

The Document as Event

Applying automatic semantic role labeling to collections of theatre reviews

Thomas Crombez

Royal Academy of Fine Arts Antwerp / St. Lucas
University College of Arts (Antwerp) / University of Antwerp
thomas.crombez@uantwerpen.be

When a theatrical performance is over, nothing remains but the memories of the participants and witnesses. At least, that is the well established cliché in the performing arts, and in theatre history as well. To investigate a theatrical event is, by definition, to investigate a historical event. It implies for that matter the study of the event's traces: the material record of costumes, props and set designs; the written scenario or script (when available); and the eyewitness accounts of performers, technicians and spectators.

For theatre historians, there is an important extra source: the critical record, consisting of reviews and essays. Historical documents of theatre criticism can bring a former performance context and its related practices (acting styles, physical communication, audience reactions) back to life. Only recently have such documents become accessible through large-scale digital collections.

This paper explores the possibilities of extracting information about historical performances from large collections of theatre reviews. In the literature, this is known as a subgenre of text-mining, namely, (semi-)automatic semantic role labeling (Gildea and Jurafsky 2002). The resulting data are an invaluable source of information for theatre history. Not only do they constitute a virtual record of past theatrical events, but they may also be used for studying processes of canonization (e.g., which names rise and fall in frequency?) and changing tendencies in the performing arts (e.g., the diminishing importance of the dramatist in the second half of the 20th century).

Two techniques from natural language processing and machine learning are crucial in the process of transforming plain-text documents into semantically marked-up data. First, all named entities should be detected and extracted from the text. Second, the extracted names have to be classified according to their specific role. In my paper, I will focus on detecting artistic roles of personal names in the context of the performing arts (i.e., roles such as dramatist, director, performer, scenographer, or critic), based on the surrounding snippets of text. Further, I will demonstrate the results and possible use-cases based on mining documents from different collections of theatre reviews, coming from Belgian newspapers in French and Dutch from the 1989–2014 period.

References

Gildea, D. and Jurafsky, D. (2002) 'Automatic Labeling of Semantic Roles', *Computational Linguistics*, vol. 28, 3, pp. 1–45.