



HAL
open science

Identifier et décrire l'hétérogénéité du français aux 17^e et 18^e siècles: le projet MACINTOSH (Missing hAlf the picture, ClassIcal NoT sO claSsical French)

Myriam Bergeron-Maguire

► To cite this version:

Myriam Bergeron-Maguire. Identifier et décrire l'hétérogénéité du français aux 17^e et 18^e siècles: le projet MACINTOSH (Missing hAlf the picture, ClassIcal NoT sO claSsical French). *Zeitschrift für romanische Philologie*, 2023, 139 (4), pp.1161-1179. 10.1515/dzmrpp-2t-022031-60-04063. halshs-04261075

HAL Id: halshs-04261075

<https://shs.hal.science/halshs-04261075>

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Myriam Bergeron-Maguire

Identifier et décrire l'hétérogénéité du français aux 17^e et 18^e siècles : le projet MACINTOSH (Missing hAlf the picture, ClassIcal NoT sO claSsical French)

<https://doi.org/10.1515/zrp-2023-0046>

Abstract: This article presents a project focusing on French private letters written during the 17th and 18th centuries. As the first French initiative to ever investigate this collection, the project aims to show how the alternative data provided by these letters can broaden the scope by filling the gaps left by traditional historical linguistics, focusing on one of French's greatest periods of expansion and shedding new light on the dynamics and mechanisms that led to the existing French and creole varieties in America and in the Indian Ocean.

Keywords: Classical French, private letters, historical sociolinguistics, French varieties, Creole varieties

Mots-clés : français classique, lettres privées, sociolinguistique historique, variétés du français, variétés de créole

1 Présentation

La standardisation du français, au sens historique du terme, tire son origine d'une tendance centripète faite de lieux récepteurs et de centres de gravité agissant comme des épicentres. Mais il est des secteurs où cette attitude se trouve concurrencée par une autre, centrifuge, qui a eu et qui aura toujours des répercussions sur

Promu par: Ce projet bénéficie du soutien de l'Agence nationale de la recherche (ANR-22-CE54-0005-01) dans le cadre du programme « Jeunes chercheuses et jeunes chercheurs » (2023–2026). **Nous remercions André Thibault ainsi que les évaluateurs de la revue de leur relecture attentive de cet article et des précieuses remarques dont ils nous ont fait bénéficier.** Cette phrase en rose devrait figurer sur la page 2 à la note 1.

Adresse de correspondance : Myriam Bergeron-Maguire, Maître de conférences, Sorbonne Nouvelle, 8 avenue Saint-Mandé Paris F-75012,
E-Mail : myriam.bergeron-maguire@sorbonne-nouvelle.fr

les comportements langagiers. Les milieux conservateurs se caractérisent par une attitude typiquement centrifuge, réticente à l'adoption de valeurs exogènes au profit de valeurs identifiées comme étant autochtones. La communication entre interlocuteurs intimes représente elle aussi un lieu par excellence de l'attitude centrifuge, où l'intimité et la complicité se manifestent notamment par l'emploi de variantes géographiquement restreintes, pouvant être éventuellement caractérisées par des restrictions sur d'autres axes (diaphasique ou diastatique).¹

C'est dans l'optique de documenter un secteur de la variation diaphasique – l'« immédiat communicatif » (Koch/Oesterreicher 2001, 586) – et la variation diatopique – les variantes régionales – que le projet MACINTOSH (Missing hAlf the picture, ClassIcal NoT sO claSsical French) a vu le jour. Il s'occupe du français des 17^e et 18^e siècles, un intervalle temporel pour lequel les caractéristiques concernées par ces deux types de variation sont relativement peu documentées, en comparaison avec les époques qui ont précédé ou avec celles qui ont suivi (Bergeron-Maguire 2018, 3 ; 2022a). Le choix de cette strate chronologique n'est pas le fruit du hasard, puisqu'il s'agit d'une période-clé pour documenter la genèse des variétés expatriées du français² et des créoles à base lexicale française, correspondant à la première grande vague d'expansion coloniale de la France vers l'Amérique et l'Océan Indien. L'analyse permettra ainsi de mieux connaître le français tel qu'il s'actualisait dans les régiolectes pratiqués par les contingents de population partis s'installer outremer au cours des 17^e et 18^e siècles.

La documentation de première main sur laquelle s'appuie le projet fait partie d'un fonds bien connu des historiens, les « Prize papers », conservé aux Archives nationales du Royaume-Uni. Les « Prize papers » ont été rassemblés par le tribunal de l'amirauté britannique (la High Court of Admiralty), auquel fut confiée la juridiction maritime jusqu'à la fin du 19^e siècle. L'une des missions de ce tribunal était de légiférer sur la guerre de course, une activité lucrative consistant à affaiblir les nations rivales en capturant leurs navires et en se saisissant de leurs cargaisons. Les marchandises saisies étaient revendues aux enchères et les recettes étaient ensuite réparties entre les corsaires, l'équipage et la couronne. La documentation transportée sur les navires interceptés, qui se chargeaient en même temps d'acheminer le courrier entre les empires et leurs colonies, était elle aussi confisquée,

1 ■■■ Mettre la phrase figurant en rose sur la page précédente ici à la place des trois carrés noirs.

2 « Les français d'Amérique sont des variétés expatriées d'une langue européenne [...]. Il s'agit plus précisément de variétés de langue maternelle transmises par des locutrices et locuteurs natifs de génération en génération, ce qui les distingue des variétés résultant de l'acquisition par la communauté linguistique d'une langue étrangère (cas de figure bien représenté au Maghreb ou en Afrique subsaharienne, où le français est une langue d'importation – par contraste, on pourrait dire que les français expatriés sont des variétés d'exportation » (Martineau/Remysen/Thibault 2022, 26–27).

puisqu'elle était requise comme pièce à conviction par le tribunal. Sans documentation, la guerre de course s'exposait à de sévères sanctions pécuniaires ou pire, à être condamnée pour piraterie.

D'après une équipe d'historiens de l'Université Carl von Ossietzky d'Oldembourg qui s'occupe de cataloguer le fonds (cf. Freist 2018–), les « Prize papers » contiendraient environ 160.000 lettres rédigées entre 1652 et 1817. D'après cette équipe, ce sont dix-neuf langues qui y seraient représentées ; lors de notre consultation du fonds, nous avons pu nous assurer personnellement de la présence de lettres écrites en français (très nombreuses), en espagnol, en portugais, en néerlandais, en allemand et, chose exceptionnelle, en basque.³ Les courriers ont été interceptés par la marine britannique sur plus de 30.000 navires qui effectuaient la traversée de l'Atlantique dans un sens ou dans l'autre ; y sont donc incluses des lettres ayant été rédigées dans les colonies nord-américaines et antillaises qui venaient d'être fondées. Les destinataires de ces courriers étaient le plus souvent des proches demeurant dans une ville portuaire de l'ouest de la France (cf. infra 3). D'autres lettres ont été envoyées depuis l'un de ces ports vers le Canada, la Louisiane ou les Antilles. Généralement brèves, les lettres comportent tout au plus deux folios.

