

JADT 2024
Mots comptés, textes déchiffrés

Tome 2

Anne Dister et Dominique Longrée (éd.)

PUL PRESSES
UNIVERSITAIRES
■ DE LOUVAIN



© Presses universitaires de Louvain, 2024
<http://pul.uclouvain.be>
Dépôt légal : D/2024/9964/20
ISBN : 978-2-39061-473-9
ISBN pour la version numérique (pdf) : 978-2-39061-474-6
Imprimé en Belgique par CIACO scrl – n° d'imprimeur : 106408b

Tous droits de reproduction, d'adaptation ou de traduction, par quelque procédé que ce soit, réservés pour tous pays, sauf autorisation de l'éditeur ou de ses ayants droit.

Couverture : Cédric Fairon

www.i6doc.com, l'édition universitaire en ligne

Distribution :
France – Librairie Wallonie-Bruxelles
46 rue Quincampoix - 75004 Paris
Tél. 33 1 42 71 58 03
librairie.wb@orange.fr

Belgique, Luxembourg et Pays-Bas – MDS Benelux
Zoning de Martinrou
Rue du Progrès 1
6220 Fleurus, Belgique
Tél. 32 71 60 05 20
service.clients@mds-benelux.com

Reste du monde – Diffusion universitaire CIACO (DUC)
Grand-Rue, 2/14
1348 Louvain-la-Neuve, Belgique
Tél. 32 10 45 30 97
duc@ciaco.com

La variation graphique intratextuelle dans les imprimés québécois, 1800-1939

Wim Remysen¹, Myriam Paquet-Gauthier²

¹Université de Sherbrooke – Wim.Remysen@USherbrooke.ca

²Université de Sherbrooke – Myriam.Paquet-Gauthier@USherbrooke.ca

Abstract

This article aims to analyze intratextual graphic variations, namely the presence, within the same text, of concurrent spellings, usually one old and one modern (e.g., *fesait/faisait*, *siège/siège*). We analyze the presence of this type of variation in a corpus comprising literary texts by Québécois authors, published before 1940. To measure the presence of intratextual graphic variation, and to better understand the role played by certain factors (publication date, publisher, text genre, etc.), we conducted a quantitative analysis based on 188 works, representing nearly 5.7 million words.

Keywords: intratextual variation, Québec printed literary works, spelling and orthograph, standardization.

Résumé

Cet article a pour objectif d'analyser la variation graphique intratextuelle, soit la présence, dans un même texte, de graphies concurrentes, anciennes et modernes (p. ex. *fesait/faisait*, *siège/siège*). Nous analysons la présence de ce type de variation dans un corpus constitué de textes littéraires parus avant 1940 et produits par des autrices et auteurs québécois. Pour mesurer la présence de variation graphique intratextuelle, et pour mieux comprendre le rôle joué par certains facteurs (date de publication, éditeur, genre textuel, etc.), nous avons mené une analyse quantitative faite sur 188 œuvres, représentant près de 5,7M de mots.

Mots clés : variation intratextuelle, imprimé littéraire québécois, graphie et orthographe, standardisation.

1. Introduction

L'histoire de l'orthographe du français au Québec a été relativement peu étudiée, une situation qui s'explique par l'absence de corpus informatiques suffisamment riches et variés permettant de contribuer à son étude. Le *Corpus de littérature québécoise* (CLIQ), créé en 2020 dans la foulée de la mise sur pied d'un nouveau fonds documentaire numérique (le Fonds de données linguistiques du Québec¹), permet dorénavant de se pencher sur cette histoire. Une première étude (Paquet-Gauthier et Remysen, à paraître) faite à partir de ce corpus a permis de montrer que certaines graphies ont connu, au Québec, des trajectoires diachroniques différentes de celles observées en Europe francophone, notamment dans le corpus Frantext. On peut en effet noter au Québec le maintien de certaines graphies déjà tombées en désuétude en Europe, parfois avec un décalage important ; par exemple, dans la période de 1860 à 1879, la graphie *collège* est encore largement prédominante dans les œuvres littéraires québécoises, alors qu'elle est tout à fait marginale dans la production littéraire européenne dès les années 1820.

¹ Ce fonds (<https://fdlq.recherche.usherbrooke.ca>) réunit plusieurs ensembles de données, textuelles et orales, constitués depuis les années 1970 en vue d'étudier et de décrire le français en contexte québécois. La fiche de présentation du corpus CLIQ est disponible à l'adresse <https://fdlq.recherche.usherbrooke.ca/corpus/corpuslitterature-quebecoise-cliq.html>.

Cette contribution propose de poursuivre cette analyse en portant cette fois-ci une attention particulière à la variation graphique intratextuelle, soit la présence, dans un même texte, de formes graphiques concurrentes, anciennes et modernes (p.ex. *fesait/faisait*, *horison/horizon*, *siège/siège*), toujours dans les imprimés québécois publiés entre 1800 et 1939. La variation intratextuelle, généralement associée aux imprimés remontant à l’Ancien Régime (Pellat, 1995 ; Baddeley et Biedermann-Pasques, 2003) ou encore aux productions de scripteurs « ordinaires » (Steuckardt, 2015; Martineau et Remyesen, 2019), n’a jamais été étudiée dans les œuvres littéraires modernes, qu’elles soient québécoises ou françaises, ni fait l’objet, à notre connaissance, d’études quantitatives. Il faut dire que l’analyse de la variation graphique dans l’imprimé, littéraire ou autre, est loin d’être une tâche aisée, en grande partie en raison de la modernisation des textes disponibles aux fins de recherche (voir à ce sujet Catach, 2004 ; Duval, 2015). En effet, les textes versés dans les bases de données et grands ensembles documentaires sont rarement disponibles dans leur format original et proviennent habituellement d’éditions plus récentes. Ces éditions font souvent peu de cas de la variation graphique, presque toujours associée à des coquilles ou à des erreurs typographiques et, partant, gommée entièrement, parfois sans qu’un appareil critique permette de la retracer. Cette pratique projette sur les textes du passé une vision tronquée, la standardisation de la graphie étant un phénomène beaucoup plus récent qu’on ne le croit généralement, et empêche de restituer avec assurance la chronologie de certains phénomènes qui ont pourtant longtemps été tout à fait courants.

