

The Effects of Data-driven News Automation on Article Comprehensibility

Audience perceptions of linguistic features and the use and volume of numbers in manually written, post-edited, and automated data stories

In today's data-heavy information environments, journalistic reporting with and about numbers is crucial, as it provides readers with the necessary information to fulfill their roles in democracy (Nguyen & Lugo-Ocando, 2016). However, information about data can only be communicated successfully if the data-driven reporting is well done; otherwise, the information “fail[s] to be adequately understood and appreciated by the audience” (Battersby in Koetsenruijter, 2011, p. 75). A relevant factor contributing to the comprehensibility of certain text types, including news articles, is their *textual composition* (Göpferich, 2009). For data-driven reporting, article composition can include the amount and presentation of numerical information (Waite, 2020; Livingston and Voakes, 2005) as well as the clarity of the writing (Meyer, 1975).

At the same time, data-driven reporting is becoming increasingly automated (Carlson, 2015). Due to their technological requirements, the automation systems change journalists' approaches to writing news texts (Diakopoulos, 2019), which can impact article composition. As a result, automatically generated data-driven articles contain linguistic features that account for all contingencies of a story depending on the data retrieved, fewer key editorial features of journalistic writing (see Caswell, 2019), and more numerical information (forthcoming). To improve their readability and comprehensibility, some journalists have started ‘post-editing’ automatically generated stories (Thäsler-Kordonouri & Barling, 2023).

The increasing use of automation for news production and the emerging practice of post-editing emphasise the importance of recognising the technology-driven production diversity of data-driven reporting—including the manually written, automated, and post-edited kind. This diversity has led to readers consuming a variety of data-driven news texts with different article compositions. In light of the relevance of article composition for readers' understanding of numbers in the news, this study investigates how they evaluate the article composition of these

three types of data-driven reporting in comparison. Furthermore, we investigate how these evaluations relate to their comprehensibility perceptions of the numbers in the articles and the article overall.

A large-scale 3 (article source: manually written, automated, post-edited) × 12 (story topics) between-subjects online survey experiment was conducted using a sample (N = 4,734) representative of UK online news consumers by age and gender. The sample of respondents was drawn from various local regions and divided into 36 treatment groups. Each treatment group was exposed to a data-driven news article that had been produced either: (1) manually by a human journalist, (2) using template-based automation, or (3) in a post-edited manner, where a human journalist had further developed the automated article. Respondents' perceptions were measured using news perception criteria developed in a qualitative pre-study based on group interviews with UK news consumers (N = 31). Several of these domains' criteria have not been used in prior research on the perception of data-driven journalism, including that produced with the help of automation.

With this study, we tackle a research gap since our knowledge about readers' perceptions of data-driven reporting—including all three production types—remains limited. Furthermore, our results may help journalists better understand how data-driven reporting should be composed to maximise its comprehensibility for readers.

References

- Carlson, M. 2015. The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority. *Digital Journalism* 3(3), pp. 416–431. DOI: 10.1080/21670811.2014.976412.
- Casswell, D. 2019. David Caswell (2019) Structured Journalism and the Semantic Units of News. *Digital Journalism*, 7(8), pp. 1134-1156. DOI: 10.1080/21670811.2019.1651665.

- Diakopoulos, N. 2019. *Automating the News: How Algorithms Are Rewriting the Media*.
Harvard University Press.
- Göpferich, S. 2009. Comprehensibility assessment using the Karlsruhe Comprehensibility
Concept. *The Journal of Specialised Translation* 11, pp. 31-52.
- Koetsenruijter AWM. 2011. Using numbers in news increases story credibility. *Newspaper
Research Journal* 32(2): 74–82.
- Livingston, C. & Voakes, P.S. 2005. *Working With Numbers and Statistics: A Handbook for
Journalists* (1st ed.). Routledge. DOI: 10.4324/9781410612960.
- Meyer, P. 1975. *Precision Journalism: A Reporter's Introduction to Social Science Methods*.
Rowman & Littlefield.
- Nguyen, A. and Lugo-Ocando, J. (2016). The state of data and statistics in journalism and
journalism education: Issues and debates. *Journalism* 17(1): 3–17. DOI:
10.1177/1464884915593234.
- Thäsler-Kordonouri, S. & Barling, K. 2023. Automated Journalism in UK Local Newsrooms:
Attitudes, Integration, Impact. *Journalism Practice* DOI:
10.1080/17512786.2023.2184413.