The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2444-8494.htm

Received 11 October 2022
Revised 2 March 2023
Accepted 3 March 2023

Exploring the use of gender-fair language by influencers

Carolina Nicolas
Departamento de Administración, Facultad de Administración y Economía, Universidad de Santiago de Chile, Santiago, Chile
Angelica Urrutia
Facultad de Ingeniería, Universidad Finis Terrae, Santiago, Chile, and
Gonzalo González
Universidad Católica del Maule, Talca, Chile

Abstract

Purpose – Explore the use of Gender-Fair Language (GFL) by influencers on Instagram.

Design/methodology/approach – The clustering methodology. A digital Bag-of-Words (BoW) Method called GFL Clustering BoW Methodology to identify whether an inclusive marketing (IM) strategy can be used. Thus, this research has a methodological and practical contribution to increasing the number of marketing technology tools.

Findings – This study is original as it proposes an inclusive digital marketing strategy and contributes with methods associated with digital transfers in order to improve marketing strategies, tactics and operations for inclusive content with a data integrity approach.

Research limitations/implications – Due to the limitations of the application programming interface (API) of the social network Instagram, a limited number of text data were used, which allowed for retrieving the last 12 publications of each studied profile. In addition, it should be considered that this study only includes the Spanish language and is applied to a sample of influencers from Chile.

Practical implications – The practical contribution of this study will lead to a key finding for the definition of communication strategies in both public and private organizations.

Originality/value – The originality of this work lies in its attractive implications for nonprofit and for-profit organizations, government bodies and private enterprises in the measurement of the success of campaigns with an IM communicational strategy and to incorporate inclusive and non-sexist content for their consumers so as to contribute to society.

Keywords Inclusive language (IL), Inclusive marketing (IM), Social marketing (SM), Gender-fair language (GFL), Digital content (DC), Marketing analytics (MA), Marketing technology (MarTech), Digital methods (DMs), Digital marketing (DM), Social networks (SN), Social inclusion (SI), Social diversity (SD)

Paper type Research paper

1. Introduction

Gender Equality (GE) is a significant problem worldwide, and the current pandemic jeopardizes many achievements in health, economy, safety and welfare (UNFPA, 2020). Previous research about Gender-Fair Language (GFL) has established that language is a variable that influences the attitudes, behavior and perceptions of society (De Lemus and

Declaration of competing interest: The authors declare that they have no conflict of interest.

Funding: This research was funded by Dirección de Investigación Científica y Tecnológica (DICYT) of University of Santiago of Chile (USACH), Chile, grant number 032393NA.

Estevan-Reina, 2021; Patev *et al.*, 2019; Heilman and Caleo, 2018; Sczesny *et al.*, 2016; Koeser *et al.*, 2015). Other studies show that GFL aims to reduce gender stereotypes and discrimination (Sczesny *et al.*, 2016; Koeser *et al.*, 2015), which is now well established. The measurement of the use of inclusive language (IL) by both Private and Governmental Influencers will assist in improving organizational policies.

The motivation of this study is to contribute to the achievement of Sustainable Development Goal Number 5, Gender Equality (GE), from Social Marketing (SM). In this context, defining DMs for the use of IL becomes relevant to the new consumption of marketing technology (MarTech), i.e. more awareness, inclusion, and a social sense of welcoming and belonging. Defining an inclusive marketing (IM) with DMs implies working with the conviction that the DM communication strategies must promote a social value that requires a change in attitude and behavior throughout society (Lin *et al.*, 2021).

Research on the use of DM and MarTech tools for Decision Support Systems are topics that have been growing in the literature (Humphrey *et al.*, 2021). Studies on the topics now reveal the benefits of using MarTech for performance management related to Customer Experience analysis. Artificial Intelligence (AI), predictive and automated learning analyses incorporate into DM (Kotler, 2021).

The questions research hypothesis as this study answers are: (1) What is the level of use of IL by the Top Instagram Influencers in Chile in 2021? (2) What words allow for determining the minimum degree of use of GFL? (3) What is the relationship between the Top Instagram Influencers in Chile in 2021 and the use of IL in DM?

The study aims to explore the use of GFL by Influencers on Instagram by proposing a Bag-of-Words (BoW) Method called GFL Clustering BoW Methodology to identify whether an IM strategy can be used. Thus, this research has a methodological and practical contribution to increasing the number of MarTech tools.

In practice, this study will help define communication strategies for DMs in public and private organizations to improve the identification of IC in social networks (SN) and can contribute to a more Inclusive Society. In addition, the study provides a methodology that supports teams creating DM by indicating which aspects of Instagram posts may generate stronger or weaker reactions from users, what being an Influencer means in terms of interactions, likes or comments per post, and what this implies in terms of engagement.

The rest of this paper is organized as follows: (2) Theoretical background, (3) Methodology, (4) Implementation of the methodology based on technological architecture, (5) Conclusions, (6) Managerial implications and (7) Limitations and potential future research directions.

2. Theoretical background

This study aims to explore the ways and degree of use of IL by Influencers in their SN (i.e. Instagram) to propose a methodology for digital content (DC) analysis as a tool that, in turn, supports SM analysis. Previous research shows how language is a mediator for the achievement of social inclusion (SI) and social diversity (Tankosić *et al.*, 2021; Piller and Takahashi, 2011). In addition, other authors have established that language is a variable that influences the attitudes, behavior and perceptions of society (De Lemus and Estevan-Reina, 2021; Patev *et al.*, 2019; Heilman and Caleo, 2018; Sczesny *et al.*, 2016; Koeser *et al.*, 2015; Koeser and Sczesny, 2014). The literature has identified a gap in the research of more complex methodologies based on SN information (Nicolas *et al.*, 2018); therefore, this study employs Text Analysis by Data Mining Algorithms (Urrutia *et al.*, 2021) to develop the GFL Clustering BoW Methodology.

2.1 GFL and social engagement (SE)

GE appears to be positively related to GFL or Gender-Inclusive Language (GIL) (Koeser *et al.*, 2015). A Web of Science (WoS) analysis shows 95 documents on the subject, with the first one on GFL, GIL and Gender-Neutral Pronouns (GNPs) published in 1993. In addition, 51% of the papers were produced in the last three years (see Figure 1).

Ample research using different experimental methodologies has confirmed the influence of linguistic forms on the access to mental representations of men and women (CEP-PIE, 2017; Stahlberg *et al.*, 2007). Sczesny *et al.* (2016) conducted a bibliographic review and highlighted the importance of implementing GFL in daily language and using it actively. Another research piece that stands out because of its number of citations is an analysis of the introduction of a third GNP in Swedish (Senden *et al.*, 2015).

Hollebeek (2011) signals that many distinct definitions exist for consumer Brand Engagement (BE). In our study, we are close to the definition of Calder *et al.* (2016), which defines BE as "a multilevel, multidimensional construct that emerges from the thoughts and feelings about one or more enrichment experiences involved in reaching a personal goal" (p. 40). Additionally, there are studies about Social Engagement (SE), such as the ones by Moezzi *et al.* (2017), which propose that SE is a way to understand, communicate and influence others. They recommend it as a data source and creative path toward SE. De Valck *et al.* (2009) believe that SN influences the behavior of its users. The SN theory points out that a human relations network can be bound to human behavior (Granovetter, 2011). Voorveld *et al.* (2018) show that social patterns in a Social Media environment enhance SE.

Commitment Measures through GFL do not appear in any consulted publications, but there are Measuring Systems in other areas. Buente *et al.* (2020) examined the SE of Betel [1] nut DC on the image-based platform through Instagram posts tagged #pugua. In retrospect, others measured SE by asking participants how likely they would like the post (through a survey) and how facial expression influenced the effects of visual sender presence.

In its most abstract form, predictive models refer to the use of mathematical tools to predict future results based on observed and assumed facts as input variables. Predicting an output includes, for example, future trends in behavior patterns (Iyer *et al.*, 2019).

Nowadays, with the extensive use of Social Media platforms, there are enormous amounts of data that are continuously generated and consumed by users, with valuable information about demographics, tastes, preferences and behaviors, which are the basis of Predictive Models (Bigsby *et al.*, 2019).

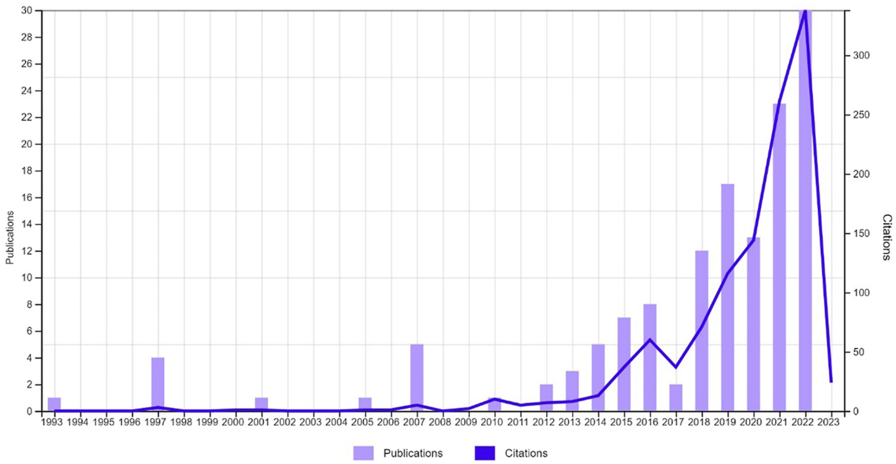


Figure 1.
Publications by year – gender-fair language (GFL)

Source(s): Web of Science (WoS), May 2022

No evidence exists in the literature about Impact Measurement when GFL is used as a DC strategy in SN regarding the SE. However, in recent years, there has been a growing number of papers about Predictive Models based on DC on Twitter – measured through retweets – specifically on sharing tweets inside one's user account. Today's retweet is the old well-known Word-of-Mouth (Ananda *et al.*, 2019). In addition, there is also a Metric to define the effectivity, popularity and influence (Nesi *et al.*, 2019; Scurlock *et al.*, 2020). Hence, knowing the motivations behind a retweet can be complex, but connecting with a Target Audience to know DC and achieve their SE.

