Analysis of Online Narratives on the Empowerment of Women Survivors of Gender-based Violence in Colombia with Natural Language Processing

Susana Martinez Restrepo, PhD.
Lina Tafur Marin
Juan Guillermo Ocio

Caroline Brethenoux
Patrick Furey
Orlando Rivera
Analysis of Online Narratives on the Empowerment of Women Survivors of Gender-based Violence in Colombia with Natural Language Processing

Authorship

Susana Martinez Restrepo, PhD.
Lina Tafur Marin
Juan Guillermo Ocio
Caroline Brethenoux
Patrick Furey
Orlando Rivera

A special thanks to Juliana Ramirez and Angélica León for their valuable comments to strengthen the analysis in this report. Also a special thanks to Alejandra Vargas Garcia (International Development Research Center), Andrew Morrison and Claudia Piras (Inter-American Development Bank) for their comments and contributions to the present study.

Style review: Maria Paulina Zuleta
Design and layout: Laura Sánchez Castro
Photographs: Unshplas, Freepik, Flick.

This work was carried out with the help of a grant from the International Development Research Center (IDRC), Ottawa, Canada. The opinions expressed here are not they necessarily represent those of IDRC or the Board of Governors.

Published agosto de 2021
Executive Summary

This study uses Big Data and Natural Language Processing (NLP) to understand 4.7 million online first-person narratives from survivors of gender-based violence in Colombia. Using sentiment and topic analysis of these narratives, written by survivors between November 2016 and February 2020, the study methodology establishes sentiment drivers to understand central aspects of the empowerment process (or lack thereof) according to Kabeer’s framework.

The results indicate that 61% of the narratives expressed a negative sentiment, 31% a neutral sentiment, and 8% a positive sentiment. According to the narratives expressing negative sentiment, the main barrier preventing survivors of gender-based violence from overcoming the negative impacts of their experience or starting a process of empowerment is scarcity of resources, primarily social and financial. On the other hand, survivors’ narratives expressing positive sentiment about their empowerment process consistently refer to the support given by friends, family, and the community. Survivors see support as a key factor enabling them to meet their basic financial and emotional needs and transform their well-being.

These results suggest that social resources are among the most important preconditions for the empowerment process of women survivors of gender-based violence. Furthermore, the results could indicate the importance of the collective as a resource for and driver of agency among survivors. This finding is consistent with the emerging importance of groups or collectives in the literature on economic empowerment, which was previously considered an individual process. The lack of specific references to government institutional support in the positive sentiment analysis, as well as the association of a negative sentiment with such institutional support, could suggest that government services are not a determining factor in overcoming the trauma of gender-based violence.

With its use of new methodologies, this study represents an important contribution to both the literature on empowerment as a process and the role of the collective as a social resource in this process. Additionally, new gender-based violence studies employing big data should be encouraged to contribute new categories, questions, and phenomena to traditional methods of analysis.
1. Introduction

According to the World Health Organization (WHO), 35% of women globally have been victims of some type of gender-based violence (GBV) in their lifetime. In Colombia, this percentage is 66%, 35% in the case of physical violence, and 8% in the case of sexual violence. An estimated 75% of this violence against women is perpetrated by their intimate partners.

Furthermore, in countries such as Colombia, only 14% of victims report these aggressions to official entities such as the police or hospitals. This reluctance to report to official authorities can be attributed mainly to stigma around, lack of knowledge of, or limited access to the Attention Routes, in addition to lack of trust in public institutions, fear of revictimization, justification of violence, or economic dependence.

According to Data2x, capturing detailed information about the lives of women and girls can be difficult using conventional data sources. Data from the National Demographic and Health Survey (DHS) and existing administrative data (from complaints, for example) do not allow understanding of the barriers that prevent women from leaving an abusive relationship, the consequences of GBV, or their empowerment process.

Big Data offers innovative opportunities to fill these information gaps. In Colombia in December of 2020, an estimated 30 million people interacted on social networks. The #MeToo movement amplified the visibility of GBV and especially of sexual harassment and violence. Thanks to millions of women sharing their stories, the movement aimed to educate society about the prevalence of sexual harassment and the need for change.

