

Master Degree in Computational Social Sciences  
2022/2023

*Master Thesis*

# Exploratory analysis of Mercadona press release content using data harvesting and data mining

---

Raquel Sánchez-Hermosilla García

Supervisor: Gonzalo Génova Fuster

Madrid, 2023

## AVOID PLAGIARISM

The University uses the Turnitin Feedback Studio program within the Aula Global for the delivery of student work. This program compares the originality of the work delivered by each student with millions of electronic resources and detects those parts of the text that are copied and pasted. Plagiarizing in a TFM is considered a **Serious Misconduct**, and may result in permanent expulsion from the University.



[Include this code in case you want your Master Thesis published in Open Access  
University Repository]

This work is licensed under Creative Commons **Attribution – Non Commercial – Non  
Derivatives**

## SUMMARY

This research aims to do an exploratory analysis of Mercadona communication, focusing on their press releases from 2013 to 2022, in order to gain insights into Mercadona communication patterns.

To do so, I am going to use web scrapping and regular expressions to get ten years of Mercadona press releases from their webpage, to later analyze press releases using the following text mining techniques word count, term frequency, distinctiveness, bigrams, trigrams and Latent Dirichlet Allocation (LDA) model for topic modeling.

Results show that Mercadona press releases are focused on placing their organizational agenda on the media, they also show how relevant Corporate Social Responsibility is for Mercadona, overall plastic reduction. Besides, press releases content demonstrate that Mercadona has an emphasis on listening and understanding their customers or “The Boss” needs, because the client is what goes first. And finally, a schema of useful words to write press releases for public relations professionals has been gathered.

Associated public repository: <https://github.com/Xermosilla/Computational-Social-Sciences-TFM.git>

## TOC

INTRODUCTION.....	1
ORGANIZATIONAL COMMUNICATION CONTEXT .....	1
RESEARCH OBJECTIVE.....	5
METHODOLOGY .....	5
Data harvesting.....	5
Text mining .....	7
RESULTS .....	8
CONCLUSION .....	24
REFERENCES.....	26

## **INTRODUCTION**

In the dynamic landscape of business, effective communication strategies play a vital role in shaping a company's brand image and market positioning. This research aims to deepen into the communication practices of Mercadona, a highly relevant brand and marketing-oriented company within the Spanish supermarket retailing sector. Moreover, Mercadona's distinctive approach to communication, characterized by its organizational or corporative communication strategy, sets it apart from its competitors who typically employ commercial communication strategies. Hence, this research seeks to analyze ten years of Mercadona's press releases to gain insights into their distinctive communication patterns, examining their press releases content.

## **ORGANIZATIONAL COMMUNICATION CONTEXT**

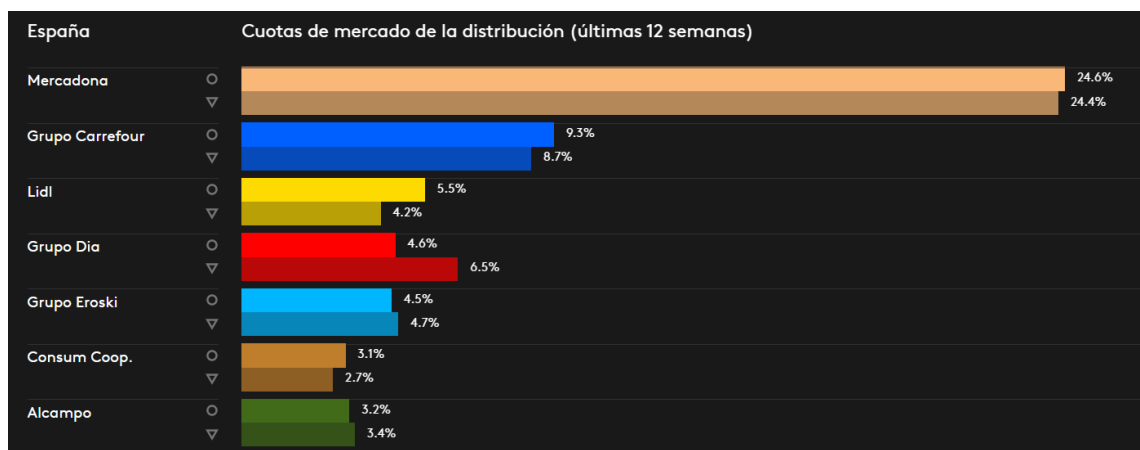
Mercadona is one of the most important and most successful Spanish brands, and the undisputed Spanish leader among retailing companies, within the supermarket segment (Marín and Lozano, 2017; Hernando et al., 2019; Blanco-Callejo and Pablo Heredero, 2019). In fact, it is “one of the great national and international references of distribution” (Hernando et al., 2019: 128), due to the fact that Mercadona is a marketing-oriented company, which means that the satisfaction of their customers is what comes first, followed by the rest of the stakeholders and in last position, the Capital.

This company's orientation towards marketing follows an unprecedented business model in Spain, "Total Quality Management", which bets on having the best products at the lowest price, placing the customer, considered the essential element of the company, at the center of their strategy and all their activities. In fact, they refer to the customer as “El Jefe” or “The Boss”, and all workers (managers, suppliers, cashiers and replenishes) have the maximum responsibility to satisfy customer needs (Hernando et al., 2019; Mercadona, 2013-2022). Moreover, Mercadona applies the principles of stakeholder theory, taking care of all their stakeholders, like consumers, employees, or suppliers, but also considering as stakeholders society itself and the environment, through their Corporate Social Responsibility activities, like reducing energy consumption or donating food (Hernando et al., 2019). "Therefore, Mercadona is a company that applies the principles of marketing and gets the benefits out of it" (Hernando et al., 2019:134).

Many facts that prove how Mercadona profits from this marketing orientation, which in fact, differentiates this organization from the rest of supermarket chains present in the Spanish retailing sector.

First, Mercadona tends to be among the ten most valuable Spanish brands in the Best Spanish Brands ranking published by Interbrand, a global brand consultancy firm known for its annual publication of Best Global Brands. Concretely, in the Best Spanish Brands ranking from 2021, last published, Mercadona is the 7<sup>th</sup> most valuable Spanish brand, and there is no other supermarket brand in this ranking, safe for Día which is in 28<sup>th</sup> position out of the thirty brands that constitute the ranking (Interbrand, 2022).

Furthermore, Kantar Market Share is a division of Kantar brand that helps companies understand their position in the market relative to their competitors, by measuring and tracking their market share. Image 1 is a screenshot of Kantar interactive graph, it shows the market share of Mercadona and its competitors, taking into account January 2019 (below bar) and January 2022 (above bar) market share. Thus, we can see that Mercadona leads and has been leading the market share of the supermarket retailing sector in Spain, with a 10% to 20% higher market share than the rest of its competitors (Kantar, n.d.).

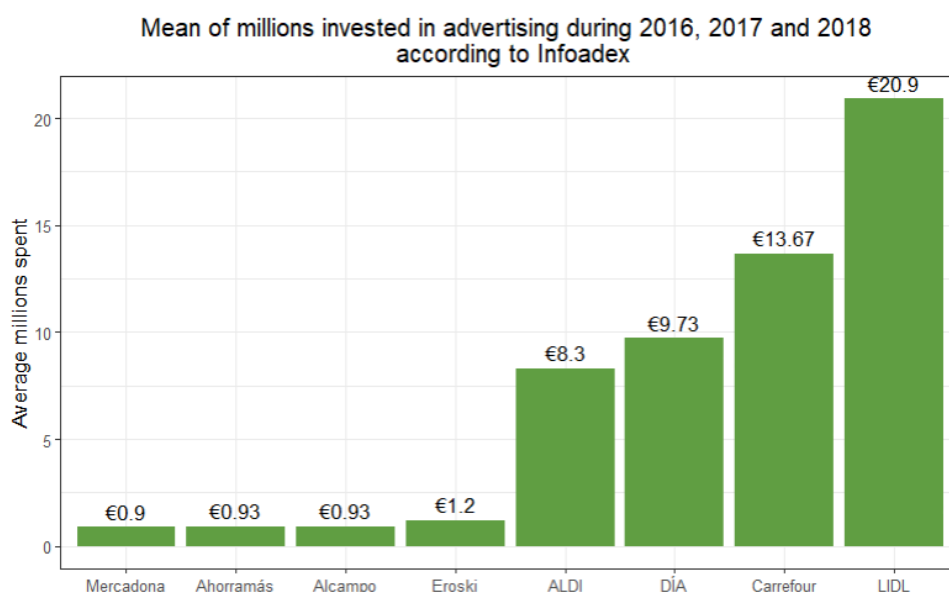


**Image 1**

*Market share of Mercadona and its competitors, comparing January 2019 and January 2022.*

Source: Kantar, n.d.

On top of that, according to Infoadex, a Spanish company specialized in the analysis of advertising investment in the Spanish market that publishes an advertisement investment ranking each year, Mercadona is the brand that spends less in advertising, when compared to its sector competitors, as graph 1 shows.



**Graph 1**

*Mean of millions invested in advertising during 2016, 2017 and 2018 according to Infoadex.*

Source: Indoadex, 2018; Infoadex 2019

Summing up, Mercadona is one of the most valuable Spanish brands and has the biggest market share in its sector, while spending so little on advertising, when compared to its competitors. This tiny investment and huge impact have its roots in the fact that Mercadona has a marketing orientation, and its communication strategy is centered on organizational or corporate communication, and not on commercial communication (Marín and Lozano, 2017; Marin et al., 2018).

