

# A Computational Analysis and Exploration of Linguistic Borrowings in French Rap Lyrics

Lucas Zurbuchen

Northwestern University

lucaszurbuchen2024@u.northwestern.edu

Rob Voigt

Northwestern University

robvoigt@northwestern.edu

## Abstract

In France, linguistic borrowings in the relatively conservative French language are an important site of cultural debate, and rap in particular is a hotspot for borrowings. In this work, we use computational methods to understand the factors that affect the prominence and prevalence of a borrowing. To do so, we manually annotate a lexicon of over 700 borrowings occurring in this context (including key aspects for each borrowing such as origin and semantic class). We analyze the prevalence of these borrowings in a newly collected corpus of over 8000 French rap song lyrics and find that there are increases in the proportion of linguistic borrowings, interjections, and Niger-Congo borrowings while terms related to the arts are decreasing in prevalence. We release our code and data to facilitate further research in this area and discuss potential future directions.

## 1 Introduction

Ever since its origin, the musical genre of rap has changed the languages it has touched, starting with placing several African American Vernacular English (AAVE) terms into the lexicons of many other English speakers (Lewis, 2023). The popularity of rap music has allowed both the genre (and opportunities for linguistic borrowing through it) to go to countries far beyond the United States, like the Western European country of France.

There, linguistic borrowing is especially intriguing, as it is particularly prone to language change through Anglicisms from the United States, linguistic minorities from countries affected by European colonialism (Paine, 2012), and European linguistic communities being in close proximity to each other – Verbeke (2017) highlights that near half of France’s and Belgium’s immigrant population comes from other European Union countries. There, the usage of linguistic borrowings in the French language is a socially complex issue. Some

Wesh, cette bitch veut mon corps, pourtant je sue comme un sumo

“Yo [Arabic borrowing], this bitch [English borrowing] wants my body, yet I’m sweating like a sumo [Japanese borrowing]”  
From “Intro (Introduced by Caspi)” by JMK\$ & Beamer

À 14 ans dans la tess, igo c’était gore  
“In the streets [Verlan] at 14 years old, dude [Spanish borrowing] that was gory”  
From “Des Nomes” by Fresh laDouille

C’est la crise au mic, fait la bise au mac  
“It’s a crisis at the mic [English borrowing], give the pimp [Argot] a kiss”  
From “Oh mama oh” by Le Classico Organisé

Figure 1: Three examples of lexical borrowings and other *argot* from the corpus of French rap lyrics collected for this work.

people and organizations value linguistic preservation and resist this change, like a forty-member organization having close ties to the French government called the *Académie Française (AF)* (Estival and Pennycook, 2011). They try to act as a “guide” for French speakers, often resisting foreign borrowings in the process (Estival and Pennycook, 2011). However, French rap poses challenges for these types of organizations, reflecting borrowings into the language that occur in different contexts of French society. Linguist and public speaking teacher Julien Barret states that linguistic borrowings through French rap have become so widespread that students he works with sometimes don’t know if they learned about a borrowed word in a rap song or from their neighborhood (Rhissi, 2021). French rap artists also help influence the popularity of various types of French *argot*, or slang, like Verlan – inverting syllables of French words to create new unusual-sounding ones, often due to motivations of identity and power to go against the status quo (Westphal, 2013). Political motivations, especially among more marginalized communities, are not uncommon in French rap; most French rappers stem from “black African,

Caribbean and North African" immigrants who use their music to highlight inequality towards their ethnic groups (Bretillon, 2014). This gives linguistic borrowings the ability to reach masses of new people alongside other political messages or implications.

As a result, what about a linguistic borrowing in French rap could make it more well-used, and thus able to reach more people, than others? In other words, it would be valuable to see which types of borrowings the French rap community can become commonplace while which types of borrowings organizations like the AF can hold back for the sake of linguistic preservation. This work will explore both if and when certain borrowings are more popular than others, contributing the following to existing research:

- Bridging the gap between related technical research, social research, and existing research on French language change through rap.
- Leveraging computation to generate corpora of rap songs and linguistic borrowings which can be used both in and beyond this paper.
- Exploring trends in the overall usage and temporal usage of linguistic borrowings, providing direction on what future research could address.

## 2 Background and Related Work

This paper aims to bridge the gap between the computational literature on temporal trends in word popularity, the causes and impact of linguistic borrowing in rap songs, and lexical borrowings in French specifically.

