que es complejo usar la misma metodología computacional para poder reproducirla en distintos casos de estudio.

Por lo tanto, objetivo de esta investigación es la de crear un protocolo computacional complementario, escalable y reusable ,basada en técnicas de procesamiento natural del lenguaje y topic modeling, para el análisis del discurso frente a un movimiento social.

### Trabajos relacionados

A diferencia de estudios como el presentado por Maier et al (2018), en donde el objetivo es explicar un método específico para uso general , el objetivo de esta investigación es desarrollar un protocolo específicamente para el análisis del cambio en el discurso frente a movimientos sociales, que permita reusabilidad.

Por otra parte frente al estudio de Prochazka, & Schweiger (2018) , en donde el objetivo es analizar los medios de comunicación basado en metodologías manuales, en esta investigación, el enfoque es computacional , dado que permite la escalabilidad.

## METODOLOGIA

Para ilustrar nuestra metodología de análisis del discurso en los medios de prensa, se tomó un caso de estudio en particular, el movimiento # MeToo ocurrido en Chile. En primera instancia, el movimiento # MeToo comienza a generar atención en el 24 de octubre de 2017, mediante un tweet de Alyssa Milano sobre las acusaciones de abuso sexual de Harvey Weinstein (Mendes et al, 2018) . A partir de esto comienzan ramificaciones del movimiento en distintas partes del mundo, incluyendo Chile. Si bien existen papers que analizan el movimiento en Estados Unidos, para Chile, son limitadas. De esta forma, el objetivo de usar # MeToo como caso de estudio, es el de analizar el impacto en la sociedad chilena, mediante en análisis de los estereotipos asociados a las mujeres en Chile, antes y después del movimiento # MeToo. Entonces, para nuestro caso de estudio, la

hipótesis sería: ¿El movimiento social # MeToo, provocó un cambio en los medios?

Tomando este caso de estudio, se implementó nuestro protocolo computacional, con el fin de obtener variables que puedan ser usadas para hacer análisis estadístico, específicamente relacionado a la distribución y a la correlación, para responder nuestra hipótesis.

### Construcción de dataset

Para esto, se construyeron dataset a partir de las palabras claves "mujer", "emprendimiento" y "ciencia", por medio de un buscador de noticias para Chile basado en Elasticsearch y mongoddb, en donde cada noticia recolectada, tiene información del medio que publicó la noticia y un timestamp asociado . Estas palabras fueron escogidas para analizar estereotipos asociados a mujeres en la sociedad. Para cada palabra clave, se recolectaron noticias durante dos periodos, cada uno de un año, partiendo desde Octubre del 2016. Para cada periodo se recolecto un total de 1600 noticias, dando un total de 9600 noticias. De esta forma, se tiene por cada palabra clave, dos periodos que ayudan a analizar el cambio en el discurso de los medios de prensa sobre las mujeres. Finalmente, una vez recolectado los 6 dataset, se filtro según su relevancia mediante los algoritmos de Elasticsearch. Junto a esto, se hizo un preprocesamiento del dataset, para eliminar las palabras que no aportan información.

### Análisis de topics

Para comprender que temáticas estaban presentes en cada dataset, se uso una metodología computacional de topic modelling llamada LDA (Latent Dirichlet Allocation). LDA, es un modelo generativo que permite explicar un conjunto de datos observados (Noticias en este caso), mediante grupos no observados (Temáticas), mediante distribuciones probabilísticas de Dirichlet (Blei, 2003). Así, se genera la distribución de topics mediante un hyperparametro alpha, y la distribución del vocabulario por topico, mediante un hyperparametro eta .Tomando esto en cuenta, ambos parametros deben ser afinados con el objetivo de tener el mejor resultado con respecto a los topics generados por

LDA. Junto a esto, para encontrar la cantidad optima de topicos, se debe calcular la perplejidad, el cual es el inverso de la verosimilitud por palabra, y minimizar la perplejidad segun la cantidad de topicos. Por otra parte, para analizar los topicos de forma cualitativa, se revisan las primeras 10 palabras segun su calculo de perplejidad por topico, para hacer un analisis y entender de que se esta hablando en ese topico. Mediante esta metodologia, se analizaron los topicos generados por LDA para cada dataset por separado, y luego en el conjunto de los dos anos por cada palabra clave, para analizar si existia un cambio en los topicos debido al movimiento # MeToo.

