

Feminist Media Studies



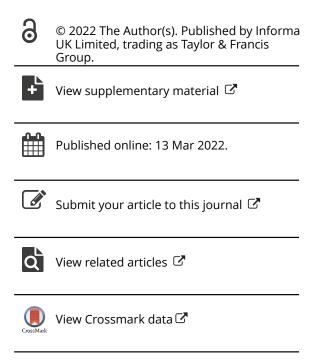
ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rfms20

Large-scale computational content analysis on magazines targeting men and women: the case of Argentina 2008-2018

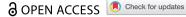
Diego Kozlowski, Gabriela Lozano, Carla M. Felcher, Fernando Gonzalez & Edgar Altszyler

To cite this article: Diego Kozlowski, Gabriela Lozano, Carla M. Felcher, Fernando Gonzalez & Edgar Altszyler (2022): Large-scale computational content analysis on magazines targeting men and women: the case of Argentina 2008-2018, Feminist Media Studies, DOI: 10.1080/14680777.2022.2047090

To link to this article: https://doi.org/10.1080/14680777.2022.2047090









Large-scale computational content analysis on magazines targeting men and women: the case of Argentina 2008-2018

Diego Kozlowski pa, Gabriela Lozano*, Carla M. Felcherc*, Fernando Gonzalezd and Edgar Altszyler (1)d,e

^aFaculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; bInterdisciplinary Institute of Gender Studies, University of Buenos Aires, Buenos Aires, Argentina; anstituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), CONICET-Universidad de Buenos Aires, Buenos Aires, Argentina; ^aMaestría en Explotación de Datos y, Descubrimiento del Conocimiento, Departamento de Computación, Universidad de Buenos Aires, Argentina; eDepartamento de Computación, FCEyN, Universidad de Buenos Aires. Instituto de Investigación en Ciencias de la Computación, Argentina

ABSTRACT

Differences in magazines content aimed specifically at women or men are a means to create and reproduce gender stereotypes. Novel computational tools allow to study differences in magazines content taking into account all available articles. In this study, we analyse the case of two Argentinian magazines published by the same publishing house over a decade (2008–2018), advertised by the publishing house as targeting women and men respectively. Using computational tools, we are able to analyse more than 24,000 articles, which would have been an impossible task using manual content analysis methodologies. With Topic Modelling techniques we identify the main themes discussed in the magazines and quantify their different frequency between magazines over time. Then, we performed a word-frequency analysis to validate this methodology and extend the analysis to other subjects. Our results show that topics such as Family, Business and Women as sex objects present an initial bias that tends to disappear over time. Conversely, in Fashion and Science topics, the initial differences are maintained. Also, we identify a considerable increase in the use of words associated with feminism since 2015 and specifically the word abortion in 2018. Furthermore, we develop a website where everyone can perform additional analysis.

ARTICLE HISTORY

Received 5 February 2021 Revised 25 January 2022 Accepted 21 February 2022

KEYWORDS

Magazines; natural language processing; digital humanities; gender bias; computational social science; topic modelling

Introduction

Representations of gender in any society cannot be understood without considering the political and cultural intersections that comprise the context in which they were produced and maintained (Butler 1998). Gender is a historical construction (Joan W. Scott 1986) deeply embedded in the use of discourse (Marisol Del-Teso-Craviotto 2006). The relevance of popular media for the production and reproduction of stereotypes in societies has been

CONTACT Diego Kozlowski 🔯 diego.kozlowski@uni.lu 🗈 Faculty of Science, Technology and Medicine, University of Luxembourg L-4364, Esch-sur-Alzette, Luxembourg

*These authors contributed equally

Supplemental data for this article can be accessed here

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

widely studied (Gaye Tuchman 1979; Janice Winship 1987). One of the most important theoretical frameworks on the role of the media as a stereotype reproductive agent is the Social Cognitive Theory (Bandura Albert, Ross Dorothea and Sheila A. Ross 1963). This theory states that behavioral patterns and attitudes can be acquired by observing symbolic models, both those observed in real life and those shown in popular media. Therefore, portrayal of women and men in stereotypical roles and frames in mass media serves as a source of generation and reproduction of gender stereotypes (Marisol Del-Teso-Craviotto 2006).

Over the years, several authors have reported the reproduction of stereotyping associations by mass media outlets such as magazines (Ervin Goffman 1979; Marín Murillo, et al. 2010), movies (Ramiro H Gálvez, Tiffenberg Valeria and Edgar Altszyler 2019; Katy Gilpatric 2010; Kimberly A. Neuendorf, et al. 2010), newspapers (M. Pilar Matud, Carmen Rodríguez and Inmaculada Espinosa 2011), radio (Martin Eisend 2010), music (Jacob S Turner 2011; Carolina Spataro 2013), and television (Martin Eisend 2010; Mallika Das 2011; Roger Desmond and Anna Danilewicz 2010; Heather J. Hether and Sheila T. Murphy 2010; Nathalie Koivula 1999). A revision on content analysis in mass media has been reported by Rebecca L. Collins 2011 and Rena M. Rudy, Lucy Popova and Daniel G. Linz 2010. The Global Media Monitoring project 2020 (Sarah Macharia 2020) shows that globally, news stories have not challenged gender stereotypes since 2005.

Another piece of evidence of stereotyping association is the existence of differential content in the media expected to be consumed by men or women specifically. This bias strengthens and perpetuates differences between the symbolic role models of genders (Marisol Del-Teso-Craviotto 2006).

The use of coding schemes for content analysis represents a standard approach for studying stereotyped roles (Kimberly A Neuendorf 2011), and for several decades, these schemes have been used to quantify elements of static categories. However, categories designed in the recent past might become outdated in the short-term, due to fast sweeping social changes (Yorgos C Zotos and Eirini Tsichla 2014). These changes require new categories to be built increasingly frequently. In addition, these categorizations need to be specifically adapted to each study, which poses a problem, as it may require careful manual labor on each specific research.