In fact, Mercadona uses marketing communication in a distinctive way when compared to its competitors (Carrefour, Día, Eroski and Lidl, among others), because it has an organizational or corporate communication strategy, which focuses on its image, its reputation, their way of acting and their dialogue with the consumer, to understand and satisfy their needs, while the rest of competitors have a commercial communication strategy, based on informing consumers about promotions or discounts on their products, a communication strategy never used by Mercadona (Marín and Lozano, 2017; Marin et al., 2018).

Moreover, Mercadona media communication strategy also differs from the rest of their competitors, in that, Mercadona does not place any kind of adverts on massive media like radio, television or newspapers, they do not distribute leaflets in mailboxes, and the only outdoor advertising they use are billboards near their selling points, to inform about its location (Marín and Lozano, 2017; Marin et al., 2018).

Additionally, the difference between commercial and corporate communication relies on the fact that commercial communication focuses directly on selling products or services, prioritizing solely economic or market interests, while corporate communication focuses on improving the organization's image and reputation, and its relationships with its stakeholders (Wilcox, 2001:525 in Palencia-Lefler and Murillo, 2009; Weil 2003 in Apolo et al., 2017).

Hence, organizational or corporate communication "can be explained as a strategic public relations planning process whose purpose is to establish a climate of trust through which the organization pursues the confluence of its corporate interests with the interests of its stakeholders<sup>1</sup>" (Pulido, 20015: 97). On top of that, organizational or corporate communication is a strategy that uses research to build internal and external communication strategies that achieve organization objectives, while building long-lasting links with the organization stakeholders, being the center of this communication the organization itself, rather than their products or services (Apolo et al., 2017).

Public relations, as any other communicational strategy, in order to be effective needs to be a conscious, organized and planned process, that anticipates opportunities and problems (Cees Van Riel, 2001: 1-2; Grunig, 2000: 35-36; Xifra 2003:36; Cutlip y Center, 2001:444 in Pulido, 20015: 97). In addition, "the success of public relations techniques lies in their ability to manage the relationship between an organization and the media in order to place an organization's corporate messages, media agenda and in the public opinion<sup>2</sup>" (Pulido, 20015: 88). In fact, placing an organization agenda in the media, is one of the main tasks of public relations, and among the most common tools to achieve this objective, the most widely used is the press release (Palencia-Lefler and Murillo, 2009).

"The press release is a brief piece of writing that uses journalistic language and imitates the traditional news format to convey information to the media. It is written with the aim of being published by the media with little modification, and its purpose is to convey the

---

<sup>1</sup> Original reference: "la comunicación organizacional puede explicarse como un proceso de planificación estratégica de las relaciones públicas cuya finalidad es establecer un clima de confianza a través del cual la organización persigue la confluencia de sus intereses corporativos con los intereses de sus stakeholders"

<sup>2</sup> Original reference: "El éxito de estas técnicas de relaciones públicas radica en la capacidad que tienen de gestionar las relaciones entre una organización y los medios de comunicación a fin de situar en la agenda mediática y la opinión pública los mensajes corporativos de la organización"

organization's point of view to the media<sup>3</sup>" (Pulido, 20015: 93). Moreover, its purpose is also to get an organization's data and facts, which benefit the organization, to reach their stakeholder's through media, based on the idea that this information is of general interest and based the fact that positive communication is more credible when it comes from a transmitter that is not the organization itself (Palencia-Lefler and Murillo, 2009).

Nevertheless, this information targets not only media, but every stakeholder directly (Liu and Zhang, 2021), and "their primary readership, however, is gradually shifting to the public" (Catenaccio, 2008 in Liu and Zhang, 2021), because currently, press releases are often posted on a company's official website, which "expands access to press releases beyond the circles of professional journalists" (Strobbe & Jacobs, 2005: 289 in Liu and Zhang, 2021). This is due to the fact that nowadays organization webpages are not only a place to buy products, but also a place to interact with the public and communicate, therefore organizations like Mercadona need strategic communication that takes into account the online environment and its possibilities (Marin et al., 2018), to achieve their corporative and public relations goals.

## **RESEARCH OBJECTIVE**

My objective is to do an exploratory analysis of Mercadona communication through their press releases from 2013 to 2022, using computing techniques such as data harvesting and text mining. The goal is to gain valuable insights through content analysis and topic identification, in order to detect commonly used words, expressions and topics, with the aim of understanding patterns and trends in Mercadona press releases over ten years.

## **METHODOLOGY**

### **Data harvesting**

Mercadona press releases compendium can be found on Mercadona webpage (Mercadona, n.d.) and can be distinguished from any other online publication on their webpage, because all press releases are tagged as "Notas de prensa", independently from having other tags. And to gather and process needed press release data, I am going to use web scraping and data harvesting techniques, along with regular expressions.

---

<sup>3</sup> Original reference: "La nota de prensa es un escrito breve que utiliza el lenguaje periodístico e imita el formato tradicional de la noticia para transmitir información a los medios de comunicación. Se redacta con el objetivo de que sea publicada por el medio con la menor manipulación posible y su finalidad es trasladar a los medios de comunicación el punto de vista de la organización"

For this research I am going to consider ten years of press releases, from 2013 to 2022, and in total there are 109 URLs of press releases. Hence, to spot possible scrapping issues, first I built a script for each of the ten years, this way I made sure every press release per year was properly scraped, and I was able to create a final script with the 109 press releases on a list.

The URLs list was created manually because I encountered some inconsistencies in the URLs I scraped. All URLs start with “https://info.mercadona.es/es/actualidad/” and finish with “/news?idCategoriaSeleccionada=1470731303671”, however when it comes to “CategoriaSeleccionada” beside it seemed that it was always the title of the press release separating each word by “-“, many variations of this schema were found. For example, when there is a “%” symbol on the title it is removed, and in some URLs random numbers are added at the end of the press release title. Hence, as many URLs do not follow a consistent format, it was not possible to generate a list of URLs automatically.

Once I read all URLs manually into a R list, I found two html paths where all the press release information wanted, title, subtitle, and body text, was. Consequently, I created a function and a loop to scrap that information from all the press release webs and I stored that information as a data frame with 109 rows, containing press releases text, and one column, a data format that will be useful for future data transformations, needed to apply data mining techniques.

Scraped text data stored in the data frame had to be processed and cleaned using regular expressions, which allowed me to get a cleaner dataset for further analysis. Concretely, for cleaning and processing data I erased the word “Mercadona” because it has many occurrences and I also removed scraped html code symbols like “\n” or “\t”. Moreover, I transformed symbols like “€” or “%” to words because those two symbols give information about the text content and context, however, the rest of the punctuation symbols were removed. In addition, I removed all digits despite the ones that are four numbers that start with 19 and 20, because years can be informative while quantities, in general, are not. Furthermore, I did some transformations like concentrating different words or different word spellings into one, like “coinnovación” and “eljefe”, due to the fact that both words are related to Mercadona strategy (Mercadona, 2013-2022), as the following reference from one press release show “the Co-Innovation Center will allow the company to collaborate and be inspired by its "Bosses", as it calls its customers



internally<sup>4</sup>” (Mercadona, 2017). And finally, to conclude the data cleaning process, I checked each press release content one by one, to make sure they were clean, and I erased some links present in a few press releases.

### **Text mining**

Firstly, I added a second column as an index to the data frame, that indicates to which year the press release belongs, in order to be able to analyse all the press releases together but also differentiating per year.

Secondly, to apply text mining techniques I tokenized press releases, breaking them into single words, and I also filtered stopwords, to remove extremely common grammar words that are not useful to analyze press releases content. In addition, all press releases words were lemmatized, a process which involves reducing each word to its base form or lemma, for instance, conjugated verbs are converted to the infinitive form, and plural nouns are converted to the singular form. This is due to the fact that if text mining was applied without lemmatization, each word would be treated independently, for example, each conjugation of the same verb, which would lead to a greater dispersion of terms and more difficulties to identify content patterns. Nevertheless, lemmatization cannot be done in R due to the lack of libraries to lemmatize in Spanish or any other language, therefore I did this step of the analysis using Python.

First text mining technique I implemented was word count, which refers to the total number of words in the press releases and how many times those words were repeated. Concretely, I will analyze the ten most used words to gain insight into the prominence of specific words within the press releases, besides, I also used word counts to visualize the use of specific words of interest across years, like “eljefe”, “juan roig”, “plástico” or “covid”, among others.

Second text mining technique was Term Frequency, used to get the most repeated words in the ten years of press releases data, but taking into account each year number of words difference. As every year has a different word length, because the number of press release and their length varies across years, results from word count are not directly comparable, because word count shows how many times a term appears in the press releases, but Term

---

<sup>4</sup> Original reference: el Centro de Coinnovación permitirá a la compañía colaborar e inspirarse con sus “Jefes”, como denomina internamente a sus clientes

Frequency complements word counts by considering the relative importance of a term within the collection of press releases.

And once I got Term Frequency, I computed TF-IDF, a numerical statistic that measures the importance of a word in a press release within a collection of press releases and it combines two metrics, Term Frequency (TF) and Inverse Document Frequency (IDF). This metric allows me to analyze distinctiveness because the words with higher TF-IDF scores are considered distinctive. This is due to the fact that those words have a high frequency within a specific year of press releases and are rare in the overall collection of press releases, hence, they are significant and important words to analyze because they represent the distinctive themes of press releases, differentiating by year.

Previously mentioned text mining techniques work with single words and concepts, however, it is also interesting to analyze phrases and expressions, so I am going to use bigrams and trigrams. Bigrams are two words that tend to occur together, and trigrams are the same but take into account three words that tend to occur together, and these techniques are helpful because visualizing the most repeated ones can provide further insight into the commonly used phrases, expressions or topics discussed in the press releases.