A substantial body of prior computational work has targeted the question of identifying when linguistic borrowings occur in context – lexical borrowing identification. Approaches to this task generally involve wordlists for all the donor and target languages and pre-processing the text to computationally search for deviations in the target language’s sound patterns. For example, Tsvetkov et al. (2015) converted Swahili text to the International Phonetic Alphabet to find derived words from Arabic, Mæhlum and Ivanova (2023) analyzed phonotactic patterns in the Siberian language of Sakha to uncover borrowings from Russian, and Miller and List (2023) needed a wordlist for every language when searching for Spanish borrowings in

indigenous Central and South American languages. However, this means that examining linguistic borrowings from as many languages as possible using these techniques quickly becomes a difficult data collection and annotation task.

Another approach to detecting lexical borrowings in French rap could be slang detection because many lexical borrowings in French rap act as slang. Approaches like Pei et al.’s (2019) using Bidirectional Long Short-Term Memory (Bi-LSTM) to detect English slang have been relatively effective, though these models are more accurate at identifying the presence of slang in text than identifying the particular slang word. Still, this method could detect French slang terms that are not linguistic borrowings, like Verlan (a type of French *argot* where syllables are inverted – “bonjour” becomes “jourbon”).

Another strain of computational research more directly related to our work examines factors influencing the popularity of lexical borrowings. One approach has been to study the survival of words in a language (regardless of if it is a lexical borrowing) in a natural selection lens, like WordWars, which found that the word’s morphology was the most important factor to its survival (with word length following) (Mohammad, 2020). Works studying changes in word meaning like Hamilton et al. (2016) are also related to our work despite not being its main objective. Keidar et al. (2022) took the approach of studying exclusively slang words relative to non-slang counterparts, finding that simply the word’s status as a slang word was the largest determining factor of its popularity. In the lens of specifically linguistic borrowings, most of the work outside of English is done on Anglicisms, like with Alvarez-Mellado’s (2020) and Stewart et al.’s (2021) work on Anglicisms in Spanish.

In the French context in particular, Chesley and Baayen (2010) studied lexical borrowings in two corpora of French newspapers 10 years apart, examining the factors determining its entrenchment into French. They found dispersion (a measure of a word’s spread in text and not simply its quantity) was a better indicator of its longevity in French than frequency, though both were important factors (Chesley and Baayen, 2010). This highlights the need to examine both, even though frequency is a more interpretable metric to analyze. Chesley and Baayen (2010) also found that languages other than English were less likely to be borrowed, but it

would be interesting to analyze this in more detail in this work through collecting borrowings with many different origins.

Qualitative research on rap lyrics demonstrates that France is not the only Western European country that has rap artists eager to spread social awareness about certain linguistic groups. Arabic has a foothold in German rap, though much of it has to do with expressing Muslim identity (Hebblethwaite, 2018). In Spain, a Galician rap group uses their language to address local issues while using Spanish and English to rap about more global ones (Loureiro-Rodríguez, 2013). Lastly, identity and anti-imperialism are a large topic of discourse in Portugal’s Kriolu rap (a Portuguese-based creole from Cape Verde) (Pardue, 2012).

To our knowledge, our study is the first to apply a computational approach to ask large-scale questions about the popularity of borrowings into music in the rap domain specifically.

### 3 Methods

#### 3.1 Data Collection

We collected a dataset of French rap lyrics, manually curated a lexicon of linguistic borrowings, and then created a dataset recording metrics of the word’s usage in the lyrics (mainly raw word count but also the number of songs and artists’ discographies that the borrowing appeared in) for each year from 1991 to 2024.

##### 3.1.1 Corpus Collection

We collected a large corpus of rap songs by querying the Spotify<sup>1</sup> and Genius<sup>2</sup> APIs for lyrics. There are existing corpora of French rap lyrics like RapCor<sup>3</sup>, but this corpus has a relatively small number of songs (1,360) and has not been updated in at least a year. Searching for songs in the Spotify API started with searching for a select number of songs within the genre of French rap (specifically querying the genre "rap français" in each request). Because Spotify’s API limits the songs one can receive in a single API request, we implemented a recursive search where we queried on the artists from a prior request (still with the constraint of the French rap genre) until we reached a depth of

15 in the search tree. Something important to consider was that Spotify’s API tended to oversample newer songs over older ones (given the increase of the popularity both of French rap since the late 20th century and of Spotify in general), but simply sampling as many songs as possible with this approach mitigated this. Nonetheless, for each song, we sampled the song name, artist(s), and release date. For each song received from Spotify’s API, we queried the Genius API to find its lyrics, filtering out noise by verifying that songs returned by the Genius API were the same as those received from the Spotify API and that the lyrics of songs were predominantly in French (using the Python language detection library Lingua<sup>4</sup>). This resulted in a dataset of 8,222 French rap lyrics from 1991 to 2024 for analysis.