Les linguistes sont peu nombreux à avoir examiné ce fonds. Les seuls à s'y être intéressés de près, à notre connaissance, sont des sociolinguistes hollandais qui ont fourni l'édition et l'analyse d'environ 2.000 lettres en néerlandais rédigées aux 17^e et 18^e siècles (Rutten/van der Wal 2014). Aucune étude linguistique d'ensemble n'a encore porté sur les lettres en français. Le présent projet, qui rassemble une équipe constituée de linguistes et de spécialistes des humanités numériques, entend étudier cette documentation inédite à partir d'un échantillon sélectionné selon des critères dont nous donnerons à présent le détail.⁴

2 Échantillon : critères de sélection

Le projet prévoit de s'appuyer sur un échantillon de 1.000 lettres faisant partie de ce fonds. Les lettres sélectionnées pour en faire partie contiennent toutes au moins une caractéristique dont on peut déterminer qu'elle fut exclue de la culture écrite à une date précoce, qu'elle n'en a jamais fait partie ou qu'elle demeure rarement attestée. Cette typologie tripartite inclut des caractéristiques diastatiquement marquées, diaphasiquement marquées ainsi que des faits diatopiquement restreints.

³ La documentation ancienne en basque est pour le moins fragmentaire et éparse (Coyos 2013, 432), ce qui rend ce petit ensemble d'environ 80 lettres extrêmement précieux.

⁴ La composition de l'équipe peut être retrouvée sur le site du projet hébergé par la TGIR HumNum : lettresoutremer.huma-num.fr [dernière consultation : 26.03.2023].

L'objectif est en outre d'obtenir un échantillon qui permette d'identifier des phénomènes linguistiques partagés – en isolant les idiolectalismes, puisque nous ne cherchons pas des faits individuels – et de les décrire le plus précisément possible en les contextualisant dans l'espace géographique et social. Cela requiert de constituer un échantillon composé de lettres écrites par des hommes et des femmes, issus de plusieurs milieux sociaux et socioprofessionnels et originaires de plusieurs lieux. Avec ses dizaines de milliers de lettres, le vaste fonds des « Prize papers » est particulièrement adapté pour répondre à ces critères, ménageant même la possibilité de construire un ensemble relativement calibré.

Deux points compromettront néanmoins la représentativité de l'échantillon. Les lettres du fonds étant le plus souvent rédigées dans des localités de l'ouest de la France ou d'outre-mer, il faudra tenir compte du fait que les caractéristiques occidentales d'une part, ou originaires du Nouveau Monde d'autre part, pourront être surreprésentées. Par ailleurs, nous prévoyons de profiter de ce que le fonds contient un nombre important de lettres rédigées par des peu-lettrés en privilégiant ces dernières pour la constitution de l'échantillon ; en effet, les corpus de français écrit de cette période n'en contiennent généralement pas, hormis quelques exceptions notables récentes (Martineau 2005 ; Ernst 2019).

Par « peu-lettré », nous entendons un scripteur présentant un faible degré de familiarité avec la culture écrite, ayant reçu une instruction sommaire, tel que le conçoivent Rutten/van der Wal (2018, 238) à propos de lettres en néerlandais tirées des « Prize papers » : « The concept refers to the extent to which language users participated in the written culture, particularly to the importance of reading and writing in everyday life. In addition, literacy and writing experience depended on schooling opportunities, and differed across gender and social rank ». L'intérêt pour les peu-lettrés s'est développé d'abord en Italie, où les écrits produits par ces individus ont été étudiés notamment pour les interférences (lexicales, grapho-phonétiques) avec les variétés dialectales italo-romanes qu'ils contiennent (Spitzer 1921). Les lettres que nous avons examinées jusqu'à présent contiennent elles aussi de telles **interférences**, mais recèlent aussi des archaïsmes du français, des néologismes du Nouveau Monde ainsi que les habituelles pratiques de l'écrit qui ne correspondent à aucune réalité phonique connue (hypercorrections, habitudes idiolectales). Le concept de « peu-lettré », ses dénominations et ses applications selon les langues ayant déjà été étudiés en détail dans une analyse remarquable (Sowada 2021, 18–24), nous nous contenterons d'ajouter à ces vues quelques considérations concernant la maîtrise des conventions écrites par nos scripteurs (cf. infra 4).

Ajouter
(galloromanes)
entre
parenthèses ici
après le mot
interférences

3 Lieux d'écriture et profils des écrivains et de leurs destinataires

Les lettres interceptées par les Britanniques au milieu de l'Atlantique (ou, le plus souvent, près des côtes) n'ont naturellement jamais pu être remises à leurs destinataires. Plusieurs lettres du fonds sont d'ailleurs toujours cachetées.⁵ La plupart des lettres ont néanmoins été ouvertes et lues par les magistrats de l'amirauté anglaise ; la guerre de course servait certes à s'enrichir en interceptant des navires marchands étrangers, mais elle était aussi un moyen privilégié pour espionner les nations rivales à travers le contenu des courriers envoyés.⁶

Les lieux d'écriture sont, dans l'immense majorité des cas, des villes de l'ouest et du nord-ouest de la France – principalement Bayonne, Bordeaux, La Rochelle, Nantes, Saint-Malo, Le Havre – ou encore des localités situées dans les jeunes colonies françaises de l'époque : à la Guadeloupe, en Martinique, à Saint-Domingue (actuel Haïti) et aux Mascareignes (La Réunion et l'île Maurice). Quelques lettres ont été écrites au Canada et en Louisiane, mais elles demeurent minoritaires, du moins en l'état actuel de nos connaissances du fonds. Précisons en outre que le lieu d'écriture d'une lettre et le lieu d'origine de l'écrivain coïncident parfois, mais pas toujours. Ce sont des indices contenus dans les courriers qui nous permettent de le déterminer, ces informations pouvant en outre être recoupées avec les lieux de destination de la lettre. Lorsqu'ils sont originaires de métropole, les auteurs proviennent d'aires urbaines ou de petites agglomérations de l'ouest et du nord-ouest de la France. Certains individus sont nés en Guadeloupe, en Martinique ou à Saint-Domingue, dans des localités peuplées de l'époque, à savoir Pointe-à-Pitre, Fort-de-France, le Cap-Français (aujourd'hui Cap-Haïtien), Port-au-Prince (et ses environs, notamment Léogâne).⁷

Souvent, les courriers sont destinés à un membre de la famille immédiate (mère, père, sœur, frère) ou sont échangés entre époux : on ne trouve pas plus intime (Koch/Oesterreicher 2001, 594). Les écrivains appartiennent en outre à des milieux socioprofessionnels relativement divers, sur lesquels on peut avoir des indications directement par le contenu des lettres ou indirectement à travers d'autres

5 La décision des Archives nationales du Royaume-Uni est de ne pas ouvrir ces lettres pour l'instant, puisqu'il est prévu de détecter leur contenu sans les décacheter par fluorescence de rayons X. Celles-ci sont ainsi exclues *de facto* de notre échantillon.

6 Une lettre du gouverneur Caylus donne ainsi en 1748 le nombre de corsaires français en mer et le nombre de leurs canons (Villiers 2014, 83).