Furthermore, a small but growing body of studies applies Predictive Models on Instagram. Some of the models are Neural Network, Convolutional Neural Network and TextCNN (Kumar and Sachdeva, 2021).

Piller and Takahashi (2011) notice that despite the SI becoming a normative framework of general politics, the ways language is used to generate SI have received little attention.

The Sentiment Analysis (SA) considers IL and GE's use levels (Koeser *et al.*, 2015). Sczesny *et al.* (2016) reviewed the literature from the time and found evidence in studies that prove that GFL is almost universal.

2.2 Sentiment analysis (SA) of Twitter data

Nagarajan and Gandhi (2019) conducted a SA with data obtained from Twitter. This study had a sample of 600 million public tweets, and they used opinion mining and automatic SA, proposed method shows a better analysis than others.

The study by Sahayak *et al.* (2015) concluded that the functions related to SN could be employed to predict the sentiment on Twitter. They found that SA for Twitter data varies in difficulty depending on the complexity of the expressions. For example, product reviews are a relatively simple field to analyze, as opposed to tweets about books, movies, art and music, from which sentiment is harder to extract.

2.3 Web data extraction systems for inline business intelligence (IBI)

Grigalis and Čenys (2013) state that nowadays, many customers do their shopping online and publicly share their user experience, opinions and purchase preferences. In most cases, users express their opinion in the form of comments or posts in forums or on SN. Customer SA is fundamental for companies to maintain a competitive advantage in delivering goods and services. Therefore, Inline Business Intelligence (IBI) solutions need access to Facebook, Twitter or any other SN to automatically identify publications about specific products, extract text, execute Natural Language Processing and understand the sentiment expressed by users.

Many big data and business intelligence (BI) software exist; the research uses tools of open access codes. One of them is the open-source Python programming language which was used to web scrape the target website (Bengfort *et al.*, 2018; Cury, 2019). Web Scraping is a technique used for automatic extraction of big data from the Internet, with many advantages: the data extracted are behavioral, the collection of datasets with millions of cases and unknown data extraction (Landers *et al.*, 2016). Another tool used was Pentaho Data Integration, business intelligence platform, which is an open-source and free tool, and its capability is world famous (Li *et al.*, 2021). It offers world-class data integration, online analytical processing (OLAP), data mining, reporting, and Extraction, Transform and Loading (ETL).

Finally, for the visualization of the results and to improve the BI process, we used Microsoft Power BI platform to create a report customized to key performance indicators (KPI) [2].

2.4 Influencers on social networks (SN)

Influencers are people who share their lives through SN and generate some degree of influence on the people who follow them called Followers (De la Piedra and Meana, 2017).

The types of influencers are as follows:

(1) Sector Specialists: They have the intuition to identify the evolutions of the sector and its different trends. They often collaborate with communication companies and organizations from different sectors.

(2) Product Specialists: They have valuable technical training for analyzing products in-depth.

(3) Niche Influencers: They have credibility with their Followers. They often advise companies.

(4) Generalist Influencers: They have Faithful Followers and write about different topics from a critical perspective. They are usually journalists or media professionals.

(5) Trend Influencers: They are specialists in their field, creative and able to revolutionize their fields to create new things.

(6) Occasional Influencers: They are high-ranking people from the cultural and political world.

(7) Reference Influencers: They hit sudden success by creating a company or brand and becoming well known.

The roles of influencers are (1) Inspiring: The Influencers need to be trusted by their Followers. To be considered, reference people in the topic, from whom there is always something new to learn. (2) Collaborator: The Influencers serve their Followers by sharing knowledge on how to stand out in the desired field. (3) Famous Star: The influencers always upload videos or photos depicting what they do at any moment. Part of their charm lies in this characteristic, as this is a way of staying close to their Followers. (4) Amplifier: The influencers are experts whom their Followers can trust. The Influencers have Faithful Followers in their DC. Potential Customers are essential to selling a product. (5) Critic: The Influencers need to know the Followers' opinions about the DC posted on the SN. For example, the purchase decisions to buy a product.

These roles in influencing another person in SN do not need to exist separately but can coexist.

3. Methodology

The GFL Clustering Methodology proposed uses a BoW of IL and non-sexist words (NSW) that we call "ILandNSW language". It may apply to Influencers of any country, not only on Instagram but any other SN like Facebook, YouTube, Twitter, TikTok or others. It has the following stages: (1) Influencer Identification, (2) Data Extraction through a Web Scraping (Landers *et al.*, 2016) code, (3) Transforming and Loading with the Pentaho Data Integration (Li *et al.*, 2021) tool, (4) Language Analysis of a Target Group using IL and NSW written by Influencers on their posts and (5) Data Visualization using the Power BI tool.

This Instagram Clustering Methodology can help

(1) To generate a Data Dictionary with a BoW of the Target Group using ILandNSW.

(2) To identify Conversation Nodes of Influencers and categorize them by the Types of Influencers.

(3) To identify whether a DC by an Influencer is complying with the ILandNSW of the Target Group.

(4) To obtain a Data Extraction process.

(5) To generate a Platform to visualize the Data Extraction process.

The following Set of Stages analyzes Instagram Influencers of any country:

Stage 1: Identifying and Selecting the Influencer Target Group.

Stage 2: Generating a GFL BoW for IL and NSW used by Influencers on Instagram.

Stage 3: Identifying and Extracting Profile and Instagram Data with Web Scraping.

Stage 4: Applying an ETL process with Web Scraping.

Stage 5: Creating, writing and saving the Profile and Instagram Data.

Stage 6: Analyzing the GFL by the Influencer Target Group.

Stage 7: Visualizing the Dashboard generated on Power BI.

Scopes: Concerning the BoW, this was created before the text analysis stage and included words belonging to the Influencer Target Group ILandNSW language. This BoW may grow if new words used in the ILandNSW language studied are discovered.

Concerning the BoW, this was created before the text analysis stage and included a small group of Spanish words belonging to the Influencer Target Group ILandNSW language used in Chile. This BoW may grow if new words used in the ILandNSW language studied are discovered. Adding new lines to the Excel file that stores the BoW and also adding other languages (English, French and so on), creating a new Excel file that stores other languages. Any difficulties you may encounter when doing something similar in another language are resolved by creating a new Excel file that stores the required words in other languages.

4. Implementation of the clustering methodology based on technology architecture

In this section, the results for the application of each proposed stage are shown in detail and the implemented Web Scraping code and dashboard are presented. The results of the same allow for answering the research questions of the study.

Stage 1: Identifying and Selecting the Influencer Target Group

In this stage, www.starngage.com shows the Top Instagram Influencers by country and year. In this study, the Country is Chile and the Year is 2020.

The Target Group is the first ten Chilean Influencers. Table 1 shows the Instagram Data. The parameters considered are Instagram Influencer Name, Followers, Following, Avg. Likes, Avg. Comments, Participation Rate and Total Publication (Posts).

Among these Influencers of the Target Group are Writers, Athletes and people from Entertainment and TV, all classified as Influencers based on their Number of Followers.

As shown in Figure 2, Arturo Vidal (@kingarturo23oficial) is the Influencer with the most significant Number of Followers, with a total of approximately 14 million, corresponding to 34.8% of Total Followers. In contrast, the Influencer with the least Followers is Raquel Calderón, with around two million, corresponding to 4.14% of Total Followers.

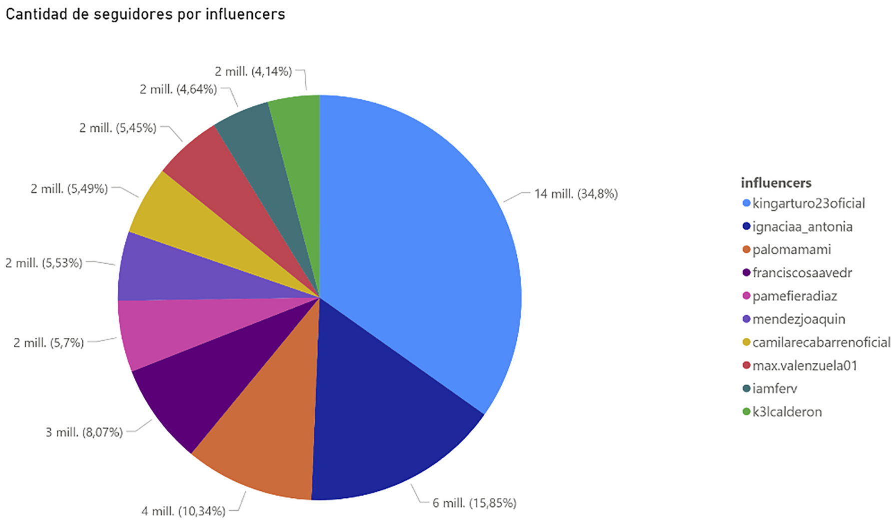
As shown in Figure 3, Accounting has 30% of the Total Followers. Three Influencers have this occupation. The occupations with the least Influencers are Lawyers, Singers and Professional Soccer Players, with one User each corresponding to 10% for each Influencer.

The Number of Followers by the Influencer and The Number of Influencers by Occupation is a Pie Chart Power BI.