Qualitative examination shows that borrowings are prevalent throughout the corpus, with three examples in Figure 1 to provide some context. Even though the research question remains only about linguistic borrowings, it’s useful to examine that these are not the only linguistic innovations in French rap – Verlan and other types of *argot* exist as well.

##### 3.1.2 Linguistic Borrowing Lexicon Collection

We manually curated a lexicon of 741 linguistic borrowings occurring in our French rap lyrics dataset.

To do this, we employed three main strategies. Our borrowings list is predominantly drawn from Wiktionary, where we exhaustively collected the terms listed on the Wiktionary page for French terms derived from other languages<sup>5</sup>. Second, we randomly sampled 20 songs from the corpus and manually examined all words in each song that did not occur in a French wordlist<sup>6</sup> since terms outside of a basic dictionary wordlist may be more likely to be borrowings. Finally, for each term collected by the above two strategies, we queried MUSE embeddings<sup>7</sup> to calculate nearest neighbors with a given threshold of cosine similarity to evaluate whether those terms should also be included.

We then used Wiktionary and its French counterpart Wiktionnaire<sup>8</sup> to manually annotate each borrowing entry in our lexicon for the origin of

<sup>1</sup><https://developer.spotify.com/documentation/web-api>

<sup>2</sup><https://docs.genius.com/#/getting-started-h1>

<sup>3</sup>[https://is.muni.cz/do/phil/Pracoviste/URJL/rapcor/index\\_en.html](https://is.muni.cz/do/phil/Pracoviste/URJL/rapcor/index_en.html)

<sup>4</sup><https://github.com/pemistahl/lingua-py>

<sup>5</sup>[https://en.wiktionary.org/wiki/Category:French\\_terms\\_derived\\_from\\_other\\_languages](https://en.wiktionary.org/wiki/Category:French_terms_derived_from_other_languages)

<sup>6</sup><https://github.com/Taknok/French-Wordlist/tree/master>

<sup>7</sup><https://github.com/facebookresearch/MUSE>

<sup>8</sup><https://fr.wiktionary.org/wiki/Wiktionnaire>

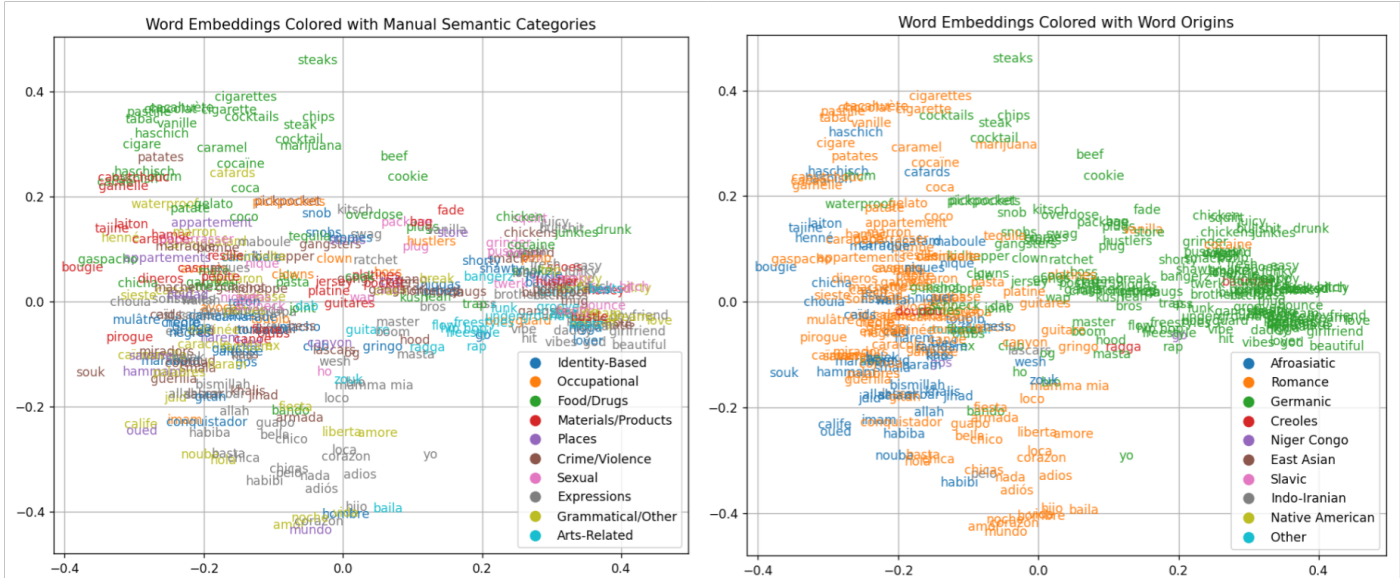


Figure 2: Collected borrowings embedded with MUSE embeddings, projected onto 2 dimensions, and colored based on manually-annotated semantic class (left) and origin (right).

