

L'emploi du passé simple et du passé composé dans les lettres aux XVII^e et XVIII^e siècles : une analyse multivariée des facteurs intralinguistiques

Ryo Nakagawa^{1*}, et Hadrien Charvat¹

¹Université Juntendo, Faculté des Arts libéraux internationaux, 113-8421, Tokyo, Japon

Résumé. Cet article examine le passé simple (PS) et le passé composé (PC) dans les lettres écrites par des scripteurs n'appartenant pas à la classe aristocratique aux XVII^e et XVIII^e siècles. Afin de déterminer quels facteurs intralinguistiques influaient sur le choix entre le PS et le PC, nous avons réalisé une analyse multivariée reposant sur un modèle de régression logistique à effets mixtes. Parmi les quatre facteurs identifiés, la co-occurrence avec une spécification temporelle du passé et la négation étaient les facteurs influant sur le choix entre le PS et le PC. La fréquence totale du PS était largement inférieure à celle du PC, ce qui suggère une certaine rareté du PS dans le langage informel aux XVII^e et XVIII^e siècles. Suite à notre analyse quantitative, nous avons examiné trois questions soulevées par les statistiques, à savoir : l'absence du PS dans la phrase négative, l'influence des marqueurs temporels, et l'ambiguïté graphémique qui empêche une distinction entre le PS et d'autres formes verbales.

1 Introduction¹

L'objectif de cet article est de décrire la distribution du passé simple (ci-après « PS ») et du passé composé (ci-après « PC ») dans les lettres aux XVII^e et XVIII^e siècles. En français contemporain, le PS est pratiquement limité au langage écrit ([1], p. 619) et devient « un temps plus ou moins défectif » ([2], p. 73), qui est une forme verbale marquée même à l'écrit ([3], p. 151). Étant donné que ce temps est plus fréquemment observé dans les textes anciens qu'en français d'aujourd'hui ([4], p. 349), il est alors naturel de se demander comment le PS est sorti du langage oral et/ou informel. Cette question du déclin du PS a fait couler beaucoup d'encre et certains ont identifié les XVII^e et XVIII^e siècles comme représentant la période charnière [5-8]. La situation durant ces deux siècles concernant la vitalité du PS et de son concurrent PC a été décrite par des études portant sur le genre épistolaire [9-11].

La contribution que notre analyse essaie d'y apporter est double. Premièrement, nous avons analysé deux sources qui n'ont pas été prises en compte, même dans la littérature récente comme l'étude de Labeau [12] : le Corpus Macintosh [13] et la transcription des

* Corresponding author: r.nakagawa.mh@juntendo.ac.jp

lettres de la famille Caillouel, cette dernière ayant été réalisée dans le cadre du projet doctoral du premier auteur de cet article. Ces sources contiennent des lettres de scripteurs populaires non répertoriées par les études antérieures, qui se penchaient sur l'écrit des nobles ou des intellectuels tels que Madame de Sévigné ou Voltaire.

Deuxièmement, nous avons réalisé une analyse multivariée, une approche qui n'est pas entreprise dans les études mentionnées ci-dessus, afin d'examiner simultanément quatre facteurs intralinguistiques potentiellement significatifs : type de sujet, spécification temporelle, négation, et transitivité du verbe. La question centrale de notre enquête est de 1) savoir si les lettres des scripteurs d'une couche plus populaire exhibent le même schéma d'emploi que les auteurs nobles et intellectuels, et de 2) identifier les facteurs linguistiques statistiquement significatifs et effectifs sur l'emploi du PS au moyen d'une analyse multivariée. À cette fin, nous avons recouru à un modèle de régression logistique à effets mixtes. Ce modèle fait partie de la famille des modèles linéaires généralisés à effets mixtes (ci-après « GLMM », soit l'abréviation du terme anglais « Generalized linear mixed model »). En pratique, l'introduction dans le modèle d'un effet aléatoire permet de prendre en considération une éventuelle hétérogénéité existant parmi les scripteurs concernant leurs habitudes d'utilisation des formes verbales. Cette hypothèse d'hétérogénéité semble pertinente dans la mesure où les occurrences du PS ou du PC ne sont vraisemblablement pas indépendantes, celles employées par un scripteur donné ayant tendance à être plus similaires que les occurrences apparaissant sous la plume d'un autre scripteur.