And in fourth place, I used Latent Dirichlet Allocation (LDA) statistical model for topic modeling, this generative model helps to uncover the underlying topics within a collection of observations, as it assumes the data used as input is a mixture of small topics and when each word appears in a document it is due to one of the topics being present within the document. Therefore, I am going to use it to identify underlying topics or themes within the collection of press releases, which would help to understand the main subjects in which Mercadona has focused over the years.

## **RESULTS**

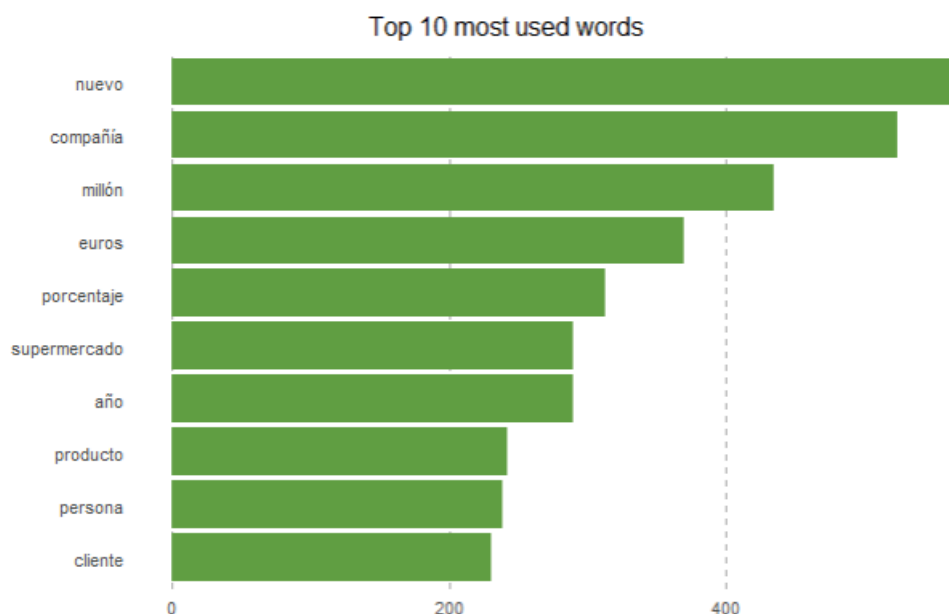
First text mining technique used was word count, which allowed to spot the ten most used words in ten years of press releases, which were:



**Graph 2**

*Top ten most used words in Mercadona press releases from 2013 to 2022.*

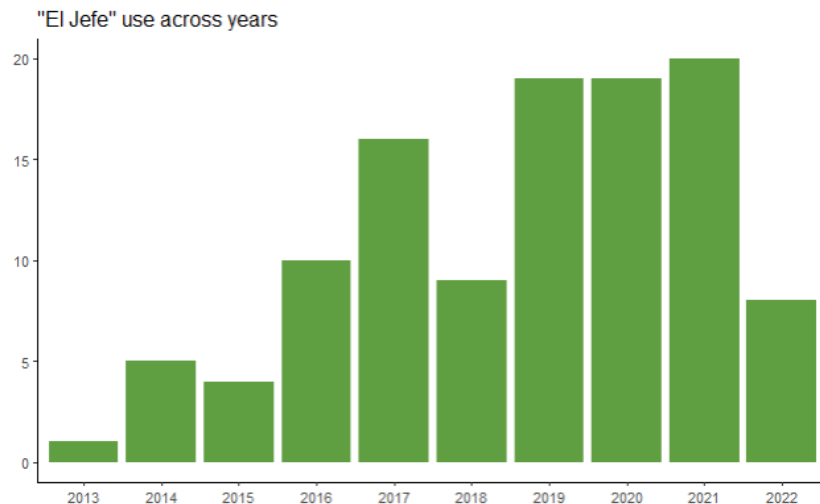
These prominent words provide valuable insights into the recurring themes of Mercadona press releases. As graph 3 shows, “nuevo” is the most used word, and it indicates new product launches or company initiatives, which seems to be the most repeated concept in their press releases communication. It is followed by “compañía”, which presence indicates a focus on the company itself, its developments, operations, activities... “Millón”, “euros” and “porcentaje” show how recurrent it is for press releases to mention economic, financial, and statistical data, and the word “año” indicates that press releases are focused on the company's annual activities, strategies, or developments. On top of that, “supermercado” and “producto” shows the relevance of this type of establishment and item for the company because they are the base of Mercadona’s business. Moreover, "customer" and "person" indicate a focus on customer care and consumer experience.



**Graph 3**

*Top ten most used words in Mercadona press releases from 2013 to 2022.*

Subsequently to perform word count and visualize the ten most used words, the yearly occurrence and the number of occurrences of certain words that are relevant in Mercadona communication, were examined. First of all, “eljefe” because it is the way in which Mercadona communication refers to the client in many official documents (Mercadona, 2013-2022), this word is present through all years and was used 111 times, which altogether with “cliente” being among the ten most used words, highlights Mercadona focus on the customer, a distinctive feature of marketing-oriented companies.

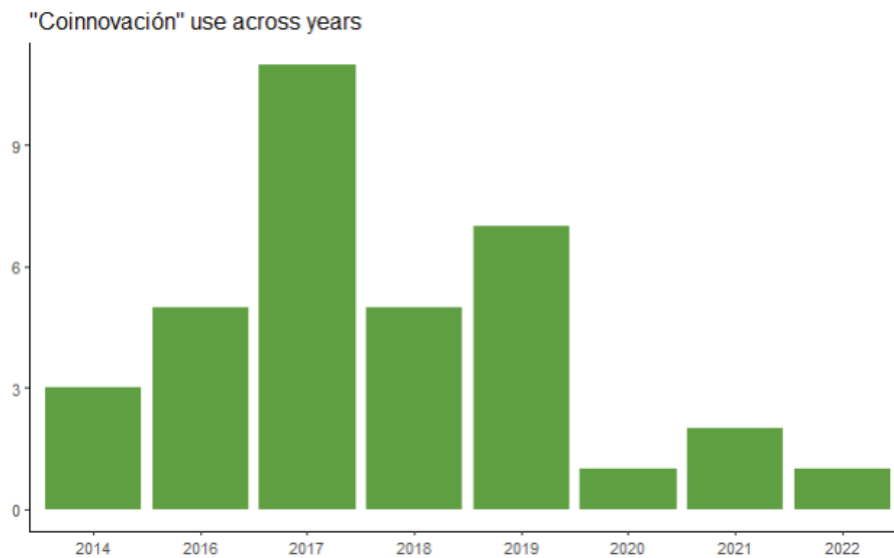


**Graph 4**

*“El Jefe” use across 2013-2022 in Mercadona press releases.*

Co-innovation is an “original and unique innovation model in the retail sector” (Blanco-Callejo and Pablos Heredero, 2019) and it refers to “centers where customers cook, clean, or use Mercadona products. Mercadona’s monitors listen to customers’ recommendations, complaints and suggestions (Hernando et al., no date:134). This word is mentioned 35 times, is not present across all years, and it was used the most in 2017 because that is when the first co-innovation center opened (Hernando et al., 2019). This highlights again Mercadona focus on customers, because despite all retailing sector companies face a dynamic environment, because customer needs and preferences change at an increasingly rapid pace, the rest of the companies of the Spanish supermarket retailing sector led this innovation to suppliers. As a consequence of co-innovation, Mercadona offers a unique

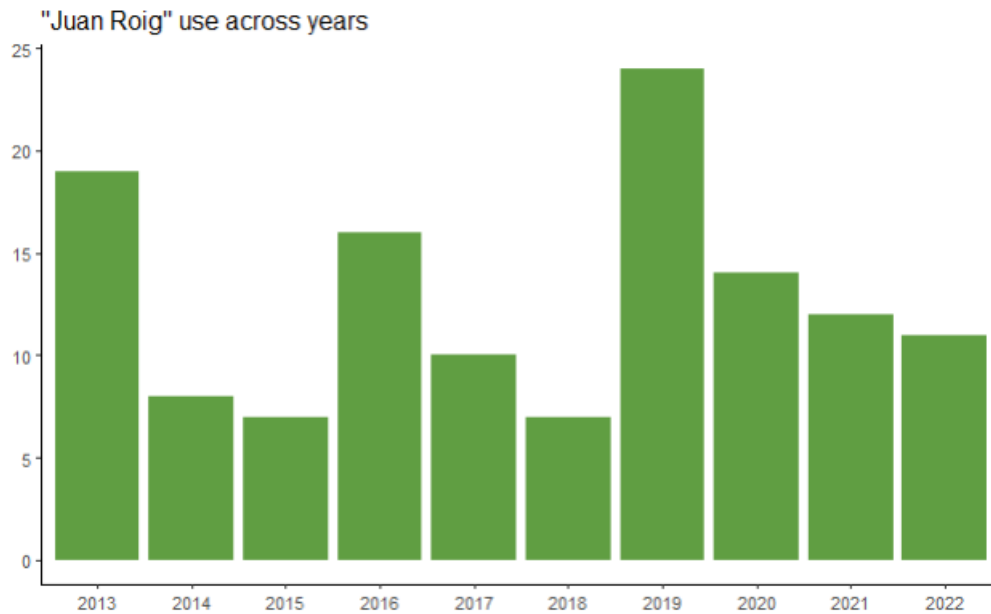
and exclusive range of products that is constantly and permanently updated, resulting in co-innovation as a key differentiating tool for competitiveness, in which the boss acquires a leading role in the development and improvement of solutions (Blanco-Callejo and Pablos Heredero, 2019).



