# TRANSFORMATION OF STATISTICAL INFORMATION INTO MEDIA CONTENT IN DATA JOURNALISM AND LANGUAGE STRATEGIES

**Vafa Zahid qizi Isgandarova**
PhD Candidate of the Doctor of Science Program at BDU
Republic of Azerbaijan, City of Baku
Vafa1.aslan@gmail.com

**Abstract:** This study examines the process of transforming statistical information into media content within the framework of data journalism and explores the role played by language strategies in this process. It emphasizes that language in data journalism is not merely a tool for conveying information, but also a constitutive element that contextualizes data, interprets it, and makes it understandable for readers. The importance of fundamental linguistic approaches-such as terminology selection, conceptual consistency, simplification, and explanatory narration-is analyzed to ensure the accurate understanding of statistical data. In addition, the study highlights that the harmony between textual narration and visual data presentations, such as graphs, tables, and infographics, plays a decisive role in the correct interpretation of data. The findings reveal that effective language strategies in data journalism are among the key factors that enhance the comprehensibility, reliability, and public impact of statistical information.

**Keywords:** data journalism, statistical information, language strategies, media content, data interpretation.

Data journalism is a contemporary journalistic approach that transforms not only the tools of news production in the digital age but also the ways in which knowledge is constructed, reality is explained, and information is presented to the public. Although this transformation is often described through technical stages such as data collection, verification, cleaning, analysis, and visualization, one of the most decisive dimensions of data journalism is the process of linguistic construction. This is because data, on its own, does not constitute a complete narrative carrying inherent meaning; rather, it becomes public knowledge only when it is selected within a specific context, interpreted, and made socially comprehensible through language [4, p. 3].

In traditional journalism, language has often been considered merely as a carrier that enables the transmission of events. In data journalism, however, the function of language is far more constitutive. Here, language is not simply a tool that conveys discovered information; it is the fundamental structure that selects, classifies, prioritizes, interprets, and renders data readable from the reader’s perspective. Numbers, tables, percentages, ratios, or trend graphs may initially appear as objective and self-explanatory structures, yet the context in which they are presented, the words used to explain them, the comparisons that support them, and the social consequences to which they are connected are largely shaped by language [1, pp. 217–218]. In other words, in data journalism, language is not an ornamental addition applied to data afterward; it is the very ground upon which meaning is produced.

At this point, the fundamental logic of data journalism should be understood as the process of transforming raw data into socially meaningful information. Raw data often appears fragmented, technical, and not directly accessible to readers. Presenting numbers, tables, or datasets alone may, in many cases, create ambiguity rather than



DOI: 10.5281/zenodo.20005908

produce understanding. This is because readers are not satisfied with answers to the question of “how many”; they also seek to understand what the number represents, under what conditions it emerged, what social consequences it signals, and why it matters. For this reason, in data journalism, statistical information can become part of a public narrative only through language. As Bounegru suggests, data journalism is part of a broader ecosystem shaped by data sites, tools, services, and applications; within this ecosystem, sharing the data behind stories and presenting sources transparently is of critical importance [4, p. 22]. However, as important as sharing data is, the way in which that data is narrated is equally significant. The public function of data largely becomes possible through processes of linguistic framing.

The approach defined by Julian Assange as “scientific journalism” also makes the role of language in data journalism more visible. In this approach, not only the outcome of a news story but also the process that leads to that outcome should be made transparent [4, p. 22]. However, methodological transparency cannot be achieved solely by sharing datasets; the process itself must be explained in an understandable manner. From which source was the data obtained? How was it filtered? According to which criteria was it classified? What types of comparisons were conducted? Within which limits was it interpreted, and what uncertainties does it contain? Opening these questions to the reader demonstrates the function of language in producing methodological transparency within data journalism.

One of the primary functions of language in data journalism is the contextualization of statistical information. Numerical data is an abstract indicator; when presented without context, it often appears incomplete, ambiguous, or open to interpretation for readers. For example, the information that unemployment in a region has risen to 12 percent may appear striking on its own; however, unless it is explained what this rate means compared to previous years, which social groups it affects, how it is positioned relative to regional averages, and what its role is within the broader economic structure, the explanatory power of the news remains limited. For this reason, in data journalism, language constructs an explanatory framework around data. This framework determines the place of numerical information within social reality and enables readers to understand data not merely as numbers but as indicators carrying meaning.

A second important function that accompanies contextualization is interpretation. In data journalism, transferring numbers into text alone is not sufficient; it is necessary to explain not only what the numbers state but also what they imply. An increase does not always signify a positive or negative development; similarly, a decrease may indicate an expected improvement, but it may also result from changes in measurement methods. Therefore, language in data journalism disrupts the raw appearance of numbers and reveals their social, political, economic, or cultural meanings. In this sense, the journalist is positioned not merely as a spokesperson for data but as its interpreter and public translator. This translation process is constructed through linguistic choices, conceptual selections, and narrative organization.

