Cluster analysis of sustainability narratives in the Spanish press during the period 2014-2019

Fernando Castelló-Sirvent

Universitat Politècnica de València

Vanessa Roger-Monzó

Universitat de València

Abstract

This study explores how sustainability is represented in Spanish media through 12,647 news items that are analysed from the Factiva® database, published between 2014 and 2019, covering 264 national newspapers and several news agencies like EFE and Europa Press, to identify recurring themes related to sustainability, innovation, and technology over six years. Utilising T-LAB Plus 2020 for textual data analysis, the study identifies significant lexical patterns within these narratives. The analysis involved 2,280 initial lexical units, narrowed to 455 for detailed examination based on frequency thresholds. The data clustering revealed three primary thematic groups: "technology in supply chains," "innovation in the university," and "tourism, city, and mobility." These clusters highlight different aspects of sustainability narratives.

The "technology in supply chains" cluster emphasises the role of digitisation, transformation, and investment in technology for modern business practices, focusing on improving efficiency and competitiveness. This cluster underscores the importance of integrating advanced technologies like blockchain and business intelligence to enhance operational capabilities and security within supply chains. The "innovation at university" cluster reveals a strong connection between academic research and societal advancement, highlighting universities' role in fostering innovation. This cluster emphasises the significance of collaborations between academia and industry, particularly in health and science, to drive research that can have a practical societal impact. The "tourism, city, and mobility" cluster focuses on sustainable urban development and integrating intelligent technologies. It examines the effects of tourism on urban environments and the need for efficient transportation systems that align with sustainability goals, emphasising energy efficiency and renewable resources.

The study concludes that these three clusters offer valuable insights into the dominant sustainability narratives in the Spanish press. Further research should consider additional media formats to understand public perceptions and narratives surrounding sustainability better. Despite the limitations of the selected media sources, these findings provide a foundation for developing new analyses and case studies related to sustainability topics in Spain. This empirical paper is exploratory and contributes to the literature on technologies

and innovations for sustainability. An important novelty of this research is to study the press and apply robust text analysis methodologies. The findings of this study contribute to the literature of the field of knowledge, advising practitioners and guiding scholars in their future research.

Introduction and purpose

This study offers evidence of the sustainability narratives published in the Spanish press between 2014 and 2019. The Factiva® database from Down Jones & Company© was used, and 12,647 news items published in the study period were analysed.

Materials and Methods

The Factiva® tool, an information database owned by Dow Jones & Company©, searches for publications. This database provides access to over 33,000 premium sources globally and includes 264 national newspapers from Spain. It also tracks information from agencies such as EFE and Europa Press.

The analysis, including references to the terms “sustainability,” “innovation,” and “technology,” is carried out between January 1, 2014, and December 31, 2019. This case study contributes to expanding the knowledge of blockchain on the green agri-food supply chain.

The subsequent analysis uses T-LAB Plus 2020, a data analysis software that identifies word patterns through statistical, graphic, and content analysis applications. T-LAB information extraction relies on two types of textual units: elementary contexts and lexical units.

Elementary contexts are text segments corresponding to syntagmatic units of one or more sentences, resulting from the linguistic corpus segmentation by T-LAB for co-occurrence computation analysis. Lexical units include a word and its lemma, which are displayed and enumerated as they appear in the corpus and classified according to custom dictionaries and semantic categories.

The initial automatic normalisation by T-LAB yields 2,280 lexical units. To ensure the reliability of statistical data, T-LAB sets a frequency threshold, with the minimum frequency threshold for this corpus set at 10. This results in selecting 455 lexical units (lemmas or keywords).

Results and Discussion

Based on these quantitative characteristics, clusters or thematic groups are classified. This tabulation identifies the topics covered in the news, their relationships, and their frequency in the media. The corpus examination yields significant thematic clusters characterised by similar lemma patterns.

The unsupervised "clustering" method (bisecting k-means algorithm) in T-LAB first performs a co-occurrence analysis, followed by a comparative study. Out of 170,694

analysed contexts, 139,925 (81.97%) are classified. A three-cluster partition is selected for its adherence to the sample.

Three clusters are identified: "technology in supply chains," "innovation in the university," and "tourism, city, and mobility." The first cluster is chosen to design a case study focusing on the impact of technology and innovation on sustainability.

