



# Estereotipos de Género en las Revistas Brando y Ohlala

Diego Kozlowski - Fernando Gonzalez - Alfredo Rolla - Jorge Federico Cubells

July 26, 2018

## 1 Introducción

En épocas donde el rol de la mujer en la sociedad está cambiando, y donde se plantea de manera constante la ruptura de estereotipos de género en todos los ámbitos de vida en sociedad, nos propusimos investigar esta temática desde la óptica de cómo los medios gráficos abordan la perspectiva de género.

Los medios gráficos son instrumentos poderosos para mantener y/o cuestionar las ideologías de género de nuestra sociedad, y nuestra hipótesis principal de trabajo es que las revistas orientadas al público de cada género se sustentan en léxicos y tópicos diferenciados.

El **Objetivo General** de este trabajo de investigación es "Analizar los estereotipos presentes en las revistas de consumo masivo y su evolución en los últimos 10 años, mediante técnicas de procesamiento de lenguaje."

Los **Objetivos Específicos** son:

1. Topic Modeling
  - Construir un modelo de TM con el corpus completo (ambas revistas) para todos los años.
  - Seleccionar tópicos que resulten de interés a la investigación.
  - Analizar la participación de cada uno de estos tópicos en cada revista en el tiempo.
2. Palabras / Listado de Palabras
  - Realizar una selección de palabras que resulte de interés seguir.
  - Ver la evolución de la presencia de estas palabras en el tiempo, de forma condicional a cada revista.
  - Analizar si la probabilidad de encontrar esas palabras condicionadas a cada revista es significativamente diferente para cada corpus.
  - Comparar con hitos históricos relacionados.

Las **Hipótesis de Trabajo** utilizadas son:

1. A lo largo del período bajo análisis, se mantiene la separabilidad de los corpus de cada revista por tópicos estereotipados, es decir que no refieren a diferencias genuinas entre el hombre y la mujer.

2. No obstante la hipótesis 1, existe un movimiento en el tiempo, donde desaparecen ciertos tópicos de contenido más marcadamente misógino y aparecen ciertos tópicos nuevos, que expresan la superación de tabúes.
3. Otros tópicos se mantienen en el tiempo, asignando roles aspiracionales diferenciados entre hombres y mujeres. Por ejemplo, como tópicos relevantes para los hombres los "negocios" y el "consumo suntuario" (aspiraciones económicas), y en la revista de mujeres se promueve la pseudociencia, como el "Horóscopo", marcando una diferenciación respecto a las formas de "superación personal".

## 2 Metodología

El software utilizado para las tareas de obtención y procesamiento de los datos fue Python 3 [10]. Dentro del software citado, las principales librerías utilizadas fueron:

- Pandas: para manejo de datos. [14]
- Numpy: de cálculos científicos para manejo de datos multidimensionales. [13]
- Scikit Learn: especializada en machine learning [1].
- NLTK: para trabajar con datos del lenguaje natural. [6]
- Gensim: software robusto y eficiente para realizar modelado semántico no supervisado a partir de textos [15].

La fórmula para el cálculo de la probabilidad de aparición de una palabra dada un a revista es:

$$P[W_i|R_j] = \frac{\#W_{ij}}{\#W_{*j}}$$

$W_i$  : Palabra en particular.

$\#W_{ij}$  :Total de esa palabra para esa revista.

$\#W_{*j}$  :Total de palabras en el corpus para esa revista.

$R_j$  : Revista j.

La fórmula para el cálculo de la probabilidad de aparición de un tópico dada una revista es:

$$P[T_i|R_j] = \frac{\sum_{d=1}^{d=m} P(T_{ijd})}{\#d}$$

$T_i$  : Tópico en particular.

$P(T_{ijd})$  : Probabilidad del tópico i para la revista j para el documento d.

$\#d$  : Número de documentos.

$R_j$  : Revista j.

Para realizar la evaluación de la significancia estadística de la asociación entre las revistas, en el caso del análisis de palabras y listas de palabras por años, se empleó el test exacto de Fisher, que permite testear dicha significancia a partir de una tabla de contingencia.

## 3 Selección y Limpieza del Corpus

Para la realización de este trabajo se utilizaron dos revistas disponibles online, cuyo target de público es respectivamente el colectivo masculino y femenino.