Between 2008 and 2018, relevant events around the expansion of women's rights and sexual diversity took place in Argentina. These series of events have been the focus of attention and debate in the mass media as well as in social media. The rise of the feminist agenda in Argentina developed between 2008 and 2018 on three levels: feminist activism, government representatives, and progressive legislation. As examples of the increased mobilization capacity of feminist grassroots organizations we can mention the "Slutwalk" march (2011), the ni una menos march protesting femicides, which gathered a million people in Buenos Aires in 2015; and the growing participation in the Encuentro Nacional de Mujeres (national women's meeting), which gathered 5000 women in 2008, and 200,000 in 2019. Changes in the government representatives were also experienced, with the election of Cristina Fernández de Kirchner (in 2008), the first woman to be elected president in Argentinian history; a change also seen in other countries in the region, like Chile (in 2006) and Brazil (in 2011). Several acts were passed by the congress as a reflection of these changes: the Human Trafficking Prevention and Punishment and Victims Assistance Act (N° 26,364) in 2008, and the Comprehensive Act on Protection to

Prevent, Punish and Eradicate Violence against Women in the Areas in which they Develop their Interpersonal Relationships (N° 26,485) in 2009. The Equal Marriage Act (2010) and the Gender Identity Act (2012) placed Argentina as an international legal rolemodel for LGBTQ+ rights. The Voluntary Pregnancy Termination Act deserves a special chapter in the history of women rights in Argentina. With the unprecedented debate that took place in 2018, both in the congress and the streets, although the act was not passed at that time, the massive debate laid the groundwork for its passing two years later. Our aim in this article is to explore the contents of two Argentinian magazines which target segmented audiences (BRANDO and OHLALÁ) within the context of the feminist movement rise in Argentina (2008-2018).

The emergence of the feminist movement transformed women and men's daily lives in Argentina in different spheres, such as family structure, child rearing, and the way of dressing and consuming. It also transformed values and cultural forms towards a more diverse and respectful society where women are not treated as objects, they can make decisions about their own bodies and participate in all professional spheres on equal terms with men. As the literature has shown, movements can influence the mass media agenda (Candi Carter Olson 2016). In this regard, we wonder if these social changes are reflected in the editorial lines of the media in Argentina. Has there been any improvement in the stereotypical representation of men and women in mass media? Have any stereotypes been preserved or reinforced during this period? These questions guide our research.

In the present study, we introduce a novel methodology for a semi-automatic analysis of gender bias in magazines content. Our corpus is based on the online articles from OHLALÁ and Revista Brando, two Argentinian magazines from the same publishing house that target women and men, respectively. The fact that they come from the same publishing house provides a unique opportunity for comparison. We apply computational techniques to identify the topics present in 24,000 articles between 2008 and 2018 from these magazines, and then, we analyse content evolution and differences between them over the years. As our methodology is able to capture a multiplicity of phenomena, in this article we focus only on those topics that we considered most relevant. Nevertheless, we created the website: http://magbias.dc.uba.ar/ where the reader can complement the present analysis with their own research questions.

The main contributions of this article are the following:

- (1) Our case study context presents a unique time and place. The decade of 2008–2018 in Argentina is marked by the emergence of the feminist movement, and fast and strong changes in cultural patterns. Because of this, we expect to see changes in the content of magazines in such a short period (historically speaking).
- (1) We conducted a case study on two magazines from the same publishing house, which allows us to monitor the "editorial effect" of the content differences.
- (2) This methodological proposal allows us to be thorough in our case study. We are able to see the evolution of topics/words over time for each magazine. We illustrate this with a selection of results, but our tool allows for the study of a plethora of hypotheses. We created a website where the readers can formulate their own research questions regarding the vocabulary changes of these magazines.

(3) Our analysis corroborates previous studies, finding biases. It also accomplishes a level of detail that allows us to see which biases remain over time, and which changed.

Methods

In this section we describe the technical approaches of content analysis used to conduct this research. Firstly, we present the compiled dataset and its characteristics. Secondly, we describe a Topic Modelling technique employed for automatic detection of topics in the dataset. Thirdly, we outline a Word Frequency Analysis used to validate the automatic content detection. Finally, we describe the dataset preparation for these technical approaches. Given that the articles are originally in Spanish, the content analysis was performed in this language. A translation into English of the selected words and topics is provided.

Data collection

The corpus compiled for this paper consists of 6,060 articles downloaded from *Brando* magazine, and 18,082 articles from *OHLALÁ* magazine, published between 2008 and 2018 (Magazine links 2018). Figure 1 shows the distribution of articles over time for each magazine. All available articles online were gathered. According to the commercial information of each magazine (Comercial LA NACION [online] 2020), *Brando* magazine's target is men between 30 and 50 years old. It has a net sale of 10,000 copies, and 4,500 subscribers. *OHLALÁ*'s target are women between 25 and 45 years old. It has a net sale of 37,500 copies and 15,000 subscribers. Also, when launched, *OHLALÁ* was presented as

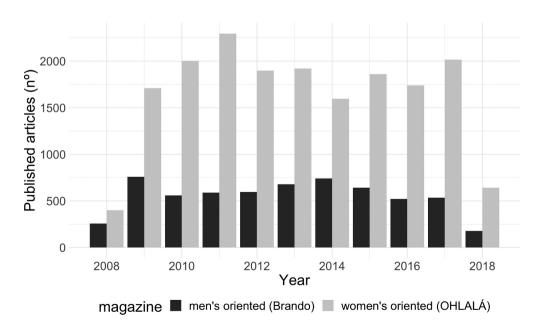


Figure 1. Articles published by year for men and women-oriented magazines, Brando and OHLALÁ.

a womens' magazine (La Nacion 2008). Nevertheless, it is important to highlight this is information provided by the publishing house, which is meant to convince potential advertisers. No public information is available regarding the actual readership of each magazine.

Topic modelling

We implemented a Topic Modelling technique, the Latent Dirichlet Allocation (LDA) (David M. Blei, Andrew Y. Ng and Michael I. Jordan 2003). The main function of LDA is to automatically identify the topics that best describe a dataset. This approach provides the opportunity of a semi-automatic analysis, as the words that describe each topic arise from the corpus itself and not from the choice of the researcher. In this sense, the researcher does not need to previously study the jargon used in the magazine in order to define the relevant topics, as the most repeated terms are automatically retrieved and thematically sorted by LDA. Since this model has been extensively explained by Blei et al. (2003), it will only be described briefly herein. Given a corpus of documents, LDA makes the assumption that there is a set of fixed topics that describe the content. LDA models each article as a distribution over topics, and each topic as a distribution over words. This can be thought of as the process of writing an article, in which the author first chooses the main topic, and then those themes condition the words that are more likely to appear in the manuscript. After processing the texts, LDA automatically computes the probability distribution for each topic on each article. Then, each topic obtained by the LDA can be represented by a list of its most probable words. Given that list, a researcher in the field needs to analyse the lists of words and assign a label to each emergent topic. For example, a list composed of ten words including children, mother, mom and father, among others, can be manually labelled as family.

