

Les arbres ne cachent pas la forêt : arbres d'inférence conditionnelle et forêt aléatoire appliqués à la variation entre « so », *alors, donc* et *(ça) fait (que)* dans le rap québécois

Emily Brooke Leavitt

Université de Sherbrooke

La contemporanéité se manifeste au Québec par la diversification en progression du français vernaculaire (Blondeau et Tremblay 2016; Blondeau 2020) entre un pôle centré à Montréal et un pôle québécois délocalisé et par la matérialisation d'espaces culturels sécuritaires à partir et au sein desquels de nouvelles identités linguistiques hybrides et plurielles se voient créées, refaçonnées et circulées. Le rap québécois en est un des plus saillants.

La présente étude a pour objectif d'examiner un cas de variation sociolinguistique préalablement démontré comme étant soumis à l'indexicalité sur le continuum vernaculaire montréalais–québécois au sein de la communauté artistique de rap québécois : l'alternance entre les formes lexicales *alors, donc, ça fait (que)* et « so » et les formes phonologiques de *fait que*, /fɛk/ et /fak/. Selon Blondeau, Mougeon et Tremblay (2019), *alors* est en déclin à Montréal, pendant que *(ça) fait que* y est en progression. « So » n'a été attesté à Montréal, uniquement à Welland (ON) où cette forme indexe une identité bilingue. Blondeau et Tremblay (2022) constatent que /fɛk/ est privilégié notamment les locutrices les plus jeunes, tandis que /fak/ se voit stigmatisé.

Nous examinons la variation entre toutes ces formes lexicales et phonologiques, et nous cherchons découvrir quelles précisions peuvent fournir des modèles sophistiqués d'apprentissage automatique concernant les interactions entre les facteurs sociaux et linguistiques dont l'influence sur l'utilisation variable de ces formes est reconnue. Nous soumettons l'ensemble de ces variantes (N=2050) retrouvées au sein du mégacorpus « RapKeb21 » (1,45 millions de mots) de chansons rap québécoises parues entre 1997 et 2021 à une analyse de variation à partir de la modélisation multivariable à effets mixtes et des techniques non paramétriques de forêt aléatoire et d'arbres d'inférence conditionnelle (voir Tagliamonte et Baayen 2012).

Nous isolons d'abord les effets significatifs de facteurs linguistiques (fonction grammaticale et cooccurrence avec la particule *là*) et extralinguistiques (âge, genre, lieu de naissance, origines géographiques et année de parution) et de la variable aléatoire de l'individu. Ensuite, nous évaluons l'ampleur relative de ces facteurs et les interactions entre eux en nous servant du langage de programmation *R* (R Core, 2020) que nous manipulons au sein d'un environnement *RStudio* (RStudio Team, 2022) à l'aide des trousseaux *lme4()* (Bates, Mächler, Bolker et Walker 2015), *partykit()* (Hothorn et Zeileis 2015), *randomForest()* (Liaw et Wiener 2002) et *randomForestExplainer()* (Paluszynska, Biecek et Jiang 2020).

Nos analyses préliminaires révèlent une influence d'origines géographiques, de lieu de naissance et de fonction grammaticale sur le choix de variante. Globalement, les artistes originaires ou nés hors de Montréal privilégient *ça fait (que)*, /fɛk/ et /fak/ en fonction discursive, alors que « so » s'avère dominer comme marqueur de conséquence particulièrement chez les Montréalais(es) aux années 2010. Nous effectuerons des analyses approfondies et en examinerons les résultats, parmi lesquels l'évolution diachronique de /fɛk/ et /fak/ et l'influence des trajectoires migratoires vers et

au sein du Canada et du Québec. Notre étude contribuera à démontrer l'application de techniques analytiques basées sur l'apprentissage automatique et leur utilité pour interpréter un cas de variation particulièrement saillant au changement en cours au Québec.

Bibliographie

- Bates, Douglas, Mächler, Martin, Bolker, Ben, et Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blondeau, Hélène. (2020). Pratiques langagières et diversité culturelle chez de jeunes Montréalais : le français dans la métropole. Dans *Attribuer un sens. La diversité des pratiques langagières et les représentations sociales*, sous la direction de Kristin Reinke (pp. 151-175). Presses de l'Université Laval.
- Blondeau, Hélène, Mougeon, Raymond, et Tremblay, Mireille. (2019). Analyse comparative de *ça fait que, alors, donc* et *so* à Montréal et à Welland: mutations sociales, convergences, divergences en français laurentien. *Journal of French Language Studies*, 29(1), 35–65. <https://doi.org/10.1017/S0959269518000169>
- Blondeau, Hélène, et Tremblay, Mireille. (2016). Le traditionnel et l'émergent : l'apport de jeunes Montréalais issus de l'immigration au français vernaculaire. *Cahiers internationaux de sociolinguistique*, 10(2), 19-45. <https://doi.org/10.3917/cisl.1602.0019>
- Blondeau, Hélène, et Tremblay, Mireille. (2022). The hidden dimensions of a change from below: Consequence markers in Montreal French. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 67(1–2), 22–52. <https://doi.org/10.1017/cnj.2022.3>
- Hothorn, Torsten, et Zeileis, Achim. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905-3909. <https://jmlr.org/papers/v16/hothorn15a.html>.
- Liaw, Andy, et Wiener, Matthew. (2002). Classification and regression by randomForest. *R News*, 2(3), 18-22. <https://CRAN.R-project.org/doc/Rnews/>
- Paluszynska, Aleksandra, Biecek, Przemyslaw, et Jiang, Yue. (2020). randomForestExplainer: Explaining and visualizing random forests in terms of variable importance. *RandomForestExplainer 0.10.1*. <https://cran.r-project.org/web/packages/randomForestExplainer/index.html>
- R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. <https://www.R-project.org/>
- RStudio Team. (2020). RStudio: Integrated Development for R. *RStudio*. Boston, MA, PBC. <http://www.rstudio.com/>
- Tagliamonte, Sali A., et Baayen, Harald R. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135-178. <https://doi.org/10.1017/S0954394512000129>