Pour cerner l’ampleur et la chronologie du phénomène de la variation graphique intratextuelle dans l’imprimé littéraire québécois, et pour mieux évaluer le rôle qu’ont pu jouer certains facteurs, nous avons mené une analyse quantitative à partir d’un corpus composé de 188 œuvres québécoises (près de 5,7M de mots). La prise en compte de données quantifiées relatives à la présence dans un même texte de graphies concurrentes nous permet *in fine* de caractériser de manière empirique la standardisation, soit l’élimination progressive de variantes graphiques en faveur d’une seule graphie de référence stable, une pratique orthographique qui émerge dans l’imprimé québécois à la fin du 19^e siècle.

2. Aspects méthodologiques

Le *Corpus de littérature québécoise* (CLIQ) est constitué de plusieurs centaines d’œuvres informatisées (en édition originale) qui sont soumises à un protocole d’édition commun, basé sur un certain nombre de principes philologiques. Si le principe de base du travail d’édition repose sur le respect du texte original, le protocole autorise néanmoins certaines interventions destinées à 1^o corriger des erreurs évidentes ; 2^o rétablir des lettres manquantes ; 3^o supprimer des répétitions de lettres ou de mots inutiles ; 4^o rétablir l’individualité de certains mots (ou parfois, à l’inverse, éliminer des espaces incongrues) ; ou encore 5^o corriger des problèmes liés aux casses inversées. Voici quelques exemples de l’ensemble de ces cas de figure :

1. *si jhardi* → *si hardi* (Gaspé_1864_Canadiens), *vingt-hiut ans* → *vingt-huit*, *La famille était* → *La famille était* (Auclair_1933_Figures_Vol1)
2. *causerait leur sépation* → *séparation* (Dupuy_1930_Laurence), *noces d’argent épiscopales* → *épiscopales* (Auclair_1933_Figures_Vol1)
3. *des des mariniers* → *des mariniers* (Bibaud_1879_Vicissitudes), *Bellechasse* → *Bellechasse* (Proulx_1888_Europe)
4. *sousl’impulsiond’une* → *sous l’impulsion d’une* (Barthe_1855_Reconquis), *la plu s explorée* → *la plus explorée* (Chopin_1913_Exil)
5. *qu’v dit* → *qu’y dit* (confusion entre les casses v/y, Bourgeois_1919_Ladébauche), *aspirations coufuses* → *confuses* (inversion u/n, D’Arles_1926_Estampes)

Sans intervention de notre part, ces données demeureraient invisibles, car certains mots seraient introuvables dans le corpus. Le protocole prévoit également qu'en présence d'un errata, les corrections qui y sont signalées soient utilisées dans la transcription (plutôt que la forme jugée fautive utilisée dans le texte). Comme le signale Duval (2015), ce travail d'édition, aussi bien balisé soit-il, comporte toujours une part d'arbitraire et il demeure impossible de saisir réellement l'intention (ou l'absence d'intention) de l'auteur ou du typographe derrière des formes que l'on peut rencontrer dans les textes. Lorsque le travail d'édition porte sur un grand nombre de textes, comme c'est le cas dans la réalisation du corpus CLIQ, certaines formes rares, et initialement jugées « déviantes », peuvent se révéler récurrentes et ultimement significatives pour l'histoire de l'orthographe. D'où l'importance de bien documenter toutes les décisions et, parfois, de reconsidérer certains choix de corrections ; c'est pourquoi nous avons recours à l'utilisation systématique de commentaires (leçons rejetées) dans les textes qui permettent de garder des traces des interventions, aussi minimales soient-elles.

Ce travail d'édition est dûment documenté au fur et à mesure que de nouvelles œuvres s'ajoutent au corpus. Ainsi, nous prenons en note, pendant l'étape de la double révision des extractions de textes, les unités lexicales qui sont écrites d'au moins deux manières différentes au sein d'un même texte. Cette approche nous a permis de dresser une liste de nombreuses variantes graphiques récurrentes dans le corpus². Dans bien des cas, les séries de graphies concurrentes incluent une forme conservatrice, c'est-à-dire associée à une période plus ancienne, et une forme innovante, attestée à date plus récente (comme *appeller/appeler*, *bayonnette/baïonnette* ou *chûte/chute*). Certaines séries incluent aussi des formes récurrentes dans notre corpus, mais qui ne semblent avoir jamais été normalisées dans les dictionnaires ou qui sont peu attestées dans les travaux consacrés à l'orthographe française (ainsi, les formes *insu* et *insçu* cohabitent dans certaines œuvres avec *inçu*). Nous utiliserons l'expression *graphies de référence* pour désigner les formes qui sont ultimement devenues normatives à l'époque moderne (à tout le moins avant la réforme de l'orthographe en 1992, comme *apparaître*, *goéland*, *québécois* et *rappeler*) et *graphies concurrentes* pour les formes alternatives qui sont attestées dans la littérature québécoise de l'époque et que la norme de référence a progressivement éliminées (comme *apparaitre*, *goëland*, *québecquois* et *rappeller*).