### Analisis de representatividad

Para entender la representatividad real de las mujeres en los dataset de noticias, se uso una metodologia de procesamiento natural del lenguaje llamada reconocimiento de entidades nombradas (RCEN). Esta metodologia , permite clasificar y separar entidades, con el fin de estructurar la informacion. En este caso, el objetivo es el de contabilizar la aparicion de las entidades separados por genero, "mujer" y "hombre", para analizar la proporcion en la cual se habla de cada genero, para cada dataset, y separado por periodo.

### Definicion de variables

Con respecto a estas metodologias computacionales (LDA, y reconocimiento de entidades nombradas), se definieron las siguientes variables a considerar al momento de hacer el analisis estadistico:

- Topico: Tematica presente en varias noticias
- Tasas de genero: Proporciones calculadas mediante RCEN.
- Nivel de granularidad: Forma de agrupar los distintos medios existentes al momento de hacer el analisis, segun diferentes criterios (localidad, ideologica politica , clasificacion segun el dueno y ecosistema completo)
- Periodo de tiempo: Ventanas discretas de un mes, agrupados en un periodo de

un ano, antes y despues del movimiento #MeToo.

Con estas variables en cuenta, se puede analizar el discurso en los medios, en varios niveles de granularidad, mediante la representatividad y las tematicas para cada dataset, para analizar holisticamente si el movimiento # MeToo, provoco un cambio en los medios de prensa.

### Analisis estadistico

Para poder analizar metodicamente si existe un cambio en la representatividad como en los topicos, provocado por el movimiento # MeToo , se hicieron analisis estadisticos de dos tipos, uno de distribucion y otro de correlacion. Para el analisis de distribucion, se usaron tecnicas de visualizacion de la informacion, para analizar la evolucion de la distribucion de la variable de los topicos, frente al tiempo. De esta forma, se puede ver cada topico, y si es que se presenta o no un cambio luego del movimiento.

Por otra parte, para analizar el cambio de representatividad, se hizo un analisis de correlacion con respecto a la tasas de genero, para analizar, en caso de que hubiese un cambio sustancial antes y despues del movimiento MeToo, si fue el movimiento el que provoco el cambio en las tasas de genero.

## REFERENCIAS

- Meyer, D. S. and Tarrow, S. 1998. 'A Movement Society: Contentious Politics for a New Century', in Meyer, D. S. and Tarrow, S. (eds) *The Social Movement Society*, pp. 1-28. Lanham, MD: Rowman & Littlefield.
- Leopold, J., & Bell, M. P. (2017). News media and the racialization of protest: An analysis of Black Lives Matter articles. *Equality, Diversity and Inclusion: An International Journal*, 36(8), 720-735. doi:10.1108/edi-01-2017-0010
- Bahamonde, J., Bollen, J., Elejalde, E., Ferrer, L., & Poblete, B. (2018). Power structure in Chilean news media. *Plos One*, 13(6). doi:10.1371/journal.pone.0197150
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3), 93-118. doi:10.1080/19312458.2018.1430754
- Prochazka, F., & Schweiger, W. (2018). How to Measure Generalized Trust in News Media? An Adaptation and Test of Scales. *Communication Methods and Measures*, 1-17. doi:10.1080/19312458.2018.1506021
- Haller, B., Dorries, B., & Rahn, J. (2006). Media labeling versus the US disability community identity: A study of shifting cultural language. *Disability & Society*, 21(1), 61-75. doi:10.1080/09687590500375416