Table 1. Cluster “technology in supply chains”

Original Lemma (Spanish)	Lemma	In Clu	In Tot	Chi^2	p-value
empresa	company	30,725	36,868	16,678.070	0.000
mercado	market	9,791	10,702	7,320.991	0.000
cliente	client	6,867	7,277	5,678.623	0.000
negocio	business	7,892	8,785	5,525.007	0.000
inversión	investment	7,916	9,604	3,988.180	0.000
digitalización	digitization	10,266	13,676	3,385.839	0.000
transformación	transformation	3,449	4,816	874.197	0.000
líder	leader	3,305	4,623	829.144	0.000
cadena	chain	1,969	2,697	554.541	0.000
business intelligence	business intelligence	1,189	1,483	527.914	0.000
distribución	distribution	1,298	1,707	450.847	0.000
automatización	automation	1,217	1,598	425.893	0.000
competitividad	competitiveness	3,552	5,713	320.485	0.000
alimentación	feeding	4,299	7,055	317.542	0.000
supermercado	supermarket	359	425	197.813	0.000
fruta	fruit	534	732	149.453	0.000
blockchain	blockchain	696	1,004	144.456	0.000
ciberseguridad	cybersecurity	447	601	138.664	0.000
Carrefour	Carrefour	162	190	92.569	0.000
productor	producer	718	1,222	34.443	0.000

In Clu: total number of elementary contexts that include that same lemma
In Tot: number of elementary contexts that include a specific lemma
Chi^2: Chi-squared test
p-value: the probability that the calculated chi-squared test value is possible given one null hypothesis. The required significance value has been set to $p<0.05$.

Source: Own elaboration

Cluster 1: Technology in Supply Chains

The cluster "technology in supply chains" (Table 1) analysis reveals significant insights into technology's role in modern business practices. The evidence available demonstrates a strong emphasis on digitisation and transformation within companies. Key terms such as "company," "market," "client," and "business" appear frequently, indicating that these entities are central to discussions about supply chains.

Investment in technology is highlighted, with "investment" and "digitalisation" being prominent. This reflects financial resources' critical role in adopting and integrating new technologies in supply chains. The presence of "business intelligence," "automation," and "blockchain" in the cluster suggests a growing trend toward utilising advanced technologies to enhance efficiency and competitiveness.

Moreover, "cybersecurity" indicates that protecting digital supply chain infrastructures is a significant concern. This study's findings underscore the importance of technology in maintaining and improving supply chain efficiency and security. The application of technology improves operational efficiency and provides a competitive edge in the market.

Table 2. Cluster “innovation at university”

Original Lemma (Spanish)	Lemma	In Clu	In Tot	Chi^2	p-value
investigación	research	10,560	13,724	24,787.770	0.000
sanidad	health	7,306	8,639	20,110.040	0.000
universidad	university	7,898	11,097	16,019.160	0.000
ciencia	science	4,365	5,593	10,418.210	0.000
sociedad	society	3,379	8,051	1,929.963	0.000
medicamento	medication	1,130	1,730	1,920.753	0.000
agricultura	farming	2,899	6,945	1,625.500	0.000
transferencia	transfer	987	1,624	1,445.932	0.000
doctorado	doctorate	409	539	923.305	0.000
matemáticas	maths	411	587	799.217	0.000
máster	master	615	1,129	705.806	0.000
innovación	innovation	10,315	40,632	314.548	0.000
universidad-empresa	university-company	35	65	39.046	0.000

In Clu: total number of elementary contexts that include that same lemma
In Tot: number of elementary contexts that include a specific lemma
Chi^2: Chi-squared test
p-value: the probability that the calculated chi-squared test value is possible given one null hypothesis. The required significance value has been set to $p<0.05$.

Source: Own elaboration

Cluster 2: Innovation at the University

The "Innovation at University" cluster (Table 2) focuses on research and development within academic institutions. The available evidence shows that terms like "research," "university," and "innovation" are predominant, suggesting a strong link between higher education and innovative activities.

Health and science are significant areas of interest, with "health" and "science" frequently mentioned. This indicates that universities are learning centres and hubs for scientific and medical advancement. The term "society" appears often, reflecting the societal impact of university-led research and innovation.

