Dynamics of Media Discourse: Timescales, Thematic Trajectories and Stochastic Models

Dario Zarcone* David Sánchez† Salvatore Micciché*

Abstract

Newspapers are among the most continuous records of historical change, yet those published before the general availability of the internet are rarely accessible in digital form, which limits quantitative work on historical media discourse. We use the complete archive of the Italian daily "La Repubblica" for the years 1985 to 2000, recovered from the CD-ROMs released in 2001 by reverse-engineering the format in which the articles were stored. The resulting corpus of about 600,000 articles covers a period of exceptional political and geopolitical change in Italy and beyond, from the final years of the First Republic through the Gulf and Kosovo wars.[1].

We first study the newspaper as a whole at different temporal aggregation scales, from daily to yearly. Time series of individual words show that the distribution of the Fano factor collapses across timescales once rescaled, so burstiness is a property of lexical dynamics, preserving the same structure at every scale. The temporal analysis in a latent semantic space reveals structure that aligns with the major events of the period and changes with the timescale of observation. The newspaper therefore appears as a sequence of regimes and transitions whose visibility depends on the timescale of observation.

We then move beyond a description of the corpus as a single object and analyse it separately for each newspaper section, representing each section in its own latent semantic space and tracking its evolution over time. While retaining the same lexical dynamics and the same pattern of rapid shifts between semantic zones during periods of change, the sections follow distinct trajectories, and the regimes and transitions found for the whole newspaper are not reproduced identically in every section.

We finally test whether these trajectories have a simple stochastic description. A random-walk description appears insufficient, while simple stochastic models account for much of the temporal dependence in the data without capturing it entirely. The results point to a media discourse organised both across scales and across sections, and to a temporal structure that such models reproduce to a large extent.

References

[1] D. Zarcone, S. Micciché, and D. Sánchez, "Detecting Historical Turning Points in Italian Media: A Complex Systems Approach to a Diachronic News Corpus," *arXiv preprint arXiv:2606.14348* (2026) (to be published in *Frontiers in Complex Systems*).

*Dipartimento di Fisica e Chimica Emilio Segrè, Università degli Studi di Palermo, Viale delle Scienze, Ed. 18, 90128 Palermo, Italy

†Institute for Cross-Disciplinary Physics and Complex Systems (IFISC), CSIC-UIB, 07122 Palma de Mallorca, Spain