---

1 Traditionally, GBV refers to gender-based physical, psychological, sexual, economic, and patrimonial violence against a person.
5 These refer to the channels a survivor would follow to make a government entity aware of her situation and which also allows them a way to access protection, justice, and psychological and economic support.
6 Revictimization is the process by which institutions and professionals in charge of providing care to the victim cause additional suffering.
10 The #MeToo movement was created by Tarana Burke in 2007 and then taken up by Alyssa Milano in October 2017 in order to uncover sexual harassment in Hollywood. This hashtag was deemed appropriate to demonstrate the prevalence of sexual harassment experienced by women. In Spanish, #YoTambién was used.
first-person stories of abuse, the conversation managed to permeate sectors of society it had not previously reached. Between October 2017 and September 2018, estimates are that the #MeToo hashtag had been shared 19 million times by women survivors on Twitter. In Colombia and Latin America, other hashtags such as #MiCuerpoDiceLaVerdad, #UnVioladorEnTuCamino, #ElMachismoMata and #NiUnaMenos went equally viral, which opened a space for discussion and for changing narratives surrounding the topic.

From a Big Data approach, social networks and other discussion forums are a growing source of information to understand people's needs and experiences. Largely thanks to the #MeToo movement, many women have been motivated to post very detailed testimonies specifically related to their experiences of GBV. Unlike the closed questions imposed by surveys (such as those in the DHS) and administrative records, online conversations are characterized by their spontaneity. Given the lack of confidence in public entities, the Internet allows survivors to report openly, not only about the victimizing event but also the critical factors (objects, entities, and strategies) that have enabled or hindered the empowerment process.

Natural Language Processing (NLP) makes it possible to study patterns in millions of conversations and to classify each narrative according to its topic or sentiment. The topic and sentiment analysis can be applied to large volumes of data, free of the restrictions imposed by manual classification of information, a process common in traditional qualitative methodologies.

The use of Big Data for GBV analysis is recent, but is starting to gain importance. Using Google mobility data, a study by DataPop and GIZ on GBV risk factors in times of COVID showed a correlation between periods of confinement (mobility restrictions) and increased number of calls to GBV hotlines in Bogota, Colombia. Similarly, using Google mobility data from New Delhi, India, Borker found that long commutes decreased the likelihood that young women would choose a high-quality university due to risk of sexual violence during displacement. Also using NLP, García-Díaz and colleagues developed a corpus of Spanish words or phrases expressing misogyny, enabling automatic and early detection of this phenomenon in social networks. Using data from the US healthcare system, Karakurt and colleagues used NLP to identify the specific health needs of GBV survivors as well as risk factors for violence or revictimization. According to Data2x, an information gap exists in the use of Big Data to understand gender issues, specifically GBV. Due to the novel nature of Big Data and NLP in the social sciences, it is important to further explore its uses and potential challenges.

---


12 Throughout the text, reference is made to the narratives of survivors and not to conversations because one conversation may contain different thematic or opinion (feeling) narratives. This makes possible a more precise classification.


This study is the second report in the series “Gender-based Violence and Women’s Economic Empowerment in Colombia.” The report studies online narratives about the empowerment process of women survivors of GBV in Colombia using Big Data and NLP techniques and answers the following questions: What barriers prevent women from leaving abusive relationships? What aspects of their lives help them overcome abusive situations? What role does the economic component play as a barrier to overcoming the impacts of GBV? What factors cohere (or not) with existing evidence on the empowerment process of women survivors? What is the potential utility of Big Data as a tool to study GBV? Can it be used to propose suitable ways to study the phenomenon and suggest public policy recommendations?

To respond to these questions, the study analyzed the sentiments and topics of 4.7 million survivors’ online narratives that occurred in Colombia between November 2016 and February 2020. Sentiment analysis is a methodology that studies the causes of the sentiments implicit in the narratives—positive, neutral, or negative—and topic analysis classifies the narratives according to their most recurrent topics.

