the authors who have their reviews published on *MyTheatreMates* share the following characteristics: They have their own personal website, they post original theatre-related content on their personal website at least once a fortnight, they can provide three professional arts references (e.g. artists they have interviewed or, if they review, producers or publicists who already regularly provide them with complimentary press tickets to shows) and they are active on Twitter.

When this subcorpus was created (September 2020 – April 2021), 52 bloggers were members of *MyTheatreMates*. We selected the blogs which had the highest number of reviews (at least 200 reviews) as well as the ones which were mainly focusing on the Londonian stage.

3.3 Overview of the Corpus

	Newspapers	Blogs
Number of sources	33	18
Number of words	8,831,160	10,364,855
Number of reviews	22781	19045

Table 1: The Descriptive statistics of each corpus (source refers to newspapers vs blog platforms).

Table 1 gives an overview of the two datasets. The corpus is available in textual format (PDFs from *Theatre Record* have been converted and manually corrected) so that NLP tools coming from Stanford could be directly applied. It is to our knowledge the first corpus collecting so many reviews of theater performances. The corpus is freely available online, on the website dedicated to this project: `Dramacritiques.com`.

4 Experiment Description

4.1 Annotation Scheme and Data Labeling

Once the data were collected, the first step consisted in labeling a random sample of reviews that could be used for training. The annotation scheme corresponds to the 8 different possible sections of a review.

The definition of these sections is based on (Fisher, 2015). In his analysis, Fisher examines the various possibilities for one critic to structure his arguments, which leads to the following 8 different categories with 8 different colours:

For this first experiment, the data were labeled by an expert with a strong background in theatre studies. This expert spent more than 15 minutes per review, or 250 hours in total, annotating the sample. Each of the sentences was carefully analyzed to propose the best category it belonged to.

Zone category	Associated colour
Introduction	Purple
Reviewer analysis	Blue
Visual and audio details	Green
Conclusion	Yellow
Performance of actors	Orange
Plot	Red
Structure of the play	Brown
Related to the audience	Grey

Table 2: Delimitation of the different zones and their colours used in the model.

However, some of the sentences could have been classified in two different categories. These cases were recorded and resolved following explicit rules to ensure the consistency of the annotation. 1000 reviews were manually annotated, which was deemed enough for training.

4.2 Data Preparation

Several preprocessing steps were applied to the corpus, following previous experiments in text zoning. Texts were first segmented into sentences, tokenized and tagged (with POS and morphological features) and empty words were removed. Named Entity Recognition and Term Frequency-Inverse Document Frequency (TF-IDF) were also applied on the corpus. Annotations were performed using Stanford tools and were then used as features for training.

In the end, more than eighty variables were created, following previous work in the domain (among others (Teufel, 1999) and (Guo et al., 2011)):

- Statistical variables: average word length, average sentence length, frequency of personal pronouns, etc.
- Tense variables: proportion of verbs in future, present and past tenses
- Grammar variables: top verbs, adjectives, superlatives, nouns, etc.
- Parts of Speech variables: position of the words and their roles in the sentences