In the present study, we fed the LDA with all the articles from both magazines and instructed the model to construct 100 topics. Given the table containing the 100 emerging topics and the 10 most probable words defining each of them, we selected and tagged the specific topics that we considered of interest regarding gender stereotypes, without prior knowledge of how they were distributed across magazines (Supplementary Table 1).

Then, for each magazine, we calculated the proportion in which each labelled topic is present for each year $(P_{M,Y}(T_i))$. As an example, this would be read as the proportion of the family topic in the OHLALÁ magazine in 2010.

This is accounted by

$$P_{M,Y}(T_i) = \frac{\sum_{d} P_{d,M,Y}(T_i)}{\# d_{M,Y}}$$
(1)

Where $P_{d,M,Y}(T_i)$ is the probability of the topic i (T_i) for document d in magazine M and year Y, and $\#d_{M,Y}$ is the number of documents for magazine M and year Y.

The main and novel contribution of Topic Modelling is the large-scale automatic identification of recurring topics in texts. In addition, it allows to quantify the presence of these topics among the magazines, thus revealing the differences present in magazines targeting women and men.

It is worth noting that not all inferred topics are interpretable (Jonathan Chang, et al. 2009). Therefore, an expert from the field is required to select the topics that she or he considers most coherent and most relevant for the hypothesis testing. Another relevant factor to consider is that the inferred topics can contain unexpected associations that arise from the discursive context of the corpus (John W Mohr and Petko Bogdanov 2013). To validate the LDA analysis in this study, we also performed a Words Frequency analysis as described below.

Words frequency analysis

We used the standard approach based on word counting. First, a list of unambiguously manually selected words was assembled to represent each topic. Second, the frequency of occurrence of these words for each year and each magazine was calculated. The frequency F of occurrences of the word-list I, in the year Y and magazine M was calculated as

$$F_{M,Y}(I) = \frac{\#W_{I,M,Y}}{\#W_{*,M,Y}} \tag{2}$$

Where $\#W_{I,M,Y}$ is the number of occurrences of words in the wordlist I in magazine M and year Y, while $\#W_{*,M,Y}$ is the total number of words in magazine M for that given year (Jean-Baptiste Michel, et al. 2011).

Finally, for each word list, we compared the time series of the D_{M.Y}(I) frequencies of both journals, and used a fisher-exact test to assess whether differences are significant between both magazines. The words that have polysemy were excluded from the lists to avoid overlapping in word meanings. Given that the vocabulary used in the magazines is composed of thousands of words, and that we are building lists of words of undefined sizes, the possible combinations of words is enormous. Therefore, there are several thousands of potential tests. This generates a statistical problem known as "multiple comparison problem", meaning that it is possible to find a combination of words that provide statistically significant differences where there are none, just by chance. This has to be avoided, otherwise it could become a source of misleading conclusions. Thus, the lists have to be carefully constructed and theoretically justified before experimentation. In this work, we construct word lists based on the Topic Modelling results, consequently avoiding this problem. We also build three lists of words that did not appear automatically in the Topic Modelling results, but we consider they pose relevant theoretical issues: the rise of the feminist movement in Argentina during the period of study, together with the abortion legalization debate in the Argentinian congress, and finally the increasing consumption of pseudoscience, represented in the use of zodiac signs terminology.

Dataset preparation

Each article was labelled with its release year t, and the magazine of origin (OHLALÁ or Brando). The stop words, like the, they, to, etc, were removed using the Spanish stop words list in the NLTK Python package (Steven Bird, Ewan Klein and Edward Loper 2009). The list of the 500 most frequent words in Spanish was reviewed and a selection of 319 words



without semantic content related to our analysis (e.g. friendship, love) was also removed from the dataset. In addition, specific words that refer to one or another magazine (like their names and derived neologisms) were also removed.

For the Topic Modelling (LDA), each word was replaced with its stem form (a short version of the word with suffixes removed (Bird et al. 2009), and after the Topic Modelling step, a de-stemming was performed. Considering that in the present study the frequency analysis was implemented as a control, the stemming and de-stemming procedure was not performed for this case.

Summary

The proposed methodology is a combination of automatic Topic Modeling, together with a word frequency analysis. We consider that this provides the following advantages:

- (1) LDA automatically discovers recurrent themes that are present in the corpus, which is beneficial for the formulation of hypotheses.
- (2) LDA automatically represents each theme according to the probability of occurrence of each word in it. This is a major advantage, because it avoids the researchers' need to build lists of words for each topic to study. The manual construction of word-lists is time-consuming and can potentially introduce new biases. The LDA generated lists of terms is contextual to the used corpus, and therefore it captures the specific jargon within the magazine. This model avoids the problem presented by Zotos et al. (2014) of outdated categories and vocabulary.
- (3) When LDA assigns the distribution over topics for each article, it considers all words present in the document, and not only the salient terms that characterize a topic. It can also assign articles to multiple topics. Therefore, LDA is able to consider words in their context.
- (4) A complementary word frequency analysis allows to validate the results obtained by the LDA, and also, to study new content that could be omitted by the LDA due to its low frequency in the texts

In the following section, we will analyse the results of these techniques.

Results

Automatic topics detection

To evaluate the results from Topic Modelling, we run the LDA model over the full corpus, i.e., both women-oriented and men-oriented magazines, for 100 topics. We extracted the ten most relevant words for each topic (Supplementary Table 1). After careful analysis of the list of emergent topics, we selected and labelled six of them to perform the following experiments based on the relevance of these topics on gender stereotypes (Table 1).

Table 1. Selected topics extracted from LDA analysis, manually assigned tags and ten most probable words automatically detected in each topic. The words underlined appeared originally in English in the articles. In italic, a translation for the words originally in Spanish is provided.

id	Assigned Tag	Top 10 words
1	Women as sex object	natalia hot ana emma romina versus diez (ten) camilo morochas (brunettes) mega
4	Business	empresa (company) redes (networks) sistema (system) comprar (to buy) productos (products) mercado (markets) traves (crossing) tecnologia (technology) permite (it allows) desarrollo (development)
7	Children	niños (kids) adultos (adult) educativo (educative) colegio (school) chiquito (tiny) padre (father) secuestro (kidnap) change sauna pegote (goop)
21	Fashion	moda (fashion) diseño (design) estilo (style) marca (brand) colección (collection) ropa (cloth) tendencia (trend) prendas (garments) rosa (pink) zapatillas (sneakers)
50	Family	hijos (children) madre (mother) mama (mom) padre (father) bebe (baby) familia (family) papa (dad) embarazo (pregnancy) regalo (gifts) años (years)
82	Science	estudio (study) problema (problem) trabajo (work) explica (explains) ley (law) medico (medic) social (social) generar (to generate) desarrollo (development) investigar (to research)

Automatic content analysis

To evaluate whether the topics addressed in each magazine have a gender related content bias, we estimated the probability of each topic for each article. Then, we compared the evolution of the topic probability over time for each magazine, using GAM smoothing (Trebor J. Hastie and Robert J. Tibshirani 1990) in a percentage scale (Figure 2).