Table 1.
Target group Instagram data, ten top Instagram influencers in Chile in 2020 (12/29/2020 www.starngage.com)

Instagram influencer name	Followers	Following	Avg. likes	Avg. comments	Participation rate	Total publication
@kingarturo23oficial	14,072,974	362	323,853	1,993	2.3%	1,719
@ignaciaa_antonia	6,411,946	272	618,341	8,181	9.70%	539
@palomamami	4,180,476	75	668,131	7,602	16,10%	144
@franciscosaavedr	3,264,034	4,740	6,335	134	0.10%	7,832
@pamefieradiaz	2,304,349	358	25,034	708	1.1%	1,872
@mendezjoaquin	2,236,564	3,635	39,119	1,887	1.8%	2,849
@camilarecabarrenoficial	2,219,578	1,271	71,138	357	3.2%	1,885
@max.valenzuela01	2,202,340	131	163,994	841	7.4%	31
@iamferv	1,877,563	423	331,493	5,209	17.9%	223
@k3lcalderon	1,675,004	2,089	33,778	770	2.0%	6,042

Note(s): www.starngage.com shows the Top Instagram Influencers by country and year. In this study, Chile and 2020
Source(s): Authors' elaboration



Source(s): Authors' elaboration

Figure 2.
Pie chart power BI of the number of followers by the influencer

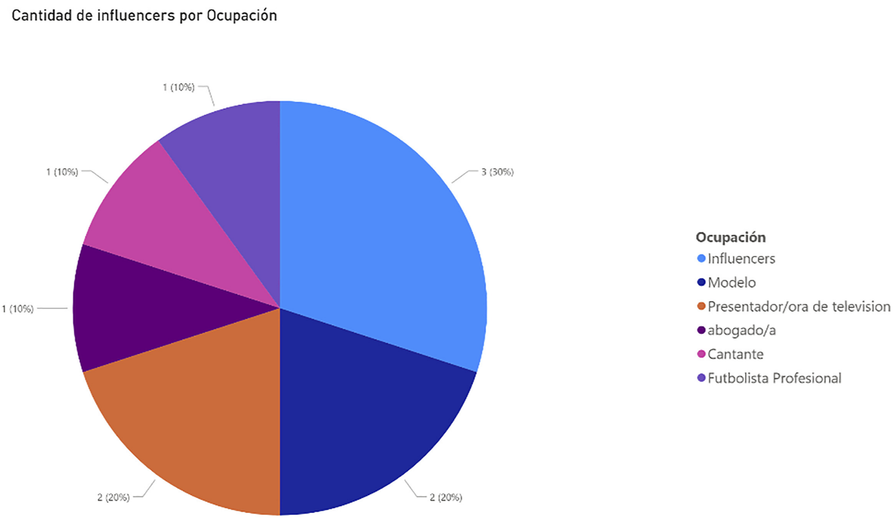
Stage 2: Generating a GFL BoW for IL and NSW used by Influencers on Instagram

In this stage, an Excel file stores the BoW that contains the IL and NSW. The column “Palabras” (Figure 4) is called “ILandNSW_words” in Stage 6.

The Identification of ILandNSW is conducted manually from Literature and comments from Instagram, Facebook and Twitter. Other sources are websites, TV and Newspapers, among others.

Stage 3: Identifying and Extracting Profile and Instagram Data with Web Scraping [3][4].

In this stage, Web Scraping will get the Instagram profile, such as the Instagram Influencer name, publication ID, publication description (what the Instagram Influencer wrote on the post), the number of comments and the number of likes per post (publications) for each Influencer of the Target Group.



Source(s): Authors' elaboration

Figure 3.
Pie chart power BI of the number of influencers by occupation

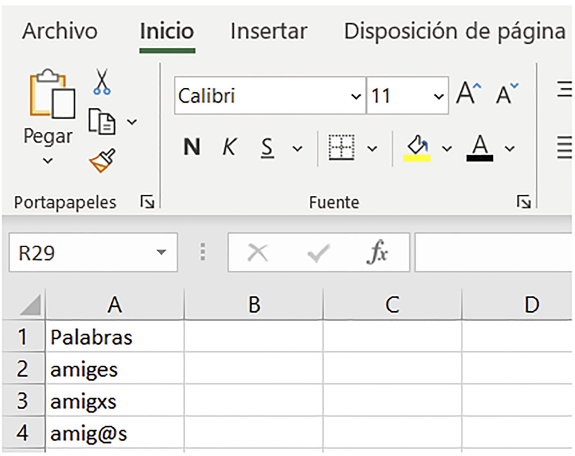
To get a public Instagram profile, a specific JSON (JavaScript Object Notation) URL [5] is implemented, see Figure 4.

The format is

profile = https://www.instagram.com/<instagram_influencer_name>/?__a=1.

In other words, to access a public influencer feed, the URL has a unique < Instagram_influencer_name > without the parameter @ that characterizes usernames on Instagram, and the parameter?_a = 1 to read all the Instagram Influencer posts. Table 2 shows the Instagram Influencer Name and the unique JSON URL for each Influencer of the Target Group.

Stage 4: Applying an ETL process with Web Scrapping



Source(s): Authors' elaboration

Figure 4.
Excel file BoW (bag-of-words) that contains the inclusive language (IL) and Non-sexist words (NSW)

In this stage, the JSON library, REQUESTS library and CSV (Código Seguro de Verificación) library extract, transform and load the profile and Instagram data for each Influencer of the Target Group using Web Scrapping. The JSON format structure to browse the "Instagram_data_not_clean" file is shown in Figure 5.

The Python Code (Figure 6) starts with import json, import requests and import csv. The array tags contain the Instagram Influencer names (see Table 2), and the array keys contain the Instagram profile.

The Instagram profile extraction from the JSON file considers the following parameters: id (Publication ID), owner (Instagram Influencer name), edge_media_to_caption [6] (text), edge_media_to_comment [7] (count) and edge_like_by [8] (count).

The Instagram data extraction from the JSON file considers the following parameters: "graphql", "user", "edge_owner_to_timeline_media", "edges" and "node".

The WRITER function csv.writer(Instagram_data_not_clean) [9][10] converts the Instagram data into a delimited strings chain with a write( ) method that writes on the CSV[11] called "Instagram_data_not_clean.csv". Therefore, once the file is created or opened, the function csv.writer( ) is employed to deliver a writer object that transforms user data into a delimited chain of characters.

Table 2.
Instagram influencer name and JSON URL for each influencer of the target group

Instagram influencer name	JSON URL
@kingarturo23oficial	https://www.instagram.com/kingarturo23oficial /?__a = 1
@ignaciaa_antonia	https://www.instagram.com/ignaciaa_antonia /?__a = 1
@palomamami	https://www.instagram.com/palomamami /?__a = 1
@franciscosaavedr	https://www.instagram.com/franciscosaavedr /?__a = 1
@pamefieradiaz	https://www.instagram.com/pamefieradiaz /?__a = 1
@mendezjoaquin	https://www.instagram.com/mendezjoaquin /?__a = 1
@camilarecabarrenoficial	https://www.instagram.com/camilarecabarrenoficial /?__a = 1
@max.valenzuela01	https://www.instagram.com/max.valenzuela01/?__a = 1
@iamferv	https://www.instagram.com/iamferv/?__a=1
@k3lcalderon	https://www.instagram.com/k3lcalderon /?__a = 1

Source(s): Authors' elaboration

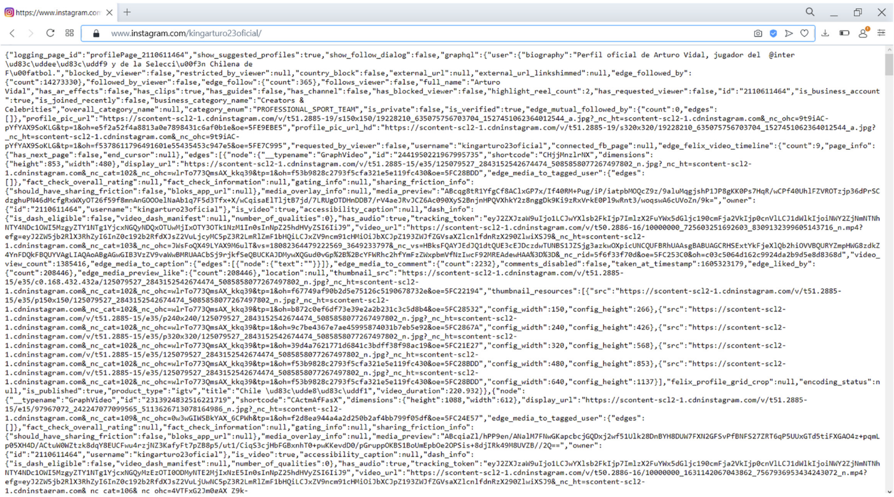


Figure 5.
Shows the instagram influencer URL in JSON format

Source(s): Authors' elaboration

The function FOR for tag in tags fetches the Instagram profile and the Instagram data from the JSON file writer = csv.writer("Instagram_data_not_clean") for each Influencer of the Target Group through the array tags that sweep the tag and extract each url of Table 2, one by one through the function get_ig_page(url) to then use them in the function "fetches" through the variable ig_data_dict.

The function def get_ig_page(url, session = None) contains two parameters, "url" that returns the content of the page, and "session" [12] that calls the requests.session( ) method and calls the requests GET method session.get(url) that extracts the value of the variable "url", i.e. the content of the Influencer page, and saves it in the variable "response".

The "response.status_code" [13] method saves the value of the status in the variable "response_status_code".

In the function IF if response_status_code = = requests.codes.ok [14]. If the value of the requests status code is equal to the value of the status code, the value of the variable "url" is assigned to the variable "response"; otherwise, the value of "response" remains as None.

The function def fetches(ig_data_dict, writer) contains two parameters, "ig_data_dict" the content of the page and "writer" the JSON file [15]. This function has two IF and one FOR.

In the function IF if ig_data_dict is not None, it calls ig_data_dict.json( ) method that completes the data crawled with graphql keyword.

