RAP AND POETRY

Elsa Bidant

Université Paris Sciences et Lettres

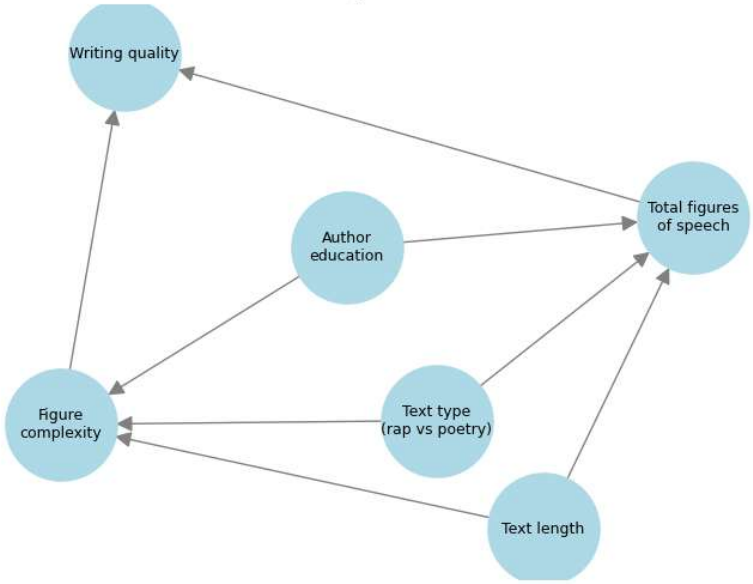
The full project report is available here: https://github.com/elsabidant/Rap-and-poetry

I. Introduction

Rap music is often perceived as poorly written and lacking literary value. This negative image is reinforced by the media's tendency to highlight only the most controversial aspects of rap lyrics, rather than their complexity and depth (Parmar, 2005). However, several recent studies challenge this perception. Rap can in fact be understood as a "poetry of performance," rooted in orality, rhythm, and social commentary (Nachtergael, 2021). Moreover, the literary quality of rap is confirmed by its frequent intertextual dialogue with canonical authors such as Charles Baudelaire, whose poetry is sampled, reinterpreted, and transformed by rap artists to assert new poetic identities (Rolland, 2021).

Considering these insights, it becomes particularly interesting to compare rap lyrics with traditional poetry. To what extent can rap lyrics be considered as rich and

Figure 1



sophisticated as traditional poetry in their use of literary devices? By analysing the use of figures of speech in both forms, this project aims to evaluate whether rap lyrics are as well-crafted, less sophisticated, or even richer than traditional poetic texts.

***Figure 1** : Directed acyclic graph representing the hypothesized causal links between text type (rap or poetry), total of figures of speech by text, writing quality, author education and text length.*

Following various readings about rap and poetry and my personal experience as a listener, I have made the following predictions: (1) rap lyrics have, at least, as many figures of speech as traditional poetry texts and (2) rap lyrics have, at least, as sophisticated figures of speech as traditional poetry texts (see **Figure 1**).

II. Data

This project relies on two main datasets. First, rap lyrics were collected from the LRFAF corpus (*Une exploration numérique du rap français depuis les années 1990*, Courson, 2024), which gathers 37,306 French rap texts along with metadata such as the artist's name and the specific sub-genre. Second, poetry texts were collected by scraping the