Les phénomènes concernés se divisent en trois grandes catégories de variation graphique : 1° les diacritiques, soit la présence/absence d'un accent (*chaûme/chaume*) ou le choix même de l'accent (*crême/crème*); 2° la segmentation des mots, soit la présence/absence d'un trait d'union ou la soudure (*audessus/au dessus/au-dessus*); 3° les éléments graphématiques, incluant le choix entre consonne simple/double (*gouffre/gouffre*), l'utilisation de lettres étymologisantes ou analogiques (*exhorbiter/exorbiter*), le recours au -s adverbial (*grâces/grâce*), la présence de lettres sous l'influence de la graphie anglaise (*comforter/conforter*), etc. Nous n'avons pas tenu compte des variations attribuables aux homophones (*ou/où*, *quand/quant*), ni de l'alternance majuscule/minuscule (*Récollets/récollets*), ni de toutes les formes de variation concernant les noms propres (*Shakespeare/Shakspere*). Nous avons aussi exclu tous les phénomènes généraux qui concernent l'orthographe grammaticale, comme l'alternance *-ois/-ais* dans les formes de

² Nous saluons la proposition de Duval (2015), qui suggère de dresser des listes des variantes graphiques que les chercheuses et chercheurs observent dans les textes plus anciens : « Ces listes permettent de discuter de l'opportunité des corrections et d'attirer l'attention sur des formes dont la récurrence pourrait amener à leur maintien dans les éditions savantes » (Duval, 2015 : 376). C'est pour cette raison que la liste dont nous nous sommes servis dans le cadre de cette étude est publiée sur le site Web du Fonds de données linguistiques du Québec à l'adresse <https://fdlq.recherche.usherbrooke.ca/cliq-orthographe>.

l'imparfait et du conditionnel ou encore l'alternance des pluriels *-ens/-ents* et *-ans/-ants* ; les données incluent en revanche l'alternance entre les formes radicales du verbe *faire* (phénomène qui relève de l'orthographe lexicale : *fes-/fais-*) ainsi que quelques phénomènes spécifiques d'orthographe grammaticale relativement faciles à repérer (comme les alternances *ayions/ayons* et *soyions/soyons*).

En tout, nous avons dressé une liste de 195 types morpholexicaux, identifiés dans notre texte en petites majuscules, dont la graphie varie fréquemment dans le corpus. Nous parlons ici de *types morpholexicaux* plutôt que de *mots*, de *lemmes* ou d'*unités lexicales* étant donné que l'analyse informatisée nous a parfois contraints à regrouper certaines formes appartenant à différentes catégories grammaticales (une limite que la lemmatisation du corpus, actuellement en cours, pourra régler puisqu'elle permettra de faire des requêtes plus complexes). Ainsi, le type FRAÎCHIR, qui a fait l'objet de deux requêtes dans le corpus (*fraîch** et *fraich**), comprend aussi bien les flexions verbales de *fraîchir/fraichir* que les formes du singulier et du pluriel de l'adjectif *fraîche/fraiche* et du substantif *fraîcheur/fraicheur*. Dans quelques rares cas, les types morpholexicaux peuvent présenter plusieurs variations graphiques ; par exemple, pour GUÈRE, la variation peut concerner non seulement l'accent (*guère, guère*), mais aussi l'absence ou la présence du *-s* adverbial (*guère, guères*).

Si cette liste, qui n'est pas constituée sur la base d'une détection automatique de formes concurrentes, ne peut en aucun cas être considérée comme exhaustive de tous les phénomènes de variation graphique présents dans les textes versés au corpus CLIQ, elle cible en revanche des variations spécifiques qui peuvent être quantifiées. Elle nous a en effet permis de relever systématiquement, grâce à une série de manipulations informatiques complexes³, le nombre d'occurrences de variation effectivement attesté pour chaque œuvre du corpus et de calculer un taux de variation intratextuelle. Ce dernier correspond au nombre de types morpholexicaux qui présentent de la variation graphique dans un texte divisé par la somme des types morpholexicaux qui apparaissent au moins 2 fois dans le texte et qui pourraient donc varier graphiquement.

À titre d'exemple, les relevés produits à la suite du traitement informatique du corpus nous permettent de savoir que l'on trouve, dans la pièce *Colas et Colinette, ou Le bailli dupé* de Joseph Quesnel, publié en 1808, 34 des 195 types morpholexicaux de notre liste. De ce nombre, 11 ne sont attestés qu'une fois et ne peuvent donc pas varier dans le texte. Pour les 23 autres types restants, on en compte 14 qui ne sont pas systématiquement écrits de la même façon (on trouve entre autres *piège* et *piège*, *vîte* et *vite*, *tems* et *temps*) alors que la graphie des 9 autres types ne varie jamais dans le texte (par exemple, l'auteur utilise toujours *hazard*, attesté deux fois). Pour ce document, le taux de variation intratextuelle s'élève donc à 60,9 % (= 14/23, soit le nombre de types morpholexicaux présentant des variantes divisé par le nombre de types attestés plus qu'une fois dans l'œuvre). Il faut préciser que le nombre de fois qu'un type morpholexical apparaît dans un texte n'entre pas dans le calcul du moment qu'il y a variation intratextuelle ; par exemple, un texte peut contenir 2 fois la forme *bled* et 3 fois la forme *blé*, ce qui correspond à une occurrence de la variation BLÉ.