The results show that survivors identify social resources (or support) as the highest priority contributor, followed by material and economic resources, to their empowerment process after GBV. Social resources (or support) refers to assistance provided by a third party, either family or community, in a way that positively affects the survivors’ fallback position. Such a resource may take the form of material, financial, emotional, and/or knowledge assistance. The social resource also can be understood as a facilitator of—or even as a prerequisite for—access to and control of economic resources to transform a woman’s situation and enable her to negotiate within and outside her home. Moreover, following the feminist economics literature, the results show that scarcity of resources is the main barrier to survivors’ empowerment process after they experience GBV. This study contributes to the literature on the use of Big Data to understand GBV, while contributing to a greater understanding of the process of economic empowerment with a focus on social and collective resources.

---

18 The online narratives written in Spanish repeatedly used the saying “salir adelante”, which draws on the idea of “moving beyond” or “overcoming”. This suggests a power within the individual to change circumstances for the better and leave behind obstacles or unhealthy situations.

19 A fallback position in this context refers to the possibility of a woman leaving an abusive relationship. Conditions that enhance a woman’s bargaining power with an abusive partner—a job, savings, or a property asset such as a house—strengthen her fallback position.

The online narratives of GBV survivors incorporate elements that contribute to an understanding of the barriers they face when trying to rebound after experiencing violence. The economic component emerges as central, even though it has been excluded from studies of GBV in disciplines such as psychology, sociology, or health sciences. Studies on GBV commonly focus on physical and emotional effects as well as on psychosocial needs.

Naila Kabeer’s conceptual framework of economic empowerment is useful for understanding the process by which women are able to strengthen their agency and achieve well-being following experiences of GBV. Kabeer frames empowerment along three dimensions: resources, agency, and achievements. Resources, a precondition for empowerment, are not only material, but also human and social, and enable a person to make strategic decisions. Evidence has pointed out that certain resources (income, savings, land ownership, and productive assets, among others) guarantee a better fallback position in a relationship and reduce the likelihood of domestic violence, since they improve women’s bargaining power in the household.

Agency is defined as the ability to set, feel capable of achieving, and act on one’s goals. In a situation of GBV, agency empowers victims to make decisions such as exiting an abusive relationship and can lead to survivors creating a business or seeking employment.

Finally, well-being achievements transform the status quo and remove gender inequalities. As Kabeer points out, achievements result from strategic decisions that positively transform women’s well-being.

21 Ibid.
24 Agarwal, “‘Bargaining’ and Gender Relations: Within and Beyond the Household.
Until a few years ago, the literature of economic empowerment focused only on the process of empowerment of women as individuals, without considering women as part of a collective or community. This approach suggests that empowerment should depend primarily on women’s individual economic, personal, and educational resources; their individual capacity to make strategic decisions; and their individual well-being. Such an individualistic conception of empowerment does not consider that, in countries like Colombia, social resources play a central role in the lives of survivors. This is due to high levels of poverty, inequality, and female unemployment; the informality of the work sector; and lack of a welfare state to provide even minimal economic and physical security. In such an environment, economic resources can come from a collective of people (family, friends, and grassroots organizations, among others), which can be called a social resource. Additionally, this collective can become a social resource to facilitate, promote, and even stimulate a woman’s strategic decisions, goals, and actions—allowing her to exit an abusive relationship and generate income in the aftermath.

Only recently has feminist economics incorporated the role of the collective in its analysis of the process of economic empowerment. In their study, Gammage, Kabeer and Van der Meulen Rodgers explain how the ability to articulate strategic needs and interests collectively in the private and public spheres is an important aspect of agency. They focus on collective agency as part of the process of empowerment, especially as a key resource for women’s fallback position. In Colombia—particularly in rural areas affected by armed conflict—the social resource can come in the form of support from other women. This support is not only a facilitator of other resources but also a key resource in the process of women’s empowerment and in the exercise of their agency. Thanks to social resources or the support from the collective, women survivors can make strategic decisions that allow them to improve their economic well-being or mental health, among other aspects of their lives.

This study presents evidence on the importance of this social resource in the process of women’s economic empowerment and demonstrates how it relates to economic resources and to the strengthening of the bargaining power and fallback position of women survivors of GBV. Using NLP to analyze the online narratives of survivors allows us to inquire about the factors that influence empowerment positively or negatively.

### 3. Data and Methodology

Using data mining techniques and NLP, we analyzed 4.7 million online narratives of GBV survivors that emerged in Colombia between November 2016 and February 2020. With NLP, a sentiment analysis was conducted to understand the barriers to and mechanisms of the empowerment process in the aftermath of GBV.