Figure 2

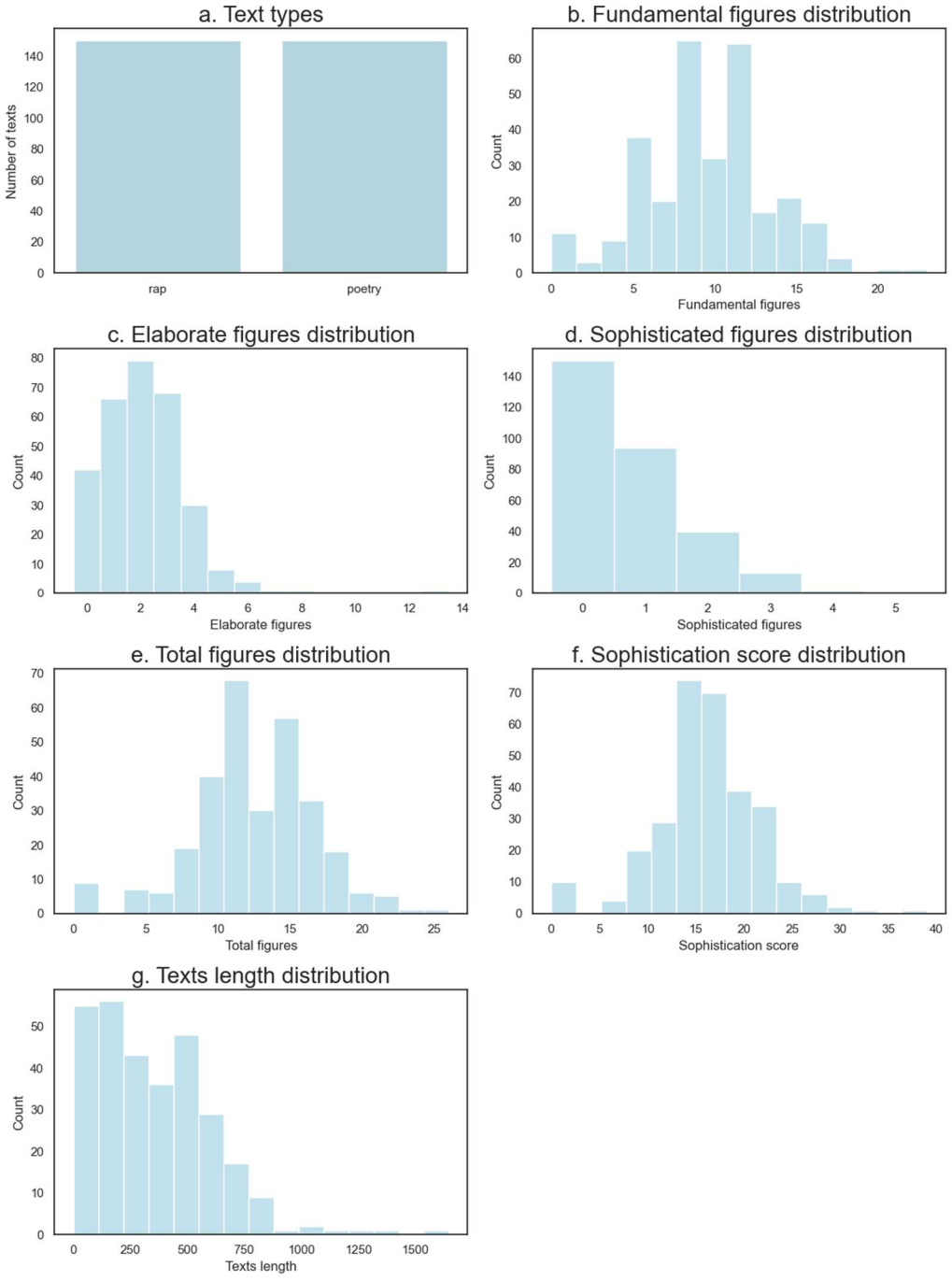


***Figure 2** : Representation of*

a. Text type

b. Fundamental figures distribution

c. Elaborate figures distribution

d. Sophisticated figures distribution

e. Total figures distribution

f. Sophistication score distribution

g. Texts length distribution

Poetica website (https://www.poetica.fr/), resulting in a corpus of 3,165 poems, each associated with its author and date of publication.

From these two large datasets, a random sample of 150 rap lyrics and 150 poems was drawn for detailed analysis. Each text was then processed using a Large Language Model (LLM) to automatically detect and extract all the figures of speech present.

The final dataset I use therefore contains 150 rap lyrics and 150 poems (N= 300), with the following information for each entry: the author, the date of publication or release, the full text, the identified figures of speech, the total number of figures of speech detected (see **Figure 2**). Moreover, figures of speech were classified in 3 categories depending on their level of complexity: fundamental (or simple), elaborate (or average) and sophisticated (or complex).