L'ensemble des taux de variation (variable dépendante) ont par la suite fait l'objet de calculs statistiques permettant d'évaluer l'incidence de différents facteurs susceptibles d'influencer la cohabitation de variantes graphiques au sein d'un même texte (variables indépendantes). Ces

³ Nous sommes redevables au travail de Benoît Mercier, analyste aux Services des technologies de l'information de l'Université de Sherbrooke et responsable du Groupe de soutien à la recherche. Nous remercions également Hugo Saint-Amant Lamy de l'Université du Québec à Rimouski pour les analyses statistiques.

facteurs concernent certaines caractéristiques relatives à la publication des œuvres, à savoir 1° l'année de publication ; 2° le lieu d'édition⁴ ; 3° le genre textuel⁵ ; 4° la présence d'un errata en fin d'ouvrage et 5° l'indice de graphies concurrentes globalement attestées dans le texte (calculée en fonction du nombre de types morpholexicaux qui n'apparaissent qu'en graphies concurrentes divisé par la somme des types qui n'apparaissent qu'en graphies concurrentes et en graphies de référence, toujours pour les 195 types morpholexicaux). L'incidence de tous ces facteurs a été mesurée par une régression linéaire multiple. Nous avons aussi analysé les résultats suivant le type de variation en cause (diacritique, segmentation ou graphématique).

3. Résultats

Le modèle retenu pour l'analyse quantitative des données est significatif ($R^2 = 0,576$; $F(10, 177) = 24,046$; $p < 0,001$) et fait ressortir trois variables indépendantes qui expliquent le taux de variation graphique intratextuelle dans les œuvres littéraires du corpus, soit l'année de publication ($\beta = -0,223$; $p = 0,003$), l'indice de graphies concurrentes ($\beta = 0,494$; $p < 0,001$) et la présence d'un errata ($\beta = 0,143$; $p = 0,006$)⁶. Les variables du genre textuel et du lieu de publication n'ont pas d'effet significatif sur le taux de variation.

3.1. L'année de publication

Les résultats présentés ci-dessous (Figures 1 et 2) montrent que la variation graphique intratextuelle diminue tout au long de la période étudiée de 1800 à 1939. Les textes les plus anciens de notre corpus, parus entre 1800 et 1819, affichent en moyenne un taux de variation graphique de 55,5 %, ce qui signifie qu'un peu plus de la moitié des types morpholexicaux attestés dans ces textes varient. La proportion moyenne de variation graphique intratextuelle s'élève à environ 25 % dans les décennies qui suivent, de 1820 à 1859. Elle baisse par la suite à 10 % dans les textes parus de 1860 à 1899. Cette baisse marquée coïncide avec la circulation de plus en plus importante du discours puriste vers le dernier tiers du 19^e siècle, porté par un nouveau mouvement de rectification langagière qui se met progressivement en place (voir Paquet-Gauthier et Remysen, à paraître).

La Figure 1 montre non seulement un recul de la variation graphique intratextuelle dans le temps, mais fait également voir que l'écart entre les œuvres dans lesquelles la graphie varie fortement et les œuvres qui présentent peu de variation graphique se resserre vers la fin du 19^e siècle, et plus particulièrement à partir des années 1910, même s'il continue à y avoir des textes qui s'écartent de la moyenne. Ce resserrement signale une stabilisation progressive des pratiques graphiques dans le corpus en faveur de l'utilisation de formes graphiques uniques, même si la variation graphique intratextuelle ne disparaît pas pour autant dans les dernières décennies de la période étudiée. D'une certaine manière, on pourrait se surprendre de la persistance de la variation graphique intratextuelle dans les premières décennies du XX^e siècle ; on verra plus loin (voir 3.5.) que les variations présentes au début du XX^e siècle sont toutefois qualitativement différentes de celles présentes au cours du XIX^e siècle.

⁴ Soit Montréal, Québec, Canada (autre), États-Unis ou Europe.

⁵ Les œuvres sont classées en fonction des catégories suivantes : fiction, art oratoire (chanson, théâtre, monologues), non-fiction (essai, biographie, récit autobiographique ou récit de voyage), poésie et recueil de textes (essentiellement des recueils d'articles journalistiques).

⁶ Lorsqu'on retient seulement les facteurs significatifs, on obtient les résultats suivants : année de publication ($\beta = -0,260$; $p < 0,001$), indice de graphies concurrentes ($\beta = 0,497$; $p < 0,001$), présence d'un errata ($\beta = 0,140$; $p = 0,006$)