**Graph 5**

*“Coinnovación” use across 2013-2022 in Mercadona press releases.*

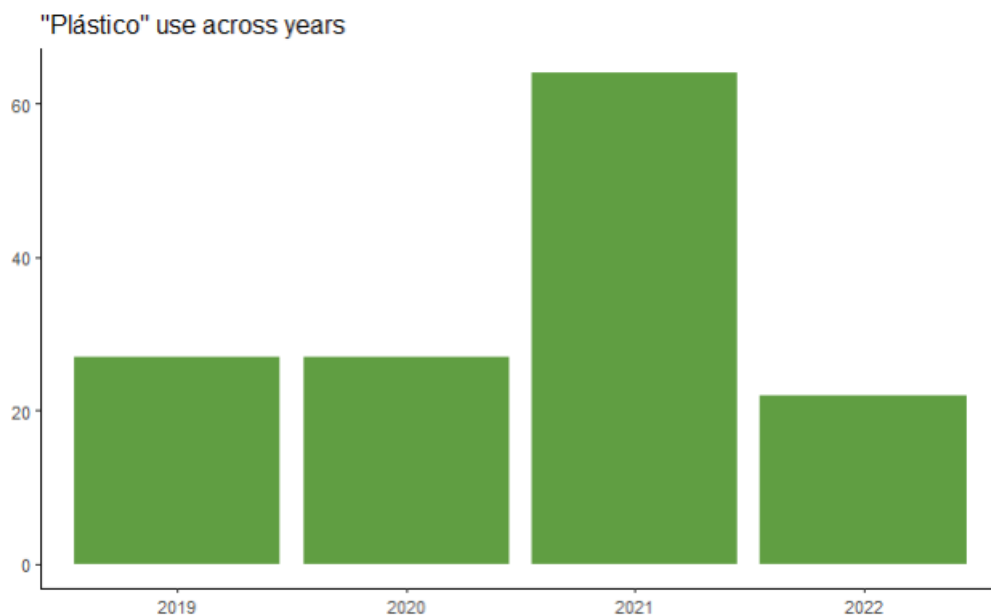
Juan Roig is the chairman and CEO of Mercadona, and his name is present in all years a total of 135 times, which highlights his prominent role and influence in the company. An importance that can also be seen in Mercadona Annual Reports, due to the fact that any Annual Report starts with a section called “Message from the Chairman”, his name appears constantly, to cherish his personal goals and improvements. next to Mercadona name, and there is plenty of pictures of him and other Roig family members (Mercadona, 2013-2022).



**Graph 6**

*“Juan Roig” use across 2013-2022 in Mercadona press releases.*

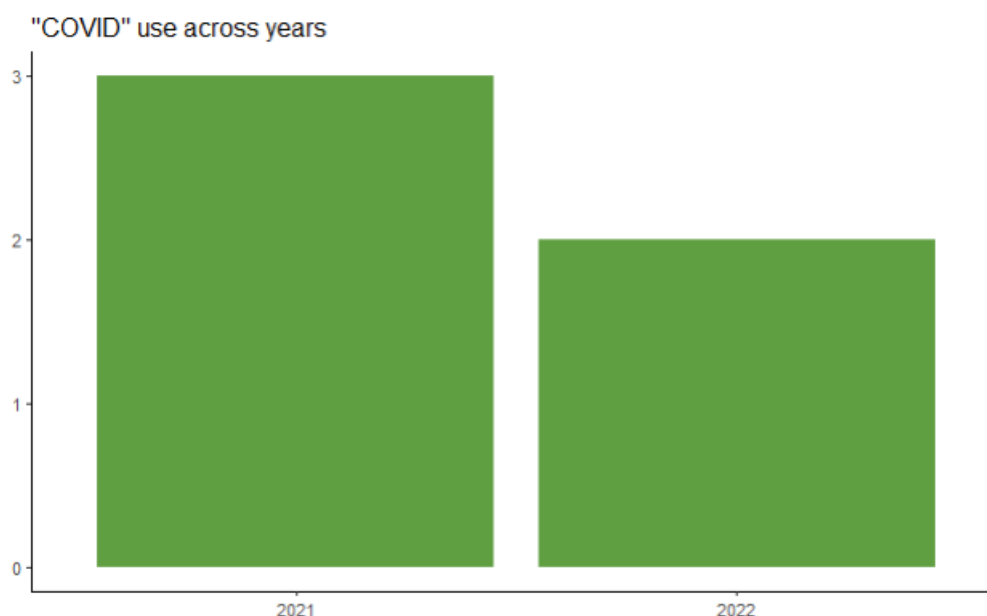
“Plástico” is used 140 times and interestingly it is only mentioned since 2019, when plastic overuse started to be an environmental concern for the population, and therefore for their stakeholders. Consequently, they started carrying out actions and measures to reduce plastic use, this shows once again Mercadona focus on listening to the client and attending to their needs and preferences, as well as their focus on Corporate Social Responsibility, because society and the environment are considered Mercadona stakeholders too. Moreover, it is interesting to see how collective concern can change a company’s communication topics, because before 2019 “plástico” was a word never used.



**Graph 7**

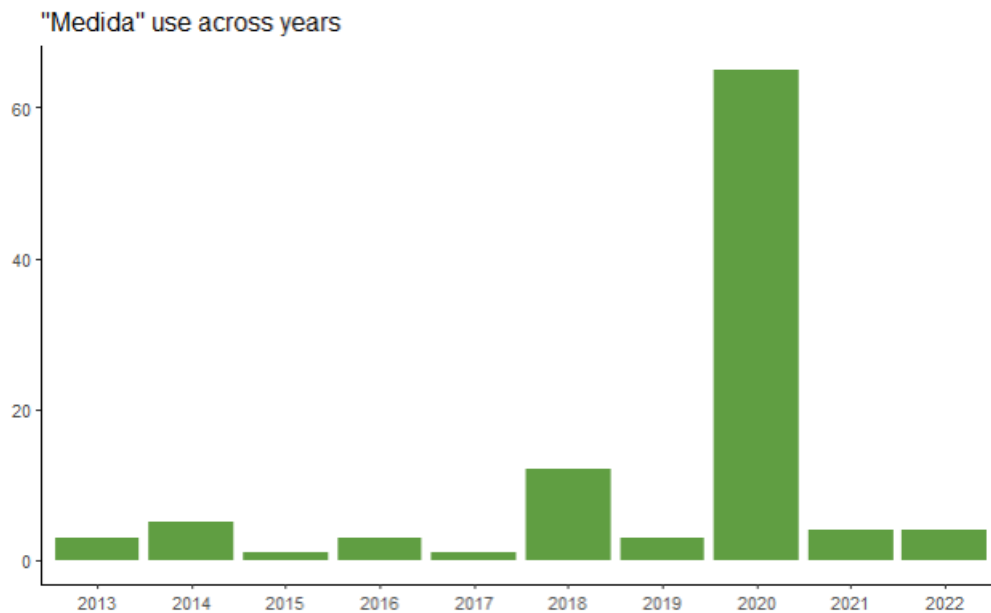
*“Plástico” use across 2013-2022 in Mercadona press releases.*

“COVID” is a word just repeated 5 times, during 2021 and 2022, and it is curious that it was used when the pandemic slowed down in 2021 and 2022, but not during 2019 and 2020 when most cases were happening. Nevertheless, “medida” was used across all years but there is a huge peak in the use of this word in 2020, because the published press releases informing about their actions to stop the coronavirus spread at their supermarkets. In fact, when I scraped all press releases, I realized that in 2019 there is no press release about COVID-19, but in 2020 there is fifteen press releases about that topic, but as graph 8 shows, those fifteen press releases do not mention “COVID” even once, but as graph 9 shows, those fifteen press releases focus on the measures or “medidas” that were needed due to the pandemic health crisis.



**Graph 8**

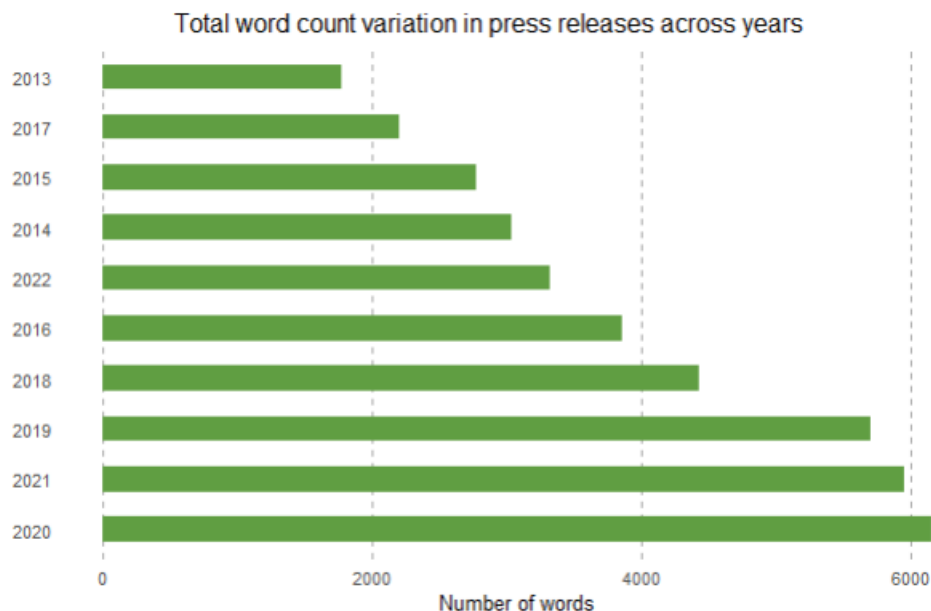
*“COVID” use across 2013-2022 in Mercadona press releases.*



**Graph 9**

*“Medida” use across 2013-2022 in Mercadona press releases.*

After using word count, I also implemented Term Frequency, because as graph 10 shows, not all years have the same number of total words, and word counts are not directly comparable. Consequently, I am using Term Frequency, because this metric displays the most common words, given each year’s word length.