The requests GET method ig_data_dict.get("graphql", None), data.get("user") and user.get("edge_owner_to_timeline_media", None) extract the values of the attributes "graphql", "user" and "edge_owner_to_timeline_media", which contains an Instagram data fetch parsing, an Instagram Influencer name, and a photo or video. Those values are stored in graphql, user and posts, respectively.

```
{
    'graphql':{
        'user':{
            'edge_owner_to_timeline_media':{
                'edges':[
                    'node':{
                    }
                ]
            }
        }
    }
}
```

Note(s): It is an attribute to fetch Instagram data with queries. A query is a GraphQL Operation that retrieves specific data from the server. The most common way to browse a GraphQL API is to use GraphiQL. GraphiQL is a tool built by Facebook (pronounced "graphical") that makes it easy to explore any GraphQL API. https://hasura.io/learn/graphql/intro-graphql/graphql-queries/
It is an attribute with all the details of photos and videos posted on Instagram.
It is a node containing zero, one, or more publications (photos and videos) on Instagram
Source(s): Authors' elaboration

Figure 6.
JSON format structure ["graphql"]["user"]["edge_owner_to_timeline_media"]["edges"]["node"] to browse the "Instagram_data_not_clean" file

In the function IF if posts is not None, it calls the requests GET method posts.get("edges",None). To browse "edges", each publication for each Influencer of the Target Group (see Table 2) is contained by a different "node". The Instagram profile extraction starts once inside the "node".

To navigate the posts (publications for each Influencer of the Target Group), the function FOR for post in posts accesses the node for each post iteration.

The variables to grab are id, owner, edge_media_to_caption, edge_media_to_comment and edge_liked_by. In other words, the Publication ID, the Instagram Influencer name, the description of the photo or video posted by the Influencer on Instagram, the number of User comments per post, and the number of User likes per post for each Influencer of the Target Group (see Table 2), respectively.

Below show the Instagram profile extraction process scraping = post.get("node", None).

Instagram Scraping for Id: The function IF if keys = = "id", a variable id stores the Publication ID for each publication. The requests GET method scraping.get(keys) extracts the value of the array keys. The dumps JSON method json.dumps(keys) decodes the value and saves it in the variable id, as shown in Figure 7.

```
import json
import requests
import csv
tags=['kingarturo23oficial','ignaciaa_antonia','palomamami','franc
iscosaavedr','mendezjoaquin','camilarecabarrenoficial','pamefierad
iaz','max.valenzuela01','iamferv','k3lcalderon']
keys=['id','owner','edge_media_to_caption','edge_media_to_comment'
,'edge_liked_by']
writer = csv.writer('instagram_data_not_clean')
for tag in tags:
        url = 'https://www.instagram.com/'+ tag +'/?__a=1'
        ig_data_dict = get_ig_page(url)
        fetches(ig_data_dict, writer)

def get_ig_page(url, session=None):
    session = session or requests.session()
    response = session.get(url)
    response_status_code = response.status_code
        if response_status_code == requests.codes.ok:
        return response
    else:
        return None
def fetches(ig_data_dict, writer):

    if ig_data_dict is not None:
        ig_data_dict = ig_data_dict.json()
        graphql = ig_data_dict.get('graphql', None)
        user = graphql.get('user')
        posts = user.get('edge_owner_to_timeline_media', None)

        if posts is not None:
            posts = posts.get('edges',None)

            for post in posts:
                scraping = post.get('node', None)
```

Source(s): Authors' elaboration

Figure 7.
Python code instagram profile web scraping

Instagram Scraping for Owner: The function IF if keys = = "owner", a variable owner stores the Instagram Influencer name (see Table 2). The requests GET methods scraping.get(keys) and keys.get("instagraminfluencername") extract the values of the array keys and the Instagram Influencer name "instagraminfluencername" in the study. The dumps JSON method json.dumps(instagraminfluencername) decodes the value and saves it in the variable owner, as shown in Figure 8.

Instagram Scraping for Edge_media_to_caption: The function IF if keys = = "edge_media_to_caption", a variable edge_media_to_caption stores the description per post of photos or videos posted by the Influencer on Instagram. The requests GET methods scraping.get("edge_media_to_caption",None), edge_media_to_caption.get("edges", None), edge.get("node", None) and node.get("text") extract the value of the attributes "edge_media_to_caption", "edges", "node" and finally the description of photos or videos posted by the Influencer in the attribute "text". The dumps JSON method json.dumps(edge_media_to_caption) decodes the value and saves the value of "text" in the variable edge_media_to_caption, as shown in Figure 9.

To navigate the edges, the function FOR for edge in edges accesses the "node" for each iteration of edge up to grab "text".

Instagram Scraping for Edge_media_to_comment: The function IF if keys = = "edge_media_to_comment", a variable edge_media_to_comment stores the number of comments per post by the User of photos or videos posted by the Influencer on Instagram. The requests GET methods scraping.get("edge_media_to_comment", None) and edge_media_to_comment.get("count") extract the value of the attributes "edge_media_to_comment" and "count", and save the value of "count", i.e. the number of comments per post, in the variable edge_media_to_comment, as shown in Figure 10.

```
if keys == 'id':
   keys = scraping.get(keys)
   id = json.dumps(keys)
```

Source(s): Authors' elaboration

Figure 8. Instagram scraping for Id. Extraction of the publication ID for each publication

```
if keys == 'owner':
      keys = scraping.get(keys)
      instagraminfluencername = keys.get('instagraminfluencername')
      owner = json.dumps(instagraminfluencername)
```

Source(s): Authors' elaboration

Figure 9. Instagram scraping for owner. Extraction of the instagram influencer name

```
  if keys == 'edge_media_to_caption':
      edge_media_to_caption    =    scraping.get('edge_media_to_caption',
None)
      edges = edge_media_to_caption.get('edges', None)
      for edge in edges:
            node = edge.get('node', None)
            edge_media_to_caption = node.get('text')
            edge_media_to_caption = json.dumps(edge_media_to_caption)
```

Source(s): Authors' elaboration

Figure 10. Instagram Scraping for Edge_media_to_caption. Extraction of the description per post of photos or videos posted by the influencer on instagram

Instagram Scraping for Edge_liked_by: The function IF if keys = = "edge_liked_by", a variable edge_liked_by stores the number of likes per post by the User of photos or videos posted by the Influencer on Instagram. The requests GET methods scraping.get("edge_liked_by", None) and edge_liked_by.get("count") extract the value of the attributes "edge_liked_by" and "count", and save the value of "count", i.e. the number of likes per post in the variable edge_liked_by, as shown in Figure 11.

Stage 5: Creating, writing and saving the Profile and Instagram Data

In this stage, once the profile data id, owner, edge_media_to_caption, edge_media_to_comment and edge_liked_by of the Instagram Influencer Target Group (see Table 2) and data extraction of graphql, user, edge_owner_to_timeline_media, edges and node ends, the Instagram data is created, written and stored in a CSV file, called "Instagram_data_not_clean.csv".

The Python code Instagram Data Web Scraping (see Figure 12) starts with import json, import re and import pandas as pd.

To see the number of data, the number of columns, the data type of each column and the number of nulls of the CSV file, a number of PANDAS functions are necessary, as shown in Figure 12.

The function read_csv("Instagram_data_not_clean.csv") loads the data file in a variable df = pd.read_csv("Instagram_data_not_clean.csv").

The function DataFrame(df) transforms df into a Data Frame and loads it in the variable df = pd.DataFrame(df).

The function df.shape gets the number of rows (registers) and the number of columns (attributes) of the Data Frame.

The function df.info gets the types of data.

The function pd.isnull(df).sum( )) gets the total of null data per attribute.

The results of PANDAS functions: Shape, Info and isNull may be observed in Table 3a and 3b.

No missing data in the "Instagram_data_not_clean.csv" file, but the column Edge_media_to_caption has emojis, extra inverted commas, accents, special characters, uppercase and lowercase. Once all of this is eliminated, the identification is made manually, and the whole column is transformed into lowercase. An Excel file is created called "data_clean.xls".

Therefore, the Instagram profile is saved in a CSV file, not clean, and then cleaned in an Excel file. As shown after the Shape, Info and isNull PANDAS results in Figure 12.

The function open("Instagram_data_not_clean.csv", "w", newline = ") receives the following parameters: a file name, a letter "w" for a file writing and a new line parameter newline = "', to add an empty line, as shown in Figure 13.

Figure 11.
Instagram scraping for Edge_media_to_comment. Extraction of the number of comments per post by the User of photos or videos posted by the influencer on instagram

```
  if keys == 'edge_media_to_comment':
      edge_media_to_comment  =  scraping.get('edge_media_to_comment',
None)
      edge_media_to_comment = edge_media_to_comment.get('count')
```

Source(s): Authors' elaboration

Figure 12.
Instagram Scraping for Edge_liked_by. Extraction of the number of likes per post by the user of photos or videos posted by the influencer on instagram

```
if keys == 'edge_liked_by':
    edge_liked_by = scraping.get('edge_liked_by', None)
    edge_liked_by = edge_liked_by.get('count')
```

Source(s): Authors' elaboration

The WRITEROW [16, 17] function writes items in a sequence, separating them by a comma character. After creating the chain, the writerow( ) function writes “six columns” on the CSV file: “id”, “owner”, “edge_media_to_caption”, “edge_media_to_comment”, “edge_liked_by” and “isILandNSWused”, for the Publication ID, the Instagram Influencer name, the description of the photo or video posted by the Influencer on Instagram, the number of User comments per post, the number of User likes per post and a numerical attribute isILandNSWused, which is “1” if the publication belonged to the IL and NSW contain in the BoW and “0” if not, for each Influencer of the Target Group (see Table 2), respectively, as shown in Figure 13.