Table 1: Basic occurrence statistics for linguistic borrowings from our wordlist in the collected corpus.

| Maximum | Minimum | Mean  | Median | STD   | IQR |
|---------|---------|-------|--------|-------|-----|
| 10362   | 0       | 133.3 | 28     | 525.3 | 71  |

the borrowing (a categorical factor, also known as the donor language), its part(s) of speech, and its semantic class. For annotating semantic class, we first experimented with clustering methods on embedding representations from MUSE and Urban Dictionary-based embeddings from Wilson et al. (2020); however, we found that in static embedding-space methods, borrowed words tended to cluster much more clearly along the lines of linguistic origin rather than meaningful semantics, as can be observed in Figure 2.

Moreover, polysemous words in the rap context tend to have a predominant sense that may differ from their most common sense in other contexts, leading to inaccurate clustering. For example, the AAVE-derived borrowing "beef" refers to a feud but is clustered near other types of food and drugs (in the upper left corner). Existing ontologies like WordNet were another alternative, but we found that hierarchies combined with polysemy would interfere with the interpretability of the analysis and might not be as fine-tuned to newer and less common borrowings into French. Therefore, we manually annotated each word with a semantic class

from an ontology we developed for this context:

- Referring to a certain identity of people
- Person – occupational
- Food/drinks/drugs
- Other inanimate material/product
- Places
- Events/materials related to conflict/crime
- Sex/sexual connotations
- Common exclamations/expressions
- Common usage/grammatical function/other
- Related to music/other arts

### 3.1.3 Recording Temporal Word Usage

We calculated word usage (through metrics of raw word count, song usage, and artist usage) for each borrowing in French over time via string matching, iterating through the entire corpus to find all instances of the word as a token separated by whitespace or punctuation. Additionally, to account for the variations in spelling that occur when transcribing lyrics, we augmented our borrowing list with all inflections and alternate spellings of each borrowing that we counted in our corpus at least 5 times. We recorded the release date of each song at a granularity of a year in order to identify broader temporal trends. At the end of the process, one gathers data denoting the desired word usage metric (like raw word count) for each year for each collected instance of linguistic borrowing. There were



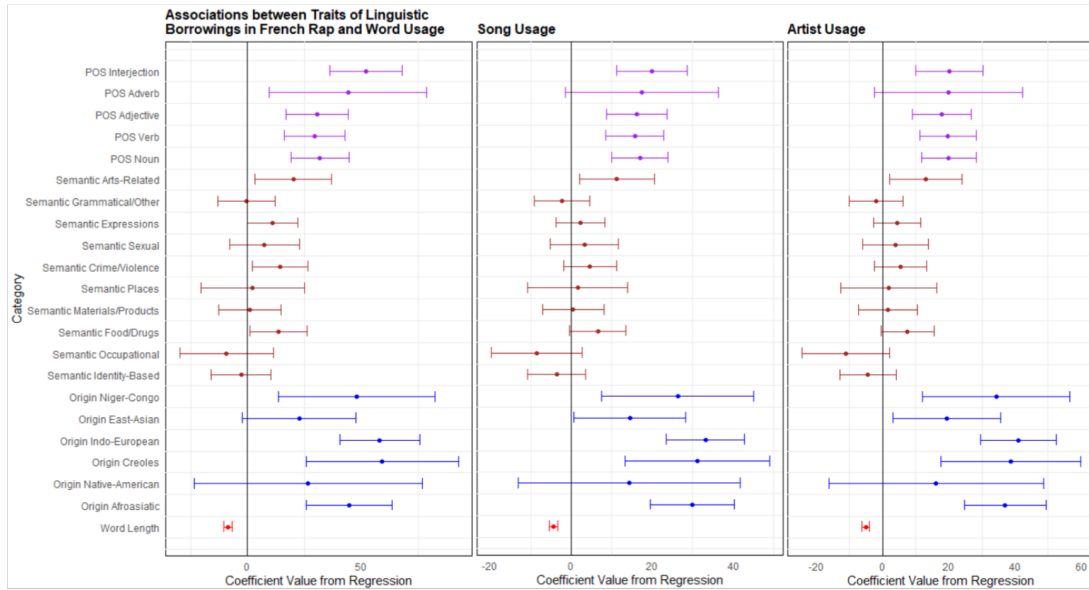


Figure 3: A plot of all variables and their coefficients from the RLM regression run on all collected borrowings (with a 95% confidence interval) with an output (left to right) of the borrowing’s raw word count, number of songs used in, and number of appearances in artists’ discographies.