7 Nous n'avons étrangement pas trouvé pour l'heure de lettres rédigées à Saint-Pierre (Martinique), qui était pourtant l'agglomération la plus peuplée de l'île avant l'éruption volcanique du début du 20^e siècle.

courriers saisis sur le même navire. Ce sont les métiers de capitaines, de marins, de commerçants, de tonneliers ou de gérants d'habitation qui sont le plus souvent mentionnés. À cette liste, on peut ajouter les négriers, qui ne se déclarent toutefois jamais comme tels. La traite négrière et le détail de ses circonstances sont bien mentionnés à quelques reprises dans les lettres, mais subrepticement.

Contrairement aux corpus de français écrit existant pour cette période, les lettres rédigées par des femmes ne sont pas en petit nombre : jusqu'à présent, environ le tiers des lettres retenues pour faire partie de l'échantillon proviennent de femmes. Celles-ci sont des veuves, des dames de compagnie, des épouses de propriétaires ou de petits gérants et commerçants en tous genres. Lorsqu'elles sont mariées à un propriétaire ou à un gérant, il n'est pas rare qu'elles prennent part à la vie de l'habitation,⁸ notamment en veillant à l'éducation des enfants, qu'il s'agisse des leurs ou de ceux des esclaves. Occasionnellement, lorsqu'elles ont quelques notions d'arithmétique, elles participent au commerce des denrées (sucre, tabac, indigo) en correspondant avec la clientèle métropolitaine.

Enfin, dans les études sur les origines des variétés du français en Amérique, l'intérêt s'est d'abord porté sur la part de la population s'étant fixée durablement dans les colonies de l'époque : une fois l'Atlantique traversé, ces individus formant le fonds stable de la démographie locale se sont installés dans les habitations pour y mener une vie rythmée par l'agriculture et l'hivernage. Or, cela ne doit pas faire oublier que la mobilité n'est pas rare dans l'histoire coloniale : plusieurs courriers mentionnent des individus provenant des îles passant une partie de leur vie en France, ayant été confiés pour leur éducation à un membre de la famille ou à une communauté religieuse en métropole (Bergeron-Maguire 2020, 51). Certains choisissaient de demeurer ensuite en France, tandis que d'autres retournaient aux îles à la demande de la famille. Inversement, les métropolitains qui venaient aux îles n'y séjournaient parfois que quelques mois ou quelques années avant de rentrer en France, le plus souvent ruinés par les révoltes des esclaves et par les guerres successives opposant les puissances coloniales dans l'aire caribéenne de l'époque. Ce va-et-vient a certainement contribué à maintenir des liens entre les jeunes colonies d'Amérique et la métropole – jusqu'à ce qu'une majorité d'entre elles passent aux mains des Anglais après la guerre de Sept Ans –, et ces individus mobiles sont assurément à considérer comme des vecteurs dans la diffusion de caractéristiques linguistiques d'origine métropolitaine. Certaines de ces caractéristiques sont, sans surprise, des diatopismes (Bergeron-Maguire 2020, 56).

⁸ Typique des systèmes coloniaux, cette dénomination renvoie non seulement au domaine agricole, mais également à l'ensemble des bâtiments et des installations qui faisaient corps avec l'exploitation rurale, selon la taille des domaines (Rézeau/Rézeau 1995, vol. 2, 358).

4 Bilinguisme des écrivains

L'usage du français comme langue écrite est tout à fait exclusif dans les lettres examinées. Néanmoins, pour plusieurs des individus à l'origine de ces courriers, on peut supposer que la langue qu'ils employaient quotidiennement n'était pas le français. Cela n'étonne guère, puisque les aires urbaines, petites et grandes, de la France occidentale et nord-occidentale des 17^e et 18^e siècles se caractérisaient toutes par une situation de bilinguisme, voire de plurilinguisme, qui touchait, à des degrés divers, l'ensemble de l'espace social (Chauveau 2011, 67).

La coexistence de systèmes linguistiques se manifeste dans plusieurs lettres où s'observent des phénomènes de contact langagier tels que des emprunts ou des calques, dont la présence est imputable à un système linguistique pouvant être plus ou moins apparenté au français selon les lieux. Certains faits de contact relevés sont dus à l'occitan, d'autres au normand ; ils peuvent être anciens, voire très anciens, et donc être passés en français régional à une date précoce ou encore être relativement récents. Dans le cas des faits de contact récents, leur statut dans les lettres est moins assuré : s'agit-il d'idiolectalismes (c'est-à-dire de faits dialectaux ponctuels « habillés en français »), de xénismes ou bien d'emprunts bien acclimatés n'ayant simplement jamais franchi la barrière médiale du code graphique auparavant ? Nous présenterons (cf. ci-dessous 6.1) un cas concret pour lequel la question reste en suspens (*mouscle*) et un autre cas dont on sait qu'il s'agit d'un emprunt stabilisé, en raison de la postérité qu'il a connue (*cagoule*).

Quant au français écrit et à ses conventions, on observe un degré de maîtrise qui varie beaucoup d'un individu à l'autre et qui n'est pas toujours corrélé avec une éventuelle appartenance sociale prestigieuse. La maîtrise est évaluée en se fondant sur la régularité de la calligraphie, sur la cohérence de l'orthographe et sa conformité avec les modèles contemporains – qu'on atteint par les imprimés, les dictionnaires et les éditions de correspondances qui ne modernisent pas la graphie – ainsi que sur la concordance entre la segmentation des mots et les unités lexicales. Évidemment, ces caractéristiques peuvent être plus ou moins bien représentées d'un scripteur à l'autre, et les pratiques peuvent être très hétérogènes : un scripteur qui présente une calligraphie remarquablement régulière peut être très peu au fait des frontières des mots graphiques et vice versa. Les profils ne se laissent ainsi pas départager aisément et peuvent être reportés, en réalité, sur un continuum où les notions de « lettré » et de « peu-lettré » n'ont pas de valeur catégorisante dans l'absolu (Branca-Rosoff/Schneider 1994, 9).

5 L'autographie des documents

La question de savoir si les lettres sont autographes ou non est un problème auquel nous sommes régulièrement confrontés. Aux 17^e et 18^e siècles, le recours à une tierce personne plus expérimentée que soi – un écrivain public, un ecclésiastique, un instituteur ou un membre lettré de la famille – n'est pas rare lorsque l'on doit donner de ses nouvelles ou s'acquitter d'une démarche administrative (Graff 1987, 173 ; Mitchell 2012, 230).

Par principe, seules des lettres dont nous aurons pu établir le caractère autographe feront partie de l'échantillon. Pour les identifier, nous nous appuyons sur des indices contextuels : une lettre autographe est aisément reconnaissable lorsque l'expéditeur mentionne le processus d'écriture (en mentionnant, par exemple, une blessure à la main gênant l'activité de rédaction) ou se présente explicitement comme étant le scripteur. Dans les lettres non autographes, le secrétaire se signale très souvent comme servant de truchement, en transmettant par exemple au destinataire ses salutations, souvent soigneusement distinguées de celles de l'expéditeur. Enfin, la calligraphie pourra elle aussi permettre de déterminer si l'on a affaire à un courrier autographe ou non : nous comparerons notamment la forme des graphèmes et les signatures d'une lettre à l'autre. Plusieurs lettres envoyées par plusieurs individus rédigées par une seule et même main peuvent être considérées comme non autographes ; la conclusion sera identique pour des lettres envoyées par une seule et même personne, rédigées par des mains différentes.