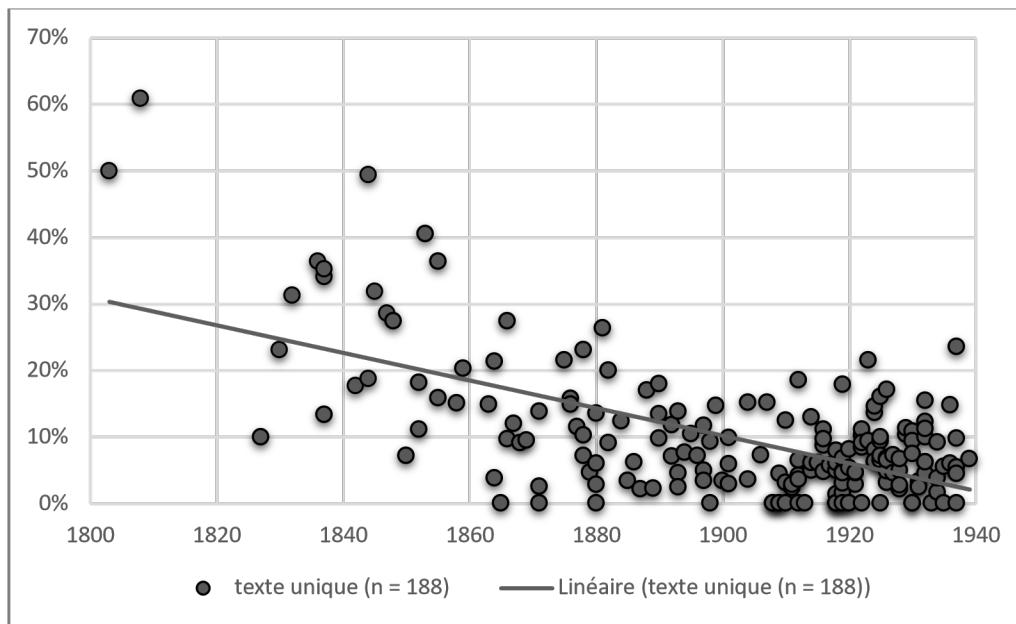


Figure 1 : Taux de variation graphique intratextuelle (en pourcentage), par texte et par année de publication

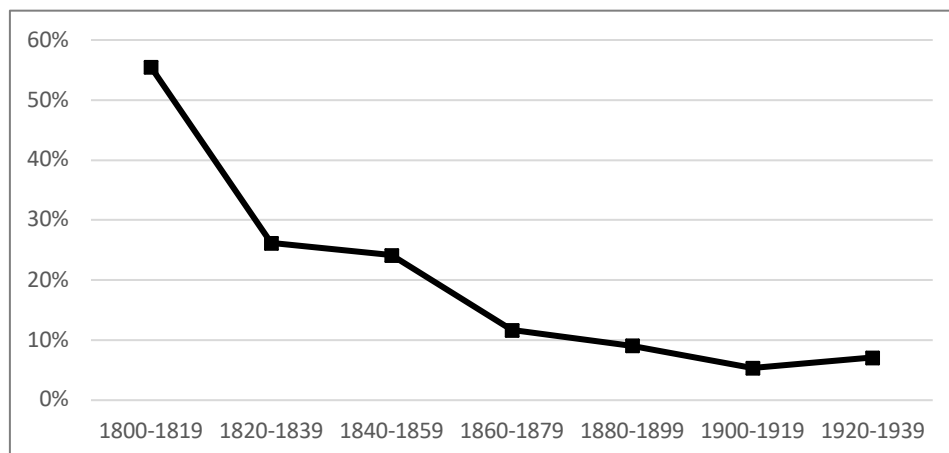


Figure 2 : Moyenne des taux de variation graphique intratextuelle, par période de 20 ans

3.2. L'indice de graphies concurrentes

L'indice de graphies concurrentes d'un texte prédit de manière significative le taux de variation graphique que l'on peut y observer. Plus particulièrement, comme le montrent les résultats illustrés dans la Figure 3, plus on relève de types qui apparaissent uniquement en graphies concurrentes (par rapport aux graphies de référence modernes), plus on relève aussi d'occurrences de variation. La variation graphique intratextuelle est donc plus importante dans les œuvres qui s'inspirent des modèles normatifs plus anciens.

Ce résultat suggère que le modèle normatif cible se précise progressivement au cours de la période à l'étude. Dans les textes publiés de 1800 à 1859, les usages graphiques sont très variés et aucune graphie concurrente ne se démarque réellement des autres par sa fréquence élevée. Pour 1860-1899, quelques graphies concurrentes se distinguent par leur présence récurrente, notamment *collège*, *privilège* et *siège*, des formes qui se sont longtemps maintenues dans les imprimés québécois sous l'influence de la prononciation et qui ne reculent qu'à partir des années 1880 (voir Paquet-Gauthier et Remysen, à paraître). Pour la période 1900-1939, les

graphies tendent à s'uniformiser : on relève un moins grand nombre de types morpholexicaux pour lesquels des graphies concurrentes sont utilisées ; en revanche, ces graphies sont plus fréquentes et ont tendance à concerner surtout des types morpholexicaux qui présentent encore aujourd'hui un certain flou orthographique, tels AULNE (*aune*), ARÔME (*arome*), CLÉ (*clef*) et GAIÉTÉ (*gaîté*) (Catach, 1995). Le modèle normatif semble ainsi se préciser jusqu'à ne permettre qu'un petit nombre de graphies concurrentes qui sont pour ainsi dire « normalisées ».