### 3.1 Selección y Obtención del Corpus

Tanto Ohlala como Brando son revistas del Diario La Nación que se encuentran disponibles online. Debido a que buscábamos con este trabajo detectar la evolución de los tópicos y palabras a lo largo de los años, y teniendo en cuenta que ambas revistas se publican desde mediados de la década del 2000, se procedió a obtener los artículos publicados.



Figure 1: Revistas de Género

La definición fue tomar los artículos de los últimos 10 años para cada revista. Para ello elaboramos un web scraper que incluye dos pasos principales:

1. **Obtención de las URLs:** Se hizo un análisis inicial de la página web, a partir de la cual se detectó la existencia de una API interna en La Nación. Dicha API permite ir navegando por los artículos históricos de las revistas mediante el botón "siguiente".

Los inputs de dicha API son:

- *Magazine:* puede ser Brando, Ohlala, Hola, Lugares, Living, Jardín, Editorial y Tecnología. Nosotros trabajaremos con las dos primeras.
  - *Número de página (p):* de la historia de los artículos donde nos encontramos. Iniciamos en la página 1.
  - *Número de artículos a devolver en cada página (c):* Seteado en 100 artículos por página.
2. **Obtención de los Documentos:** el resultado es un diccionario de python con la siguiente estructura:

*Url: [Magazine, Fecha, Titulo, Texto]*

Donde:

**Url:** Link al artículo recuperado.

**Magazine:** Nombre de la revista.

**Fecha:** Fecha normalizada ISO-8601 (YYYY-MM-DD).

**Título:** Título del artículo, limpio de código html y caracteres especiales.

**Texto:** Texto del artículo, limpio de código html y caracteres especiales.

Como resultado de este web scraping obtuvimos en principio 7.779 artículos en Brando y 22.309 artículos en Ohlala.

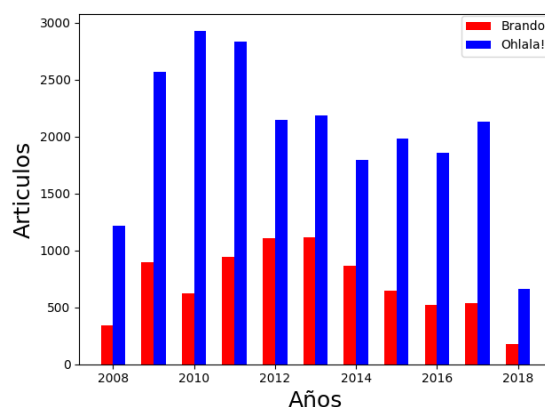


Figure 2: Número de Artículos por Revista

Aclaremos que en principio ese es el número de revistas, ya que veremos que luego de la limpieza dicha cantidad se redujo considerablemente debido a la existencia de artículos vacíos (ejemplos: encuestas, imágenes sin texto, etc.).

### 3.2 Limpieza del Corpus

Los pasos seguidos fueron:

1. **Carga y Join de Corpus:** Se leyeron los corpus de Brando y Ohlala y luego se realizó el join.
2. **Normalización de Fechas:** Solamente nos interesa el año de publicación de cada artículo, por lo tanto, se extrajo dicha información de las fechas y se la incorporó como feature.
3. **Limpieza:**
  - Se armaron dos listados, uno de palabras y otro de frases, que serían excluidos del corpus. Las palabras son aquellas ambiguas y comunes en nuestro lenguaje, y otras que pueden sesgar el análisis (ejemplo ohlaleando, brando). Las frases son especialmente Disclaimers que aparecían al final de la mayoría de los artículos (ejemplo: "Los comentarios publicados son de exclusiva responsabilidad de sus autores...", "Enviar un comentario implica la aceptación del Reglamento.").
  - Eliminación de Stopwords del Español.
  - Remoción de acentos.
  - Reemplazo de signos de puntuación por un espacio. Esto es así porque de no incluir el espacio hacía que las palabras se unieran.
  - Llevar a minúsculas todas las palabras.
  - Reemplazo de todos los valores numéricos por la palabra "NUM".
4. **Stemming:** recorte de las palabras de manera de generalizarlas. Ejemplo: Biblioteca y Bibliotecario a Bibliotec
5. **Destemming:** Implica llevar cada "stemm" a una palabra que tenga más sentido al leerla. Esto se hace asignando la palabra completa que mayor frecuencia tenga en cada stemm. Bibiotec a Biblioteca, asumiendo que esta aparece con mayor frecuencia que Bibliotecario.
6. **Eliminación de Artículos Vacíos:** el total de artículos se reduce de cerca de 30 mil a 24.142, repartidos en 6.060 artículos en Brando y 18.082 en Ohlala.