6 Premiers résultats

Afin de donner un aperçu des matériaux contenus dans l'échantillon (en cours de constitution à l'heure actuelle), nous présentons dans ce qui suit quelques-unes des caractéristiques relevées au fil des dépouillements effectués. S'ajoutent aux phénomènes mentionnés ici ceux qui ont été traités ailleurs, qui touchent le lexique, la morphologie (nominale et verbale) et la syntaxe (Bergeron-Maguire/Greub 2020 ; Bergeron-Maguire 2021 ; 2022b).

6.1 Caractéristiques lexicales

Nous signalerons dans cette section deux diatopismes typiques du sud-ouest de la France. Ceux-ci se lisent dans une même lettre adressée en 1778 depuis La Tremblade (situé non loin de La Rochelle) par la « veuve Elber » à son gendre Pierre Bourdron, voilier de profession, parti pour Saint-Domingue (aujourd'hui Haïti).

- (1) *je ta voit des **mouscle** elle ne sont pas mauvesse au si que des suitre des **cagoule** dans un petict pa nies* 'je t'envoie des moules elles ne sont pas mauvaises aussi que des huîtres [et] des cagouilles [escargots] dans un petit panier'
1778, La Tremblade, Veuve Elber à son gendre Pierre Bourdron (HCA 30/32, Temporary Sorting Box 105)

6.1.1 *mouscle* s.f. 'mollusque bivalve comestible, **moule**' à mettre en italique

La forme *mouscle* pour *moule* est bien enregistrée dans la lexicographie française (Furetière 1690–Trévoux 1752, von Wartburg 1966 in FEW 6/3, 261b, mŪscŪlus I 2 a), mais tout pousse à croire que nous avons affaire en réalité à deux types lexicaux distincts, l'un [mukl] – qui est celui qui est attesté ici (1) – et l'autre [mul]. Dans les deux cas, <s>, depuis longtemps amuï, est purement graphique. La variante [mukl] se trouve dans les glossaires dialectaux occidentaux, et notamment saintongeais, depuis 1780 jusqu'au 20^e siècle. Elle figure en outre dans une source nantaise qui est un recueil de français régional daté de 1820 (FEW 6/3, 261b, mŪscŪlus I 2 a). L'attestation relevée par nous pourrait suggérer que l'aire de ce type dialectal emprunté en français régional s'étendait jadis sur environ 200 kilomètres au sud, depuis Nantes jusqu'à La Tremblade. Mais, en l'absence d'attestations supplémentaires en dehors des sources dialectales, on ne peut pas non plus écarter la possibilité qu'il puisse s'agir d'un fait dialectal (donc, d'un fait ponctuel d'emprunt francisé) ayant exceptionnellement franchi la barrière médiale de l'écrit sous la plume de cette veuve.

adapter la
taille
des
caractères

à mettre en
italique

6.1.2 **cagoule** s.f. 'mollusque terrestre à coquille arrondie en spirale, **escargot**' à mettre en italique

Ce substantif est un emprunt à l'occitan – où le type lexical existe à date ancienne (aocc. *cogolha*, non daté, von Wartburg 1945 in FEW 2, 1453a, cŪcŪlla II 1) – relevé dans la lexicographie en 1611 dans le dictionnaire de Cotgrave, où il figure sans marque diatopique.⁹ Il s'agit pourtant sans conteste d'un diatopisme, au vu de sa rareté dans la documentation et de l'aire moderne qu'il occupe (dans le sens 'escargot petit-gris, *Helix aspersa*) : *cagouille*, selon la forme graphique moderne, est em-

9 Le dictionnaire de Cotgrave contient un nombre de diatopismes qui ne sont pas systématiquement signalés comme tels. La préface (1611, 4), rédigée en français par un dénommé L'Oiseau de Tourval, souligne le grand intérêt bibliographique de l'auteur pour les « dialectes », notamment : « Et t'assure que si on l'eut voulu croire, il fût encore apres à se tourmenter, pour trouver la signification de telz mots, qui, possible, ne seront jamais plus ouyz en ce monde, (quoy que luz) et dont, ie croy, il n'y a personne qui ait ouy parler depuis cent ans, que luy ; tant sa curiosité a été grande et exacte à lire toute sorte de livres, vieuz et nouveaux, et de tous noz dialectes ».

blématique des Charentes (DRF). Le lexème réapparaît près de trois siècles plus tard, en 1877, dans le *Supplément* du dictionnaire de Littré, et figure ponctuellement dans la lexicographie contemporaine, toujours accompagné d'étiquettes géolinguistiques plus ou moins appropriées (cf. DRF s.v. *cagouille*).

Le lexème s'est dérégionalisé en passant dans un vocabulaire technique et dans celui de la mode. D'abord, il a pris, par métaphore, dans le vocabulaire maritime, le sens de 'motif d'ornementation formé d'un enroulement en spirale, volute d'un vaisseau', présent dans toute la lexicographie française du 17^e au 20^e siècles (cf. TLF s.v. *cagouille*). Sa popularité aux Antilles pour désigner un ornement porté par les femmes dans leurs cheveux lui vaut également de figurer dans le Larousse vers la fin du 19^e siècle : « Entom[ologie] Nom vulgaire d'un petit ver luisant des Antilles, que les femmes de ce pays mettent comme ornement dans leurs cheveux » (Larousse 1868¹⁰ s.v. *cagouille*).¹¹

6.2 Caractéristiques grapho-phonétiques

6.2.1 Représentations écrites de consonnes finales prononcées

Les courriers examinés jusqu'à présent contiennent plusieurs attestations pouvant témoigner tantôt de la prononciation de consonnes finales, tantôt de leur amuïssement (Bergeron-Maguire/Greub 2020). En voici une suggérant la restitution d'un [t] contenue dans une lettre adressée en 1778 depuis La Tremblade par Anne Elbert à son époux se trouvant en mer vers Saint-Domingue :

- (2) *mr madchaux et mademoislle vimaux a fer lagroche avan la minuite osi que un un bataille et mdemoiselle gorgi gille au non faite autain* 'Monsieur Madchaux et Mademoiselle Vimaux à faire la grâce [?] avant la minuit aussi que un bataille et Mademoiselle Gorgi Gilles en ont fait autant'
1778, La Tremblade, Anne Elb[ert] à son mari Pierre Bourdron à bord du navire *Le Prudent* (HCA 30/32, Temporary Sorting Box 105)

La restitution d'un [t] final dans la prononciation sous l'influence de la graphie (Buben 1935) est ancienne. Le phénomène a caractérisé le français du centre-ouest de la France aux 16^e et 17^e siècles (Chauveau 2008), d'où il s'est exporté outre-Atlantique.¹²

¹⁰ La datation du volume 3 est fondée sur la chronologie établie par Rétif (1973, 111).