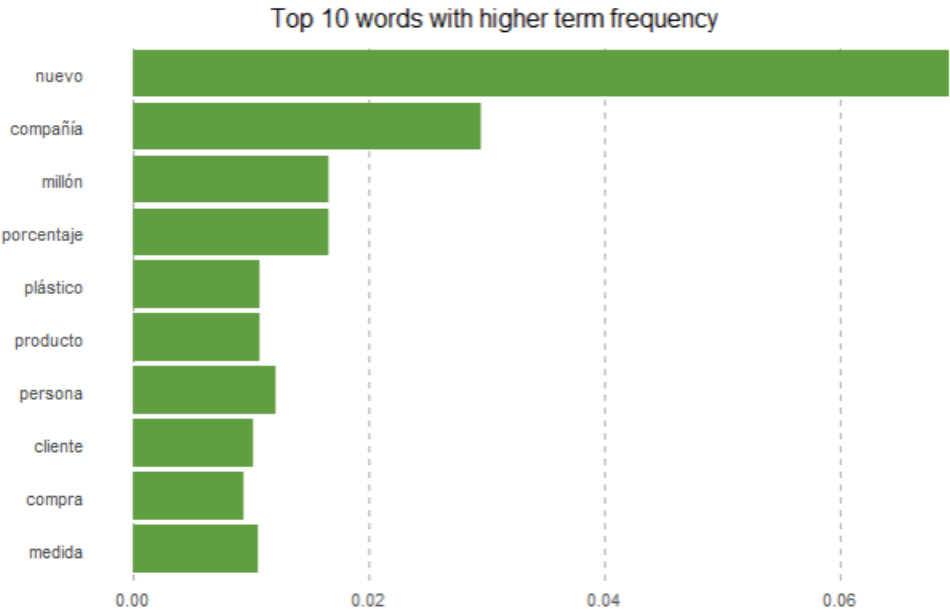
**Graph 10**

*Total word count variation in Mercadona press releases from 2013 to 2022.*

Using Term Frequency, we can see in graph 11, that out of the top ten used words, computed using word count, seven words match the top ten Term frequency results “nuevo”, “compañía”, “millón”, “porcentaje”, “producto”, “persona”, and “cliente”.

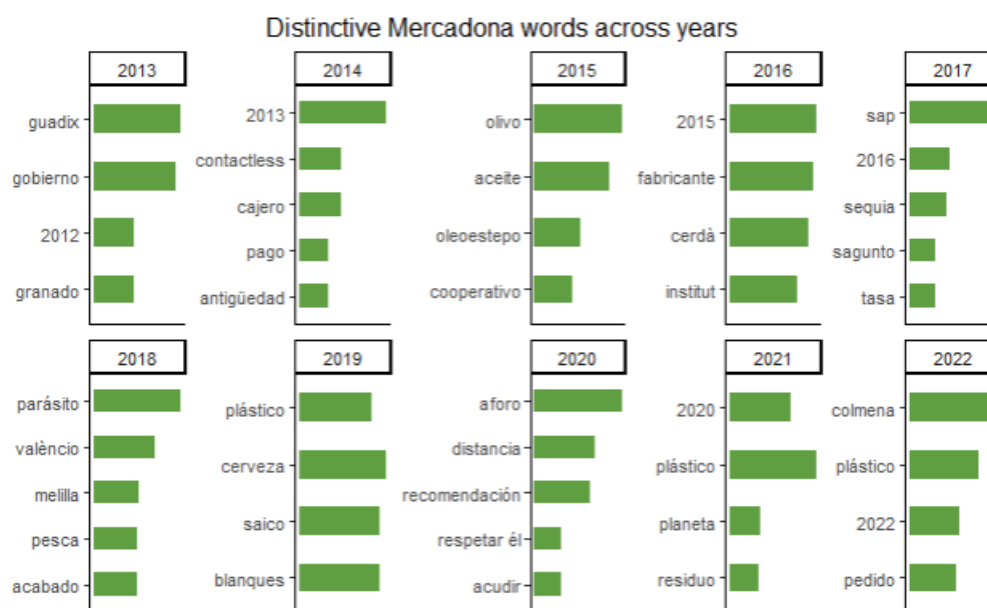


However, it is interesting that when considering each year’s number of words, the most frequent words keep being “nuevo” and “compañía”, but with a huge difference when compared to any other word, “nuevo” is the most frequent one. Moreover, new words such as “plástico” or “medida” appear, because when using word count and taking into account the absolute total number of words, they are not repeated that many times, but as seen in graph 7 and 9 respectively, those two are words repeated many times, but in or since specific years, not across all press releases.



**Graph 11**  
*Top ten words with higher term frequency in Mercadona press releases from 2013 to 2022.*

Following Term Frequency results, I computed TF-IDF metric, to visualize distinctive words across years, in order to provide valuable insight on the most significant and representative terms associated with each year press releases.



**Graph 12**

*Distinctive words across years in Mercadona press releases from 2013 to 2022.*

Many distinctive words are places like Guadix, Cerdà, Sagunto, Melilla or Blanques, probably because during those years implementations or innovations were carried out in those places, but those locations are never mentioned again. The same happens with years, because 2013, 2014, 2016, 2017 and 2021 are among the four most distinctive words, probably because is common to refer to data from the previous year as a point of comparison or to provide context about economic, financial, and statistical data. Despite 2022, where the year itself is a distinctive word.

Moreover, other organizations or companies, like “SAP”, “Institut”, “Oleoestepo”, or “Saico” are distinctive in certain years, a finding that follows the tendency spotted by Liu and Zhang, that “the companies’ collaboration with other businesses and their newly launched products or services” (Liu and Zhang, 2021: 3) is a common topic among Walmart and other companies press releases, and despite their differences, Mercadona and Walmart are successful retail companies that regularly issue press releases.

Furthermore, analyzing distinctiveness year by year, we can see the most significant and representative topics of those years. In 2014 contactless payment was an important theme, in 2015 most important topic was olive oil, in 2017 it was Iberian Peninsula drought, in 2018 anisakis parasite and fishing issues, in 2020 COVID-19 pandemic concerns and subsequent measures, and in 2021 environmental concerns. Besides, it is interesting that since 2019 plastic is a commonly used but also distinctive word, a tendency broke in

2020, probably due to the fact that 2020 press release communication was focused on health concerns and measures about COVID-19 pandemic.

Later on, I computed bigrams and trigrams, to visualize the ten most repeated ones, to gain insight into the recurring expressions and topics in the press releases.



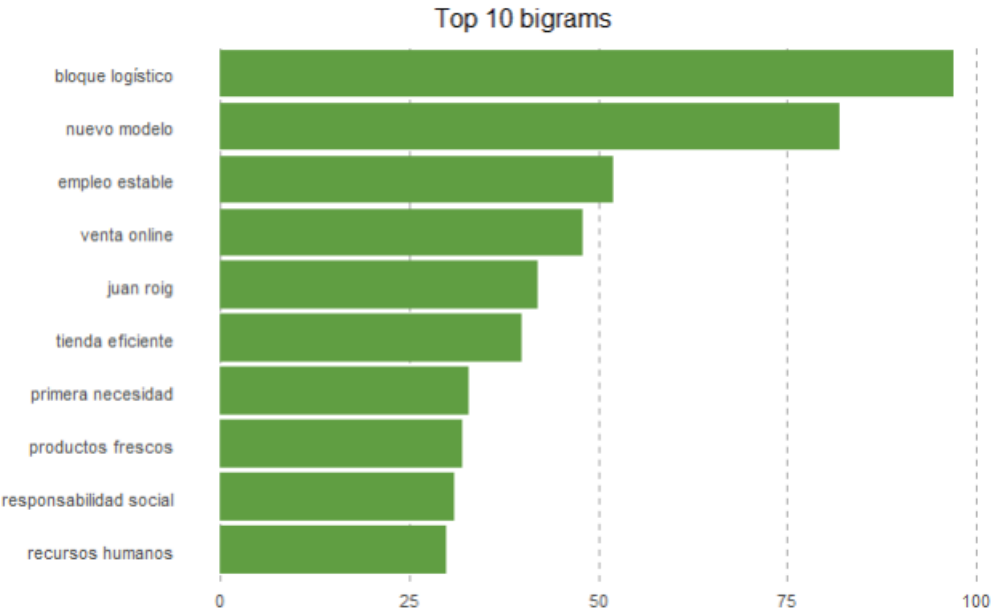
**Graph 13**

*Top ten bigrams in Mercadona press releases from 2013 to 2022.*



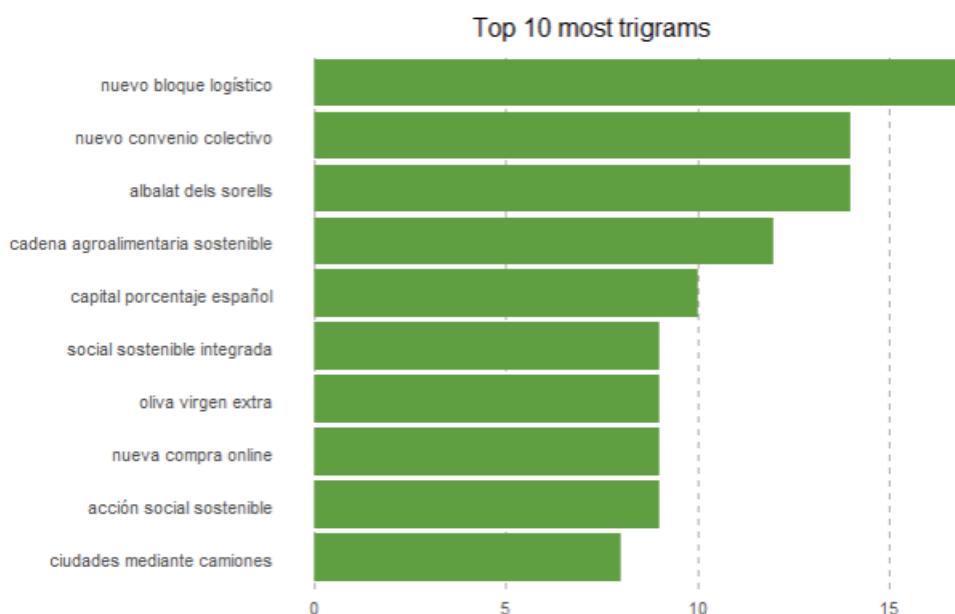
**Graph 14**

*Top ten trigrams in Mercadona press releases from 2013 to 2022.*



**Graph 15**

*Top ten bigrams in Mercadona press releases from 2013 to 2022.*



**Graph 16**  
*Top ten trigrams in Mercadona press releases from 2013 to 2022.*

Six main topics or themes can be identified using the visualization of the ten most common bigrams and trigrams in Mercadona's press releases. The first recurrent topic observed is the emphasis on logistics and innovation, as depicted by "bloque logístico", "nuevo bloque logístico", "ciudades mediante camiones", "tienda eficiente" and "nuevo modelo", these bigrams and trigrams suggest a strategic focus on optimizing logistic infrastructures, enhancing supply chain operations, and redefining the business model to achieve heightened efficiency and reliability.