Additionally, the cluster highlights the importance of collaboration between universities and industries, as indicated by the term "university-company." This collaboration is essential for the practical application of research and for fostering innovation that can be commercialised. The findings suggest universities are crucial in driving innovation and contributing to societal development through research and industry partnerships.

Table 3. Cluster "tourism, city and mobility"

Original lemma (Spanish)	Lemma B	In Clu	In Tot	Chi^2	p-value
ciudad	city	11,667	14,183	21,272.250	0.000
turismo	tourism	12,082	15,976	18,514.190	0.000
energía	energy	12,042	17,731	14,483.150	0.000
sostenibilidad	sustainability	18,355	32,421	13,881.460	0.000
eficiencia	efficiency	6,517	10,022	6,992.158	0.000
inteligente	smart	5,095	7,806	5,504.582	0.000
movilidad	mobility	3,508	4,662	5,253.776	0.000
agua	water	4,556	6,947	4,976.223	0.000
urbano	urban	2,871	3,601	4,855.442	0.000
destino	destination	3,245	4,551	4,311.005	0.000
climático	climate	2,391	3,245	3,414.658	0.000
renovable	renewable	2,860	4,479	2,913.852	0.000
smart-cities	smart-cities	2,622	4,214	2,498.114	0.000
cambio	change	4,941	9,920	2,418.519	0.000
transporte	transport	2,636	4,318	2,388.328	0.000
futuro	future	4,707	9,851	1,979.582	0.000
emisión	emission	1,850	3,082	1,599.514	0.000
eléctrico	electric	2,321	4,211	1,572.240	0.000
medioambiental	environmental	2,252	4,152	1,451,996	0.000

OMT	UNWTO	605	738	1,079.951	0.000
infraestructura	infraestructure	2,034	4,174	914.770	0.000
residuo	residue	1,576	3,145	782.376	0.000
gobernanza	governance	494	748	545.364	0.000
ferrocarril	ralway	310	676	109.870	0.000
COP25	COP25	99	154	102.136	0.000
AVE	AVE	136	296	48.559	0.000
Utor	Utor	44	76	34.289	0.000
Baleària	Baleària	75	164	26.269	0.000
corredor	corridor	93	214	26.137	0.000

In Clu: total number of elementary contexts that include that same lemma
In Tot: number of elementary contexts that include a specific lemma
Chi^2: Chi-squared test
p-value: the probability that the calculated chi-squared test value is possible given one null hypothesis. The required significance value has been set to $p<0.05$.
Source: Own elaboration

Cluster 3: Tourism, City, and Mobility

The "tourism, city, and mobility" cluster (Table 3) emphasises the interconnections between urban development, tourism, and transportation. The evidence points to a strong focus on sustainability and smart cities, with "sustainability," "smart," and "city" being key terms.

Tourism's impact on urban environments is a significant theme, as indicated by terms like "tourism," "destination," and "urban." The emphasis on "mobility," "transport," and "efficiency" reflects the need for efficient transportation systems to support both tourism and city life.

Environmental considerations are also crucial, with terms like "energy," "renewable," "emission," and "environmental" appearing frequently. This underscores the importance of sustainable practices in urban planning and tourism management.

This study's findings suggest that integrating sustainable practices and intelligent technologies is essential for the future development of cities and tourism industries.

In summary, this cluster analysis provides valuable insights into the interconnected themes of technology in supply chains, innovation in universities, and urban development.

These findings highlight the critical roles of technology, research, and sustainability in shaping modern society.

Conclusions

This research allows us to understand the underlying narratives in the Spanish press on sustainability topics from 2015 to 2019. The findings report the three main clusters of these narratives. The main narratives are articulated around (a) technology in supply chains, (b) innovation at universities, and (c) tourism, city, and mobility.

The evidence found allows researchers to develop new analyses and case studies based on the results of this study.

This research has limitations. Despite being extensive and widely used in scientific research, the media database could contain biases from the selected media. The analysis of the press may be partial, given that agenda-setting influences its publication priority. Future research should extend its analysis to other media, such as radio or television and social networks, that control the behaviour of public opinion.