

Données sur le rap québécois

Par Laurent Corbeil

Étant donné que nous nous sommes mis en équipe avec des étudiants de Polytechnique, nous voulions sortir un peu des sentiers battus. Plutôt que de « faire parler » de simples archives publiques, le domaine culturel était, à mon sens, une zone peu quantifiée dans la sphère journalistique. [L'article](#) de Matt Daniels sur le site [pudding.cool](#) faisait le portrait exhaustif du vocabulaire dans le hip-hop américain. L'intégralité des données provenait du site [Genius.com](#), qui regroupe les paroles de milliers d'artistes.

Actuellement, le rap québécois fait de plus en plus sa place sur la scène culturelle. Plusieurs facteurs reliés à l'avancée des technologies numériques ont déclenché l'ascension fulgurante de cette industrie présente depuis plus de 30 ans. Nous avons déterminé que ce sujet trouvait sa pertinence dans un champ d'analyse linguistique.

Le jargon propre à chaque rappeur reflète souvent celui de leurs adeptes. Cette appréhension nous permet donc de spéculer sur les tendances et les changements linguistiques d'un vaste public. Bien que plusieurs chansons de rappeurs québécois figurent sur Genius, force est d'admettre qu'un grand bassin d'artistes n'y figurait pas. Michaël s'est chargé essentiellement d'élaborer du script pour recueillir les paroles des artistes concernés. Étant donné que Genius possède un API permettant d'extraire les données, le [code](#) se trouvait sur le web après quelques recherches. L'extraction du contenu musical de 52 artistes a été réparti en 52 fichiers .json afin de bien isoler chaque profil d'artiste et son contenu.

Au départ, nous avons jugé qu'un mot, dans ce contexte, ne serait pas soumis aux normes linguistiques imposés par les différentes langues exprimées dans les chansons. Un mot inventé de toute pièce reste une donnée pertinente. Nous avons tout simplement convenu qu'un mot serait repérable aux espaces et aux apostrophes les séparant.

Nos collègues de polytechnique ont réussi tout de même à dissocier les mots français et anglais des données extraites. J'ai remarqué une homogénéité statistique à travers les 52 artistes québécois dans leur utilisation globale de l'anglais et la vaste majorité se trouve entre 20% et 45%. Le mélange de ces deux langues a donc été ma cible principale dans le cadre de ce reportage.

Aucune trace numérique ou physique

Sommairement, la sous-représentation du hip-hop québécois sur le site Genius a été la limite flagrante de cette cueillette. De prime abord, Michael et moi avons convenus que cette marge d'erreur serait acceptable. Or, il s'est avéré que, non seulement une majorité de rappeurs québécois n'y figuraient pas, mais également que plusieurs chansons n'avaient pas été numériser pour ceux qui avait un profil. J'ai contacté plusieurs maisons de disques afin d'obtenir les paroles de leurs artistes, qu'elles soient numérisées ou non. Malheureusement, beaucoup de rappeurs québécois dorénavant inactifs n'ont pas écrits, a priori, les paroles de leurs chansons. Ce manque de représentativité ce reflète dans notre reportage, car Genius est le seul site sur lequel nous avons moissonnée des données.