Another frequent topic in Mercadona's press releases is Corporate Social Responsibility (CSR), because among the top ten bigrams and trigrams there are concepts that show Mercadona commitment to social responsibility and sustainable practices, such as "responsabilidad social", "cadena agroalimentaria sostenible", "social sostenible integrada" and "acción social sostenible". In fact, "the ethical behavior of companies has become an issue of major concern to their management and control bodies, public administrations and society in general"<sup>5</sup> (Blanco y Gutiérrez, 2008:185). Following Porter and Kramer classification (Portner and Kramer, 2022 in Blanco y Gutiérrez, 2008:185) there are three levels of Corporate Social Responsibility importance within a company, and the third and top one is when "social responsibility and business ethics are integrated

<sup>5</sup> Original reference: "el comportamiento ético de las empresas se ha convertido en un tema que preocupa de forma importante a sus órganos de dirección y control, a las administraciones públicas y a la sociedad en general"

into management, strategy and business competitiveness, becoming an instrument that can provide competitive advantages in the marketplace<sup>6</sup>” (Blanco y Gutiérrez, 2008:185), and this is the case of Mercadona, whose Corporate Social Responsibility “has achieved outstanding economic results and very positive levels of ethical behaviour, recognized nationally and internationally<sup>7</sup>” (Blanco y Gutiérrez, 2008:186).

Actually, both topics, Corporate Social Responsibility and logistics and innovation, are closely related, because “Mercadona is involved in various improvement activities to reduce its environmental impact through logistics optimization, energy savings and waste management, even introducing efficient store models that result in energy savings to reduce its environmental footprint (Hernando et al., 2019).

Besides, another frequent topic observed is online sales as shown by “venta online” and “nueva compra online”, which shows Mercadona efforts to adapt to the growing trend of online shopping and expand its online purchasing options, because “the company was a latecomer to e-commerce<sup>8</sup>” (Interbrand, 2022).

One more recurrent topic in Mercadona press releases is customer needs satisfaction because the concepts “productos frescos”, “primera necesidad” and “oliva virgen extra” depict Mercadona emphasis on offering high-quality basic necessities to their customers.

The fifth common topic observed is employee welfare, as the following bigrams and trigrams show “empleo estable”, “recursos humanos” and “nuevo convenio colectivo”. This is due to the fact that “one of the most important pillars in Mercadona’s success is their workers, their philosophy is that workers must be satisfied in order to satisfy customers. In fact, “besides permanent positions and wages above the sector average, they have performance related bonuses, and the company makes a high investment in training” (Hernando et al., 2019: 135).

Moreover, the last observed common topic is Mercadona itself as a company, the trigram “capital porcentaje español” suggest discussions about the percentage of the Spanish capital in the company and highlights the significance of Spanish ownership in its operations, which according to Marin and Lozano is a common topic in Mercadona

---

<sup>6</sup> Original reference: “la responsabilidad social y la ética de los negocios se integra dentro de la gestión, la estrategia y la competitividad empresarial, convirtiéndose en un instrumento que puede otorgar ventajas competitivas en el mercado”

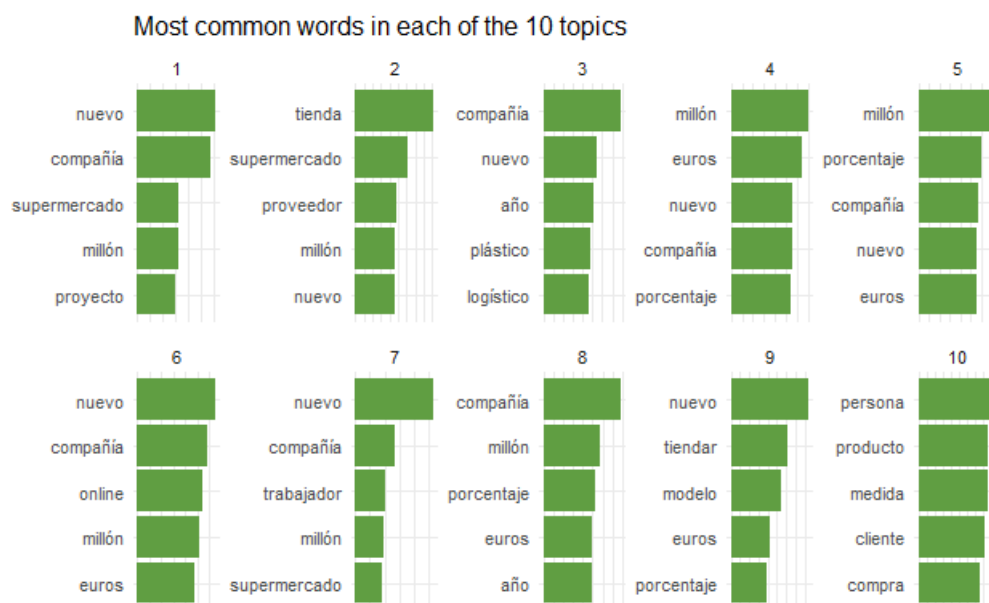
<sup>7</sup> Original reference: “ha permitido alcanzar unos resultados económicos sobresalientes y unos niveles de comportamiento ético muy positivos, reconocidos a nivel nacional e internacional”

<sup>8</sup> Original reference: “la compañía llegó tarde al comercio electrónico”

webpage (Marín and Lozano, 2017:24-25). Besides, there is the bigram “Juan Roig”, the chairman and CEO of Mercadona, among the top ten bigrams in Mercadona press releases, which shows his prominent role and influence in the company. And finally, the trigram “Albalat dels Sorrels” refers to “Ciudad Mercadona” a megacomplex project that started in 2014 and is scheduled to be completed by 2025, that will house Mercadona’s new headquarter (Guevara, 2023).

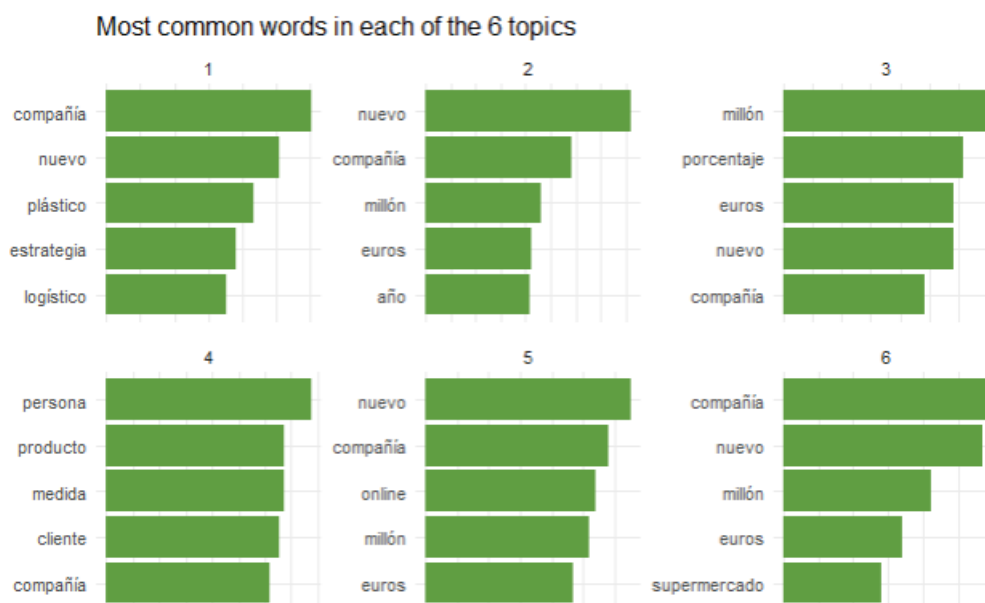
Finally, the last text mining technique applied was Latent Dirichlet Allocation (LDA) model, for which I needed to transform the data into a term matrix. Later, press releases sparsity which was 74%, was examined. This number indicates that approximately 74% of the elements in the document-term matrix are zeros, and a zero value means that a particular term does not appear in a specific document. Hence, this sparsity value might be due to the fact that the press releases vocabulary, understood as a set of unique words, is relatively large compared to the number of press releases.