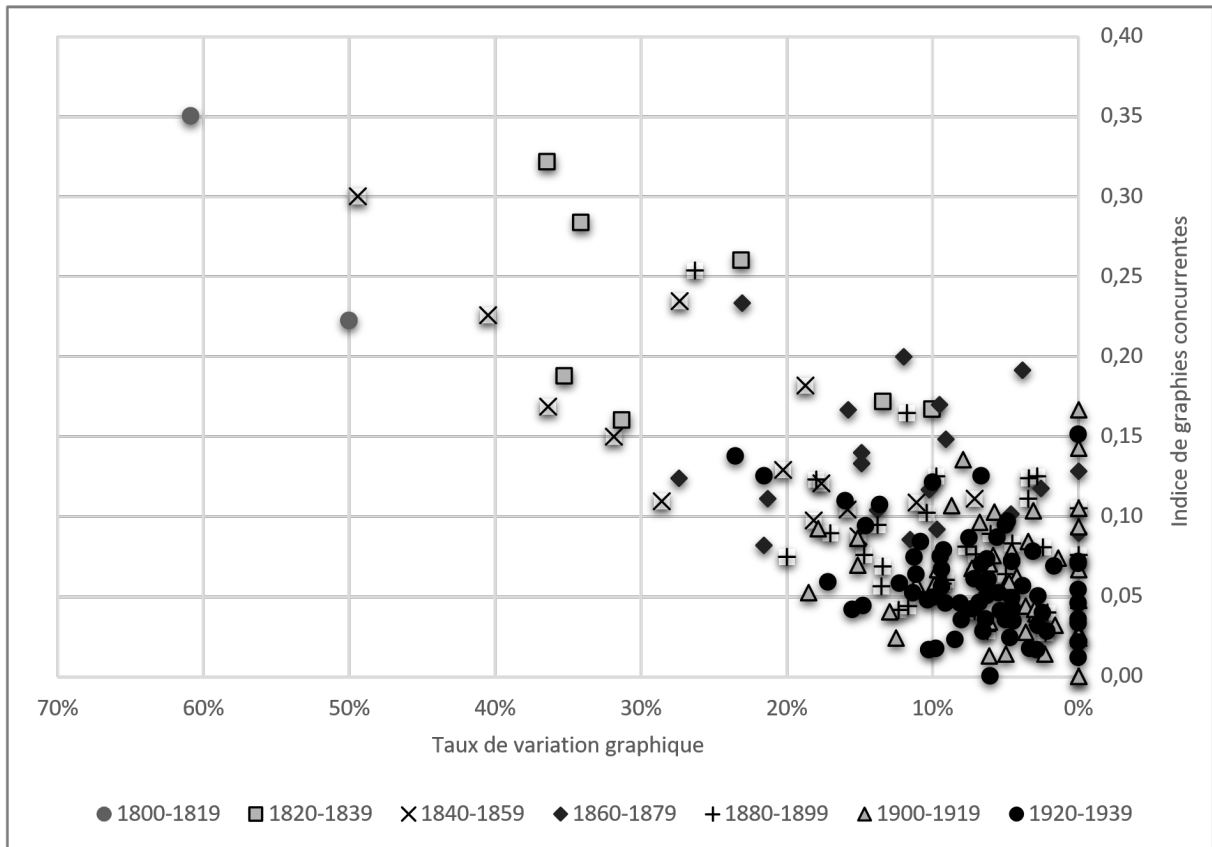


Figure 3 : Relation entre le taux de variation graphique intratextuelle et l'indice de graphies concurrentes, par texte et par période de 20 ans

3.3. La présence d'un errata

Même s'ils sont peu nombreux ($n = 20$), certains textes de notre corpus contiennent des errata qui signalent des corrections. En général, ces errata servent à rectifier des erreurs d'inadvertance qui se sont glissées dans la préparation du texte imprimé, comme des fautes d'accord ou des mots homophones (cela dit, il arrive que les modifications soient plus substantielles et concernent des choix lexicaux ou des structures de phrase). À l'occasion, les corrections signalées dans les errata concernent d'anciennes graphies qui ne semblent plus jugées conformes à la norme, dont certaines apparaissent dans la liste de types morpholexicaux étudiés ici : Chauveau corrige par exemple *inçu*, *contrecarèrent* et *duement* en *insu*, *contrecarrèrent* et *dûment* (Chauveau_1853_Guérin), Gaspé signale qu'il faut lire *sanglota* et non *sanglotta* dans son texte (Gaspé_1866_Mémoires) et Rouleau mentionne que *giffle* doit être remplacé par *gifle* (Rouleau_1901_Légende).

De manière générale, les textes qui contiennent des errata présentent des taux de variation graphique plus élevés que les œuvres qui n'en contiennent pas. Il semble donc y avoir dans ces textes un flottement qui peut s'interpréter comme le résultat d'une transition vers la mise en place d'un nouveau modèle normatif, mais dont les contours ne sont pas toujours bien définis

et que la correction ponctuelle de quelques formes jugées inappropriées ne permet pas d'atteindre. Devant le « grand nombre de fautes » qui se sont glissées dans les épreuves de *Charles Guérin : roman de mœurs canadiennes*, Chauveau (ou l'éditeur ?) précise d'ailleurs que l'errata ne permet que de signaler « les principales » fautes. Dans certains cas, les corrections peuvent même introduire de la variation lorsqu'elles proposent, mais pas systématiquement, d'insérer des graphies de référence dans les textes qui ne contiennent à l'origine que des graphies concurrentes.

3.4. *Le genre textuel et le lieu d'édition*

L'analyse n'a pas permis de conclure à un effet significatif pour deux variables indépendantes, le genre textuel et le lieu d'édition. Cela ne signifie pas pour autant que ces facteurs ne jouent aucun rôle dans les choix en matière de graphie qui sont faits par les auteurs et les éditeurs des textes faisant partie du corpus analysé. Il convient en effet de souligner certaines limitations de notre corpus qui ont pu jouer sur les résultats. Ainsi, le genre *recueil de textes* est pratiquement absent dans le corpus jusqu'à la décennie 1860. En outre, la répartition par lieu d'édition n'est pas équilibrée tout au long de la période. Des 37 textes publiés à Québec, 27 l'ont été avant 1900 (soit 73 %), tandis qu'on ne relève que 35 textes pour la même période sur les 128 publiés au total à Montréal (soit 27 %) ; de plus, les textes publiés en dehors de ces deux centres sont peu nombreux dans le corpus. Cela signifie que pour certaines périodes, le nombre d'œuvres pour ces facteurs n'est pas suffisamment élevé pour permettre de mesurer des effets.