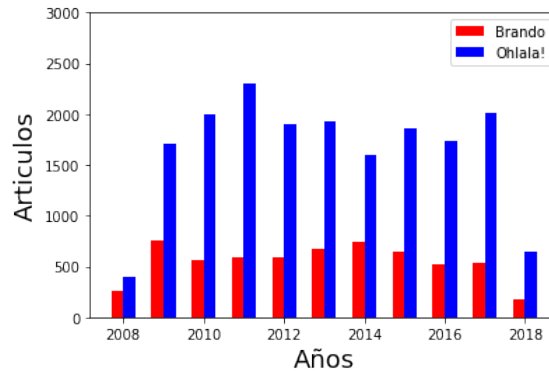


Figure 3: Número de Artículos por Revista post Limpieza

Como un primer análisis exploratorio buscamos conocer cuáles eran las palabras que mejor distinguían a cada revista.

Para ello armamos una wordcloud por revista, filtrando aquellas palabras con una probabilidad mayor a 0.001 y con un odd superior a 2.

Definimos  $\text{odd}$  como el cociente entre la probabilidad que la palabra  $W_i$  aparezca en la Revista 1, dividido la probabilidad que dicha palabra aparezca en la Revista 2.

$$\text{Odd} = \frac{P(W_i | Revista_1)}{P(W_i | Revista_2)}$$



Figure 4: Wordcloud por Revista. Palabras que distinguen

Notamos que en el caso de la revista masculina, las palabras que la distinguen son Vino, Mundo, Modelo, Argentina, Ciudad, Gente, Cocina, Auto, Sistema, Equipo, Marca, Mujer/es entre otras. Para el caso de la revista femenina son Vida, Día, Tiempo, Pelo, Casa, Color, Moda, Podemos, Todas, Amor, Piel, entre otras.

Es pertinente aclarar que la palabra "vs" (abreviación de versus) aparece seguido en artículos anteriores a 2011 en Brando, y proviene de notas cuya tematica era la comparación de modelos mujeres. Esto se evidencia con la aparicion de palabras como "hot", "mujeres" y "modelo" en el wordcloud.

Esto nos da pautas de los temas que se podrían estar tratando en ambas revistas, con algunos más relacionados a Autos, Negocios, Sociedad y Política, Tecnología (Brando), y otros más relacionados a Moda, Cuidado Personal, Casa y Familia (Ohlala).

## 4 Análisis de Tópicos

## 4.1 Aplicación del Modelo

Para detectar los tópicos del corpus se aplicó el modelo estadístico generativo denominado Latent Dirichlet Allocation (LDA) desarrollado por David Blei et al (2003) [2].

Con este modelo a partir de la matriz Término-Documento se obtuvieron dos matrices: la matriz Término-Tópicos y la matriz Tópicos-Documentos. En la primera tendremos la distribución de las palabras en los distintos tópicos y la segunda la distribución de los tópicos en los documentos.

Los parámetros usados por este modelo son:

- *Input*: corpus luego de la limpieza.
- *Máximo de Iteraciones*: se definió en 50.
- *Método de Aprendizaje*: seteado en "Online" que se corresponde al Método Variacional Bayesiano.
- *Semilla*: se definió una semilla para que arroje los mismos resultados para poder repetir los experimentos.
- *Número de Tópicos*: definido en 100 tópicos.

Una vez entrenado el modelo, se lo aplicó sobre el corpus para obtener las matrices Término-Tópicos y Tópicos-Documentos. Con este procedimiento pudimos conocer qué tópicos son tratados en cada documento de las revistas, y cuáles son las palabras más representativas de cada uno de los 100 tópicos.