4,563,577 total tokens in the lyrics, with 98,740 instances of borrowings identified, and Table 1 shows basic statistics about the distribution of raw word counts for recorded borrowings.

### 3.2 Data Analysis

We analyzed the data in two ways. First, we identified associations between the overall frequency of usage of a borrowed word and our four independent variables by using a regression-based approach. Then, we explored temporal trends in the usage of different types of linguistic borrowings in French rap by fitting smoothed curves to the data.

In the first case, we ran a Robust Linear Model (RLM) regression on the collected data. Using this model specifically instead of standard linear regression was beneficial mainly because of the data’s potential sensitivity to outliers, which is difficult to prevent (because of factors like smaller sample sizes in borrowings from less common languages or in lyrics from earlier rap songs). For each recorded linguistic borrowing, the four parameters that served as independent variables for the regression were its length (in characters), origin of borrowing, semantic class(es), and part(s) of speech, with all except the first being dummy-coded as categorical variables. We ran three models to examine the impacts of these variables on three different operationalizations for frequency of usage: the total number of times that the word appeared

in the collected lyrics, the total number of songs containing the word, and the total number of artists having the word in their lyrics. To reduce the number of potentially noisy dimensions in the model, we removed parameters corresponding to language groups with an insufficient number of samples (less than 5 in this case) before running the model. We used a standard significance threshold of  $\alpha = 0.05$ .

For the second method we visualized usage of linguistic borrowings over time with smoothed lines of best fit. Specifically, the main dependent variable of interest was the proportion of borrowings of the said category relative to all borrowings from the collected data (based on the metric of raw word count). We used both Linear Model (LM) and Locally Estimated Scatterplot Smoothing (LOESS) methods to find correlations, using the LM method for relationships with a stronger observed linearity.

## 4 Results

### 4.1 Regression Analysis

Figure 3 shows a forest plot for the output of the three regression models we ran, which identified several interesting trends. Perhaps the clearest relationship was with word length, with there being a significant negative association between the length of the word and all three word usage output variables. This is both intuitive and confirms trends identified in prior work such as [Mohammad](#)

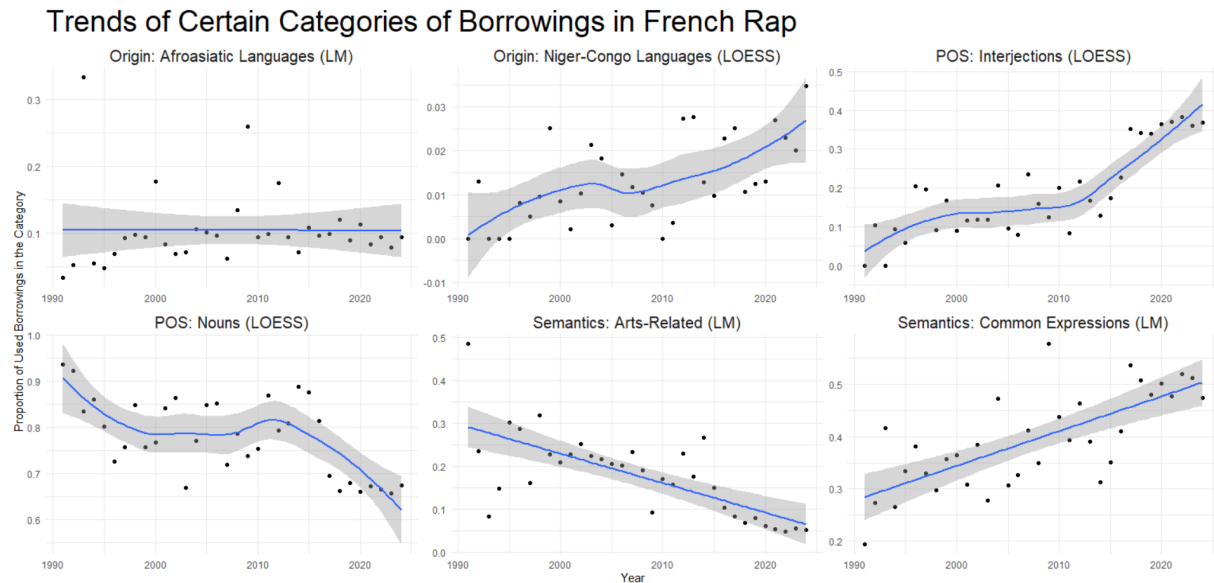


Figure 4: Using the LM and LOESS methods from ggplot’s `geom_smooth`, we find that the use of Niger-Congo languages, interjections, and common expressions as borrowings have grown over time, Afroasiatic languages as borrowings has remained consistent, and nouns and arts-related terms as borrowings have declined over time.