#### 3.1. Data Profile

We collected 6.8 million online narratives about GBV that emerged in Colombia during the three years and four months between November 2016 and February 2020, focusing on the narratives of women survivors. Using NLP techniques, we analyzed these narratives to understand the factors that influence women’s economic empowerment positively or negatively.
2020. Most of the narratives (4.8 million) were from women (90%) while the remaining ten percent were from men (Figure 1). Due to the objective of this study, the conversations were filtered to consider only the narratives of women GBV survivors.

As Figure 3 shows, most of the information came from topical sites (33%) and message boards (32%). Social networks represented only 16% of the forums used for discussion of GBV. The highest participation in online narratives by geographic area was concentrated in the departments with the largest population and urban development (Bogota, 24.6%; Antioquia, 12.5%, and Valle del Cauca, 8.9%). These departments have the highest concentration of population with Internet access (Figure 4).

32 A span of three years and four months was determined for data collection in order to capture the increase in discussions and narratives about GBV related to the #MeToo movement.

33 A topical site is a web page that focuses on specific thematic content.

34 A message board is an online page where public discussion occurs in the form of messages. It differs from a chat room because the messages are usually longer and it is possible to have a moderator in the discussion. For example, Reddit is made up of message boards that focus on specific topics.
3.2. Vocabulary Identification and Data Collection

To perform a Big-Data analysis based on NLP, we identified the colloquial lexicon used in Colombia to talk about GBV\(^{35}\) by reviewing qualitative studies with original narratives of women in rural and urban areas of Colombia\(^{36}\). We then created a list of vocabulary related to GBV based on this review. Subsequently, we used data scraping technologies (crawlers\(^{37}\) and web scraping\(^{38}\)) to extract and classify information from the web.

The collection of online narratives was based on four parameters. First, we considered only narratives about GBV. Second, narratives came from all over the Internet. Third, the narratives occurred between November 2016 and February 2020, following the #MeToo/#Yotambién movement and other trends such as #MiCuerpoDiceLaVerdad and #UnVioladorEnTuCamino\(^{39}\). Fourth, the IP addresses of users’ narratives were used to ensure that we worked with the constituency of the Colombian territory.

NLP is a field of artificial intelligence that aims to approximate human understanding of natural language through computer programs\(^{40\ 41}\). NLP is used to infer and disambiguate the meaning of data in a text by establishing patterns and trends\(^{42}\). This type of data processing does not aim to replicate traditional qualitative analysis, but to analyze millions of online conversations in an automated way\(^{43}\).

To capture the sentiment and topic of online conversations or posts, the study divided these into narratives\(^{44}\) consisting of sentences or ideas\(^{45}\). The

---

\(^{35}\) A vocabulary of 181 terms used colloquially throughout Colombia was identified based on qualitative studies on GBV. This vocabulary refers to topics such as violence, sex work, sexual orientation, sexual violence, sexual harassment, psychological violence, physical violence, murder, drug trafficking, control, conflict, alcohol, and abortion.

\(^{36}\) Some examples include documents from the National Center of Historical Memory (CNMH) with interviews that describe in detail the use of women’s bodies to establish dominance between armed actors and communities—as in the case of the Portete Bay massacre or sexual violence in Arauca, for example.

\(^{37}\) Crawlers are computer programs that automatically review online documents.

\(^{38}\) Web scraping is a technique used to extract information from websites automatically and analyze it.

\(^{39}\) According to the Pew Research Center, between October 2017 and September 2018, the hashtag #MeToo was estimated to have been shared 19 million times by women survivors on Twitter. In Colombia, between 2017 and 2020, GBV narratives were shared more than 5,000 times a day.


\(^{42}\) Ibid.

\(^{43}\) Johnson, “How the Statistical Revolution Changes (Computational) Linguistics.”

\(^{44}\) This occurs with the use of a deep sequential learning method that analyzes the sequence of words so as not to lose the context, establishing the narrative more accurately.

study used a deep sequential learning\textsuperscript{46} method, which analyzes word sequence to maintain context, to establish the narrative more accurately.