## 4.2 Resultados

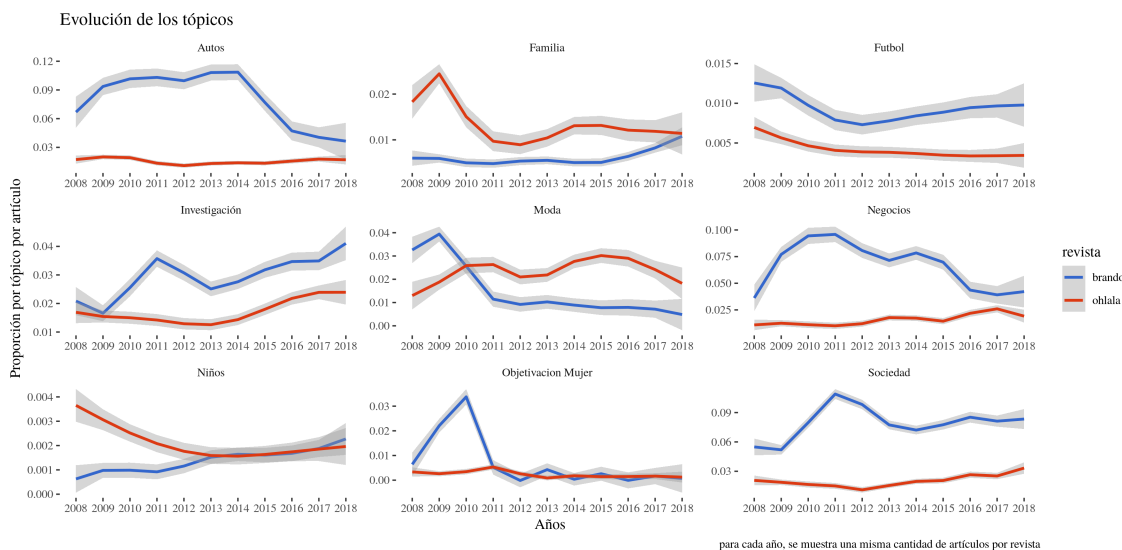
### 4.2.1 Tópicos Relevantes

Analizando las 10 palabras más importantes de cada tópico, y atendiendo a las hipótesis de trabajo definidas, seleccionamos 9 tópicos.

Nro Tópico	Nombre Asignado	Top 10 Palabras
1	Objetivación	natalia hot ana emma romina versus diez camilo morochas mega
4	Mujer	
5	Negocios	empresa redes sistema comprar productos mercado traves tecnologia permite desarrollo
7	Autos	version auto marca modelo presente equipo llego motor argentina electrica
21	Niños	ninos adultos educativo colegio chiquito padre secuestro change sauna pegote
50	Moda	moda diseno estilo marca coleccion ropa tendencia prendas rosa zapatillas
71	Familia	hijos madre mama padre bebe familia papa embarazo regalo anos
82	Sociedad	anos argentina mundo historia pais publico primera habia libro mundial
93	Investigación	estudio problema trabajo explica ley medico social generar desarrollo investigar
	Fútbol	juego jugar futbol cancha programa ganas copa radio pelota gol

Table 1: Tópicos Seleccionados

#### 4.2.2 Evolución en el Tiempo



En cuanto al tópico **"Objetivación de la Mujer"** notamos que claramente a los inicios de ambas revistas, Brando utilizaba este tópico en mayor proporción que Ohlala. Aunque también notamos que a partir de 2012 los niveles bajaron considerablemente, quizás con un cambio editorial acorde a las nuevas épocas.

El tópico de **"Niños"** estaba claramente mayor representado en Ohlala, como una temática de mayor interés para las mujeres. Sin embargo hacia 2013 se nota que las proporciones de este tópico en ambas revistas se asemeja, incluso tocándose en mayor medida en Brando a partir de 2017.

Los tópicos **"Negocios"**, **"Futbol"** y **"Sociedad"** son mayormente tratados en Brando a lo largo del tiempo. Sin embargo en Negocios vemos cómo la proporción en Brando cae de niveles de 0.075 a 0.040, aunque aún por encima de Ohlala. Suponemos que Negocios es un tópico mayormente abordado en Brando porque se asocia al hombre con emprendedor, hombre de negocios. Con lo cual en principio se vería corroborada la hipótesis 3. Las temáticas asociadas a "Sociedad" (política, historia, etc.) también se tocan más en Brando, probablemente asociado al estereotipo de hombre como principal actor en la vida política y social de un país. Y finalmente Futbol con marcada tendencia masculina, asumiendo una preferencia por este deporte por encima de las mujeres.