(2020).

The Afroasiatic, Creoles, Indo-European, and Niger-Congo language families all showed positive associations with uptake. Only one semantic class had a significant positive coefficient in all regressions – terms related to music and the arts. This suggests that in general borrowings related to these semantic classes that enter French rap lyrics are more likely to receive uptake than those from other semantic classes. Furthermore, all parts of speech except for adverbs had significantly positive correlations in the regressions, though all the confidence intervals overlap with each other, so the prevalence of borrowings of one part of speech over another is inconclusive with the regression models.

The trends identified were broadly similar between the word-, song-, and artist-level models, suggesting there are not dramatic differences between these three operationalizations of usage for borrowing.

## 4.2 Temporal Analysis

Examining the proportion of all collected linguistic borrowings relative to every word in the lyrics provided some interesting findings (see Figure 5). We find that the overall usage of all borrowings in general has been increasing over time – essentially doubling since the 1990s to a current level of 2% of all words in the lyrics. We use this relative proportion to normalize all other figures because it prevents any false interpretation of results that

simply reflect this overall trend.

Examining languages of origin this way, Niger-Congo and Afroasiatic languages exhibit interesting trends. Borrowings from Niger-Congo languages increased the most rapidly, getting gradually more popular up to the present day (see top-middle in Figure 4). A possible explanation for this is that Sub-Saharan Africa regions are front-runners in population growth (Uni, 2019), which includes regions that France historically colonized like the Democratic Republic of the Congo. On the other hand, borrowings from Afroasiatic languages have stayed both substantial (at around 10% of total borrowings) and consistent over time (see top-right in Figure 4), suggesting that Afroasiatic languages, many of which are from Arabic, have been a staple in linguistic borrowing usage in French rap since the beginning.

With semantic classes, the largest finding was that arts-related terms have had a linear decrease as a proportion of all borrowings (see bottom-middle in Figure 4) since the start while common expressions have had a linear increase (see bottom-right in Figure 4). This potentially indicates a major shift in song topics or even style since then.

Another trend we identify that could indicate a stylistic change is that the proportion of borrowings that are interjections has been increasing rapidly after 2010 (see top-right in Figure 4) while the inverse has been happening to nouns (see bottom-left in Figure 4). This could be because many interjec-

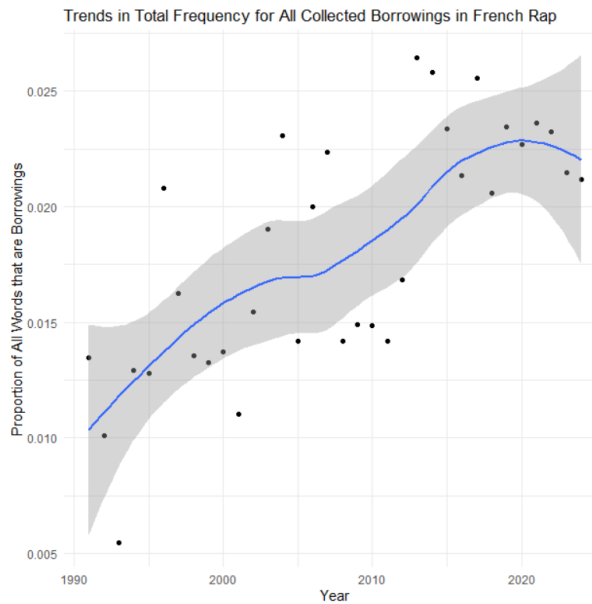


Figure 5: The overall frequency of linguistic borrowings in French rap is increasing over time (using LOESS method from ggplot’s `geom_smooth`).

tions in rap songs, at least in the United States, act as ad-libs, whose role has become steadily more important since 2010 for new subgenres of rap like mumble rap (Waugh, 2020), so it would not be surprising if this stylistic trend moved over to France as well.

## 5 Discussion

This paper finds that a linguistic borrowing’s length is a likely determining factor in its popularity while the origin, semantic class, and part of speech of a borrowing can all influence its overall usage. Furthermore, certain categories of borrowings have experienced both ups and downs in their popularity over the years relative to each other, hinting at trends involving demographics or musical style. Most importantly, the number of borrowings has been increasing over time, showing the increasing influence that linguistic borrowings are having on the French language. All code and data are available for open-source use on this project’s GitHub repository<sup>9</sup>.