*Sentiment analysis* studies narratives by focusing on their emotional polarity, which can be positive, negative, or neutral. During the training process, the NLP model learns to associate narratives with their corresponding sentiment classification based on language use patterns. This allows an analysis of the causes of these emotional states or *sentiment drivers*\textsuperscript{47}. *Sentiment drivers*, in turn, allow identification of the elements—such as financial and economic resources, the community, fulfillment of primary necessities, etc.—that have been critical to overcoming the difficulties imposed by GBV. After classifying the narratives according to their underlying sentiment, they are then disaggregated by topics (*topic analysis*) to understand elements related to *agency*.

<table>
<thead>
<tr>
<th>Table 1. Sentiment analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis</strong></td>
</tr>
<tr>
<td>![Positive]</td>
</tr>
<tr>
<td>![Negative]</td>
</tr>
<tr>
<td>![Neutral]</td>
</tr>
</tbody>
</table>

*Source:* Own analysis and elaboration.


\textsuperscript{47} Sentiment drivers are factors that trigger positive, negative, or neutral feelings, which also help to understand the root of the feeling in question. An example of a sentiment driver is the support provided by people that produces a positive sentiment.
3.4. Representativeness and Data Biases

It is important to keep in mind that these data are not representative of all women survivors of GBV in Colombia. These data come from spontaneous narratives published online, which implies the invisibility of people who do not share their stories online; who do not have Internet access, a smartphone, a computer, or tablet; and who generally are excluded from digital tools and assets. Online narratives do not necessarily represent the most vulnerable GBV survivors: rural and older women who lack the connectivity or digital skills to share their stories. These data also possibly exclude the voices of women in abusive situations who may be unable to share their stories because of safety concerns or lack of access. Likewise, the testimonies of women who are unaware of having been victims of violence, due to the normalization of GBV, also may be excluded. Finally, GBV within the context of the armed conflict may be unrecognized since, in order to label a narrative within this thematic axis, a conversation must explicitly reference this context.\footnote{For example, the conversation must explicitly state that the act of GBV was perpetrated by an actor in the armed conflict. A preliminary analysis showed that 9\% of survivor narratives were about the armed conflict.}

4. Sentiment Analysis: Barriers to and Strategies for Empowerment

The sentiment analysis seeks to understand the barriers to and mechanisms for regaining agency after GBV. Of 4.7 million narratives, 61\% (2.8 million) had a negative sentiment, 31\% (1.45 million) had a neutral sentiment, and only 8\% had a positive sentiment.\footnote{The classification of narratives according to their polarity—positive, negative, or neutral sentiment—divides them into established categories such that 100\% of the narratives are assigned to a sentiment.} This suggests that survivors’ narratives focus on problems and barriers that prevent them from coping with and overcoming the episode of violence.
Although the analysis does not delve into the neutral narratives (31%), it is important to highlight that they demonstrate the growing importance of digital resources as a means for survivors of GBV to work through doubts, as well as to find professional and institutional help and guidance.

4.1. Negative Sentiment Analysis of Survivors’ Narratives

The negative sentiment analysis (Figure 6) highlights elements that hinder a survivor’s process of empowerment after the violent event. These represent the majority of online GBV survivors’ narratives (61%). From the narratives, five types of barriers were identified: *scarcity of resources* (23%), *stigmatization* (social alienation resulting from being a GBV survivor) (21%), *emotional impact* (20%), *lack of control* (loss of confidence in one’s ability to achieve goals) (20%), and *physical integrity and vulnerability* (perception of physical insecurity) (16%). The importance of *resource scarcity* and *emotional impact* are consistent with the bargaining power and fallback position literature of authors such as Agarwal and Kabeer.

---

**Figure 5. Sentiment analysis of survivors’ narratives**

![Sentiment analysis graph](image)

Source: Own analysis and elaboration.

---

50 Agarwal, “‘Bargaining’ and Gender Relations: Within and Beyond the Household.”

I didn’t have a peso. I didn’t earn anything. I had debts and that was how he abused me. He blackmailed me with that; he knew I couldn’t do anything.”

Every time I approached someone to try to talk about what happened to me, I felt judged, like people didn’t believe me, like it was my fault.... at one point I decided that I didn’t want to talk to anyone anymore, if in the end all they did was judge me.”