Stage 6: Analyzing the GFL by the Influencer Target Group

In this stage, once the Instagram profile of the Influencer Target Group (see Table 2) is clean and loaded in an XLS file, as well as the BoW that contains the IL and NSW. The two of them, “data_clean.xls” and “BoW.xls”, are used as parameters in a function read_excel(“data_clean.xls”) and read_excel(“BoW.xls”) that loads the Excel data files in a variable file_clean = pd.read_excel(“data_clean.xls”) and file_BoW = pd.read_excel(“BoW.xls”). And then, it transforms into a Data Frame: DataFrame(file_clean) and DataFrame(file_BoW) that are loaded in the variables df_file_clean = pd.DataFrame(file_clean) and df_file_BoW = pd.DataFrame(file_BoW), respectively, as shown in Figure 13.

Thus, the XLS files contain the clean and lowercase text of the publications per post of the Influencer Target Group “data_clean.xls” (i.e. the Data Frame df_file_clean) and the IL and NSW, the ILandNSW language “BoW.xls” (i.e. the Data Frame df_file_BoW).

A numerical variable called isILandNSWused is defined, which initializes from zero.

The function FOR for i in df_file_clean.index swipes the Data Frame df_file_clean by its index to enter the column “edge_media_to_caption” for each row iteration of i. The value is stored in the variable df_caption_post.

The function IF if nan verifies the result of the function isCaptionPostNull(df_caption_post). If the result is “TRUE”, the post contains a “NaN’value” and therefore is ignored since the function re.search( ) cannot process these values. If the result is “FALSE”, the function check_ILandNSW(df_file_BoW,df_caption_post) is called, through which a value from one to zero is obtained if this is coincident or not with a BoW in the post. Therefore, the value of the function check_ILandNSW(df_file_BoW,df_caption_post) is assigned to the variable isILandNSWused; otherwise, the value of isILandNSWused remains as 0.

Then, with all the results, the row analyzed is saved through the function writer.writerow( ), as shown in Figure 13.

(a) Results for number of rows (publications) and number of columns (attributes)

Number of rows	Number of columns
120	5

(b) Results for data type and total null data

Attribute	Data type	Total null data
Id	Object	0
Owner	Object	0
Edge_media_to_caption	Object	0
Edge_media_to_comment	Int64	0
Edge_liked_by	Int64	0

Source(s): Authors’ elaboration

Table 3. CSV file “Instagram_data_not_clean.csv” Shape, Info and isNull results

```
  import json
  import re
  import csv
  import pandas as pd

  df=pd.read_csv('instagram_data_not_clean.csv')
  df=pd.DataFrame(df)
  df.shape
  df.info
  pd.isnull(df).sum())
  data_not_clean = open('instagram_data_not_clean.csv', 'w',newline='')
  writer = csv.writer(data_not_clean)

writer.writerow(['id','owner','edge_media_to_caption','edge_media_to_commen
t','edge_liked_by','isILandNSWused'])

  file_clean = pd.read_excel('data_clean.xls')
  df_file_clean = pd.DataFrame(file_clean)
  file_BoW = pd.read_excel('BoW.xls')
  df_file_BoW = pd.DataFrame(file_BoW)

  isILandNSWused = 0
  for i in df_file_clean.index:
      df_caption_post = df_file_clean['edge_media_to_caption'][i]
        nan = isCaptionPostNull(df_caption_post)
        if nan:
           isILandNSWused = 0
           pass
        else:
            isILandNSWused = check_ILandNSW(df_file_BoW,df_caption_post)

  writer.writerow([df_file_clean['id'][i],df_file_clean['owner'][i],
df_file_clean['edge_media_to_caption'][i],df_file_clean['edge_media_to_comm
ent'][i],df_file_clean['edge_liked_by'][i], isILandNSWused])

  def check_ILandNSW(BoW, caption_post):
      isBoWinCaptionPost = 0
      for ILandNSW_word in df_file_BoW['ILandNSW_word']:
          if re.search(ILandNSW_word, caption_text):
              isBoWinCaptionPost = 1
          else: pass
      return isBoWinCaptionPost
  def isCaptionPostNull(text):
      return text != text
```

Figure 13.
Python code instagram data web scraping

Source(s): Authors' elaboration

The function def check_ILandNSW(BoW, caption_post) has two parameters BoW and caption_post. The first represents the words belonging to the BoW. The second post was extracted from the Influencer post.

A numerical variable called isBoWinCaptionPost is defined, which initializes from zero. This variable has the function of defining whether there are ILandNSW words from the Data Frame df_file_BoW in the text of the column "edge_media_to_caption" of the Data Frame df_file_clean.

Then, a FOR for ILandNSW_word in df_file_BoW["ILandNSW_words"] sweeps the column "ILandNSW_words" of df_file_BoW["ILandNSW_words"] and extracts each IL and NSW "ILandNSW_word" of them. The function re.search(ILandNSW_word, caption_text) verifies whether that ILandNSW_word is contained or not in caption_text. If ILandNSW_ word is contained, a value of one is assigned to the variable isBoWinCaptionPost; otherwise, the value remains zero. Finally, this function returns the value of the variable isBoWinCaptionPost (0 or 1), as seen in Figure 13.

The function def isCaptionPostNull(text)function verifies whether the text is or not "NaN". As the "NaN" values are not the same as any other value, not even to the same "NaN", one value returns a comparison with itself. If it is equal to itself, it returns a "TRUE", and if not, a "FALSE".

Stage 7: Visualizing the Dashboard generated on Power BI

When the previous step ends and the data analysis is complete, different visualizations are conducted using the tool Microsoft Power BI. These visualizations seek to respond to the different management indicators and observe how data behave. Some of the results of this step are shown below.

The number of publications by language type. To calculate this indicator, the option "pie chart" is used, in which publications that comply or do not with the ILandNSW language are differentiated by percentage. "Inclusive" language type means that the language belongs to the ILandNSW language, while "non-inclusive" means the contrary, as shown in Figure 14.

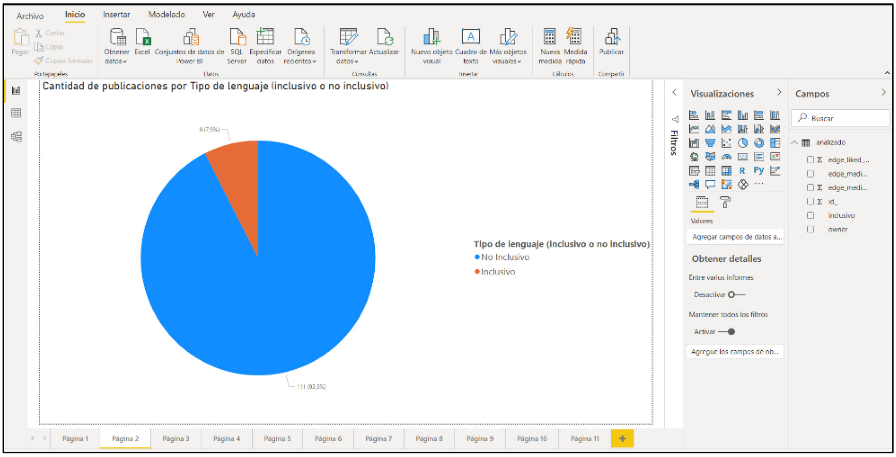
The number of publications by Influencers and language type. The option "grouped bar chart" is employed to calculate this indicator, which shows two columns per Influencer. These columns show the number of publications by language type. The red column represents the publications that do not comply with the ILandNSW language, and the blue is those that comply with it. See Figure 15.

The number of likes. A visualization called "Tarjeta" (card) shows the total number of likes in all publications employed, as seen in Figure 16.

The number of comments. A visualization called "Tarjeta" (card) shows the total number of comments in all publications employed, as seen in Figure 17.

The number of publications. A visualization called "Tarjeta" (card) shows the total number of publications employed, as seen in Figure 18.

Publications with more likes. To calculate this indicator, the option "Table" is employed, which consists of a list with a description written by Influencers on each post, the user who wrote the publication and the number of likes of the same. This list is arranged in decreasing order by the number of likes, as shown in Figure 19.