Compte tenu de ces limitations, il serait opportun d'envisager d'autres pistes d'analyse supplémentaires. Il serait notamment instructif de se pencher sur les œuvres qui contrastent avec la moyenne de variation graphique observée dans les autres œuvres contemporaines, soit parce que les œuvres affichent un taux considérablement plus élevé que la moyenne, soit parce qu'elles contiennent une proportion moins élevée de variation graphique, et de s'interroger sur les possibles causes des écarts observés. Pourquoi, par exemple, le taux de variation graphique dans *Les anciens Canadiens* (paru en 1864, 21,3 %) et dans les *Mémoires* (1866, 27,4 %) de Philippe Aubert de Gaspé (père) est-il plus élevé que la moyenne, qui est de 11,1 % pour les œuvres parues entre 1860 et 1879, alors qu'il n'y a pratiquement pas de variation graphique dans *Trois légendes de mon pays* (1871, 2,6 %) de Joseph-Charles Faucher ? Faut-il y voir des choix différents adoptés par les éditeurs de ces livres (Georges-Édouard Desbarats, dans le cas de Gaspé, et Beauchemin & Valois, dans le cas de Faucher) ? Ces questions dépassent le cadre de l'analyse quantitative présentée ici, mais méritent qu'on s'y intéresse davantage.

3.5. *Catégories de variation*

Les 195 types morpholexicaux retenus dans cette analyse présentent des variations graphiques que nous avons regroupées en fonction de trois catégories, comme nous l'avons vu plus haut. La variation attribuable à l'utilisation des diacritiques est de loin la catégorie la plus importante puisqu'elle représente 124 des phénomènes de variation étudiés dans le corpus. Les phénomènes concernant la segmentation des mots ou les autres éléments graphématiques correspondent respectivement à 15 et à 56 types différents.

On observe au fil du temps un phénomène similaire à celui déjà signalé pour le type et la fréquence de graphies concurrentes (voir 3.2.) : en effet, dans les œuvres les plus anciennes publiées au 19^e siècle, la variabilité est très diverse, c'est-à-dire que les types morpholexicaux dont la graphie varie sont très différents d'un texte à l'autre. En revanche, dans les textes plus récents, notamment ceux qui sont parus après 1900, la variation se concentre davantage sur quelques types morpholexicaux récurrents, attestés dans un plus grand nombre de textes

différents. La variation graphique intratextuelle n'évolue donc pas seulement quantitativement, mais aussi qualitativement. La figure 4 illustre ce phénomène à l'aide de l'évolution du ratio type-token pour les trois catégories de variations. Ce ratio, calculé par période de 20 ans, correspond au total des *types*, soit le nombre de différents types morpholexicaux en variation observés, divisé par le total des *tokens*, soit le nombre de textes où chaque type est attesté. Plus ce ratio est élevé, plus le type est attesté dans un nombre limité de textes seulement.

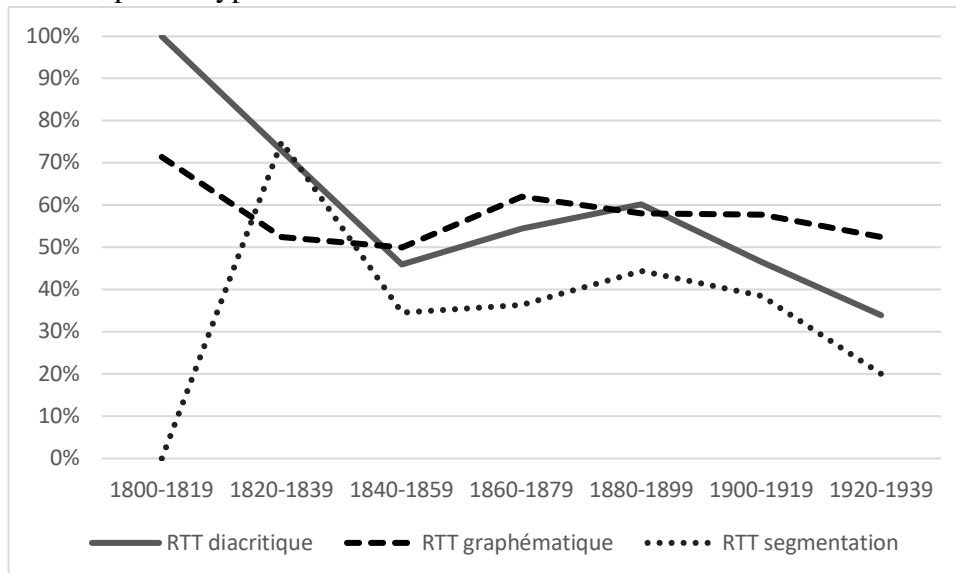


Figure 4 : Évolution du ratio type-token (RTT), par catégorie de variation et par période de 20 ans

On observe, en particulier pour les variations touchant les diacritiques et la segmentation, une baisse de ce ratio. Ainsi, si les ratios concernant ces deux catégories sont élevés pour les œuvres parues dans les premières décennies du corpus (avant 1840), ils connaissent, vers la fin du siècle, une diminution importante, soit exactement au même moment qu'émerge un discours sur la qualité de la langue au Québec. À partir de 1900, les types morpholexicaux qui entrent en variation sont de moins en moins nombreux et principalement de deux genres : des adverbes composés dont la segmentation normalisée peut être floue (comme AU-DELÀ, PAR-DESSUS, TOUT À COUP) ainsi que des mots contenant un diacritique optionnel (dont CONNAÎTRE, COTEAU, GAIÉTÉ, PARAÎTRE), dont la présence peut être attribuable, mais pas toujours, à une prononciation québécoise (*côteau* rappelle la prononciation [ko:to] courante au Québec).