Cabe resaltar el comportamiento del tópico **"Familia"**. Durante la historia es mayormente tratado en la revista femenina, aunque vemos que en 2018 dicha diferencia de proporciones desaparece. La familia pasa de ser un tema poco tocado en Brando (valores cercanos a 0.005) a un tema que comienza a incorporarse a la línea editorial (0.011). Aquí podría estar indicando el cumplimiento de las hipótesis 1 y 2, ya que la familia era más asociada a la mujer (rol de madre, ama de casa) pasando luego a ser un tema cada vez más compartido (se observa el movimiento en el tiempo planteado en H3).

Dos temas marcadamente de preferencia de cada uno de los géneros son los **"Autos"** para los hombres y la **"Moda"** para las mujeres (aunque aquí vemos que a los inicios entre 2008 y 2010 era un tema mayormente tocado por la revista de hombres).

EL tópico asociado a **"Investigación"** ha visto incrementada su participación en ambas revistas a lo largo de los años, aunque la brecha entre la revista de hombres y mujeres también se hizo mayor.

## 5 Análisis de Palabras y Listas de Palabras

### 5.1 Selección de Palabras/Listado de Palabras

Para la selección del listado de palabras a observar en cada revista y en el tiempo, nos basamos en:

1. Listados de Palabras en páginas web que hablaran sobre los temas que nos interesaban para este trabajo
2. Listados elaborados por el equipo de investigación

Para la temática "**Feminismo**" elegimos los listados de palabras de los links:

- Diccionario Feminista [9]
- Feminismo de la 'A' a la 'Z' [7]
- Glosario feminista para principiantes [12]

Allí encontramos palabras como Cosificación, Machismo, Micromachismo, Falocentrismo, Feminismo, entre otras.

Para la temática "**Sexista**" o maneras machistas de referirse a las mujeres el link usado fue de "Lenguaje Sexista para describir mujeres" [3]. Las palabras relevantes son por ejemplo Bruja, Frígida, Puta, entre otras.

En cuanto a "**Ciencia**" consultamos la página web "100 palabras relacionadas con tecnología" [5] con palabras como Algoritmo, Científico, Computación, Economía, entre otras.

Para el grupo de palabras relacionadas a "**Negocios**" consultamos el artículo "Diccionario emprendedor ¡50 términos que debes conocer!" [11] con palabras como Capital, Riesgo, Emprendedor, Marketing, Innovación, entre otras.

Para lo "**Trendy**" o tendencia, elaboramos un listado propio con términos como Tablet, Iphone y Gaming.

Para "**Moda**" usamos el artículo "Diccionario de palabras utilizadas en la moda" [8] con palabras como Escote, Chic y Outfit.

Otra temática de interés es "**Horóscopo**", donde en un listado de elaboración propia incluimos palabras como Horóscopo, y los signos del zodíaco como Tauro, Aries y Géminis.

Finalmente para los "**Tabúes**" elaboramos un listado propio, con términos como Masturbación, Orgasmo, Menstruación y Swinger.