I just felt that my life was shit, that it was better for me to kill myself once and for all, that I wouldn’t be able to do anything for myself or anyone else. My mom would have taken care of the Chinese guy and that would have been it.”

I had never felt so sad in my life, never. It was really like that person had drained me of everything: the will to live, to go to work, to go out... I just wanted it all to end.”

You don’t know how fucked up it is to try to keep life going after a rape. Everywhere you go there are enemies and it doesn’t work that way. Now imagine being responsible for an entire family in those conditions”.

Source: Own analysis and elaboration.

**Stigmatization** exacerbates the physical and mental impacts of GBV and increases the likelihood of remaining in violent relationships. Stigma often excludes survivors from entering the labor market or continuing their studies. The source of stigmatization may be the family, society, or the institutions that are part of the Attention Routes and that provide services to survivors. According to the Oxford Poverty & Human Development Initiative (OPHI), *stigmatization* by institutional services not only creates a barrier to accessing such services, but also becomes a risk factor for future cases of GBV. In fact,


55 According to the Colombian Ministry of Health, the Attention Route for victims of GBV is understood as the set of articulated actions that seek to guarantee the protection of victims, their recovery, and the restitution of their rights. Each case requires a different Attention Route depending on the conditions of the victim and the case presented. Ministerio Nacional de Salud y Protección Social y Profamilia. “Encuesta Nacional de Demografía y Salud”.
according to Scheer et al., women who have been exposed to gender-based stigma are less likely to receive help through institutions such as the police, which may explain why the rate of reporting to official sources is so low in Colombia (14%).

Consistent with the literature, stigmatization, emotional impact, perceived lack of control and physical integrity, and vulnerability are common barriers to accessing social resources (support). These factors, central in survivors’ narratives, can be understood within the framework of Amartya Sen’s capabilities approach, as they limit women survivors’ agency and reduce their possibilities to “be and do” what they value.

**Scarcity of Resources**

We conducted a topic analysis of the narratives focused on scarcity of resources. The main topics found were: social resources (31%), financial resources (27%), institutional protection (15%), human capital (access to formal or informal education) (14%), and mobility (both physical displacement and social mobility) (11%). Financial resources and human capital are central to the literature on fallback position and bargaining power.

Lack of economic or financial resources can be a cause or an effect of violence, considering that women with scarce resources are more likely to be victims of GBV because of their limited bargaining power. Especially in developing countries, economic insecurity makes marriage, or any union with a man, of interest to women because of men’s easy access to and control over economic resources. Such resources also become a reason to stay in a violent relationship. Moreover, repeated incidents of abuse can impair women’s ability to obtain and retain employment.

![Figure 7. Survivors’ narratives about scarcity of resources](source.png)

### Figure 7. Survivors’ narratives about scarcity of resources

Source: Own analysis and elaboration.

---


58 Agarwal. “‘Bargaining’ and Gender Relations: Within and Beyond the Household.”


61 Ibid.


It is important to highlight three novel aspects of this topic analysis.

First, in the feminist economics literature, empowerment is usually understood as an individual process. The role of social resources has been consistently overlooked. In Colombia, there is a growing body of literature on the role of grassroots women's organizations in supporting survivors in the absence of, or lack of trust in, state services. Grassroots organizations, as a social resource, strengthen the process of economic empowerment by enabling access to other crucial resources that can help women overcome their experiences of GBV. Given this, rebuilding community ties is central to interrupting the continuum of violence.

Second, the negative sentiment analysis mentions lack of institutional protection. The absence of mention of institutional protection in the positive sentiment analysis suggests that survivors lack confidence in, or do not seek or receive, the expected support from the government's Attention Routes. This coincides with other studies that reference the absence of adequate institutional support. Other studies find that this institutional support frequently revictimizes survivors due to inadequate training of officials who attend to women accessing these services. Mistrust of institutional services also results from a high level of impunity for officials, which deprives survivors of access to justice and protection. It is important to further investigate these results using other methodologies to understand the roots and consequences of this phenomenon.