¹¹ Ce sens a vraisemblablement été obtenu par transfert co-hyponymique : l'appellation de l'escargot est recyclée pour désigner le ver luisant.

¹² Pour certains lieux, on ne peut pas exclure qu'il puisse s'agir d'un phénomène de conservatisme, sans qu'il soit nécessaire de supposer une influence de la graphie ou une influence citadine, comme

Cette caractéristique phonétique n'affecte pas tout le lexique. Pour le type lexical *minuit*, les exemples sont relativement continus et anciens : la forme graphique *minuite*, attestée dès 1536 sous la plume du poète poitevin Jean Bouchet (Gdf = von Wartburg 1953 in FEW 7, 215b, *nōx* I 1), indique sans aucun doute une prononciation avec [t] final. Les formes dialectales occidentales relevées au 19^e siècle suggèrent également une réalisation de la consonne finale (cf. Loches *ménuitte*, Saintonge *minœt*, etc. ; FEW 7, 215b). Outre-Atlantique, cette réalisation phonétique est attestée chez des témoins canadiens nés au début du 20^e siècle, mais elle y était déjà en perte de vitesse dans les années 1970 (ALEC 6, 1717). Aux Antilles, seules les îles les plus archaïsantes ayant été coupées de la France à une date précoce connaissent **cette** la prononciation avec consonne finale (ALPA 197 ; Thibault/Avanzi à paraître).

6.2.2 Notations du résultat [ɛ] issu de l'ancienne diphtongue [ei]

C'est probablement du résultat [ɛ] issu de l'ancienne diphtongue [ei] dont témoigne la forme *neyies* part. passé 'noyé' attestée dans une lettre écrite en 1778 à La Tremblade par Marie Logelle à son fils :

- (3) *je te dirais que monsieur mercier sest perdu et il set neyies le pauvre homme* 'je te dirai que Monsieur Mercier s'est perdu et il s'est noyé le pauvre homme'
1778, La Tremblade, Marie Logelle à son fils Pierre Bourdron à bord du navire *Le Prudent* (HCA 30/32, Temporary Sorting Box 105)

Caractéristique du domaine occidental, ce trait s'est diffusé outre-mer, où il se maintient encore de nos jours dans quelques lexèmes, en français canadien et en créole antillais : *drèt* adj./adv. 'droit', *fret* adj. 'froid' (DECA I/1 ; Thibault 2011, 12), ce dernier aussi présent aux Mascareignes (sans [t] final, DECOI I/2).

Dans les lettres du corpus, le trait a été relevé en outre pour des formes du subjonctif du verbe *être* (3^e pers. sg. *say* 'soit'), pour des formes au présent de l'indicatif du verbe *croire* (1^{re} pers. sg. *cre* 'crois') ainsi que pour le pronom personnel tonique de la 1^{re} personne du singulier (*me* 'moi') (Bergeron-Maguire/Greub 2020).

aux îles Anglo-Normandes, en Vendée et dans quelques petites îles de la côte atlantique, cités par Chauveau (2008, 79).

6.3 Caractéristique morphologique

Un phénomène de morphologie grammaticale qui est extrêmement fréquent dans nos lettres est l'emploi du grammème *point* comme adverbe de négation. Compte tenu de son déclin avéré au 18^e siècle (au profit de *pas*) et de l'aire restreinte qui le caractérise au siècle suivant (Picardie et Ouest de la France en Europe ainsi que le Nouveau Monde, cf. Martineau 2005, 195 ; Thibault 2020, 423–424), il faut probablement considérer les attestations dans les extraits (4), (5) et (6) comme des maintiens à mettre au compte de la région d'origine de l'autrice.

- (4) *je te dirais que nous ne savons **point** la ou il et est [sic] le pauvr barraux* 'je te dirai que nous ne savons point là où il est le pauvre Barreau'
1778, La Tremblade, Marie Logelle à son fils Pierre Bourdron au Cap Français à Saint-Domingue (HCA 30/32 Temporary Sorting Box 105)
- (5) *il est dans la fregade la jeu non coumand par monsieur des touche nous ne savons **point** la ou il vont* 'il est dans la frégate la Jeunon commandée par Monsieur Des Touches nous ne savons point là où ils vont'
1778, La Tremblade, Marie Logelle à son fils Pierre Bourdron au Cap Français à Saint-Domingue (HCA 30/32 Temporary Sorting Box 105)
- (6) *nous ne savons **point** si il sont a brest* 'nous ne savons point s'ils sont à Brest'
1778, La Tremblade, Marie Logelle à son fils Pierre Bourdron, au Cap Français à Saint-Domingue, HCA 30/32 Temporary Sorting Box 105)

6.4 Caractéristique syntaxique

Nous terminerons ce bref aperçu par un trait syntaxique. En français contemporain oral, le double marquage du sujet est courant (Bauche 1929, 154 ; Thibault 2020, 422), mais les exemples anciens ne sont pas légion. Dans les correspondances examinées, le fait franchit régulièrement la barrière médiale de l'écrit, souvent dans des dislocations à droite (Bergeron-Maguire 2022b), comme on peut l'apercevoir dans cet exemple :

- (7) *je te dirais que nous ne savons point la ou **il** et est [sic] **le pauvr barraux*** 'je te dirai que nous ne savons point là où il est le pauvre Barreau'
1778, La Tremblade, Marie Logelle à son fils Pierre Bourdron au Cap Français à Saint-Domingue (HCA 30/32 Temporary Sorting Box 105)

7 Méthodologie : transcription, édition, annotation

Les caractéristiques relevées doivent pouvoir être attestées dans plus d'une lettre, pour des raisons méthodologiques et sociologiques. D'un point de vue méthodologique, l'objectif du projet n'est pas de relever des hapax et des idiolectalismes, mais d'identifier des caractéristiques partagées appartenant aux régiolectes pratiqués par les premiers contingents francophones venus en Amérique. Comme les lettres examinées recèlent un nombre important de traits attestés pour la première fois, il est souhaitable que ceux-ci, rares ou absents ailleurs dans la documentation, soient attestés plus d'une fois, idéalement chez plus d'un scripteur. D'un point de vue sociologique, certaines caractéristiques peuvent être corrélées avec des facteurs sociaux comme l'origine géographique, l'appartenance à un milieu socioprofessionnel ou à une couche sociale. Il n'est pas sûr que nous parvenions à tirer des conclusions définitives au sujet de ces variables à partir d'un échantillon de 1.000 lettres, relativement réduit par rapport aux dizaines de milliers de lettres françaises du fonds.¹³ Mais nous estimons qu'il sera à tout le moins possible de dégager certaines tendances. Pour fixer les idées sur les résultats qu'on peut en attendre, une étude portant sur la liaison dans un sous-échantillon composé de cinquante lettres a permis d'identifier quinze lettres dans lesquelles la liaison avait été transposée par des moyens n'ayant jamais fait partie de la tradition écrite (imprimée ou manuscrite) issue des milieux les plus éduqués (Bergeron-Maguire 2021). Le ratio observé, où l'on distingue trente-cinq lettres se conformant à la culture écrite hautement codifiée et quinze lettres s'en éloignant radicalement, suggère que pour certaines variantes, l'identification d'une stratigraphie sociale peut être envisageable.