4. Discussion et conclusion

L'analyse présentée dans cet article montre bien que la pratique de l'imprimé (à tout le moins littéraire) au Québec laisse une place certaine à la variation graphique intratextuelle tout au long du 19^e siècle et pendant les premières décennies du 20^e siècle. En même temps, les résultats font voir, dès les années 1860, une standardisation progressive des usages graphiques, signe que la norme cible se précise au fil du temps. L'analyse montre par ailleurs qu'en plus de diminuer au fil du temps, la variation graphique a tendance à se spécialiser, dans la mesure où les types morpholexicaux dont la graphie varie concernent un nombre de plus en plus limité de cas, mais sont attestés dans un plus grand nombre de textes. La standardisation qu'on observe dès la deuxième moitié du 19^e siècle est rendue possible grâce à l'arrivée et à la circulation d'ouvrages de référence plus récents qui deviennent alors plus facilement accessibles. Elle est aussi le résultat de la professionnalisation de l'édition à partir de cette période. L'impression locale est en effet apparue tardivement au Québec : interdite à l'époque de la Nouvelle-France, ce sont les Britanniques qui introduisent l'impression au pays dans le dernier tiers du 18^e siècle,

mais le marché du livre sera longtemps un marché d'importation. La production locale se développe à partir de la fin des années 1830 seulement, en même temps qu'émerge peu à peu une littérature nationale (à ce sujet, voir respectivement Alston et Bowslaugh, 2004 ; Biron, Dumont et Nardout-Lafarge, 2010). La circulation de la norme de référence n'est donc pas le seul facteur en cause ; certaines caractéristiques propres au développement du marché du livre au Québec peuvent expliquer des tendances que l'on observe dans la pratique graphique de l'imprimé québécois. La professionnalisation plus lente de la fabrication des livres ou encore le cumul des tâches (tant pour l'auteur que pour le libraire-éditeur-imprimeur) chez les premiers artisans du livre (voir Leroux, 2005 ; Vincent et Luneau, 2022) explique sans doute pourquoi la pratique de l'orthographe varie autant jusqu'au dernier tiers du 19^e siècle. Ainsi, G. H. Cherrier, éditeur du roman *Charles Guérin*, ressent le besoin d'écrire dans son *Avis de l'éditeur* :

La plus grande difficulté consiste dans la mise à l'œuvre de la publication. Les choses n'en sont pas encore rendues au point que nos auteurs puissent faire exclusivement leur métier d'auteur. Les affaires de leur profession (presque tous exercent une de ces profession[s], qu'on est convenu d'appeler *libérales*) les empêchent de pouvoir surveiller l'impression et faire réussir l'édition une fois lancée. Ce qui leur fait défaut, en un mot, c'est le libraire, c'est l'éditeur. (G. H. Cherrier, « Avis de l'éditeur », dans Pierre-Joseph-Olivier Chauveau, *Charles Guérin : roman de mœurs canadiennes*, p. iv-v)

D'autres travaux sont nécessaires pour apporter un éclairage sur le rôle joué par les différents acteurs du milieu de l'imprimerie dans la généralisation de la norme orthographique moderne, tant au Québec qu'en Europe francophone.

Bibliographie

- Alston S. et Bowslaugh J. (2004). Analyse statistique des premiers imprimés canadiens. In Fleming P., Gallichan G. et Lamonde Y. (Eds), *Histoire du livre et de l'imprimé au Canada : des débuts à 1840*. Montréal : PUM, 94-97.
- Biron M., Dumont F. et Nardout-Lafarge É. (2010). Histoire de la littérature québécoise. Montréal : Boréal.
- Baddeley S. et Biedermann-Pasques L. (2003). Histoire des systèmes graphiques du français (IX^e-XV^e siècles) : des traditions graphiques aux innovations du vernaculaire. *La Linguistique*, 39 (1), 3-34.
- Catach N. (2004). *L'orthographe*. Paris : PUF.
- Catach N. (Ed.) (1995). *Dictionnaire historique de l'orthographe française*. Paris : Larousse.
- Duval F. (2015). Les éditions de textes du XVII^e siècle. In Trotter D. (Ed.), *Manuel de la philologie de l'édition*. Berlin : De Gruyter, 369-393.
- Leroux É. (2005). Le métier d'imprimeur au Québec : 200 ans d'évolution (1794-1960). *Documentation et bibliothèques*, 51 (2), 107-116.
- Martineau F. et Remyssen W. (2019). Bouleversements sociaux et normes orthographiques : l'exemple du Régime anglais dans l'histoire du français québécois. In Dufter A., Grübl K. et Scharinger T. (Eds), *Des parlars d'oïl à la francophonie : contact, variation et changement linguistique*. Berlin : De Gruyter, 271-298.
- Paquet-Gauthier M. et Remyssen W. (à paraître). Les usages graphiques dans les œuvres littéraires québécoises d'avant 1940. *Le Français Moderne*.
- Pellat J.-Ch. (1995). Norme et variation orthographique au XVII^e siècle. *Scolia*, 3, 245-260.
- Steuckard A. (Ed.) (2015). *Entre village et tranchées : l'écriture de Poilus ordinaires*. Uzès : Inclinaison.
- Vincent J. et Luneau M.-P. (2022). *Dictionnaire historique des gens du livre au Québec*. Montréal : PUM.