

News Genres: the potential and problems of automatic genre classification for newspapers and television news

Kim Smeenk, Frank Harbers & Marcel Broersma (University of Groningen)

This paper discusses the potential and problems of automatic detection of genres in newspapers and television news through machine learning. Genre is an important characteristic of newspaper articles that can tell us a lot about the underlying discourse on what journalism is in a certain period and how it should be practiced according to journalists. As such, we study the historical development of genre in newspapers and television between the 1960s and 1995 to offer a novel approach to study how journalism developed in this period. This provides us with new insights in how Dutch journalism depillarized and moved towards critical autonomy. We first examine how genre has been conceptualized in journalism and television studies and how it can be operationalized for automatic classification via machine learning. Subsequently, we discuss the opportunities and pitfalls of our machine learning approach for media history research, and genre research specifically.

Genre and journalistic discourse

This paper is based on the research that is currently done in the CLARIAH research project on transparent automatic genre classification (NEWSGAC). NEWSGAC focuses on genre as a key category in the construction of social reality through the news. Genre can be defined as “language use in a conventionalized communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms” (Handford 2010, p. 258). Focussing on genre allows us to move beyond day-to-day news events and tap into the underlying structure of news coverage (Broersma 2007, 2010). Studying this historically sheds light on how journalism’s modes of expression have transformed and how professional norms and practices have emerged and evolved (Harbers 2014).

Classifying genre

Studying and comparing news genres in television news and newspapers from a historical perspective is a demanding task and comes with several fundamental challenges. Firstly, the sheer quantity of available material makes it a daunting task to do empirical research. This is exactly why researching automatic genre classification is such a valuable path to explore. Secondly, to examine genre over time and between different media, a standardized categorization of genre is necessary. However, genre is a typical example of a ‘latent category’, which needs considerable (human) interpretation. Genres are ideal-typical discursive constructs: their textual manifestations do not necessarily match the characteristics of these constructs perfectly, nor can they always be neatly delineated from other genres. It is even more challenging to analyse genres over time because they change or fade away while new genres do emerge (Bateman et al. 2014; Mittel, 2004). Finally, genres do not necessarily manifest themselves similarly in television news and in newspapers because they are closely related to the material and production characteristics of the medium. Genre is therefore operationalized differently in journalism studies and television studies.

Bibliography

- Bateman**, J.A., Evangelisti Allori, P. & Bathia, V.K. (2014). 'Introduction'. In: Evangelisti Allori, P., Bateman, J. & Bhatia, V.K. (eds.). *Evolution in Genre: Emergence, Variation, Multimodality*. Bern: Peter Lang. pp. 9-16
- Broersma**, M. (2007). *Form, Style and Journalistic Strategies. An introduction*. In: Idem Ed., *Form and Style in Journalism. European Newspapers and the Representation of News, 1880-2005*. Leuven, Paris and Dudley: Peeters, pp. ix-xxix.
- Broersma**, M. (2010). *Journalism as Performative Discourse. The Importance of Form and Style in Journalism*. In: V. Rupa Ed., *Journalism and Meaning-making: Reading the Newspaper*. Cresskill, N.J.: Hampton Press, pp. 15-35.
- Handford**, M. (2010). What can a corpus tell us about specialist genres. In: 'o Keeffe, A. & McCarthy, M. Eds., *The Routledge Handbook for Corpus Linguistics*. New York: Routledge.
- Harbers**, F. (2014). 'Between Personal Experience and Detached Information. The development of reporting and the reportage in Great Britain, the Netherlands and France, 1880-2005'. Groningen, s.i. [PhD Dissertation]
- Mittel**, J. (2004). 'A Cultural Approach to Television Genre Theory'. In: Allen, R.C. & Hill, A. (eds.). *The Television Studies Reader*. London & New York: Routledge. pp. 171-181.