Ces considérations d'ordre méthodologique entraînent le besoin d'atteindre, dans l'échantillon, l'ensemble des attestations pour une même caractéristique, ce qu'une interface d'interrogation du corpus permettra à terme de faire. Les lettres seront d'abord transcrites à partir de reproductions numériques réalisées par des membres du projet aux Archives nationales du Royaume-Uni, en recourant au logiciel eScriptorium.¹⁴ L'interface d'eScriptorium permet de transcrire manuellement tout en maintenant le lien entre le texte transcrit et l'image du document. Le logiciel produit automatiquement une version encodée TEI de la transcription effectuée, à laquelle il est ensuite possible d'ajouter des métadonnées. Pour chacune des lettres,

¹³ Certaines de ces données peuvent en outre être difficiles à identifier chez les scripteurs.

¹⁴ L'interface d'eScriptorium est mise à disposition par l'Institut national de recherche en sciences et technologies du numérique (INRIA) : <<https://escriptorium.paris.inria.fr>> [dernière consultation : 26.03.2023].

on rassemblera les informations suivantes : cote de classement, destinataire, destinataire, lien (familial, amical, etc.) entre les deux, provenance, destination, date, anthroponymes et toponymes mentionnés. Une identité propre sera attribuée à chacune des lettres du corpus en se fondant sur les métadonnées rassemblées : cote du fonds, nom de famille du destinataire, année, recto (F) et verso (B), ex. H32Dupaty1742F.

L'interface d'eScriptorium permet en outre d'afficher une édition diplomatique et une édition interprétative. L'édition diplomatique sera réalisée sans aucune modernisation graphique, préservant les segmentations de mots et les retours à la ligne. La ponctuation et les diacritiques, généralement rares, seront maintenus lorsqu'ils apparaissent dans les lettres et ne seront pas ajoutés lorsqu'ils n'y figurent pas. La méthode diplomatique a vocation à rester le plus fidèle possible aux textes, ménageant la possibilité d'effectuer des analyses du code écrit (graphématiques, par ex.). L'édition interprétative, par définition plus interventionniste, permettra de rendre le corpus interrogeable pour le futur outil d'interrogation du corpus. Elle comprendra un appareil critique, un résumé du contenu de la lettre et une interprétation du contexte de la correspondance, des notes visant à désambiguïser certains passages ou à rapporter des observations sur des pratiques idiolectales, par exemple.

Dans l'objectif de rendre le corpus interrogeable, il est ensuite prévu de le tokeniser, de le lemmatiser et de l'annoter au moyen d'un jeu d'étiquettes morphosyntaxiques. La tokenisation sera accomplie manuellement au moyen d'un logiciel d'édition XML. Une fois tokenisé, le corpus sera lemmatisé automatiquement avec la plateforme de lemmatisation LGeRM,¹⁵ qui sera entraînée et adaptée à nos objectifs, aboutissant à un outil propre au projet, LGeRM PeuL (sigle confectionné à partir de « Peu-Lettrés »). La lemmatisation permettra de relier chacun des tokens définis à l'étape précédente à une entrée de la nomenclature du *Dictionnaire du Moyen Français* (DMF 2020), un ouvrage dont la fiabilité est reconnue et qui est disponible en ligne. Même s'il est vrai que la période couverte par le DMF (1330–1500) ne correspond pas à celle que nous étudions, des tests préliminaires effectués avec LGeRM ont révélé que sa nomenclature était bel et bien adaptée à nos besoins en termes de précision dans l'attribution automatique des lemmes et d'acceptabilité. Cela suggère d'ailleurs que les lettres présentent un certain degré d'archaïsme par rapport à la documentation (imprimée, essentiellement littéraire) dont on dispose à l'heure actuelle pour la période des 17^e et 18^e siècles. Il est fort probable, en revanche, qu'une

15 LGeRM : plateforme de lemmatisation de la variation graphique des états anciens du français élaborée par Gilles Souvay, CNRS-ATILF/Université de Lorraine : <<http://stella.atilf.fr/LGeRM/plateforme/>> [dernière consultation : 01.04.2023].

partie des tokens du corps ne trouve pas de lemme correspondant dans le DMF (on pense en particulier aux néologismes apparus en lien avec le besoin de nommer des réalités provenant du Nouveau Monde) ; pour ces cas, un lemme sera créé en se fondant sur d'autres dictionnaires (étymologiques) reconnus comme le *Trésor de la langue française* (TLF) ou le *Französisches Etymologisches Wörterbuch* (FEW). Si un signe linguistique donné n'apparaît dans aucun des dictionnaires précédemment mentionnés, un nouveau lemme sera créé. En outre, pour chacun des lemmes créés, la forme choisie reposera, lorsque cela sera possible, sur des variantes proches pouvant être relevées dans des bases textuelles telles que Frantext. Une fois la lemmatisation complétée, les résultats seront vérifiés manuellement. Ensuite, les différentes sources des lemmes seront comparées en termes de proportions (pour chacun des dictionnaires) et une attention particulière sera portée aux lemmes entièrement nouveaux : dans quelle mesure avons-nous dû forger de nouveaux lemmes, ne trouvant pas de correspondant dans les dictionnaires (étymologiques) usuels du français et des dialectes galloromans ? Une liste générée automatiquement pourra répondre rapidement à cette question, permettant de mesurer la nouveauté apportée par les lettres éditées et leur intérêt pour notre connaissance de l'histoire du lexique français.¹⁶ Nous pouvons en outre prédire que plusieurs de ces nouveaux lemmes représenteront des caractéristiques régionales ou dialectales, n'apparaissant pas ou n'apparaissant que peu dans l'écrit formel hautement codifié.

Le corpus lemmatisé sera ensuite annoté à l'aide d'un jeu d'étiquettes morphosyntaxiques inspiré des entités définies dans le jeu d'étiquettes Cattex mis au point pour la Base de français médiéval (BFM) développée à l'ENS de Lyon (cf. Guillot-Barbance/Heiden/Lavrentiev 2017).¹⁷ Ce protocole d'annotation permettra d'effectuer des analyses lexicales étendues ou des recherches en lien avec la morphologie flexionnelle ou la syntaxe.

Le corpus édité sera rendu accessible en ligne en libre accès. Les transcriptions seront disponibles en ligne au format TEI sur la plateforme digitale TEI Publisher, qui permet une réutilisation et une pérennité des données. TEI Publisher présente en outre l'avantage de permettre d'adapter l'affichage du texte en utilisant un code relativement simple et donne la possibilité de télécharger les transcriptions au format PDF. La collection des textes édités sera également déposée en fin de projet sur l'entrepôt de données Nakala, afin de donner la possibilité à la communauté de

¹⁶ Cette nouveauté sera toutefois sous-estimée, dans la mesure où les lettres pourront attester des unités lexico-sémantiques absentes des dictionnaires consultés (nouveaux sens de vocables déjà répertoriés), sans qu'il n'ait été besoin de forger de nouveaux lemmes.