We found these results by collecting over 8,000 songs with Spotify and Genius and analyzing the trends of the usage of 700 words over time with regression models and temporal analysis. These findings are compelling for several reasons. Firstly, they illustrate the dynamic linguistic environment

of French rap, showing the influences of both linguistic minorities and organizations like the *Académie Française* on the French language as it is. It also studies linguistic borrowing with a large breadth and depth on the donor languages that make their way into French rap lyrics. Lastly, French rap provides a small window into how music can change the languages it encounters in a region undergoing demographic change.

### 5.1 Limitations

This research has a number of inherent limitations. One limitation is that our lexicon of borrowings was manually curated and therefore potentially limited the scope of words that could be analyzed. We found that computational lexical borrowing detection methods, especially for French in particular, were not sufficiently robust to be used off-the-shelf, and manual annotation was necessary for our key variables. Determining the primary origin of a borrowing was sometimes difficult due to complex etymologies, so models that can do this accurately in future work may provide a path for greater scalability. Annotation was done by the first author, who is a native speaker of English rather than French and born after 2000, making it possible that biases towards Anglicisms and newer words entered the labeling process in spite of our efforts to have consistent and objective labelling protocols.

Because of rap’s growth in France, fewer songs were available for older dates than for newer ones. This meant our data was noisier in the earlier years than later ones, preventing us from reliably pooling word usage metrics in a granularity of less than a year. Another consideration is the use of the Spotify and Genius APIs, which are potentially imperfect mechanisms for obtaining a balanced sample of French rap. The collection of a song’s lyrics in this paper occurs under the assumption that a song is recorded both on Spotify and Genius, which removes any songs that are not on the streaming platform or that got removed from it, which also likely disproportionately affects older songs.

Lastly, the use of Spotify’s API made it more challenging to evaluate song popularity (a big variable in evaluating the popularity of a borrowing), as it only outputs song popularity on a scale from 0 to 100. We attempted workarounds like calculating video viewership of songs through YouTube’s API<sup>10</sup>, but this seemed error-prone, so we leave

<sup>9</sup><https://github.com/ljz112/CLResearch>

<sup>10</sup><https://developers.google.com/youtube/v3>

the task of identifying relationships with external measures of popularity for future work.

## 5.2 Future Directions

There are many possible ways to build on this research in the future. As mentioned above, one direction could be to analyze word usage over time relative to external information about the French rap songs in which they are used. For example, it would be interesting to analyze if the popularity of a song affects a word's usage over time, like if a popular song with a borrowed word triggers its increased popularity. This might be accompanied by work on creating contextual embeddings of French rap lyrics as a means towards scalable measures for semantic classification in this setting beyond manual annotation. Examining if French linguistic borrowings act similarly in rap songs in other European francophone countries, like Switzerland or Belgium, is another potentially valuable path.

Additionally, one could look at the external feature of how popular or frequently used a borrowing is in its home language compared to French. Though this would require a larger data collection process, this could provide insight on if there are borrowings that are used more in French than in their source language, or vice versa. Analyzing external factors like ethnic backgrounds of French rappers in tandem with linguistic borrowing usage would also help contextualize the sense of identity that a linguistic borrowing can convey. This could also help guide ethical debates on if the widespread use of linguistic borrowings indicates inevitable linguistic change or creates harm like cultural appropriation or increased difficulty for a linguistic minority to distinguish themselves.

Future work could also explore whether it is possible to predict from the landscape of lyrics at a given point in time whether a particular borrowing is likely to increase or decrease in usage in both the short- and long-term future. Some notable research that could provide direction in findings on both popularity and models to use are the work of [Kitayama et al. \(2020\)](#) predicting the popularity of an online petition given the headlining text and image, [Lamprinidis et al. \(2018\)](#) examining the popularity of newspaper headlines, or even [Donnelly and Beery's \(2022\)](#) work inside music, evaluating the sentiment of music through social media comments using Large Language Models.

## 6 Acknowledgements

We would like to thank the people who contributed valuable input on this work, including Marcelo Worsley, Haoqi Zhang, and three anonymous reviewers. We would like to also extend our thanks to the Computer Science Department, Office of Undergraduate Research<sup>11</sup>, and Bienen School of Music at Northwestern University as well as the ACL Student Research Workshop for their financial and academic support of this research.