Our results show a significantly higher rate of appearance of the topic business in the men-oriented magazine compared to the women-oriented one, although this gap shrunk over time. The opposite occurred with the children topic, which at the beginning of the

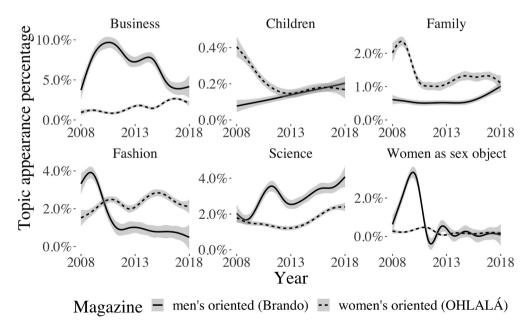


Figure 2. Topic appearance percentage over time. Each colour represents a magazine. GAM smoothing of the observations (Hastie et al. 1990) with 95% confidence intervals in grey.

analysed period presented much more frequent appearance in the women's magazine, but around 2013 this difference disappeared. The same trend was presented by the topic family, although the differences remained detectable until 2018.

The fashion topic had a particular behavior over the years. While it begins with greater representation in the men's magazine, a decreasing tendency was observed over time. In 2010 this topic was equally approached in both magazines, and then it continued to be overrepresented in the woman-oriented magazine without gap closure in 2018. One of the topics with more abrupt changes in the rate of appearance was women as sex objects, which had a very recurrent appearance in the early years of the men's magazine, but after peaking in 2011, an abrupt decrease is observed, and it practically disappeared. The topic science begins with equal presence in both magazines. However, its appearance in the men-oriented magazine increased around 2010, generating a difference between the magazines that was sustained over time.

Overall, this automatic analysis displayed several differences in the content of these magazines. It was observed that most of the evaluated topics changed over the years, in the women as much as in the men-oriented magazine. In addition, while some of the topics biases were reverted over time, some other topics remain associated with one of the two magazines.

Validation of automatic content analysis

Since the topics inferred by the LDA may contain unexpected associations that come from the discursive context of the corpus (Mohr et al. 2013), we supported these results performing a Word Frequency Analysis on the dataset.

To perform the frequency analysis, we manually selected non-ambiguous words that defined each of the six topics of interest (Table 2).

For each topic we considered the ten words resulting from the LDA analysis (Table 1) and excluded the ambiguous or non-representatives. For the science topic we use only the word science itself, because we consider that the words within this topic might be found in other contexts unrelated to science.

Figure 3 shows the evolution of the relative word frequency compared to the total words used in each magazine each year (Equation 2). In the majority of cases, the frequency of occurrence of the most representative words of each topic reproduced the trend observed in Figure 1.

Table 2. Words selected to represent each topic in the frequency analysis. The words underlined were used in English. In italic, a translation for the words used in Spanish.

Topic	Words
Family	hijos (children), madre (mother), mamá (mom), padre (father), bebé (baby), familia (family), papá (dad)
Children	niños (kids), adulto (adult), colegio (school)
Business	empresa (company)
Fashion	ropa (cloth), diseño (design), estilo (style)
Science	ciencia (science)
Women as sex objects	hot, morocha (brunette)

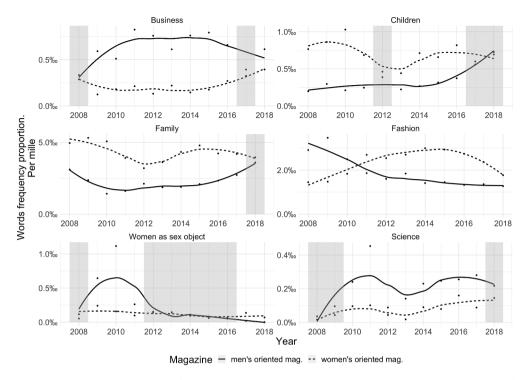


Figure 3. Frequency evolution of the words in the word-lists. Solid lines are a LOESS smoothing of observations. The years with non-significant differences between the magazines are shaded in grey.

The exception to this is the frequency analysis of words associated with *children* topic, whose behavior is more similar to the *family* topic than to its own topic. This result can be interpreted as either that this topic is not coherent enough or that we labelled the topic erroneously. This is an example of how this verification step is important when using Topic Modelling.

Figure 4 shows the frequency evolution of the word-list containing the word horoscope (horóscopo) and non-polysemic zodiac signs [horoscope (horóscopo), Taurus (Tauro), Aries, Gemini (Géminis), Scorpio (Escorpio), Sagittarius (Sagitario), Capricorn (Capricornio), Aquarius (Acuario), Pisces (Piscis) and Virgo]. We also show the frequency evolution of the word abortion (aborto), and the frequency evolution of words associated with feminism [feminist (feminista), feminists (feministas) and feminism (feminismo)].

The frequency of words associated with *horoscope* is very high and keeps increasing over the years in women-oriented magazine. On the contrary, it remains at low values in the men-oriented magazine (Figure 4a). The word *abortion* appears very infrequently in both magazines, with the exception of 2018, when the frequency of this word is significantly higher in the women-oriented magazine than in the men-oriented magazine (Figure 4b). Finally, regarding the words associated with *feminism*, an increase is observed in the appearance of them in both magazines since 2015. However, these words were more represented in the magazine oriented to women than in the one oriented to men (Figure 4c).

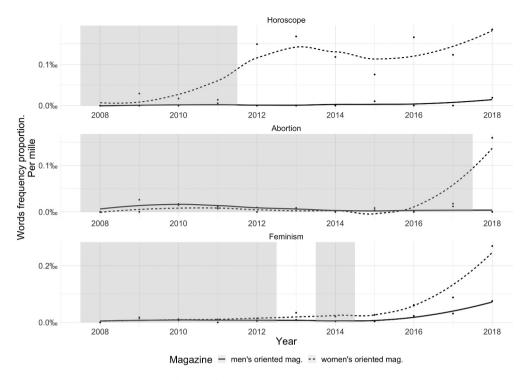


Figure 4. Frequency evolution of words for a: Horoscope, b: Abortion and c: Feminism. Solid lines are a LOESS smoothing of observations. The years with non-significant differences between the magazines are shaded in grey.