¹⁷ La liste des étiquettes morphosyntaxiques de Cattex (2009) est disponible sur le site de la BFM de l'ENS de Lyon : <<http://bfm.ens-lyon.fr/spip.php?article176>> [dernière consultation : 26.03.2023].

télécharger les données rassemblées et de les réorganiser pour d'autres objectifs de recherche.¹⁸

Enfin, le corpus pourra être interrogé sur la plateforme d'interrogation BlackLab, utilisée par le *Corpus latin Antiquité et Antiquité tardive lemmatisé* (Clérico 2020) ainsi que par le projet hollandais *Letters as loot* (Rutten/van der Wal 2008–2013), portant sur des lettres en néerlandais tirées du fonds des « Prize papers ».¹⁹ Le formulaire d'interrogation permettra d'interroger le corpus à partir d'une forme exacte, d'un lemme, d'une catégorie grammaticale ou de plusieurs des paramètres précédents, au moyen d'un langage CQL et d'un processus de recherche CQP. Un outil de recherche avancée permettra d'atteindre des résultats par l'intermédiaire des métadonnées du corpus, en sélectionnant une valeur prédéterminée dans un menu déroulant ou en effectuant une recherche par mots-clés. L'affichage des résultats inclura un cotexte de vingt mots après et avant le pivot (forme exacte ou lemme). Les métadonnées seront également visibles dans l'affichage des résultats, et un lien individuel pour chacune des lettres permettra d'atteindre l'image numérisée de la lettre disponible sur le portail des « Prize papers » mis au point par l'équipe du *Prize Papers Project* de l'Université d'Oldenbourg.²⁰ Nous travaillons à l'heure actuelle en collaboration avec cette équipe afin de relier nos bases de données, qui s'en trouveront mutuellement enrichies.

8 Conclusion

La méthodologie adoptée pour le projet MACINTOSH demeure somme toute très classique : édition numérique, annotation, outil d'interrogation. Hormis la confection d'un outil de lemmatisation propre à la documentation étudiée (LGeRM PeuL), la nouveauté apportée par le projet au plan des humanités numériques demeure assez réduite.

Ce sont plutôt, tel que nous sommes à présent en mesure de l'appréhender, les nouvelles données contenues dans cette documentation de première main qui pourront véritablement conduire à un saut qualitatif dans notre connaissance du français de cette époque. Les lettres examinées contiennent une grande quantité de phénomènes ayant fini par faire souche outre-mer dans une ou plusieurs variétés

¹⁸ Gérées par la TGIR Huma-Num, les données déposées dans Nakala sont stockées sur des serveurs hébergés en France : <<https://nakala.fr>> [dernière consultation : 28.03.2023].

¹⁹ BlackLab, open source corpus search engine, developed by the Dutch Language Institute, <<https://inl.github.io/BlackLab>> [dernière consultation : 26.03.2023].

²⁰ Le portail, régulièrement enrichi, est accessible à l'adresse suivante : <<https://portal.prizepapers.de/index>> [dernière consultation : 26.03.2023].

de français ou de créole, pour lesquels nous ne disposions parfois d'aucune donnée directe à ce jour, hormis, dans certains cas, des attestations contemporaines très récentes. Ces particularités ayant connu une postérité jusqu'à nos jours, défiant le temps et les mouvements démographiques, devaient certainement à l'époque couvrir des aires géographiques considérables et avoir été fréquentes non seulement dans le français parlé des strates socialement dominées de la population. Leur exclusion du répertoire français revêtu de toutes ses caractéristiques canoniques (graphiques, lexicales, morphosyntaxiques) a eu pour effet de les placer hors de notre vue.

Nous espérons que les résultats de ce projet permettront d'observer ces caractéristiques avec la dernière précision et d'en dresser un portrait aussi détaillé que possible, de façon à enrichir sensiblement notre connaissance du français et de son diasystème à une période de grande diffusion de son histoire.

9 Bibliographie

ALEC = Dulong, Gaston/Bergeron, Gaston, *Le parler populaire du Québec et de ses régions voisines. Atlas linguistique de l'Est du Canada*, 10 vol., Québec, Gouvernement du Québec, 1980.

ALPA = Le Dù, Jean/Brun-Trigaud, Guylaine, *Atlas linguistique des Petites Antilles*, 2 vol., Paris, CTHS, 2011–2013.

Bauche 1929 = Bauche, Henri, *Le langage populaire : grammaire, syntaxe et dictionnaire du français tel qu'on le parle dans le peuple de Paris avec tous les termes d'argot usuel*, Paris, Payot, 1929.

Bergeron-Maguire, Myriam, *Le français en Haute-Normandie aux 17^e et 18^e siècles. Le témoignage des textes privés et documentaires*, Strasbourg, Éditions de linguistique et de philologie, 2018. exposant

Bergeron-Maguire, Myriam, *Entre Saint-Domingue et Angers : le français d'une créole d'Ancien Régime*, in : Martineau, France/Remysen, Wim (edd.), *La parole écrite : des peu-lettrés aux mieux-lettrés*, Strasbourg, Éditions de linguistique et de philologie, 2020, 51–67.

Bergeron-Maguire, Myriam, « [D]es navir qui ne vienderons poien ta la pointe » : quelques témoignages remarquables pour l'histoire de la liaison en français, *Revue de linguistique romane* 85 (2021), 75–99.

Bergeron-Maguire, Myriam, *Pour une histoire de la langue « par en bas » : textes documentaires (privés) et variation des langues dans le passé*, *Linx* 85 (2022), DOI : <https://doi.org/10.4000/linx.9304> (= 2022a).

Bergeron-Maguire, Myriam, *Sur la correspondance privée d'une peu-lettrée de Honfleur (1792) : inventaire linguistique*, *Linx* 85 (2022), DOI : <https://doi.org/10.4000/linx.9509> (= 2022b).

Bergeron-Maguire, Myriam/Greub, Yan, *Qui n'entend qu'une cloche n'entend qu'un son. Le français classique des lettres interceptées durant les guerres de course atlantiques*, Actes en ligne du Congrès Mondial de Linguistique Française, 2020, DOI : <https://doi.org/10.1051/shsconf/20207802001>.

Branca-Rosoff, Sonia/Schneider, Nathalie, *L'écriture des citoyens, une analyse linguistique de l'écriture des peu-lettrés pendant la période révolutionnaire*, Paris, Klincksieck, 1994.

Buben, Vladimír, *Influence de l'orthographe sur la prononciation du français moderne*, Bratislava, Universum, 1935.

Chauveau, Jean-Paul, *Configurations géolinguistiques et histoire des français expatriés : quelques exemples de consonnes finales*, in : Baronian, Luc/Martineau, France (edd.), *Le français d'un continent à l'autre. Mélanges offerts à Yves Charles Morin*, Québec, Presses de l'Université Laval, 2008, 77–92.