Source(s): Authors' elaboration

Figure 14.
The number of publications by language type

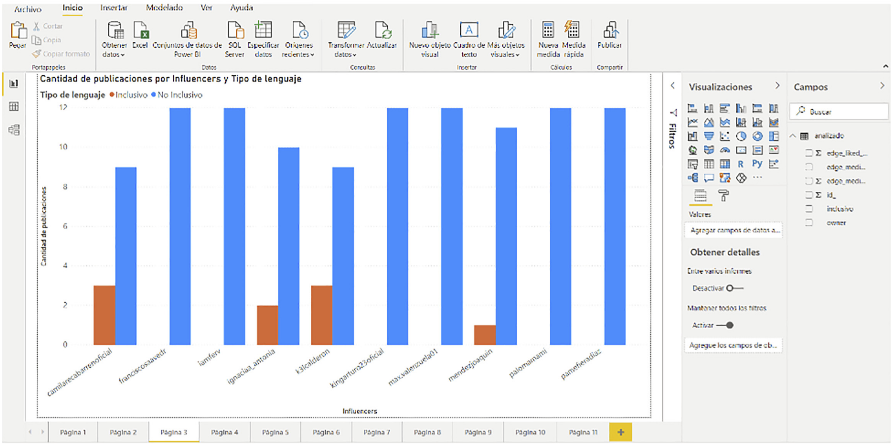


Figure 15.
The number of publications by influencers and types of language

Source(s): Authors' elaboration

Source(s): Authors' elaboration

Figure 16.
The number of likes

Source(s): Authors' elaboration

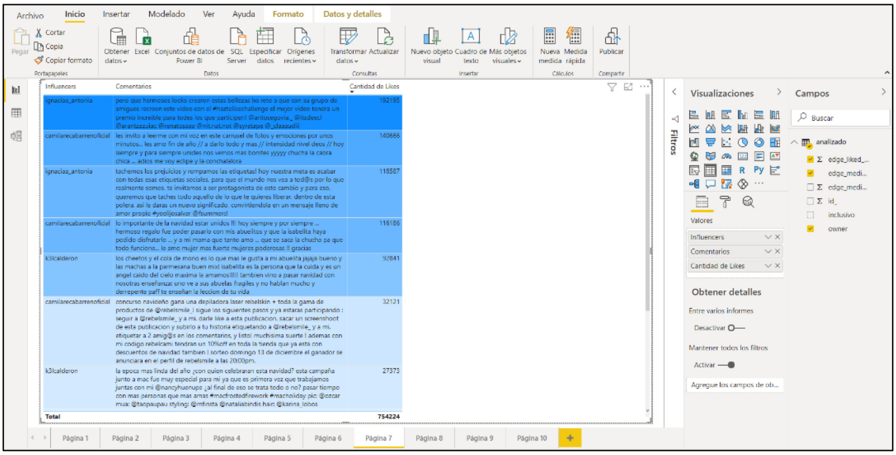
Figure 17.
The number of comments

Source(s): Authors' elaboration

Figure 18.
The number of publications

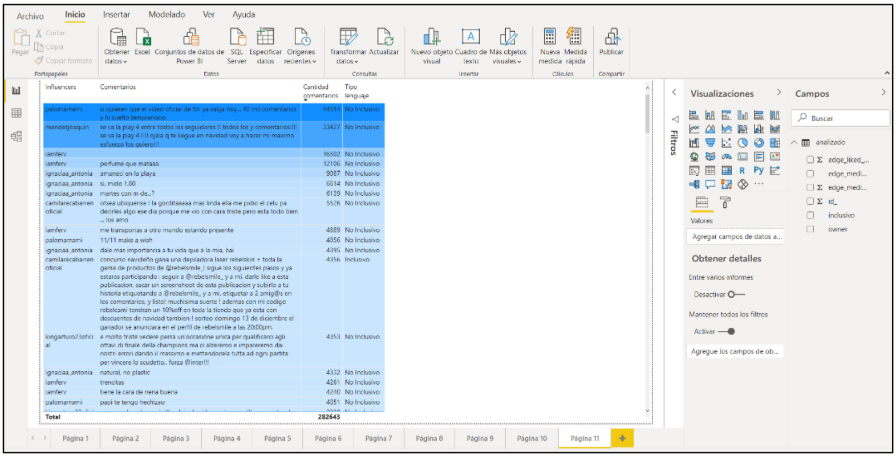
Publication with more comments. To calculate this indicator, the option "Table" is employed, which consists of a list with a description written by Influencers on each post, the user who wrote it and the number of likes of the publication. This list is arranged in decreasing order by the number of comments, as shown in Figure 20.

The number of likes and comments made by Influencers. To calculate this indicator, the option "grouped bar and line chart" is used to show the relationship between the number of likes and the number of comments by an Influencer. See Figure 21.



Source(s): Authors' elaboration

Figure 19. Publications with more likes



Source(s): Authors' elaboration

Figure 20. Publications with more comments

Finally, the dashboard is generated after creating the charts, which aims to provide a global overview of all management indicators analyzed in the study. Figure 22 shows that the dashboard created allows for visualizing the results in such a way that they are more interactive or intuitive and for visualizing the diverse relationships that may exist between data; for example, between the number of likes and comments.

The dashboard contains two panels that allow for implementing filters to constrain data. If a field of these filters is selected, the data supplied to the other charts are automatically constrained to comply with the selected filter. At a glance, the "pie chart" shows a clear difference between the number of publications that meet the ILandNSW language, with only 7.5% of total publications complying with the ILandNSW language under study and 92.5% not complying. Thus, it is concluded that this type of language is not usually employed by Influencers in their publications on the SN Instagram.

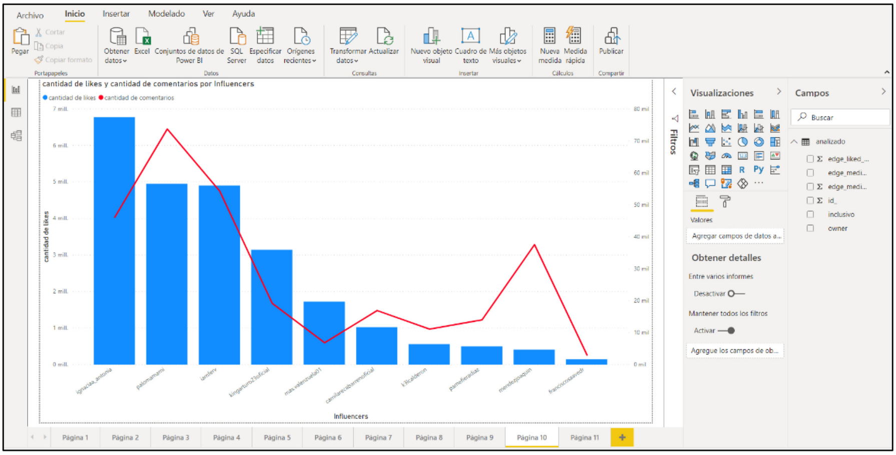


Figure 21.
The number of likes and comments made by Influencers

Source(s): Authors' elaboration

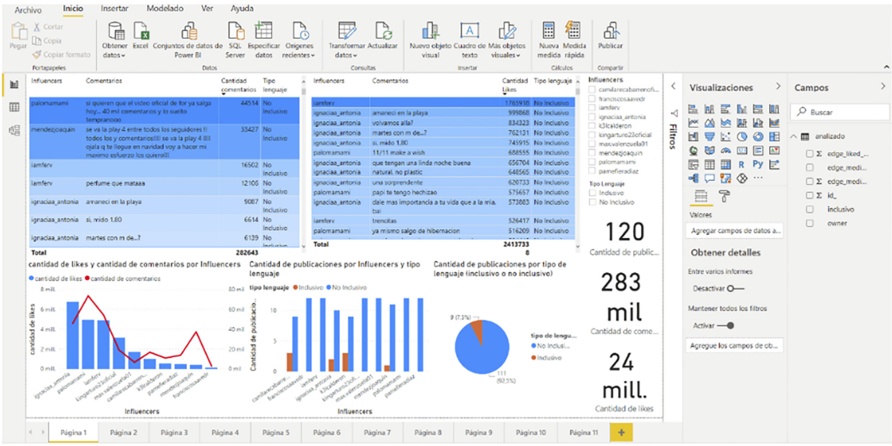


Figure 22.
Dashboard with management indicators analyzed in the study

Source(s): Authors' elaboration

Only four of the ten influencers occasionally use the ILandNSW language under study in their publications. Of these, the ones who use this ILandNSW language the most are Camila Recabarren (@camilarecabarrenoficial) and Raquel Calderón (@k3lcalderon), each of them with three posts, followed by Ignacia Antonia (@ignaciaa_antonia) with two and finally Joaquín Méndez with only one. In addition, of these four influencers, Camila Recabarren declares herself to be a feminist who supports the LGBT movement, but this is not reflected in her publications since more than 50% of her posts do not use IL and NSW.

The top places in the likes ranking are occupied by the posts of Fernanda Villalobos (@iamferv), Ignacia Antonia (@ignaciaa_antonia) and Paloma Castillo (@palomamami). Here, it may be observed that in the top ten rankings, there is no publication using the ILandNSW language under study. Additionally, a clear relationship between the number of likes and the types of language used can be observed: publications without the IL and NSW receive the most likes from followers. If the filter "type of language" is applied. In that case, the top publications for the type of language (belonging or not to the ILandNSW language under study) can be obtained, in which the difference between the number of likes among publications can be observed.

The number one post in the publication's ranking does not belong to the ILandNSW language studied and was written by Fernanda Villalobos (@iamferv), having 1,765,918 likes. In turn, the top publication that contains this language is made by Ignacia Antonia (@ignaciaa_antonia), which has 192,195 likes. Thus, it is clear that there is a significant difference in the number of likes between posts not including and including the ILandNSW language. Meanwhile, it is observed in Figure 23 that the publications with ILandNSW language made by Ignacia Antonia (@ignaciaa_antonia), who has the post belonging to the ILandNSW language with the most likes, have more likes than those without the ILandNSW language studied.

A relationship is also found between the number of comments and the Type of Language: publications with more comments are those not belonging to the ILandNSW language. This can be observed in a ranking similar to the number of likes. As in the previous ranking, if the number of publications belonging to the ILandNSW language under study is analyzed, there is one publication by Camila Recabarren (@camilarecabarrenoficial) with 4,356 comments, while if looking at the ranking of publications not belonging to the ILandNSW language, there is one post by Paloma Castillo (@palomamami) that has 44,514 comments. Therefore, a significant difference in the number of comments between the top post of the different types of language may be observed, as shown in Figure 24.

In sum, more publications must comply with the ILandNSW language, and these publications also have more reactions from Users. Therefore, it is inferred that the use of IL and NSW still needs to be more widespread in the community of the SN Instagram since its use is not constant but only occasional. In addition, the posts (publications) that use the ILandNSW language under study receive little reactions from Users.