Discussion

The concept of gender identity is an operational term that intends -in a political process-to enlarge visibility and legitimacy (Butler 1998). Gender, says Butler, "is not always constituted coherently or consistently in different historical contexts, because gender intersects with racial, class, ethnic, sexual and regional modalities of discursively constituted identities" (Butler 1998).

Several studies have analysed differential content in gender-oriented magazines. One of the first texts that made a comparison of female and male press is Alice E. Courtney and Sarah W. Lockeretz 1971, which analysed seven magazines that targeted both male and female readers. Over the years, these analyses were performed mainly by manual content quantification, presenting a limitation on the number of articles that were included.

Using a Topic Modelling technique, we were able to successfully analyse 24,000 articles from two magazines, one targeting men and the other targeting women, in a period of 10 years. By complementing this model with Word Frequency Analysis, we were able to automatically identify five relevant topics (business, family, women as sex objects, fashion and science). This method allows to discover themes that are present in the corpus and it represents each theme according to the occurrence probability of each word in it. Unlike the classic method based on static word lists, the words that represent the topics are specific to the historical moment and the environment, thus solving concerns raised by Zotos et al. (2014) about the effects of strong cultural changes.

Our analysis shows that there are differences in the appearance frequency of topics between both journals, and that these differences were consistent with well-studied gender stereotypes that reinforce traditional social roles (Murillo et al. 2010).

Our findings are aligned with the literature regarding oversexualization of women (Edward Downs and Stacy. L. Smith 2010), portrayal of women as responsible for family matters (Mallika Das 2011; Scott Coltrane and Michelle Adams 1997), association of science and business to men and not to women (Matud et al. 2011; Courtney et al. 1971) and association of fashion with women (Murillo et al. 2010; Deana B. Davalos, Ruth A. Davalos and Heidi S. Layton 2007; José I Baile, María Gabino-Campos and Ana Laura Pérez-Lugoa 2020).

Despite the fact that there are studies which characterize horoscopes content in magazines targeting women and girls (Edson C. Tandoc and Patrick Ferrucci 2014), little research has been carried out to address the difference between media targeting men and women regarding content associated to horoscope and astrology.

From a static point of view, our analysis confirms the presence of content bias, already described in the literature. However, especially in such a hectic period for Argentinian society, the longitudinal approach revealed that the content bias does change over time. The ultimate reasons that motivate these changes in the studied magazines are only fully comprehensible when knowing the inner context of the newsroom, which is beyond the scope of this article. Nevertheless, a possible hypothesis is that in the heat of the events mentioned above, women and men-oriented magazines modified their contents in an attempt to reach the readership that have new ideas, values and forms of consumption.

Our results showed that there was a strong tendency to close the content distribution gap between the magazines on topics such as business, family, and women as sex objects. The latter topic fell abruptly between 2011 and 2012. The timing of this change coincides with the rise of the feminist movement in the Argentinian political arena, starting with the first "Slutwalk" in 2011 (G. F. Díaz and C. Pastor 2012). These results are in line with the hypothesis that the feminist movement and the struggle for women rights in Argentina has, at least partially, succeeded in condemning the objectification of women and, therefore, the media has made decisions that respond to these new rules.

Since 2015 the presence of feminism in magazines targeting women has increased sharply. Such phenomenon coexists with the rise of the #NiUnaMenos movement and its presence in the media. This movement protests against gender violence and femicides (Muireann O'Dwyer 2016; María C. Davies 2017; C. Garibotti María and M. Hopp Cecilia 2019). In 2018, a jump is observed in the use of the word abortion in the magazine aimed at women, reflecting the uprising debate on abortion legalization in Argentina that was discussed in Congress that year (Mirna Lucaccini, Luca Zaidán and Mario Pecheny 2019).

Regarding the family topic, our finding is aligned with the signs observed by Marshall David, et al. (2014) of a change in the representation of fatherhood in magazine advertisements. On the other hand, the fashion and science gap shows no signs of narrowing. In addition, the gap in the content associated with the horoscope shows an increase when compared to the beginning of the analysed period. The persistence of the science and horoscope gap is in line with the ongoing debates in local feminisms. Androcentric bias in science has been debated by the Argentinian feminist movement before our period of analysis (Diana Maffia 2007; Diego Kozlowski, et al. 2022). These problems in scientific production have split waters within the local feminist movement. Feminist researchers and university students have begun to discuss gender bias in science, in order to promote diversity in research. But, at the same time, other local activists have started to question the role of science itself, and instead they have proposed to recover ancestral knowledge, traditionally excluded from science, such as the horoscope (Agostina Mileo 2020). As we can see, pseudo-sciences are not unanimously criticised by the feminist movement, and therefore mass-media magazines are unlikely to modify their content related with these topics in the foreseeable future.

Performing a word frequency analysis as a control turned out to be a good complement to Topic Modelling. The Word Frequency Analysis of the topic children did not show similar behavior to that of Topic Modelling, so we can disregard this topic from the analysis.

Overall, our work presents a novel application of methods that proves to be useful in analysing large amounts of articles with computational tools and that contribute to the field of gender bias analysis in magazine content.

Limitations

The methodological framework proposed in this study focuses on the textual evidence of gender stereotypes in magazines. This approach does not consider the non-textual evidence, such as images and other formats. The visual content on magazines plays a vital role in gender stereotypes reproduction, especially in advertising (Susan G. Kahlenberg and Michelle M. Hein 2010; Mager John and James G. Helgeson 2011; Kyoungtae Nam, Guiohk Lee and Jang-Sun Hwang 2011). An approach that considers both the textual and non-textual dimensions could be enriching in this sense. The methodology developed in this article should be complemented with a method that can automatically analyse images (Kelvin Xu, et al. 2015; Andrei Karpathy and Li. Fei-Fei 2015). Nevertheless, such endeavor is beyond the scope of the present article.

Regarding Topic Modelling, it automatically captures the most frequent topics, and it is worth considering that some of these might be difficult to contextualize and comprehend. Therefore, results interpretation and topic curation by an expert from the field is required, making the method not fully automatic. Furthermore, relevant issues or the hypothesis to be tested may not be well represented in the automatically generated topics when they are not recurrent subjects in the articles (e.g. horoscope, abortion and feminism), and thus, the analysis of word frequency would be a useful complement.