- Chauveau, Jean-Paul, *Le français d'un rural manceau d'Ancien Régime*, in : Manzano, Francis (ed.), *Mémoires du terrain. Enquêtes, matériaux, traitement des données. Actes du colloque tenu à l'Université Jean Moulin de Lyon, 12 et 13 mars 2009*, Lyon, Publications du Centre d'Études Linguistiques, 2011, 67–95.
- Clérice, Thibault, *Corpus latin Antiquité et Antiquité tardive lemmatisé*, 2020, DOI : <http://doi.org/10.5281/zenodo.4337146>.
- Cotgrave, Randle, *A dictionarie of the French and English tongues*, Londres, Islip, 1611.
- Coyos, Jean-Baptiste, *Le basque*, in : Kremnitz, Georg (ed.), *Histoire sociale des langues de France*, Rennes, Presses Universitaires de Rennes, 2013, 427–437.
- DECA = Bollée, Annegret/Fattier, Dominique/Neumann-Holzschuh, Ingrid (edd.), *Dictionnaire étymologique des créoles français d'Amérique*, 4 vol., Hamburg, Buske, 2017–2018.
- DECOI = Bollée, Annegret (ed.), *Dictionnaire étymologique des créoles français de l'Océan Indien*, 4 vol., Hamburg, Buske, 1993–2007.
- DMF 2020 = Martin, Robert/Bazin, Sylvie (edd.), *Dictionnaire du Moyen Français (DMF 2020)*, Nancy, ATILF-CNRS/Université de Lorraine, 2020, <<http://www.atilf.fr/dmf/>>. [dernière consultation : 03.05.2023]
- DRF = Rézeau, Pierre (ed.), *Dictionnaire des régionalismes de France. Géographie et histoire d'un patrimoine linguistique*, Bruxelles, De Boeck/Duculot, 2001.
- Ernst, Gerhard, *Textes français privés des XVII^e et XVIII^e siècles*, Berlin, De Gruyter, 2019. **exposant**
- FEW = Wartburg, Walther von, et al., *Französisches Etymologisches Wörterbuch. Eine darstellung des gallo-romanischen sprachschatzes*, 25 vol., Bonn/Heidelberg/Leipzig/Bâle, Klopp/Winter/Zbinden, 1922–2002.
- Frantext = ATILF, *Outil de consultation de ressources informatisées sur la langue française « Frantext »*, Nancy, ATILF-CNRS/Université de Lorraine, 1998–, <<https://www.frantext.fr/>> [dernière consultation : 03.05.2023].
- Freist, Dagmar (ed.), *The Prize Papers in-depth cataloguing project*, Akademie der Wissenschaften zu Göttingen/Carl von Ossietzky Universität Oldenburg, 2018–, <www.prizepapers.de/> [dernière consultation : 28.03.2023].
- Gdf = Godefroy, Frédéric, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, 10 vol., Paris, Vieweg, 1880–1902. **exposant**
- Graff, Harvey J., *The legacies of literacy. Continuities and contradictions in Western culture and society*, Bloomington, Indiana University Press, 1987.
- Guillot-Barbance, Céline/Heiden, Serge/Lavrentiev, Alexei, *Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique*, *Diachroniques 7* (2017), 168–184.
- Koch, Peter/Oesterreicher, Wulf, *Langage oral et langage écrit*, in : Holtus, Günter/Metzeltin, Michael/Schmitt, Christian (edd.), *Lexikon der romanistischen Linguistik*, vol. I:2, Tübingen, Niemeyer, 2001, 584–627.
- Larousse, Pierre, *Grand dictionnaire universel du XIX^e siècle*, 15 vol., Paris, Larousse, 1867–1890. **exposant**
- Litttré, Émile, *Dictionnaire de la langue française. Supplément*, Paris, Hachette, 1877.
- Martineau, France, *Perspectives sur le changement linguistique : aux sources du français canadien*, *Canadian journal of linguistics/Revue canadienne de linguistique 50* (2005), 173–213.
- Martineau, France/Remysen, Wim/Thibault, André, *Le français au Québec et en Amérique du Nord*, Paris, Ophrys, 2022.
- Mitchell, Linda C., *Teaching grammar and composition through letter writing in seventeenth- and eighteenth-century England*, in : Dossena, Marina/Del Lungo Camiciotti, Gabriella (edd.), *Letter writing in Late Modern Europe*, Amsterdam/Philadelphia, Benjamins, 2012, 229–250.
- Rétif, André, *Essai de datation du « Grand Dictionnaire » de Pierre Larousse*, *Cahiers de lexicologie 22* :1 (1973), 108–112.

- Rézeau, Dominique/Rézeau, Pierre, *De la Vendée aux Caraïbes. Le journal (1878–1884) d'Armand Massé, missionnaire apostolique*, 2 vol., Paris, L'Harmattan, 1995.
- Rutten, Gijsbert/van der Wal, Marijke, *Letters as loot research project*, Leiden, Universiteit Leiden, 2008–2013, <<https://www.universiteitleiden.nl/en/research/research-projects/humanities/letters-as-loot-towards-a-non-standard-view-on-the-history-of-dutch#tab-1>> [dernière consultation : 26.03.2023].
- Rutten, Gijsbert/van der Wal, Marijke, *Letters as loot. A sociolinguistic approach to seventeenth- and eighteenth-century Dutch*, Amsterdam/Philadelphia, Benjamins, 2014.
- Rutten, Gijsbert/van der Wal, Marijke, *Dutch private letters from the seventeenth and eighteenth centuries. The linguistic experiences of the lower and middle ranks*, in : Steffen, Joachim/Thun, Harald/Zaiser, Rainer (edd.), *Classes populaires, scripturalité et histoire de la langue. Un bilan interdisciplinaire*, Kiel, Westensee-Verlag, 2018, 227–249.
- Sowada, Lena, *Schreiben im Ersten Weltkrieg. Französische Briefe und Tagebücher wenig geübter Schreiber aus der deutsch-französischen Grenzregion*, Berlin/Boston, De Gruyter, 2021.
- Spitzer, Leo, *Italienische Kriegsgefangenenbriefe. Materialien zu einer Charakteristik der volkstümlichen italienischen Korrespondenz*, Bonn, Peter Hanstein, 1921.
- Thibault, André, *Le français dans les Antilles : présentation*, in : Thibault, André (ed.), *Le français dans les Antilles : études linguistiques*, Paris, L'Harmattan, 2011, 11–28.
- Thibault, André, *Analyse linguistique des traits phonographiques et morphosyntaxiques de la correspondance d'une femme de soldat en Bretagne romane (1915–1917)*, in : Carles, Hélène/Glessgen, Martin-D. (edd.), *Les écrits des Poilus. Miroir du français au début du XX^e siècle*, Strasbourg, Éditions de linguistique et de philologie, 2020, 389–438.
- Thibault, André/Avanzi, Mathieu, *Temps, espace et politique : le témoignage des français expatriés et des créoles sur la diffusion des innovations impulsées par le centre*, **Revue de linguistique romane, à paraître, juin 2023. remplacer par : Revue de linguistique romane 87 (2023), 23-73.**
- TLF = Imbs, Paul/Quemada, Bernard (edd.), *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle (1789–1960)*, 16 vol., Paris, CNRS/Gallimard, 1971–1994.
- Villiers, Patrick, *La guerre de course et la défense de la Martinique pendant la guerre de Succession d'Autriche*, in : Plouvier, David (ed.), *Défense et colonies dans le monde atlantique : XV^e–XX^e siècle*, Rennes, Presses universitaires de Rennes, 2014, 81–94.