Another important role of language in transforming statistical information into media content is the production of simplification and accessibility. One of the fundamental challenges encountered in data journalism is that technical and statistical expressions, when transferred directly into news language, may render texts difficult to understand for non-expert readers. However, the public function of data journalism is

not only to produce knowledge but also to circulate it and make it readable for different social groups [3, pp. vii–viii]. For this reason, simplification in data journalism does not mean oversimplification; rather, it refers to presenting complex information more clearly without distorting its accuracy. The central challenge here lies in maintaining a balance between scientific precision and communicative clarity.

During the simplification process, language functions as an interface that brings technical terminology closer to everyday understanding. Expressions such as “an 18 percent increase,” “median income level,” “standard deviation,” “correlation,” “data distribution,” or “statistical significance” may be clear within specialized fields, yet they are not always equally understandable to general audiences. Therefore, in data journalism, language operates as a form of conceptual translation mechanism. Technical concepts are explained using more familiar words, abstract ratios are supported with concrete examples, large numbers are adapted to everyday scales, and a narrative flow is established that readers can cognitively follow. For instance, public expenditures expressed in millions become much more understandable when presented alongside per capita costs or comparisons with previous years. In this way, language increases the comprehensibility of data without eliminating its complexity.

However, the simplification process also involves ethical and epistemological risks that must be carefully considered. Excessive simplification may render the nuances of data invisible; when certain details are lost, readers may encounter information that appears clearer but is ultimately less complete. Therefore, constructing plain language in data journalism is not the same as constructing simplistic language. Plain language makes complexity understandable, whereas simplistic language may ignore complexity altogether. A strong academic and professional language in data journalism must establish a balance precisely between these two dimensions. It should enable readers to follow the text with ease while preserving conceptual accuracy.

Within this framework, terminology selection and conceptual consistency occupy a central position in language strategies in data journalism. This is because every concept used in a data-driven text does not merely convey technical information; it also shapes the framework through which readers perceive the issue. There are significant differences between terms such as “increase” and “surge,” “decrease” and “decline,” “risk” and “probability,” or “mean” and “median.” Although these distinctions may appear minor at the linguistic level, they are highly decisive in the production of meaning. An incorrectly chosen term may exaggerate the social impact of data or, conversely, make it appear less significant than it actually is. Therefore, in data journalism, terminology is not merely a matter of word choice; it is also a matter of analytical accuracy and ethical responsibility.

Conceptual consistency is equally critical at this point. Using the same concept in different meanings across various parts of a text undermines the credibility of data-driven reporting. For example, expressions such as “poverty rate” in one section, “low income” in another, and “economic disadvantage” elsewhere—used interchangeably without clear distinctions—may blur conceptual boundaries. In a professional and academic data journalism text, however, the meaning of each concept must be clearly defined, and this consistency must be maintained throughout the text.

Another defining dimension of language in data journalism is conceptualization. Statistical data gains social meaning only when it is associated with specific concepts. An unemployment rate, for instance, is not merely an economic indicator; it may also

be linked to broader conceptual domains such as social exclusion, class inequality, regional development disparities, or the effectiveness of public policy. Similarly, education data does not only reflect academic achievement; it may also make visible issues such as equality of opportunity, regional distribution, gender-based disparities, or social mobility. For this reason, conceptualization in data journalism is one of the fundamental processes that transform numerical information into a component of social thought. Language stands at the center of this process, as the concepts through which data is approached, the theoretical framework within which it is presented, and the social problems with which it is associated are all determined at the discursive level.

In this context, language in data journalism performs not only an explanatory but also a framing function. The words used in a news text do not leave data in a neutral position; rather, they present it within a specific hierarchy of priorities. For example, data on violence against women may be framed as a “family issue,” a matter of “gender inequality,” a “public safety concern,” or a “human rights violation.” Depending on which of these frames is employed, the reader’s perception of the same data will differ. The same dataset may produce different public effects when expressed through different linguistic frames. Therefore, in data journalism, language is not merely reflective but constitutive in nature. Providing numbers objectively is not sufficient; it is equally important to consider the discourse within which these numbers circulate.

Another prominent dimension of language strategies in data journalism is the capacity to construct narratives. In the process referred to as data storytelling, language does not simply list data as isolated pieces of information; rather, it brings them together within a specific logic, rhythm, and flow. An effective data story, much like a strong written narrative, follows a structure of introduction, development, and conclusion. It first confronts the reader with a question, then unfolds the layers of that question through data, and ultimately leads to a meaningful conclusion. To establish such a structure, language must be both information-rich and fluid. In data journalism, the text occupies a position between a technical report and a narrative. It is neither purely a numerical record nor solely a literary flow; instead, it creates a distinctive mode of expression that balances both qualities.

At this point, sentence structure, rhythm, and paragraph organization also gain importance. Sentences constructed repeatedly with the same patterns may make data-driven texts appear mechanical and artificial. A strong academic media language, however, requires at times explanatory long sentences, at times emphatic short ones, and at times comparative structures. Such variation enhances the vitality of the text and helps sustain reader attention. Particularly in data journalism texts, where numerical density is high, linguistic monotony may complicate the reading experience. Therefore, narrative fluency is not merely an aesthetic concern; it is directly related to the comprehensibility of data.