Third, 11% of the negative sentiment narratives mention mobility in the context of scarcity of resources, and mobility remains understudied as a barrier to empowerment. According to Patel et al., insecurity in cities and its effects on women are not well documented. Patel's study suggests that suffering violence in public spaces affects women's access to education and healthcare, their ability to socialize outside the home, and their ability to secure income. Big Data can contribute to further study of this variable. For example, mobility data from Google established the relationship between street harassment and women's educational decisions, demonstrating that women may choose an educational institution based on the availability and safety of transportation routes.

4.2. Positive Sentiment Analysis of Survivors' Narratives

The positive sentiment analysis (representing only 8% or 376,000 of the narratives studied) highlights concrete elements that have helped survivors overcome incidents of GBV—that is, leave a violent relationship, start a business, continue working, or seek help. Survivors' narratives identify four main aspects that have impacted them positively after GBV: support—social resources provided by others (43%), personal achievement—control over one's own decisions (26%), plans and goals—objectives that transcend the episode of violence (19%), and material and economic resources (12%).
Support is the most prominent component of the narratives with a positive sentiment, even beyond personal achievements and plans and goals, which feminist literature considers relevant factors since they indicate strategic decisions and a transformation of women's well-being\textsuperscript{73,74}. Moreover, material and economic resources are a precondition for empowerment\textsuperscript{75}.

**Figure 8. Narratives with a positive sentiment**

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>43%</td>
</tr>
<tr>
<td>Achievements</td>
<td>26%</td>
</tr>
<tr>
<td>Plans and Goals</td>
<td>19%</td>
</tr>
<tr>
<td>Material &amp; Economic Resources</td>
<td>12%</td>
</tr>
</tbody>
</table>

My best friend never left me alone. From the moment I told her, she accompanied me, helped me with everything, and never let me feel worse than I already did. Without her, I would never have thought of a future. Today we have a small business together.*

Once I began to feel better about myself, to understand that this was not my fault, that it was his fault, I could start thinking about the future, about other things, about my projects.*

…it’s a matter of starting to move on to other things; life doesn’t end there. Studying and learning to do things for ourselves is what allows us to continue to grow.*

One of the first things I did after this mf beat me was to go to the bank and ask for a loan. I got into debt, but I was able to get out of that place as soon as possible.*

Source: Own analysis and elaboration.

Empowerment derives partially from the availability of material resources, since women with personal income or assets have a better fallback position, decreasing their risk of being victims of domestic violence. However, the narratives point out that social resources are necessary to improve fallback position and, consequently, bargaining power. Support, a social resource, is a precondition for access to and control of economic and material resources\textsuperscript{76}.

\textsuperscript{73} Kabeer, “Resources, Agency, Achievements: Reflections on the Measurement of Women’s Empowerment.”

\textsuperscript{74} Agarwal, “‘Bargaining’ and Gender Relations: Within and Beyond the Household”.

\textsuperscript{75} Kabeer, “Resources, Agency, Achievements: Reflections on the Measurement of Women’s Empowerment.”

\textsuperscript{76} Agarwal, “‘Bargaining’ and Gender Relations: Within and Beyond the Household.”
Support: Family and Community

The greatest share of narratives with positive sentiment focus on support (43%). A topic analysis of the variable of support shows that narratives focus on primary needs (29%), financial support (27%), and emotional support (17%), as well as on practical support—e.g., knowledge about women’s rights, care routes, and strategies, among others (27%) (Figure 9).

This suggests the need to restructure social bonds to help women overcome the experience of GBV. The support women receive from the collective—community, family, and friends—lays the groundwork for their empowerment process in both a collective and individual way. According to the narratives, a solid support network allowed survivors to access other resources and thus make strategic decisions, such as continuing their education, starting a business, leaving a violent relationship, finding a job, or denouncing their aggressor, as well as improving their fallback position and strengthening their bargaining power.

