

Text Zoning of Theater Reviews: How Different are Journalistic from Blogger Reviews?

Mylène Maignant¹, Gaëtan Brison², Thierry Poibeau¹

¹Laboratoire LATTICE (CNRS & ENS-PSL & Université Sorbonne nouvelle)
1, rue Maurice Arnoux, 92120 Montrouge, France

firstname.lastname@ens.psl.eu

²Institut Polytechnique de Paris, 5 Av. Le Chatelier, 91764 Palaiseau

gaetan.brison@ip-paris.fr

Abstract

This paper aims at modeling the structure of theater reviews on contemporary London performances by using text zoning. Text zoning consists in tagging sentences so as to reveal text structure. More than 40 000 theater reviews going from 2010 to 2020 were collected to analyze two different types of reception (journalistic vs digital). We present our annotation scheme and the classifiers used to perform the text zoning task, aiming at tagging reviews at the sentence level. We obtain the best results using the random forest algorithm, and show that this approach makes it possible to give a first insight of the similarities and differences between our two subcorpora.

1 Introduction

Since 2010 in England, a wave of blogs written by authors coming from various horizons has arisen on the Internet. Students, theater professionals but also mere amateurs began publishing their own theater reviews. These new independent voices in the digital space progressively redefine the shape of classic journalistic criticism. Although discreet, they offer a new vision of the history of Londonian theaters. By doing so, it sets itself apart from the canon of mainstream journalism.

The emergence of this digital culture triggered a lot of controversies on the status of the review as a literary object. Michael Billington, reviewer for *The Guardian* since 1971, states that a blog ‘is more like an informal letter: a review, if it’s to have any impact, has to have a definable structure.’ For Danielle Tarento, co-founder of the Menier Chocolate Factory, ‘a lot of people [bloggers] are not ‘proper writers.’ At the other hand of the spectrum, some of these bloggers claim the stylistic singularity of their publications. In the description of Exeuntmagazine.com for instance, the editors

claim that: ‘Exeunt believes in making beautifully written, *experimental*, fierce and *longform* writing about theatre available for free.’

A review is traditionally organized according to several sections: an introduction, a presentation of the plot, a few lines on the stage, etc. In order to compare the two subcorpora, it is first necessary to segment the reviews into textual zones corresponding to these thematic sections. We assume the two subcorpora will share the same zones, as they are all about theater, but the content of the zones may differ from one subcorpus to the other: e.g., the two communities may not focus on the same aspects of the plays. From a technical point of view, this experiment is also an opportunity to test the robustness and relevance of text zoning across different domains. Text zoning has been mainly used to segment scientific texts so far, but can this technique also be used in the humanities? Can it be used for performance reviews, where critics do not follow a fixed structure, contrary to scientific writing?

This short paper is structured as follows. We first give a brief overview of text zoning. We then present our corpus, the different features and machine learning techniques used for the task. We then comment our results and give some hints on the way these could be used to get a better understanding of the content of the corpus and the differences between the two communities at stake (official critics vs amateur bloggers).

2 Previous Work

The notion of text zoning was first introduced by Simone Teufel in her PhD (Teufel, 1999). Teufel was targeting the automatic analysis of scientific papers. In this context, argumentative zoning refers to the ‘rhetorical status of a sentence with respect to the communicative function of the whole paper.’ It is for example quite useful to distinguish ‘back-