## References

2019. [9.7 billion on Earth by 2050, but growth rate slowing, says new UN population report.](#)
- Elena Alvarez-Mellado. 2020. [An annotated corpus of emerging anglicisms in Spanish newspaper headlines.](#) In *Proceedings of the The 4th Workshop on Computational Approaches to Code Switching*, pages 1–8, Marseille, France. European Language Resources Association.
- Chong J. Bretillon. 2014. 'Ma Face Vanille': White rappers, 'Black Music', and race in France. *International journal of Francophone studies*, 17(3):421–443.
- Paula Chesley and R. Harald Baayen. 2010. [Predicting new words from newer words: Lexical borrowings in French.](#) *Linguistics*, 48(6).
- Patrick Donnelly and Aidan Beery. 2022. [Evaluating large-language models for dimensional music emotion prediction from social media discourse.](#) In *Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022)*, pages 242–250, Trento, Italy. Association for Computational Linguistics.
- Dominique Estival and Alastair Pennycook. 2011. [L'Académie Française and Anglophone language ideologies.](#) *Language Policy*, 10(4):325–341.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. [Diachronic word embeddings reveal statistical laws of semantic change.](#) In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.
- Benjamin Hebblethwaite. 2018. Arabic Lexical Borrowings in German Rap Lyrics: Religious, Standard and Slang Lexical Semantic Fields. *Delos*, 31:113–125.

<sup>11</sup>The study resulting in this presentation was assisted by a Conference Travel Grant from the Office of Undergraduate Research which is administered by Northwestern University's Office of the Provost. However, the conclusions, opinions, and other statements in this presentation are the author's and not necessarily those of the sponsoring institution.



- Daphna Keidar, Andreas Opedal, Zhijing Jin, and Mrinmaya Sachan. 2022. [Slangvolution: A causal analysis of semantic change and frequency dynamics in slang](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1422–1442, Dublin, Ireland. Association for Computational Linguistics.
- Kotaro Kitayama, Shivashankar Subramanian, and Timothy Baldwin. 2020. [Popularity prediction of online petitions using a multimodal DeepRegression model](#). In *Proceedings of the The 18th Annual Workshop of the Australasian Language Technology Association*, pages 110–114, Virtual Workshop. Australasian Language Technology Association.
- Sotiris Lamprinidis, Daniel Hardt, and Dirk Hovy. 2018. [Predicting news headline popularity with syntactic and semantic knowledge using multi-task learning](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 659–664, Brussels, Belgium. Association for Computational Linguistics.
- Miles Marshall Lewis. 2023. [How hip-hop changed the English Language Forever](#).
- Verónica Loureiro-Rodríguez. 2013. “if we only speak our language by the Fireside, it won’t survive”: The cultural and linguistic indigenization of Hip Hop in Galicia. *Popular Music and Society*, 36(5):659–676.
- Petter Mæhlum and Sardana Ivanova. 2023. [Phonotactics as an aid in low resource loan word detection and morphological analysis in sakha](#). In *Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)*, pages 111–120, Tórshavn, the Faroe Islands. Association for Computational Linguistics.
- John E. Miller and Johann-Mattis List. 2023. [Detecting lexical borrowings from dominant languages in multilingual wordlists](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2599–2605, Dubrovnik, Croatia. Association for Computational Linguistics.
- Saif M. Mohammad. 2020. [WordWars: A dataset to examine the natural selection of words](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3087–3095, Marseille, France. European Language Resources Association.
- Skye Paine. 2012. The quadrilingual vocabulary of French rap. *Multilingualism in Popular Arts*, 3(1):48–69.
- Derek Pardue. 2012. [Cape verdean creole and the politics of scene-making in Lisbon, Portugal](#). *Journal of Linguistic Anthropology*, 22(2).
- Zhengqi Pei, Zhewei Sun, and Yang Xu. 2019. [Slang detection and identification](#). In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 881–889, Hong Kong, China. Association for Computational Linguistics.
- Lina Rhissi. 2021. [D’Aya Nakamura à PNL: Comment les artistes musicaux-les transforment la langue française](#).
- Ian Stewart, Diyi Yang, and Jacob Eisenstein. 2021. [Tuiteamos o pongamos un tuit? investigating the social constraints of loanword integration in Spanish social media](#). In *Proceedings of the Society for Computation in Linguistics 2021*, pages 286–297, Online. Association for Computational Linguistics.
- Yulia Tsvetkov, Waleed Ammar, and Chris Dyer. 2015. [Constraint-based models of lexical borrowing](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 598–608, Denver, Colorado. Association for Computational Linguistics.
- Martin Verbeke. 2017. Represent your origins: An analysis of the diatopic determinants of non-standard language use in French rap. *International journal of Francophone studies*, 20(3):209–236.
- Michael Waugh. 2020. “every time i dress myself, it go motherfuckin” viral’: Post-verbal flows and memetic hype in Young Thug’s mumble rap: Popular music.
- Kelsey Quinn Westphal. 2013. [Teuf Love: Verlan in French rap and beyond](#).
- Steven Wilson, Walid Magdy, Barbara McGillivray, Kiran Garimella, and Gareth Tyson. 2020. [Urban dictionary embeddings for slang NLP applications](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4764–4773, Marseille, France. European Language Resources Association.