Influencers	Comentarios	Cantidad Likes	Tipo lenguaje
ignaciaa_antonia	amaneci en la playa	999868	No Inclusivo
ignaciaa_antonia	volvamos alla?	834323	No Inclusivo
ignaciaa_antonia	martes con m de...?	762131	No Inclusivo
ignaciaa_antonia	si, mido 1,80	745915	No Inclusivo
ignaciaa_antonia	que tengan una linda noche buena	656704	No Inclusivo
ignaciaa_antonia	natural, no plastic	648565	No Inclusivo
ignaciaa_antonia	una sorprendente	620733	No Inclusivo
ignaciaa_antonia	dale mas importancia a tu vida que a la mia, bai	573883	No Inclusivo
ignaciaa_antonia	hoy me hice un picnicdisfrute tanto mi hamburguesa light de vacuno#hamburguesaslacrianza	434516	No Inclusivo
ignaciaa_antonia	pero que hermosos looks crearon estas bellezas lxs reto a que con su grupo de amigues recreen este video con el #haztelikechallenge el mejor video tendra un premio increible para todes los que participen! @antosegovia._ @itsdeecl @arantzazuiac @renatosaaa @nit.nat.not @syretape @_claaaudiii	192195	Inclusivo
ignaciaa_antonia	si quedas atrapada en una isla ¿que harias? #salvajes @primevideolat	189528	No Inclusivo
ignaciaa_antonia	tachemos los prejuicios y rompamos las etiquetas! hoy nuestra meta es acabar con todas esas etiquetas sociales, para que el mundo nos vea a tod@s por lo que realmente somos. te invitamos a ser protagonista de este cambio y para eso, queremos que taches todo aquello de lo que te quieres liberar, dentro de esta polera. asi le daras un nuevo significado, convirtiendola en un mensaje lleno de amor propio #yoelijosalvar @fsummercl	118587	Inclusivo
Total		6776948	

Source(s): Authors' elaboration

Figure 23. Likes ranking of Influencer @ignaciaa_antonia

Influencers	Comentarios	Cantidad comentarios	Tipo lenguaje
palomamami	si quieren que el video oficial de for ya salga hoy... 40 mil comentarios y lo suelto tempranooo	44514	No Inclusivo
mendezjoaquin	se va la play 4 entre todos los seguidores !! todos los y comentarios!!!! se va la play 4 !!!! ojala q te llegue en navidad voy a hacer mi maximo esfuerzo los quiero!!!	33427	No Inclusivo
iamferv		16502	No Inclusivo
iamferv	perfume que mataaa	12106	No Inclusivo
ignaciaa_antonia	amaneci en la playa	9087	No Inclusivo
ignaciaa_antonia	si, mido 1,80	6614	No Inclusivo
ignaciaa_antonia	martes con m de...?	6139	No Inclusivo
camilarecabarren oficial	otsea ubiquense ! la gorditaaaaa mas linda ella me pidio el celu pa decirles algo ese dia porque me vio con cara triste pero esta todo bien ... los amo	5526	No Inclusivo
iamferv	me transportas a otro mundo estando presente	4889	No Inclusivo
palomamami	11/11 make a wish	4856	No Inclusivo
ignaciaa_antonia	dale mas importancia a tu vida que a la mia, bai	4395	No Inclusivo
camilarecabarren oficial	concurso navideño gana una depiladora laser rebelskin + toda la gama de productos de @rebelsmile_! sigue los siguientes pasos y ya estaras participando : seguir a @rebelsmile_ y a mi. darle like a esta publicacion. sacar un screenshoot de esta publicacion y subirlo a tu historia etiquetando a @rebelsmile_ y a mi. etiquetar a 2 amig@s en	4356	Inclusivo
Total		282643	

Source(s): Authors' elaboration

Figure 24.
Comment ranking of Influencer @palomamami

5. Conclusions

Regarding the Instagram influencers, after the identification and selection stages, only ten influencers from the top Influencers in Chile were identified, which is the critical point in this study. In total, data from 120 publications were extracted: Instagram Influencer name, publication ID, publication description (what the Instagram Influencer wrote on the post), the number of comments and the number of likes per post (publications) were obtained. Then, a numerical attribute isILandNSWused was added, which was "1" if the publication belonged to the IL and NSW and "0" if not.

Concerning the BoW, this was created before the text analysis stage and included words belonging to the Influencer Target Group ILandNSW language. This BoW may grow if new words used in the ILandNSW language studied are discovered. As for language, the BoW is contrasted with the content written by the Influencers to analyze if its words are contained in it.

Regarding the research questions of this study:

What is the level of use of IL among influencers in Chile on Instagram?

With the GFL Clustering Methodology implemented and the use of tools for processing the obtained data (Stages 1, 2, 3, 4 and 5), only 9 out of 120 publications were found to contain the ILandNSW language of the Influencer Target Group. After visualizing the information (Stages 6 and 7), it may be observed that of 120 posts made by Influencers on the SN Instagram, only 7.5% incorporated the IL and NSW in the descriptions written by diverse Influencers.

What words allow for determining the minimum degree of use of GFL? What is the relationship between Influencers and the use of ILandNSW language in digital marketing?

According to the results of this study, this type of ILandNSW language needs to be integrated into the various actions posted commonly on SN Instagram.

It is also observed that the publications with the most likes are comments that do not contain IL and NSW. Therefore, publications with this type of language could attract Chile's Instagram influencers more.

In future work, the database (128 ILandNSW words) of the BoW of IL and NSW needs to be expanded to deepen the identification of the studied language since currently there are only guidelines for using this ILandNSW language rather than a list of words belonging to it. Creating an ILandNSW dictionary in other languages is also necessary for the future, as well as working with a larger pool of Influencers. Additionally, a web tool should be created to allow organizations to assess the use levels of ILandNSW language.

6. Managerial implications

The originality of this work lies in its attractive implications for nonprofit and for-profit organizations, government bodies and private enterprises in the measurement of the success of campaigns with an IM communicational strategy and to incorporate inclusive and non-sexist content for their consumers to contribute to society. Additionally, this study contributes guidelines for DC creators who aim to conduct strategic management with a market orientation, as well as for organizations who desire both to contribute to society through strategies for business communication and annual report and to define their online communication, product commercialization and branding strategies to generate a positive attitude toward the brand. Data could be retrieved from influencers as they are of free access regardless of the people involved.

7. Limitations and potential future research directions

Due to the limitations of the API of the SN Instagram, a limited number of text data were used, which allowed for retrieving the last 12 publications of each studied profile. In addition, this study only includes the Spanish language and is applied to a sample of influencers from Chile.

Another matter that could be addressed by further work is the implementation of an algorithm in Machine Learning to identify and classify ILandNSW language in such a way that complete sentences are retrieved, and their meaning is analyzed to know if what Influencers write genuinely belongs to the ILandNSW language under study.

Notes

1. Betel nut is the seed of the fruit of the areca palm. It is also known as areca nut. The common names, preparations and specific ingredients vary by cultural group and individuals who use them. https://adf.org.au/drug-facts/betel-nut/
2. Power BI: Microsoft Power Platform, https://powerbi.microsoft.com/
3. Social media scraping collects data from social media platforms such as TikTok, Instagram, Facebook, Twitter and the like. Usually, it is done automatically, using ready-made scraping software or custom-built scrapers. It is possible to scrape many data points like followers, likes and the number of views or shares, to name a few. https://proxyway.com/guides/what-is-social-media-scraping
4. Social media scraping provides a great way to collect valuable data for research or commercial purposes. Moreover, Instagram is the most lucrative platform today. However, it is also tricky to scrape due to technical and legal challenges. https://proxyway.com/guides/how-to-scrape-instagram
5. A JSON URL (an acronym of JavaScript Object Notation) is a file format that enables stock data, and this is through this URL, your Custom Counter will be able to display a number. https://help.smiirl.com/article/133-what-should-i-do-to-enable-my-custom-counter-to-display-a-number
6. Contains the description of what the Instagram Influencer wrote on the post (publication).

7. Contains the number of comments per post by the User of photos or videos posted by the Influencer on Instagram.
8. Contains the number of likes per post by the User of photos or videos posted by the Influencer on Instagram.
9. Return a writer object responsible for converting the user's data into delimited strings on the given file like object. *CSV file* can be any object with a write( ) method. If *CSV file* is a file object, it should be opened with newline = ''. https://docs.python.org/3/library/csv.html
10. This function in the CSV module returns a writer object that converts data into a delimited string and stores it in a file object. The function needs a file object created with an open( ) function and with write permission as a parameter. Every row written in the file issues a newline character by default. The newline = '' to prevent an additional line between rows. https://www.knowledgehut.com/tutorials/python-tutorial/python-csv
11. The CSV (Comma Separated Values) format is the most common import and export format for spreadsheets and databases. https://docs.python.org/3/library/csv.html
12. The Requests Session object allows to persist specific parameters across requests to the same site. To get the Session object in Python Requests, it is necessary to call the requests.session( ) method. The Session object can store such parameters as cookies and HTTP headers. Google: How do I use Session object in Phyton Requests?
13. A status code informs of the status of the request. For example, a 200 OK status means the request was successful, whereas a 404 NOT FOUND status means that the resource it was looking for was not found. Google: What is status code in Phyton?
14. Do requests not consider a 304 as "Ok"? A property called "Ok" in the Response object returns True if the status code is not a 4xx or a 5xx. Response.ok returns True if status_code is less than 400; otherwise, False. Python requests are generally used to fetch the content from a particular resource URL. Whenever it makes requests to a specified URL through Python, it returns a response object. Now, this response object would be used to access certain features such as content, headers, etc. This article revolves around how to check the response.ok, out of a response object. https://stackoverflow.com/questions/22494794/does-requests-codes-ok-include-a-304
15. The JSON format structure to browse "Instagram_data_not_clean" file is ["graphql"] ["user"] ["edge_owner_to_timeline_media"] ["edges"] ["node"]. See Figure 5.
16. This function writes items in a sequence (list, tuple or string) separating them by comma character. https://www.knowledgehut.com/tutorials/python-tutorial/python-csv
17. CSV in Python adds an extra carriage return on Windows. https://stackoverflow.com/questions/3191528/csv-in-python-adding-an-extra-carriage-return-on-windows

References

Ananda, A.S., Hernández-García, Á., Acquil-Natale, E. and Lamberti, L. (2019), "What makes fashion consumers 'click'? Generation of eWoM engagement in social media", *Asia Pacific Journal of Marketing and Logistics*, pp. 1-57, doi: 10.1108/APJML-03-2018-0115.

Bengfort, B., Bilbro, R. and Ojeda, T. (2018), *Applied Text Analysis with Python*, O'Reilly Media.