## 5.2 Evolución en el Tiempo

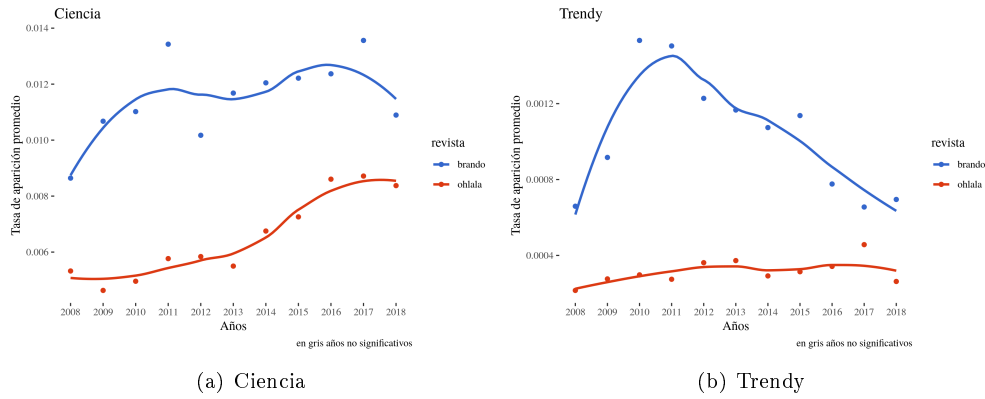


Figure 5: Palabras relacionadas a Ciencia y Trendy

En cuanto a las palabras asociadas a **"Ciencia"** y **"Trendy"** notamos que son mayormente utilizadas en la revista de orientación masculina con una diferencia significativa a lo largo de los años.

Sin embargo, la revista de orientación femenina comenzó a utilizar léxico asociado a Ciencia, con lo cual podríamos estar ante evidencia de deconstrucción de estereotipos, dándole más relevancia a estos temas para las mujeres (H2).

Las palabras asociadas a Trendy están usándose en menor proporción a lo largo de los años en Brando.

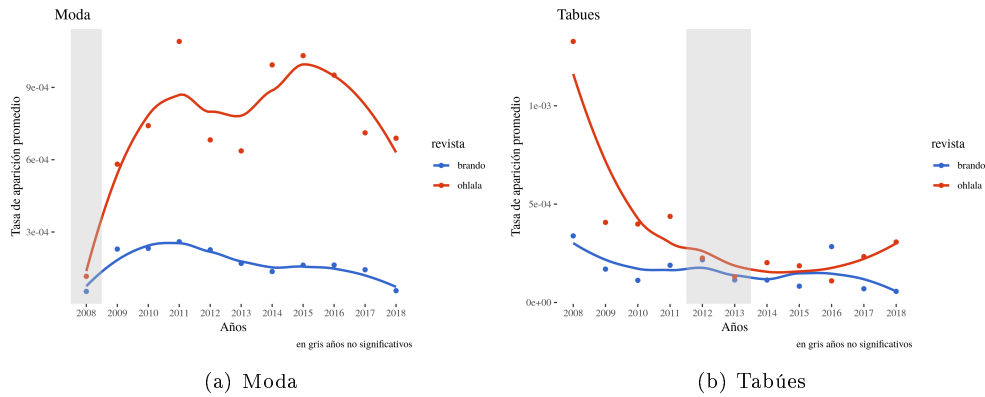


Figure 6: Palabras relacionadas a Moda y Tabúes

Observando las palabras asociadas a **"Moda"** al contrario de lo observado en los gráficos anteriores, vemos que son palabras mayormente utilizadas en la revista femenina. Excepto el año 2008 las diferencias son siempre significativas.

Moda se mantiene siempre en niveles altos en Ohlala, y declinando en el tiempo en Brando pero usándolas en baja proporción en esta última.

Las palabras sobre **"Tabúes"**, eran ampliamente usadas en Ohlala por sobre Brando en los primeros años, y si bien se mantienen en mayor uso en dicha revista con diferencias significativas (excepto 2012, 2013 y 2016), los "gaps" son menores.

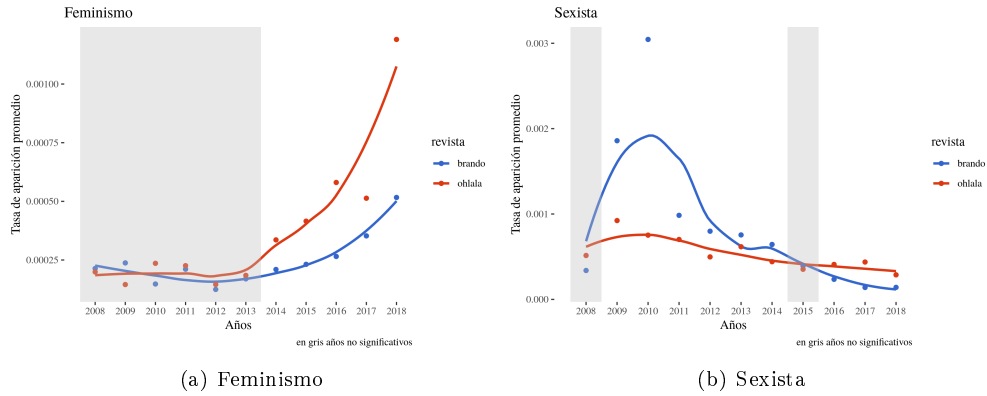


Figure 7: Palabras relacionadas a Feminismo y Sexista

Referido las palabras asociadas a **"Feminismo"** y **"Sexista"**, claramente notamos un cambio de comportamiento a lo largo de los años (H2).