Consequently, I started setting the number of topics to ten, due to the fact that there is ten years of press releases, but due to unsatisfactory results I changed to six the number of categories because that was the number of themes I spotted when using bigrams and trigrams. However, as those results were not satisfactory either, I set four as the number of topics, and I realized, as graph 17, 18 and 19 show, that independently from the number of categories used, there were some words repeated across many topics, and these words were present on the top ten most used words when word count was applied.



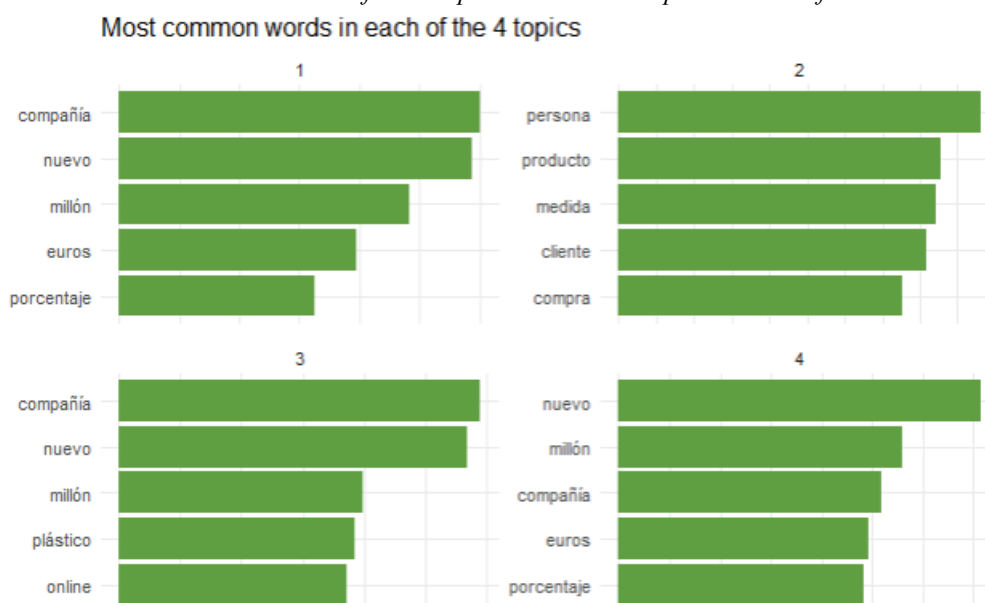
**Graph 17**

*Five most common words in each of the 10 topics in Mercadona press releases from 2013 to 2022.*



**Graph 18**

*Five most common words in each of the 6 topics in Mercadona press releases from 2013 to 2022.*



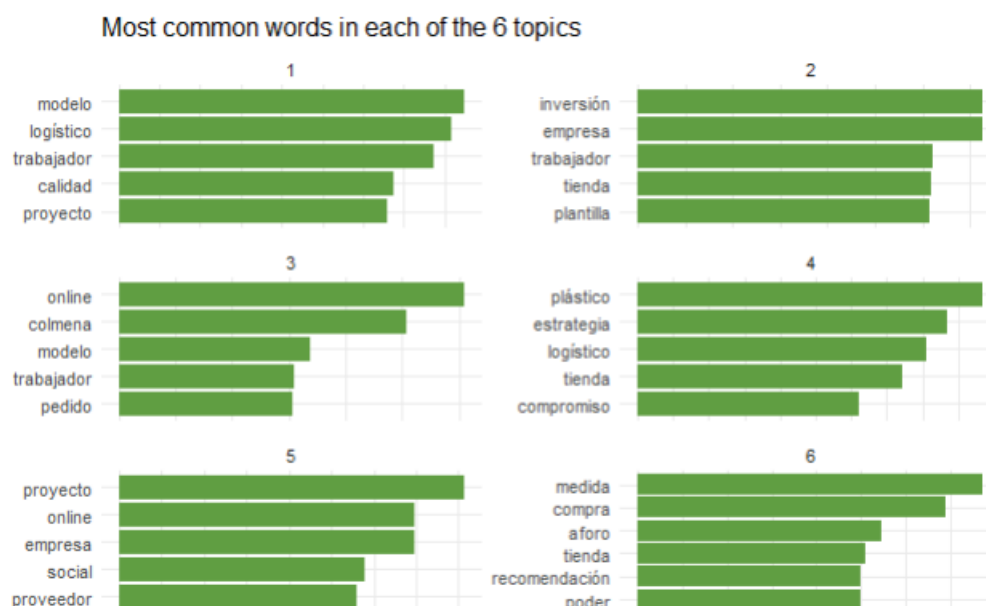
**Graph 19**

*Most common words in each of the 4 topics in Mercadona press releases from 2013 to 2022.*

Therefore, I removed the top ten most used words because they are not specific or distinct enough to identify and differentiate specific topics, due to them being too generic or common in the Mercadona press releases, which is also an interesting finding because those words are common language used in Mercadona press releases, and they could be considered part of the basic structure of Mercadona press releases.

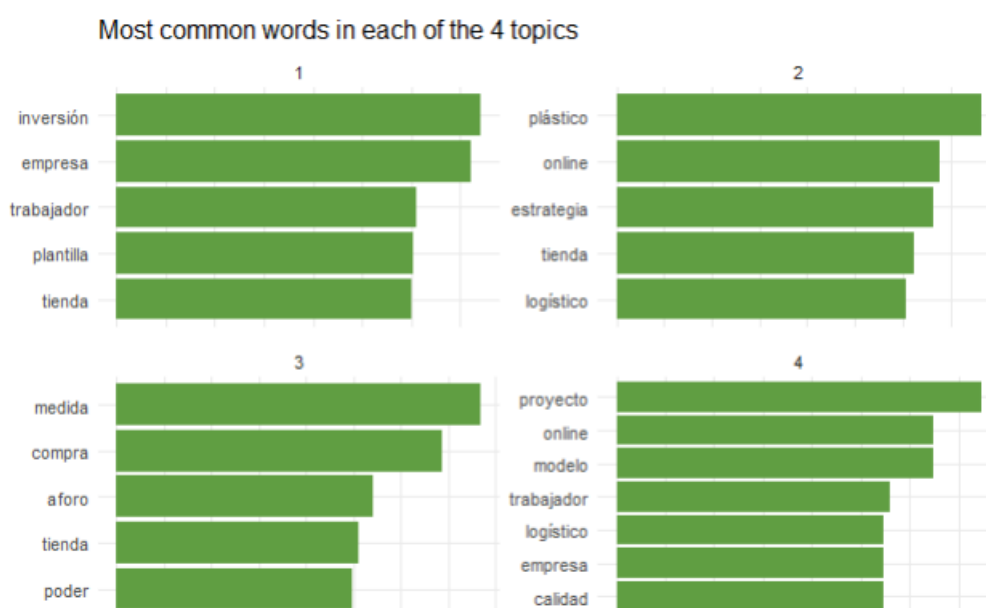
After removing the top ten common words LDA algorithm was applied again, and as graph 20 and 21 show, I realized that independently from the number of categories set, 6

or 4, there were new words repeated across many topics that needed to be removed "logístico", "empresa", "online", "tienda", "modelo", "trabajador".



**Graph 20**

*Most common words in each of the 6 topics, after removing top ten most used words in Mercadona press releases from 2013 to 2022.*



**Graph 21**

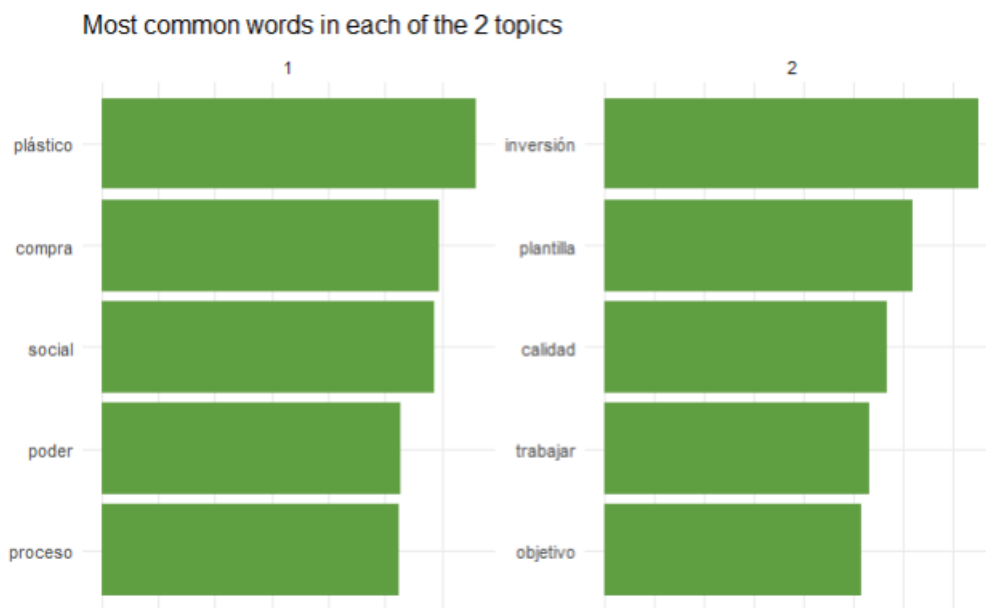
*Most common words in each of the 4 topics, after removing top ten most used words in Mercadona press releases from 2013 to 2022.*

Those five words that are repeated in most of the categories at the same time are not specific or distinct enough to differentiate topics, because they are generic or common in Mercadona press releases, and therefore, shape Mercadona press release basic structure.



Hence, these words were also removed to focus on identifying more unique and informative terms.