Reader orientation is another fundamental principle of language strategies in data journalism. Not every reader possesses the same level of data literacy, conceptual knowledge, or media consumption habits. For this reason, language should be shaped according to the knowledge level of the target audience. In texts aimed at general readers, explanatory and plain expressions, examples, everyday analogies, and clear comparisons should be emphasized, whereas content intended for expert audiences may adopt a higher degree of technical conceptual density. However, regardless of the target audience, one fundamental principle remains unchanged in data journalism: the text

should not exclude readers; rather, it should invite them into the data. A text fulfills its public function not when readers feel inadequate in the face of data, but when they feel capable of interpreting it.

Within this framework, question-based narration constitutes a highly effective linguistic technique in data journalism. Anticipating and answering questions that may arise in the reader’s mind-such as “What does this data show?”, “Why is this important?”, “How does this change affect me?”, or “Why did this number increase?”-facilitates the process of interpretation. In this way, the text becomes not merely an informative structure but one that accompanies the reader’s cognitive progression. This approach is particularly important in news stories dealing with complex social data, as it encourages readers to engage actively with the text and to interpret numerical information thoughtfully rather than passively.

Another significant dimension of language in data journalism is its capacity to generate trust. As Paul Bradshaw notes, data-like any source-must be approached critically; attention should be given to how data shapes and constrains the story [4, p. 3]. The communication of this critical approach to readers is made possible through language. When journalists clearly include in the text the source of the data, the methods of collection, its limitations, and potential problems, they do not merely present the outcome to the audience; they also reveal how that outcome was produced. In this way, language occupies a central role in establishing credibility in data journalism.

The way numerical information is represented in media language is also directly related to linguistic strategies. The same dataset may be presented as an absolute number, a percentage, a per capita value, a regional comparison, or a change over time. The choice among these forms of representation is not merely technical but discursive, as each directs the reader’s attention to different aspects of the issue. For example, the statement “10,000 people were affected” and the statement “7 out of every 100 people were affected” may be based on the same dataset, yet they produce different perceptions in readers. Therefore, in data journalism, language determines which form of representation makes the data more meaningful. The strength of the narrative often derives not from the data itself but from the way in which the data is presented.

The relationship between visual data presentation and linguistic narration also holds particular importance in data journalism. Graphs, tables, maps, and infographics visually simplify complex datasets; however, these visuals require explanatory language in order to be interpreted correctly. A graph gains meaning only when accompanied by its title, caption, axes, data labels, and references within the text [2, pp. 110–112]. In cases where explanatory language is weak, visual data may be misinterpreted or create an overly simplified impression. As Bradshaw suggests, data provides the opportunity to tell complex stories through infographics [4, p. 2]; however, for this potential to be effective, a strong discursive alignment must be established between text and visuals. For this reason, text–visual coherence in data journalism is not merely an aesthetic concern but a cognitive and semantic necessity.

With the development of digital media environments, new linguistic practices have also emerged in data journalism. Websites, social media platforms, mobile screens, and interactive data interfaces each present environments that require different forms of discourse. While long-form web analyses allow for explanatory and in-depth language, social media content emphasizes short, striking, and attention-grabbing expressions. In mobile-compatible content, language is typically clear, economical, and quickly

readable in order to fit dense information into small screens. Directive expressions used in interactive data dashboards demonstrate that language in such contexts does not merely construct narratives but also regulates user behavior. Expressions such as “Compare,” “Filter,” “Select your region,” and “Click to view details” reveal language’s capacity to manage interaction with data.

When examined specifically in the context of social media, the language of data journalism is reshaped within the logic of the attention economy. Here, numerical information is often presented through short headlines, visual cards, infographic fragments, brief videos, and impactful sentences. This situation requires language to become both shorter and denser. However, as brevity increases, the risk of losing context also rises. For this reason, in data-driven content produced for social media, language must be not only striking but also careful. Exaggerated or sensational expressions may undermine the credibility of data. Consequently, digital language strategies in data journalism must establish a balance between speed and accuracy, as well as between attention-grabbing expression and clarity.

The democratic function of language in data journalism should also be emphasized. The publication of open data alone does not automatically result in democratization. Data must be transformed into a form that citizens can understand, compare, and discuss. At this point, language serves as a bridge between technical knowledge and public awareness. Data journalism has the potential to transform citizens from passive recipients of information into active participants who question, connect ideas, and develop interpretations. However, the realization of this transformation depends on the inclusive, explanatory, and accessible use of language.

**Reference list**

1. Cairo, A. *The Functional Art: An Introduction to Information Graphics and Visualization* / A. Cairo. – Berkeley: New Riders, 2013. – 384 p.
2. Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M. & Hullman, J. The Science of Visual Data Communication: What Works // *Psychological Science in the Public Interest*. – 2021. – Vol. 22(3). – P. 110–161.
3. Nolan, D. & Stoudt, S. *Communicating with Data: The Art of Writing for Data Science* / D. Nolan, S. Stoudt. – Oxford: Oxford University Press, 2021. – 288 p.
4. Gray, J., Bounegru, L. & Chambers, L. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News* / J. Gray, L. Bounegru, L. Chambers. – Sebastopol, CA: O’Reilly Media, 2012. – 242 p.