Feminismo en los años iniciales se usa en bajas proporciones en ambas revistas y las diferencias son no significativas. Es a partir de 2014 que toman relevancia en ambas, aunque en mayor medida en Ohlala, seguramente relacionado a un cambio cultural.

Las palabras sexistas eran mayormente usadas en Brando, con diferencias significativas versus ohlala (excepto 2008 y 2015). Sin embargo en 2016 las curvas se cruzan y dichas palabras son usadas en mayor proporción en ohlala.

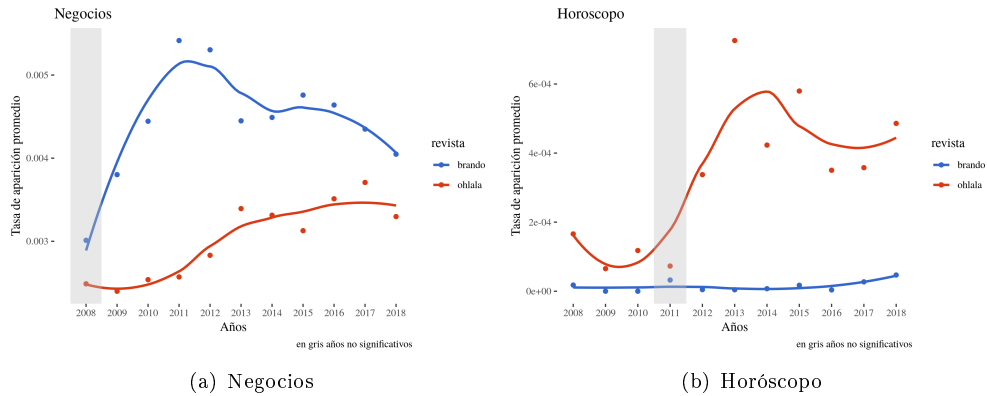


Figure 8: Palabras relacionadas a Negocios y Horóscopo

Abordamos ahora las palabras asociadas a **"Negocios"** y **"Horóscopo"**.

En Horóscopo, excepto el año 2011, las diferencias son siempre significativas y orientadas hacia el público femenino. Inicia con niveles bajos de utilización en Ohlala, aunque luego de 2011 se convierte en un tema recurrente en la magazine femenina.

Negocios en cambio (excepto 2008) siempre muestra diferencias significativas con clara orientación al público masculino. Aunque también, como ocurría en Ciencias, se devela un comportamiento ascendente en las proporciones de uso en ohlala, aunque siempre por debajo de Brando.

A partir de este último análisis podemos afirmar el cumplimiento de la H3 *"Otros tópicos se mantienen en el tiempo, asignando roles aspiracionales diferenciados entre hombres y mujeres."* donde citábamos los ejemplos que acabamos de tratar.

## 6 Conclusiones

A través de este trabajo de investigación pudimos abordar el objetivo general, analizando dos revistas de consumo masivo con orientación a públicos femenino y masculino, mediante de procesamiento de lenguaje.

Para el contraste de las hipótesis inicialmente planteadas utilizamos dos vías de análisis. Mediante los tópicos pudimos comprobar que hay temas que usualmente son planteados desde las líneas editoriales como propios del interés de cada género. Tal es el caso de Autos, Fútbol, Investigación, Negocios y Sociedad para los hombres, y Familia, Moda y Niños para las mujeres. Esto confirmaría la Hipótesis 1.

En cuanto a la Hipótesis 2, notamos que temas como Familia y Niños, antes mayoritariamente orientados al público femenino, con el pasar de los años fueron convirtiéndose en temas tratados por ambas editoriales, acorde al mayor rol asumido por el hombre en el seno familiar.

Analizando los Listados de Palabras, corroboramos que hay temas como Ciencia, Trendy y Negocios que son mayormente utilizados en la revista masculina, y otros como Horóscopo y Moda en las femeninas.