On the word frequency approach, it is important to note that conducting many tests might point to misleading conclusions due to the problem of multiple comparisons. Therefore, the lists need to be carefully constructed and theoretically justified before experimentation, highlighting once again the relevance of interdisciplinary work among computer and social scientists in the experimental design and interpretation.

Future research directions

In the present research, our data sources were two magazines from the same publishing house. In order to have a complete understanding of gender roles in magazines, it would be interesting to extend data sources to a more general corpus. Also, the analysis focuses on Argentinian magazines. Replicating these methods on magazines from different countries would allow international comparison of gender stereotypes. Our study covers the years 2008 to 2018. In future work, we hope to expand this range to the present. As mentioned above, our analysis could be enriched with analogous methods for describing the visual elements of articles.

Practical implications

In the last few years, gender bias awareness has increased in Argentinian society and media This increase is portrayed in the results of our work for some of the analysed topics, such as women as sex objects, family, and business. However, certain differences between the magazines targeting men or women persist and have not shown evolution on topics such as fashion, science, and horoscope during the analysed period. Therefore, not only do our results show the topics that are currently under discussion in Argentinian society, but they also might highlight those that have not yet emerged in public debate. Both the proposed methodology and the results themselves contribute to the open discussion. Furthermore, as the methodological framework proposes a semi-automatic way to track the evolution of gender stereotypes in magazines, providing quantitative evidence on this problem might be an interesting and useful tool for the local feminist movement. We created the website: http://magbias.dc.uba.ar/ where everyone can complement the present analysis with their own unique research questions. Another putative application of this method could be to analyse and compare different sections within a newspaper or to evaluate biases resulting from journalists' gender.

Conclusion

The goal of this study was to use Data Science and Natural Language Processing techniques to compare the content of a women-oriented magazine with that of a men-oriented magazine produced by the same publishing house over a decade (2008–2018). Both magazines contained more than 24,000 articles in total, and, although there are several studies that use traditional methods to quantify the content of women and men-oriented magazines, as far as we know, ours is the first study that compares magazines in such a large dataset.

Our approach was to use Topic Modelling to identify the main topics discussed in the magazines, and then, quantify the different presence of these topics differs between magazines over time. This method allowed us to discover content present in the articles and to represent the topics according to the words that are most associated with them. Also, a validation step was carried out by performing a Word Frequency Analysis. This test validated five of the six topics: business, family, women as sex objects, fashion and science. We created a website (http://magbias.dc.uba.ar/) where everyone can formulate their own research questions regarding the vocabulary changes of these magazines.

Overall, our work contributes to the longstanding field of gender content analysis in mass media (Kuipers Giselinde, Elise Van Der Laan and Elisa A.G. Arfini 2017; Rebecca L. Collins 2011; Rudy et al. 2010; Murillo et al. 2010; Martin Eisend 2010; Rosalind Gill 2009; Gunther Kress and Theo van Leeuwen 2006; Nathalie Koivula 1999) but with a novel methodology that highlights the relevance of interdisciplinary work and shows how computational tools are useful resources to analyse large amounts of content among platforms.

New advances from computer science have proven to be a powerful tool to increase the research scope in social sciences. Large-scale data analysis allows researchers to cover the vast sources of information generated by our society nowadays more thoroughly, providing new insights and bringing new answers to longstanding research questions. They can show general tendencies that help us guide our research. In the case of Natural Language Processing, this type of analysis can help us create broad descriptions of our corpus and formulate hypotheses. These new methods do not replace human work, but they do provide further tools to analyse reality.

Note

1. http://magbias.dc.uba.ar/

Declarations

Acknowledgments

The authors want to thank Alfredo Rolla and Jorge Federico Cubells for their contributions in the initial steps of this project, and Aldana Almaraz for her editing support that helped us finish this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data deposition

The data used in this article is available at 10.5281/zenodo.4479919

Funding

This work was supported by the Doctoral Training Unit Data-driven computational modelling and applications (DRIVEN), which is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

Notes on contributors

Diego Kozlowski holds a degree in Economics, from the School of Economics at the University of Buenos Aires, and a Master's Degree in Data Mining and Knowledge Discovery, from the School of Exact and Natural Sciences at the University of Buenos Aires, Argentina. He is currently a PhD student at the Doctoral Training Unit DRIVEN (https://driven.uni.lu/), where he works on implementing computational methods to answer social sciences questions —an emerging field of study known as Computational Social Science, with special focus on the study of science and inequalities. Email ID: diego.kozlowski@uni.lu

Gabriela Lozano Rubello holds a degree in Communication Science from the School of Political and Social Sciences at the National Autonomous University of Mexico, and a Master's Degree in Gender Studies from the Metropolitan Autonomous University in Mexico City. She is currently a PhD student at the Interdisciplinary Institute of Gender Studies at the University of Buenos Aires, where she focuses her research on academic profession labor conditions in Argentina and Mexico. Email ID: gabriela.lozano.rubello@gmail.com

Carla M. Felcher holds a degree in Biology, specialised in Molecular Biology from the School of Exact and Natural Sciences at the University of Buenos Aires, Argentina. She is currently a PhD student at the Institute of Physiology, Molecular Biology and Neurosciences (http://www.ifibyne.fcen.uba.ar/), where she focuses her research on the molecular mechanisms of breast cancer development. Carla has also been involved in several science communication and popularization projects aimed at the scientific and non-scientific community. Email ID: carfelcher@gmail.com

Fernando Gonzalez holds a degree in Marketing from the Argentine University of Business (UADE) and is currently a student in the Master's Degree course on Data Mining and Knowledge Discovery, from the School of Exact and Natural Sciences at the University of Buenos Aires. He is currently working in the industry as a machine learning engineer in the areas of automatic speech recognition and natural language understanding. Email ID: fernandoezequiel512@gmail.com

Edgar Altszyler holds a Ph.D. in Physics from the University of Buenos Aires (UBA), and a Bachelor's and a Master's degrees in Physics from the Balseiro Institute. He is currently a researcher at the Computational Sciences Research Institute (UBA-CONICET) and an associate professor of Natural Language Processing in the Master's Degree Course on Data Mining at the University of Buenos Aires. His research focuses on the application of Natural Language Processing, Data Science and Machine Learning techniques to the study of social and cultural behavior, psychology, and health data. E-mail: edgaralts@gmail.com

ORCID

Diego Kozlowski (b) http://orcid.org/0000-0002-5396-3471 Edgar Altszyler (b) http://orcid.org/0000-0002-9397-8168

References

Albert, Bandura, Ross Dorothea, and Sheila A. Ross. 1963. "Imitation of Film-mediated Aggressive Models." *The Journal of Abnormal and Social Psychology* 66 (1): 3–11. doi:10.1037/h0048687.