It is important to note that support in many instances can be found online. According to the narratives, online communities such as women’s groups help survivors find people in similar situations to provide support ranging from listening, guiding, and finding professional help to financial assistance. Information and communication technologies can provide innovative approaches to addressing GBV from a community perspective. For example, Ramirez et al. made use of WhatsApp to create focus groups and conduct research activities around GBV. When the research activities concluded, these WhatsApp groups were organically transformed into support groups for the members. The same can occur on digital platforms. However, more information is needed about their use and effectiveness in designing and implementing strategies among women survivors of GBV.
5. Conclusions and Recommendations

The sentiment and topic analysis of narratives of women survivors of GBV with NLP is innovative. It allows us to understand the process of empowerment of survivors from their perspectives in an unstructured and spontaneous way. It also surfaces emerging themes or variables that we can study further using traditional methodologies such as in-depth interviews, focus groups, and surveys.

The negative sentiment analysis showed that scarcity of resources is the greatest barrier to women moving forward after surviving GBV. Stigmatization, lack of control, and emotional impact closely follow scarcity of resources as presenting significant barriers. The topic analysis around the variable of scarcity of resources showed that lack of social and financial resources and lack of institutional protection are the main issues related to such scarcity.

The positive sentiment analysis identified support as the main element enabling women to leave situations of violence. Equally crucial were achievements, plans and goals, and material and economic resources. The topic analysis conducted to understand support from a survivors’ perspective showed that family and community support allows women to meet basic needs and access economic resources.

The results of this study show the significance of social and economic resources in the process of empowerment after GBV. It is important that, in addition to existing services—for example, access to justice and psychosocial support—the economic component becomes a central part of programs led by foundations and grassroots organizations that seek to assist women survivors, as well as of the Attention Routes put in place by the government. The online narratives demonstrate the importance of considering social resources such as support in designing and implementing strategies to promote women’s empowerment processes after experiencing GBV.

In the same vein, the absence of the role of the state in the positive sentiment analysis and its mention in the negative sentiment analysis can be telling in terms of the quality of services provided or the distrust they inspire in survivors. This evidence also emerges in the findings of Martínez-Restrepo, Ramírez, Castillo and colleagues. Likewise, Ramírez and Castillo found that, in rural areas affected by the armed conflict, grassroots organizations provide most of the psychosocial care and they, rather than government Attention Routes, are most likely to assist in the process of economic empowerment. Gammage, Kabeer and Van der Meulen Rodgers even argue that, in order to improve the conditions of employment and entrepreneurship among women, it is necessary to direct material and economic resources to women’s grassroots organizations, as well as to unions and cooperatives. However, since the data in this analysis are from spontaneous narratives gathered from Big Data repositories, we cannot assume that an element or strategy does not exist if it is not mentioned. Nevertheless, it is possible to assume that such an element is not of significant importance for the survivors to mention it in the online narratives.

The Internet constitutes a rich source of information to understand the needs of survivors. At the same time, it can become a tool to address GBV. Thanks to the spontaneity that characterizes these conversations,

83 Martínez-Restrepo, et al., “El continuum de violencias basadas en género en el contexto del conflicto armado colombiano y su relación con el empoderamiento económico de las sobrevivientes.”
84 Ramírez & Castillo León, “El Trabajo de Las Organizaciones de Mujeres En Las Regiones En El Marco de La Covid-19.”
85 Gammage, Sarah; Kabeer, Naila & Van der Meulen Rodgers, Yana. “Voice and Agency: Where Are We Now?”
they can include elements that do not commonly appear in structured survey data or semi-structured interviews or focus groups. These techniques also allow studying these patterns at scale, meaning from millions of qualitative data (online narratives).

This Big Data analysis is not intended to replace existing qualitative methods or econometric regressions. On the contrary, there are many alternatives to enrich these traditional analytical methods. New categories of analysis coming from spontaneous conversations of unstructured data that cannot be approached by traditional methodologies.

It is necessary to study the role of public institutions—in Colombia, the Attention Routes—in the process of empowerment of GBV survivors. Likewise, lack of mobility, stigma, and lack of control emerged as themes in Big Data analysis and should be studied further using other methodologies.

There is great potential in the use of Big Data and its analysis with NLP to understand phenomena related to gender equality, women’s empowerment, and GBV. This study is part of a new generation of analysis. Analyzing millions of data found in conversations or spontaneous narratives allows us to continue understanding the process of women’s empowerment and its relationship with GBV.
Analysis of Online Narratives on the Empowerment of Women Survivors of Gender-based Violence in Colombia with Natural Language Processing