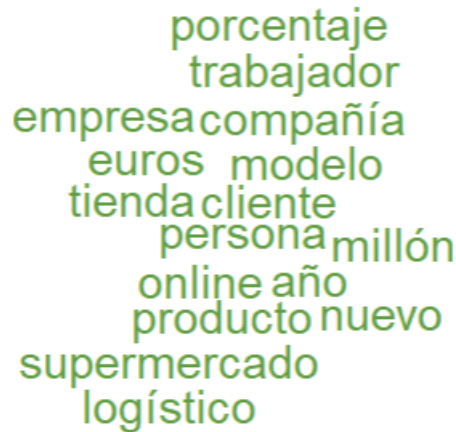
As a result of removing the words that were repeated across topics at the same time, and doing an iterative work of changing LDA statistical model parameters and arguments, I concluded that there is two main topics in Mercadona press releases, as graph 22 shows. One of the identified topics refers to the company itself, its activities and characteristics, and the other identified topic refers to its Corporate Social Responsibility, referring its sustainable and social compromise.



**Graph 22**

*Five most common words in each of the 2 topics, after removing top ten most used words among other too frequent words, in Mercadona press releases from 2013 to 2022.*

Furthermore, another interesting conclusion can be dragged from using LDA model, because the words I had to remove during the process are common language used in Mercadona press releases, hence, they are the basic structure of Mercadona press releases. Thus, these words can be used by any public relations professional as a guideline to write press releases, despite "modelo", "logístico", "online", "supermercado", "tienda" and "producto", because those words are specific to Mercadona business model. Therefore, when using words displayed in graph 23 as a schema to create press releases, those specific Mercadona business words should be changed to the company the public relations professional is working for, specific business model characteristics and terms.



**Graph 23**

*Removed words from LDA statistical model.*

## **CONCLUSION**

As a result of analyzing ten years of Mercadona press releases using data harvesting and data mining techniques, I have reached the following conclusions regarding Mercadona communicational patterns and trends.

The first conclusion is that Mercadona press releases have a clear focus on placing their organizational agenda on the media. This is a statement supported by press release theory but also by my results because “nuevo” and “compañía” are the two most repeated words, but also the most frequent ones, which emphasize Mercadona's efforts to highlight their latest developments and innovations. Besides, “millón” and “porcentaje” are also among the most repeated and frequent words, plus “euros” and “año” are among the most repeated words, which indicates a strong emphasis on providing transparent and credible financial and statistical data about year-term performance. In fact, distinct locations, different organizations or different year references were among the distinctive words yearly, which again shows Mercadona effort to provide credible information about their business activities. On top of that, Juan Roig was referenced many times every year as graph 6 shows and depicts the fact that Juan Roig is among the most frequent bigrams. Therefore, as all these widely used terms are related to their past and future company activities, we can conclude that Mercadona press releases focus on explaining their company agenda. Furthermore, the top ten used bigrams and trigrams found are related to Mercadona agenda, and even both topics identified using LDA, the company itself and Corporate Social Responsibility, are also related to Mercadona agenda.

Second conclusion is how relevant Corporate Social Responsibility (CSR) and the principles of stakeholder economy are for Mercadona. As commented above, one of the topics identified using LDA is CSR, and also, CSR was one of the topics identified using the ten most common bigrams and trigrams. Moreover, it is highly interesting that among all their CSR activities, plastic reduction is the most characteristic one, because “plástico” is among the most frequent words. Besides, graph 12 shows, were distinctive words across the years were displayed, and graph 7 shows, where the use of “plastic” word was analyzed across the years, plastic is a word that started to be used in 2019 till 2022, a total count of 104 times, being a distinctive word during all those years, despite 2020, when communication was centered around COVID-19 health crisis. Hence, CSR is highly relevant for Mercadona, and plastic reduction is their most characteristic CSR activity.

The third conclusion is that in Mercadona the customer or “The Boss” is what goes first, which is a defining feature of marketing-oriented companies, but also a statement supported by this research results. As graph 4 shows, the created word using the regular expressions “eljefe”, which is the word that Mercadona uses to refer to the client, is repeated 135 times and its presence across all the years, and “cliente” itself, is among the most used and frequent words, in graph 11. Moreover, another created word “coinnovación” is present for most of the years, and this word refers to centers where customers are listened to by Mercadona in order to improve their products and services. Besides, customer needs satisfaction was one of the topics found when analysing the ten most used bigrams and trigrams. Likewise, across the years, some of the distinctive words in graph 12, refer to trendy topics in those years, like contactless payment on 2014, Iberian Peninsula drought on 2017, “plástico” since 2019 and COVID-19 in 2020. In fact, all these distinctive words show that Mercadona press releases try to cover topics of general interest for their customers, and the rest of their stakeholders and even society itself, which altogether with the importance of co-innovation concept, the wide use of the “eljefe” and “cliente”, and also customer needs satisfaction being a common topic, depicts that the client or “The Boss” is what comes first for Mercadona.

And finally, the last conclusion was not part of my initial objective, but while iteratively working with LDA model, I had to remove a total number of sixteen words because they were too commonly used and were not helpful to distinguish topics, hence, these words have been considered basic words that shape the structure of Mercadona press releases. And as a consequence, those words can be used as a schema of what terms and topics can

be included in press releases, taking into account that a few of them are related to Mercadona business, and should be switched to characteristic words of the business we are creating press releases for, a finding that can be helpful for public relations professionals.

## REFERENCES

- Apolo D., Báez V., Pauker L. and Pasquel G. (2017): “Gestión de Comunicación Corporativa: consideraciones para el abordaje de su estudio y práctica” *Revista Latina de Comunicación Social*, volumen 72, pp: 521-539. DOI:10.4185/RLCS-2017-1177
- Blanco-Callejo, M. and de Pablos-Heredero, C. (2019) “Co-innovation at Mercadona: a radically different and unique innovation model in the retail sector” *Journal of Business and Retail Management Research (JBRMR)*, volume 13(4), 326, pp: 326-341. DOI:10.24052/JBRMR/V13IS04/ART-30
- Blanco M. and Gutierrez S. (2008) “Un modelo de gobierno corporativo que permite un comportamiento ético: el caso Mercadona” *M@n@gement*, volume 11(2), pp: 161-189
- Guevara, Á. (2023) “Ciudad Mercadona, el megacomplejo que acogerá su nueva sede en Albalat del Sorrels” in *arquitecturayempresa*. Available at: <https://arquitecturayempresa.es/noticia/ciudad-mercadona-el-megacomplejo-que-acogera-su-nueva-sede-en-albalat-del-sorrels> (Accessed: June 09, 2023).
- Hernando J., Colvin-Díes J. and Enrique J. (2019) “Mercadona a successful business case” *Academicus International Scientific Journal*, volume 20(20), pp 128-144. DOI:10.7336/academicus.2019.20.10
- Infoadex (2018) *Estudio Infoadex de la inversión publicitaria en España 2018*, Madrid. (Access provided by the library of the Faculty of Information Sciences, Universidad Complutense de Madrid)
- Infoadex (2019) *Estudio Infoadex de la inversión publicitaria en España 2019*, Madrid. (Access provided by the library of the Faculty of Information Sciences, Universidad Complutense de Madrid)
- Interbrand (2022) “Mejores marcas españolas.Mercadona”[PDF ]. Available at: <https://interbrand.com/madrid/best-spanish-brand/mercadona/> (Accessed: March 12, 2023).
- Kantar (n.d.) *Cuotas de mercado de la distribución* [Online] Available at: <https://www.kantar.com/es/campaigns/cuotas-de-mercado-de-la-distribucion> (Accessed: March 26, 2023)

- Liu S. and Zhang J. (2021). "Using Metadiscourse to Enhance Persuasiveness in Corporate Press Releases: A Corpus-Based Study" *SAGE Open*, volume 1(11), pp:1-11. DOI: 10.1177/21582440211032165
- Marín Dueñas, P. P., & Lozano Fernández, J. (2017) "La comunicación de marketing en la empresa de distribución española Mercadona" *Retos*, 13(VII), 9-26. ISSN impreso: 1390-6291; ISSN electrónico: 1390-8618. DOI:10.17163/ret.n13.2017.01
- Marín Dueñas, P. P., Mier-Terán Franco, J. J., & Lozano Fernández, J. (2018) "La comunicación en las empresas de distribución alimentaria en España: un análisis de las herramientas online y offline" *Revista de Comunicación de la SEECI*, XXII(45), 55-73. DOI:10.15198/seeci.2018.45.55-73
- Mercadona. (n. d.) "Sala de Prensa: Hemerotecas". Available at: <https://info.mercadona.es/es/conocenos/sala-de-prensa/hemeroteca?idCategoriaSeleccionada=1470731303723&tipoVisualizacion=grid&numPagina=1> (Accessed: June 3, 2023)
- Mercadona (2013-2022) "Sala de Prensa: Memorias anuales" [PDF]. Available at: <https://info.mercadona.es/es/conocenos/sala-de-prensa/memorias-anuales/memoria-anual-2021> (Accessed: March 12, 2023).
- Mercadona (2017) "Mercadona abrirá su primer centro de Coinnovación en Portugal" Available at: <https://info.mercadona.es/es/actualidad/mercadona-abrira-su-primer-centro-de-coinnovacion-en-portugal/news?idCategoriaSeleccionada=1470731303671> (Accessed: June 9, 2023).
- Palencia-Lefler, Manuel y Murillo, Ignacio (2009) "Utilización de la nota de prensa para comunicar campañas institucionales" in Fuentes, J. R., Ruiz, F. X., Pineda, A., & Latorre, O. P. (2009) *Questiones publicitarias. Revista internacional de comunicación y publicidad* (3), pp. 246-267. DOI:10.5565/rev/qp.118
- Pulido M. (2015) "Técnicas de relación con la prensa en la gestión de la comunicación organizacional: el caso del Consejo Regulador del Brandy de Jerez" *Questiones publicitarias*, volume 1(20), pp: 87-100. DOI:10.5565/rev/qp.10