Baile, José I, María Gabino-Campos, and Ana Laura Pérez-Lugoa. 2020. "Analysis of the Aesthetic Stereotypes of Women in Nine Fashion and Beauty Mexican Magazines." *Revista Mexicana de Trastornos Alimentarios* 7 (1): 40–45.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, .

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Butler, Judith . 1998. "Actos performativos y constitución del género: Un ensayo sobre fenomenología y teoría feminista." *Debate feminista* 18: 296–314. doi:10.22201/cieg.2594066xe.1998.18.526.

Carter Olson, Candi. 2016. "#bringbackourgirls: Digital Communities Supporting Real-world Change and Influencing Mainstream Media Agendas." *Feminist Media Studies* 16 (5): 772–787. doi:10.1080/14680777.2016.1154887.

Chang, Jonathan, Gerrish Sean, Wang Chong, Jordan L Boyd-Graber, and David Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems* 22: 288–296.

Collins, Rebecca L. 2011. "Content Analysis of Gender Roles in Media: Where are We Now and Where Should We Go?" Sex Roles 64 (3–4): 290–298. doi:10.1007/s11199-010-9929-5.



- Coltrane, Scott, and Michelle Adams. 1997. "Work-family Imagery and Gender Stereotypes: Television and the Reproduction of Difference." *Journal of Vocational Behaviour* 50 (2): 323–347. doi:10.1006/jvbe.1996.1575.
- Comercial LA NACION [online]. Accessed 23 November 2020. http://comercial.lanacion.com.ar/pages/productocomercial.aspx?cat=revista
- Courtney, Alice E., and Sarah W. Lockeretz. 1971. "A Woman's Place: An Analysis of the Roles Portrayed by Women in Magazine Advertisements." *Journal of Marketing Research* 8 (1): 92–95. doi:10.1177/002224377100800114.
- Das, Mallika. 2011. "Gender Role Portrayals in Indian Television Ads." Sex Roles 64 (3–4): 208–222. doi:10.1007/s11199-010-9750-1.
- Davalos, Deana B., Ruth A. Davalos, and Heidi S. Layton. 2007. "III. Content Analysis of Magazine Headlines: Changes over Three Decades?" *Feminism & Psychology* 17 (2): 250–258. doi:10.1177/0959353507076559.
- David, Marshall, Davis Teresa, Margaret K Hogg, Schneider Tanja, and Alan Petersen. 2014. "From Overt Provider to Invisible Presence: Discursive Shifts in Advertising Portrayals of the Father in Good Housekeeping, 1950–2010." *Journal of Marketing Management* 30 (15–16): 1654–1679. doi:10.1080/0267257X.2014.945471.
- Davies, María C. 2017. "Nos Están Matando [They are Killing Us]: Feminist Movements' Influence in Argentina and Chile. Zenith." *Undergraduate Research Journal for the Humanities at the University of Kansas* 2 (1): 165–182. doi:10.17161/1808.23868.
- Del-Teso-Craviotto, Marisol. 2006. "Words that Matter: Lexical Choice and Gender Ideologies in Women's Magazines." *Journal of Pragmatics* 38 (11): 2003–2021. doi:10.1016/j. pragma.2005.03.012.
- Desmond, Roger, and Anna Danilewicz. 2010. "Women are On, but Not In, the News: Gender Roles in Local Television News." Sex Roles 62 (11): 822–829. doi:10.1007/s11199-009-9686-5.
- Díaz, G. F., and C. Pastor. 2012. "Tu mamá también: Apropiaciones de la Marcha de las Putas en Argentina." Actas del 2° Congreso Interdisciplinario sobre Género y Sociedad: Lo personal es político 1 (1).
- Downs, Edward, and Stacy. L. Smith. 2010. "Keeping Abreast of Hypersexuality: A Video Game Character Content Analysis." Sex Roles 62 (11–12): 721–733. doi:10.1007/s11199-009-9637-1.
- Eisend, Martin. 2010. "A Meta-analysis of Gender Roles in Advertising." *Journal of the Academy of Marketing Science* 38 (4): 418–440. doi:10.1007/s11747-009-0181-x.
- Gálvez, Ramiro H, Tiffenberg Valeria, and Edgar Altszyler. 2019. "Half a Century of Stereotyping Associations between Gender and Intellectual Ability in Films." *Sex Roles* 81 (9–10): 643–654. doi:10.1007/s11199-019-01019-x.
- Garibotti María, C., and M. Hopp Cecilia. 2019. "Substitution Activism: The Impact of #metoo in Argentina." In #metoo and the Politics of Social Change, edited by B. Fileborn. Cham: Loney-Howes R. Palgrave Macmillan.
- Gill, Rosalind. 2009. "Beyond the 'Sexualization of Culture' Thesis: An Intersectional Analysis of 'Sixpacks', 'Midriffs' and 'Hot Lesbians' in Advertising." *Sexualities* 12 (2): 137–160. doi:10.1177/1363460708100916.
- Gilpatric, Katy. 2010. "Violent Female Action Characters in Contemporary American Cinema." Sex Roles 62 (11–12): 734–746. doi:10.1007/s11199-010-9757-7.
- Giselinde, Kuipers, Elise Van Der Laan, and Elisa A.G. Arfini. 2017. "Gender Models: Changing Representations and Intersecting Roles in Dutch and Italian Fashion Magazines, 1982–2011." *Journal of Gender Studies* 26 (6): 632–648. doi:10.1080/09589236.2016.1155435.
- Goffman, Ervin. 1979. Gender Advertisements. Macmillan International Higher Education.
- Hastie, Trebor J., and Robert J. Tibshirani. 1990. Generalized Additive Models. Vol. 43. CRC press.
- Hether, Heather J., and Sheila T. Murphy. 2010. "Sex Roles in Health Storylines on Prime Time Television: A Content Analysis." Sex Roles 62 (11–12): 810–821. doi:10.1007/s11199-009-9654-0.
- John, Mager, and James G. Helgeson. 2011. "Fifty Years of Advertising Images: Some Changing Perspectives on Role Portrayals along with Enduring Consistencies." Sex Roles 64 (3–4): 238–252. doi:10.1007/s11199-010-9782-6.