Más allá de poder mostrar un interés de hombres y mujeres a ciertos temas, también es cierto que dichos temas son de interés para el género opuesto, muchas veces con igual intensidad.

En cuanto a la hipótesis 2, con el listado de palabras Sexistas notamos claramente una evolución a la baja en su utilización en la revista masculina, probablemente asociada a un cambio cultural. Para destacar es el tema del Feminismo, donde a partir de 2014 se nota un amplio crecimiento en su proporción de utilización, acorde al surgimiento de fuerzas sociales tendientes a lograr la igualdad de trato y oportunidades para el colectivo femenino.

Con respecto a la hipótesis 3, corroboramos cómo el estereotipo de las líneas editoriales tienden a hablarles de Negocios a los hombres, y del Horóscopo a las mujeres. Suponemos que esto podría tener relación con un pensamiento de tintes machistas, donde es el hombre quien emprende y lidera los negocios, relegando a la mujer a un rol pasivo y librado a la suerte o destino.

Este trabajo puede servir de puntapié inicial para futuras investigaciones que amplíen el abanico de revistas de género analizadas, y que al ser de sumo interés para la sociedad en general, puede servir para motivar las discusiones necesarias en este sentido.

Se puede acceder al material de este trabajo ingresando a través del link de github que contiene los códigos y principales resultados obtenidos [4].

## References

- [1] A. Bakharia. Topic modeling with scikit learn. 2016.
- [2] Michael I. Jordan David M. Blei, Andrew Y. Ng. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [3] Fucsia.co. Diccionario sexista: 10 términos que discriminan. <https://www.fucsia.co/belleza-y-salud/bienestar/articulo/lenguaje-sexista-para-describir-mujeres/51880>.
- [4] Koslowski-Gonzalez-Rolla-Cubells. Material de soporte. [https://github.com/DiegoKoz/gender\\_stereotype\\_magz](https://github.com/DiegoKoz/gender_stereotype_magz).
- [5] Joaquín Moreno. 100 palabras relacionadas con tecnología. <https://jackmoreno.com/2017/07/21/100-palabras-relacionadas-con-tecnologia/>.
- [6] J. Perkins. Python 3 text processing with nltk 3 cookbook. *Packt Publishing Ltd.*, 2014.
- [7] Carlota E. Ramírez. Feminismo de la 'a' a la 'z': todas las palabras que deberías saber. [https://www.huffingtonpost.es/2018/02/06/feminismo-de-la-a-a-la-z-todas-las-palabras-que-deberias-saber\\_a\\_23348909/](https://www.huffingtonpost.es/2018/02/06/feminismo-de-la-a-a-la-z-todas-las-palabras-que-deberias-saber_a_23348909/).
- [8] Rocío Reyes-Ewest. Diccionario de palabras utilizadas en la moda. <http://www.modapreviewinternational.com/2013/09/diccionario-de-palabras-utilizadas-en-la-moda-fashion-glosario/>.
- [9] Glenda Rosado. Diccionario feminista: 16 conceptos y palabras esenciales que todos debemos conocer. <https://www.vix.com/es/poder/178423/diccionario-feminista-16-palabras-esenciales-y-sus-significados>.

- [10] Drake F. L. Rossum G. V. Python reference manual. page 9117–9129, 2006.
- [11] Luis Sandoval. Diccionario emprendedor ¡50 términos que debes conocer! <https://www.emprender-facil.com/es/diccionario-emprendedor/>.
- [12] Beatriz Serrano. Glosario feminista para principiantes. [https://www.buzzfeed.com/beatrizserranomolina/vocabulario-glosario-feminista-para-principiantes?utm\\_term=.yt9503B95#.xsl9lYwj9](https://www.buzzfeed.com/beatrizserranomolina/vocabulario-glosario-feminista-para-principiantes?utm_term=.yt9503B95#.xsl9lYwj9).
- [13] Varoquaux G. van der Walt S., Colbert S.C. The numpy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13:22–30, 2011.
- [14] McKinney W. and Team. Pandas - powerful python data analysis toolkit. 2015.
- [15] Sojka P. Řehůřek R. Gensim - statistical semantics in python. *EuroScipy, (Paris)*, 8:25–28, 2011.