III. Method

To test my predictions, I performed two multivariate linear regressions. The first models the total number of figures of speech in each text, based on two explanatory variables: the type of text (rap or poetry) and the length of the text (measured in number of words). The second uses the same predictors to model the sophistication score, for text *i*, computed as follows:

$$Sophistication\ score_i = (number\ of\ fundamental\ figures_i) * 1 + (number\ of\ elaborate\ figures_i) * 2 + (number\ of\ sophisitcated\ figures_i) * 3$$

Both regressions use the following formula:

$$Y_i = \beta_0 + \beta_1 Text\ type_i + \beta_2 Text\ length_i + \varepsilon_i$$

With:

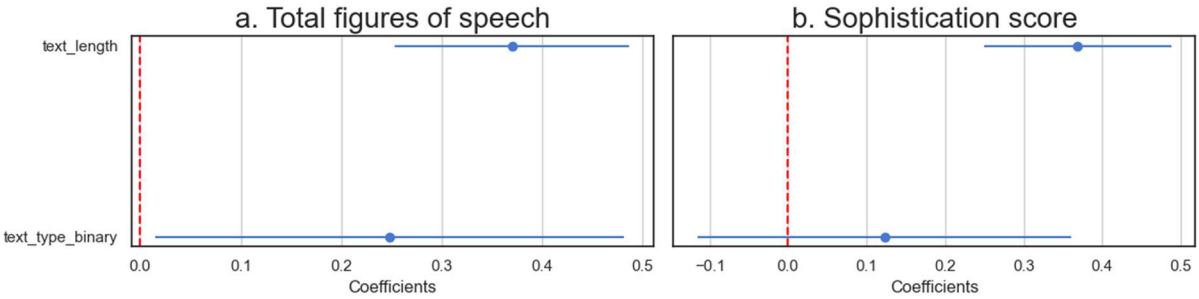
Y_i	Either the total number or the sophistication score of figures of speech for text *i*
$Text\ type_i$	Text type for text *i*, where 0 = poetry and 1 = rap
$Text\ length_i$	Number of words in text *i*
ε_i	Error term for text *i*, drawn from a normal distribution

IV. Results

The first model (R^2 = 0.197) indicates (see **Figure 3**) that text length has a positive and significant effect on the total number of figures of speech (β = 0.370, $p < 0.001$), while text type has a positive and significant effect (β = 0.248, p = 0.037). The second model (R^2 = 0.162) indicates that text length also has a positive and significant effect on the sophistication score (β = 0.369, $p < 0.001$), and that text type has a positive and non-

significant effect (β = 0.123, p = 0.312). In both cases, the R^2 values are lows, suggesting that the models explain a modest proportion of the variance. This indicate that while text

Figure 3



length and type contribute to the outcomes, other unmeasured factors likely play a substantial role.

V. Conclusion

In line with my initial predictions, the results indicate that rap lyrics contain more figures of speech than traditional poetry texts. But they do not use significantly more sophisticated rhetorical devices. These findings suggest that while rap may exhibit a higher density of stylistic figures, the difference in sophistication is less robust.

It is important to note, however, that these findings depend on how literary sophistication is operationalised. The model used here captures quantity and complexity of figures of speech but does not take into account the sound patterns or musicality that are important components of rap and poetry texts. An analysis using more powerful LLMs would provide a more detailed understanding of the qualitative differences between rap and poetry texts.

***Figure 3**: Representation of a. Regression coefficients for total figures of speech, b. Regression coefficients for sophistication of figures of speech.*

Moreover, potential confounding factors such author education were not tested (due to a lack of data) and could partially explain the observed differences. If the author of a text has received a better education than another author, it is possible that he or she will use more figures of speech and that these will be more sophisticated.

VI. References

Rolland, Nina. « Dans la rue et le bitume ne poussent que les fleurs du mal : jouer et déjouer Charles Baudelaire et sa poésie dans le rap français », *Itinéraires. Littérature, textes, cultures*, 2020-3 | 2021, p. 1-10. DOI : 10.4000/itineraires.9377. URL : https://journals.openedition.org/itineraires/9377

Parmar, Priya. « Cultural Studies and Rap: The Poetry of an Urban Lyricist », *Taboo*, Spring-Summer 2005, p. 5-14. URL: https://www.academia.edu/4064933/Cultural_Studies_and_Rap_The_Poetry_of_an_Urban_Lyricist

Nachtergael, Magali. « Le rap, une poésie de performances », *Itinéraires. Littérature, textes, cultures*, 2020-3 | 2021, p. 1-6. DOI : 10.4000/itineraires.9369. URL : https://journals.openedition.org/itineraires/9369