Kahlenberg, Susan G., and Michelle M. Hein. 2010. "Progression on Nickelodeon? Gender-role Stereotypes in Toy Commercials." Sex Roles 62 (11–12): 830–847. doi:10.1007/s11199-009-9653-1.

Karpathy, Andrej, and Li. Fei-Fei 'Deep Visual-semantic Alignments for Generating Image Descriptions.' In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3128–3137. 2015.

Kelvin, Xu, Ba Jimmy, Ryan Kiros, Kyunghyun Cho, Courville Aaron, Ruslan Salakhudinov, Zemel Richard, and Bengio Yoshua. 'Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.' In International conference on machine learning, pp. 2048–2057. PMLR, 2015.

Koivula, Nathalie. 1999. "Gender Stereotyping in Televised Media Sport Coverage." Sex Roles 41 (7–8): 589–604. doi:10.1023/A:1018899522353.

Kozlowski, Diego, Vincent Larivière, Cassidy R Sugimoto, and Thema Monroe-White. 2022. "Intersectional Inequalities in Science." *Proceedings of the National Academy of Science* 119 (2): e2113067119. doi:10.1073/pnas.2113067119.

Kress, Gunther, and Theo van Leeuwen. 2006. *Reading Images: The Grammar of Visual Design*. 2nd edn ed. London: Routledge.

Lucaccini, Mirna, Luca Zaidán, and Mario Pecheny. 2019. "Qué nos dice el debate sobre aborto en 2018 sobre la clase política y el espacio público en la Argentina." In *Miradas feministas sobre el derecho*, edited by Aluminé Moreno, Diana Maffía, and Patricia Laura Gómez, 245–263. Buenos Aires: Editorial jusbaires.

Macharia, Sarah (2020). Who Makes the News? the Global Media Monitoring Project 2020. Retrieved 11 January 2022, from ttps://whomakesthenews.org/gmmp-2020-final-reports/

Maffia, Diana. 2007. "Epistemología feminista: La subversión semiótica de las mujeres en la ciencia." Revista Venezolana de Estudios de la Mujer 12 (28): 63–98.

Magazine links. Accessed in 2018: Brando: https://www.lanacion.com.ar/revista-brando;Ohlalaurl: https://www.lanacion.com.ar/revista-ohlala

Matud, M. Pilar, Carmen Rodríguez, and Inmaculada Espinosa. 2011. "Gender in Spanish Daily Newspapers." Sex Roles 64 (3–4): 253–264. doi:10.1007/s11199-010-9874-3.

Michel, Jean-Baptiste, Yuan Kiu Shen, Aviva P Aiden, Veres Adrian, Matthew K Gray, Joseph P Pickett, and Steven Pinker. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–182. doi:10.1126/science.1199644.

Mileo, Agostina (2020). OPINIÓN: En el cielo las estrellas, en el feminismo las espinas – Economia Femini(s)ta. Retrieved 13 July 2021, from https://economiafeminita.com/feminismoyastrologia/

Mohr, John W, and Petko Bogdanov. 2013. "Introduction—Topic Models: What They are and Why They Matter." *Poetics* 31 (6): 545–569. doi:10.1016/j.poetic.2013.10.001.

Murillo, Marín, F. María, Armentia Vizuete, José Ignacio, and María Ganzabal Learreta. 2010. "Claves de la construcción de género en las revistas femeninas y masculinas: Análisis cuantitativo." Estudios sobre el Mensaje Periodístico 16: 259–289.

Nacion, La. 2008. 'Salió OH! LaLá, para la mujer moderna', April 3. Accessed 23 June 2021. https://www.lanacion.com.ar/cultura/salio-oh-lala-para-la-mujer-moderna-nid1000864/

Nam, Kyoungtae, Guiohk Lee, and Jang-Sun Hwang. 2011. "Gender Stereotypes Depicted by Western and Korean Advertising Models in Korean Adolescent Girls' Magazines." Sex Roles 64 (3–4): 223–237. doi:10.1007/s11199-010-9878-z.

Neuendorf, Kimberly A. 2011. "Content analysis—A Methodological Primer for Gender Research." Sex Roles 64 (3–4): 276–289. doi:10.1007/s11199-010-9893-0.

Neuendorf, Kimberly A., Thomas D Gore, Amy Dalessandro, Patricie Janstova, and Sharon Snyder-Suhy. 2010. "Shaken and Stirred: A Content Analysis of Women's Portrayals in James Bond Films." Sex Roles 62 (11–12): 747–761. doi:10.1007/s11199-009-9644-2.

O'Dwyer, Muireann. 2016. '#niunamenos; Standing up to Femicides and 'Machismo' in Argentina.' Gender and Politics (blog). Accessed October 2020 https://genderandpoliticsucd.wordpress.com/ 2016/02/03/niunamenos-standing-up-to-femicides-and-machismo-in-argentina/.

Rudy, Rena M., Lucy Popova, and Daniel G. Linz. 2010. "The Context of Current Content Analysis of Gender Roles: An Introduction to a Special Issue." *Sex Roles* 62 (11–12): 705–720. doi:10.1007/s11199-010-9807-1.



Scott, Joan W. 1986. "El género: Una categoría útil para el análisis histórico." Historical Review 91 (5): 1053-1075. doi:10.2307/1864376.

Spataro, Carolina. 2013. "Las tontas culturales: Consumo musical y paradojas del feminismo." Revista Punto Género 3: 27-45. doi:10.5354/0719-0417.2013.30265.

Tandoc, Edson C., Jr, and Patrick Ferrucci. 2014. "So Says the Stars: A Textual Analysis of Glamour, Essence and Teen Vogue Horoscopes." Women's Studies International Forum 45: 34-41. doi:10.1016/j.wsif.2014.05.001.

Tuchman, Gaye. 1979. "Women's Depiction by the Mass Media." Signs: Journal of Women in Culture and Society 4 (3): 528-542. doi:10.1086/493636.

Turner, Jacob S. 2011. "Sex and the Spectacle of Music Videos: An Examination of the Portrayal of Race and Sexuality in Music Videos." Sex Roles 64 (3-4): 173-191. doi:10.1007/s11199-010-9766-6. Winship, Janice. 1987. Inside Women's Magazines. Rivers Oram Press/Pandora List.

Zotos, Yorgos C, and Eirini Tsichla. 2014. "Female Stereotypes in Print Advertising: A Retrospective Analysis." Procedia-social and Behavioral Sciences 148: 446-454. doi:10.1016/j.sbspro.2014.07.064.