Bigsby, K.G., Ohlmann, J.W. and Zhao, K. (2019), "Keeping it 100: social media and self-presentation in college football recruiting", *Big Data*, Vol. 7 No. 1, pp. 3-20, doi: 10.1089/big.2018.0094.

Buente, W., Rathanayake, C., Neo, R., Dalisay, F. and Kurihara, H. (2020), "Tradition gone mobile: an exploration of", *Betelnut on Instagram. Substance Use and Misuse*, Vol. 55 No. 9, pp. 1483-1492, doi: 10.1080/10826084.2020.1744657.

Calder, B.J., Isaac, M.S. and Malthouse, E.C. (2016), "How to capture consumer experiences: a context-specific approach to measuring engagement: predicting consumer behavior across qualitatively different experiences", *Journal of Advertising Research*, Vol. 56 No. 1, pp. 39-52, doi: 10.2501/JAR-2015-028.

CEP-PIE (2017), "Guia de lenguaje inclusivo del CEP-PIE", available at: http://www.cep-pie.org/wp-content/uploads/2017/11/Guía-lenguaje-inclusivo-CEP-PIE.docx.pdf

Cury, R.M. (2019), "Oscillation of tweet sentiments in the election of João Doria Jr. for mayor", *Journal Big Data*, Vol. 6, p. 42, doi: 10.1186/s40537-019-0208-1.

De la Piedra, E.S. and Meana, R.J. (2017), "Redes sociales y fenómeno influencer. Reflexiones desde una perspectiva psicológica. Miscelánea Comillas", *Revista de Ciencias Humanas y Sociales*, Vol. 75 No. 147, pp. 443-469.

De Lemus, S. and Estevan-Reina, L. (2021), "Influence of sexist language on motivation and feelings of ostracism", *International Journal of Social Psychology*, Vol. 36 No. 1, pp. 61-97, doi: 10.1080/02134748.2020.1840230.

De Valck, K., Van Bruggen, G.H. and Wierenga, B. (2009), "Virtual communities: a marketing perspective", *Decision Support Systems*, Vol. 47 No. 3, pp. 185-203, doi: 10.1016/j.dss.2009.02.008.

Granovetter, M. (2011), "Economic action and social structure: the problem of embeddedness", in Granovetter, En M. and Swedberg, R. (Eds), *The Sociology of Economic Life*, Routledge.

Grigalis, T. and Čenys, A. (2013), "State-of-the-art web data extraction systems for online business intelligence", *Informacijos Mokslai*, Vol. 64, pp. 145-155, doi: 10.15388/Im.2013.0.1595.

Heilman, M.E. and Caleo, S. (2018), "Combatting gender discrimination: a lack of fit framework", *Group Processes and Intergroup Relations*, Vol. 21 No. 5, pp. 725-744, doi: 10.1177/1368430218761587.

Hollebeek, L. (2011), "Demystifying customer brand engagement: exploring the loyalty nexus", *Journal of Marketing Management*, Vol. 27 Nos 7-8, pp. 785-807, doi: 10.1080/0267257X.2010.500132.

Humphrey, W., Laverie, D. and Muñoz, C. (2021), "The use and value of badges: leveraging salesforce trailhead badges for marketing technology education", *Journal of Marketing Education*, Vol. 43 No. 1, pp. 25-42, doi: 10.1177/0273475320912319.

Koeser, S. and Sczesny, S. (2014), "Promoting Gender-Fair Language: the Impact of arguments on language use, attitudes, and cognitions", *Journal of Language and Social Psychology*, Vol. 33 No. 5, pp. 548-560, doi: 10.1177/0261927X14541280.

Koeser, S., Kuhn, E.A. and Sczesny, S. (2015), "Just reading? How gender-fair lenguage triggers readers' use of gender-fair forms", *Journal of Language and Social Psychology*, Vol. 34 No. 3, pp. 343-357, doi: 10.1177/0261927X14561119.

Kotler, K.S. (2021), *Marketing 5.0: Technology for Humanity*, John Wiley & Sons, Books.

Kumar, A. and Sachdeva, N. (2021), "Multimodal cyberbullying detection using capsule network with dynamic routing and deep convolutional neural network", *Multimedia Systems*, Vol. 28 No. 1, doi: 10.1007/s00530-020-00747-5.

Landers, R.N., Brusso, R.C., Cavanaugh, K.J. and Collmus, A.B. (2016), "A primer on theory-driven web scraping: automatic extraction of big data from the Internet for use in psychological research", *Psychol Methods*, Vol. 21 No. 4, pp. 475-492, doi: 10.1037/met0000081.

Li, J., Xian, G., Zhao, R., Huang, Y., Kou, Y., Luo, T. and Sun, T. (2021), "RDFAdaptor: efficient ETL plugins for RDF data process", *Journal of Data and Information Science*, Vol. 6 No. 3, pp. 123-145, doi: 10.2478/jdis-2021-0020.

Lin, C., Chen, Y., Chiang, J. and Zhang, Y. (2021), "Do 'little emperors' get more than 'little empresses'? BoyGirl gender discrimination as evidenced by consumption behavior of Chinese households", *Marketing Science*, Vol. 40 No. 6, pp. 1123-1146, doi: 10.1287/mksc.2021.1302.

Moezzi, M., Janda, K. and Rotmann, S. (2017), "Using stories, narratives, and storytelling in energy and climate change research", *Energy Research and Social Science*, Vol. 31, pp. 1-10, doi: 10.1016/j.erss.2017.06.034.

Nagarajan, S.M. and Gandhi, U.D. (2019), “Classifying streaming of Twitter data based on sentiment analysis using hybridization”, *Neural Computing and Applications*, Vol. 31, pp. 1425-1433, doi: 10.1007/s00521-018-3476-3.

Nesi, P., Pantaleo, G., Paoli, I. and Zaza, I. (2019), “Assessing the reTweet proneness of tweets: predictive models for retweeting”, *Multimedia Tools and Applications*, Vol. 77, pp. 26731-26396, doi: 10.1007/s11042-018-5865-.

Nicolas, C., Urrutia, A., Valenzuela, L. and Gil-Lafuente, J. (2018), “Systematic mapping on social media and its relation to business”, *European Research on Management and Business Economics*, Vol. 24 No. 2, pp. 104-113, doi: 10.1016/j.iedeen.2018.01.002.

Patev, A.J., Dunn, C.E., Hood, K.R. and Barber, J.M. (2019), “College students’ perceptions of Gender-Inclusive Language use predict attitudes toward transgender and gender nonconforming individuals”, *Journal of Language and Social Psychology*, Vol. 38 No. 3, pp. 329-352, doi: 10.1177/0261927X1881593.

Piller, I. and Takahashi, K. (2011), “Linguistic diversity and social inclusion”, *International Journal of Bilingual Education and Bilingualism*, Vol. 14, pp. 371-381, doi: 10.1080/13670050.2011.573062.

Sahayak, V., Shete, V. and Pathan, A. (2015), “Sentiment analysis on twitter data”, *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, Vol. 2 No. 1, pp. 178-183.

Sczesny, S., Formanowicz, M. and Moser, F. (2016), “Can Gender-Fair Language reduce gender stereotyping and discrimination?”, *Frontiers in Psychology*, Vol. 7 No. 25, pp. 1-7, doi: 10.3389/fpsyg.2016.00025.

Senden, M., Back, E.A. and Lindqvist, A. (2015), “Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior”, *Frontiers in Psychology*, Vol. 6, 893, doi: 10.3389/fpsyg.2015.00893.

Scurlock, R., Dolsak, N. and Prakash, A. (2020), “Recovering from scandals: twitter coverage of oxfam and save the children scandals”, *International Journal of Voluntary and Nonprofit Organizations*, Vol. 31 No. 3, doi: 10.1007/s11266-019-00148-x.

Stahlberg, D., Braun, F., Irmen, L. and Sczesny, S. (2007), “Representation of the sexes in language”, *Social Communication*, pp. 163-187.

Tankosić, A., Dryden, S. and Dovchin, S. (2021), “The link between linguistic subordination and linguistic inferiority complexes: english as a second language migrant in Australia”, *International Journal of Bilingualism*, Vol. 25 No. 6, pp. 1782-1798, doi: 10.1177/13670069211035561.

UNFPA (2020), “UNFPA”, available at: https://lac.unfpa.org/es/temas/igualdad-de-género-y-derechos-humanos

Urrutia, A., Rojo, F., Nicolas, C. and Ahumada, R. (2021), “Applying data mining on customer relationship management system to discover forgotten effects”, *Journal of Intelligent and Fuzzy Systems*, Vol. 40 No. 2, pp. 1783-1794, doi: 10.3233/JIFS-189185.

Voorveld, H.A.M., van Noort, G., Muntinga, D.G. and Bronner, F. (2018), “Engagement with social media and social media advertising: the differentiating role of platform type”, *Journal of Advertising*, Vol. 47 No. 1, pp. 38-54. doi: 10.1080/00913367.2017.1405754.

Further reading

Iyer, R., Zheng, R., Li, Y. and Sycara, K. (2019), “Event outcome prediction using sentiment analysis and crowd wisdom in microblog feeds”, *arXiv preprint arXiv*, Vol. 1 No. 9, doi: 10.48550/arXiv.1912.05066.

Maass, A., Suitner, C. and Merkel, E. (2013), “Does political correctness make (social) sense?”, in Forgas, J.P., Vincze, O. and László, J. (Eds), *Social Cognition and Communication*, Psychology Press, pp. 331-346.

Media and Social Media Advertising (2018), "The differentiating role of platform type", *Journal of Advertising*. doi: 10.1080/00913367.2017.1405754.

Power, B.I., "Microsoft power platform", available at: https://powerbi.microsoft.com/

Corresponding author
Angelica Urrutia can be contacted at: aurrutia@uft.cl