Dans ce qui suit, nous rapportons, d'abord dans la section 2, l'état de la littérature sur la diachronie du PS avec une attention particulière à sa concurrence avec le PC. À travers cette revue de la littérature, nous démontrerons l'apport du corpus épistolaire à une meilleure compréhension du déclin du PS. Puis, nous présenterons notre méthode de recherche dans la section 3. Y sont traités la structure de corpus, le protocole de la collecte de données, et un aperçu du procédé d'une analyse avec le GLMM. La fréquence du PS et celle du PC ainsi que les résultats de l'analyse avec le GLMM seront ensuite présentés dans la section 4, puis discutés dans la section 5 en prenant compte les descriptions des études précédentes et la question de la graphie. La section 6 est la conclusion.

2 État de la littérature

En français contemporain, le PS est presque cantonné au langage écrit ([1], p. 619). Si ce temps est toujours « passablement utilisé », il peut être qualifié comme étant « défectif » pour beaucoup de locuteurs puisque l'éventail des verbes employés au PS ou le choix de la personne grammaticale sont biaisés ([2], p. 73). À l'oral, la valeur d'aoriste est plus souvent exprimée avec un PC ([1], p. 619).

La dominance du PC n'était cependant pas toujours présente dans l'état antérieur de la langue. Yvon rapporte que le PS (ou « prétérit parfait » dans sa terminologie) est plus fréquent dans la *Vie de saint Alexis* (XI^e s.) avec 174 occurrences contre 53 occurrences du PC ([14], pp. 245-246)². Martin calcule la proportion du PS et du PC parmi les verbes conjugués du passé dans neuf textes du genre théâtral en moyen français d'un côté, et des journaux au XX^e siècle de l'autre. Selon ses données, la fréquence du PS a chuté de 47,11% en moyen français à 16,57% en français moderne, alors que le PC a gagné du terrain de 7,74% en moyen français jusqu'à occuper 46,91% des cas au XX^e siècle ([4], pp. 348-349). Les données fournies par Wilmet offrent une autre perspective. Afin d'examiner l'influence de la présence ou non d'une détermination temporelle, Wilmet a exploré les emplois de ces deux formes dans le théâtre comique en moyen français et a relevé 1178 cas du PS et 1898 cas du PC ([15], pp. 275-276). Les données de Martin et celles de Wilmet divergent quant à l'implantation du PC dans le genre théâtral, mais elles s'accordent à suggérer que l'usage du PS était plus répandu en moyen français qu'en français d'aujourd'hui.

Le déclin diachronique du PS a suscité l'intérêt de chercheurs (voir entre autres GGLF [1], section 38.1.2.1 ; [4, 9, 12, 20-23]). L'ouvrage récent de Labeau [12] a examiné les données fournies par les études antérieures sur ce sujet afin de situer l'évolution du PS dans le schéma à quatre stades proposé par Harris [24] pour les langues romanes. Selon Harris, la forme simple exprime d'abord non seulement le prétérit mais également le « present perfect » ou le « past with present relevance », alors que la forme composée dénote un état résultant d'un événement dans le passé (Stade I). La forme composée commence ensuite à signifier un événement dans le passé ayant un lien quelconque avec le moment d'énonciation (Stade II). Cet emploi en tant que « present perfect » est approprié, écrit Harris, lorsque ce même événement est marqué par l'aspect duratif ou répétitif. Au Stade II, la forme simple hérite de la plupart des fonctions du Stade I. Puis, la forme simple étant limitée à la fonction du prétérit, la forme composée développe sa fonction « archétypale (archetypal) » du « present perfect », qui fait référence au passé ayant sa pertinence dans le présent (Stade III). Enfin, la forme simple devient un temps limité au registre formel ou bien absent de l'usage, au profit de la forme composée qui assume non seulement la fonction du « present perfect » mais aussi celle du prétérit (Stade IV). Le français standard d'aujourd'hui a atteint ce stade IV ([24], pp. 49-50). Comme Labeau le note, ce schéma réduit illégitimement l'évolution du PS à sa concurrence avec le PC, tout en négligeant les autres formes verbales qui peuvent remplacer le PS ([12], pp. 38, 196). Cela dit, le schéma harrissien saisit la tendance générale ([12], pp. 38), et en ce qui concerne la distribution des formes simple et composée, le moment décisif de transition vers la dominance du PC se trouve entre le Stade III et le Stade IV, soit la période allant du XVI^e siècle jusqu'au XVIII^e siècle ([12], pp. 66-75).

Au niveau sémantico-pragmatique, la fonction du PC se superpose sur celle du PS : si le focus est mis sur l'événement en question dans le passé plutôt que sur son résultat présent, le sémantisme du PC se rapproche de celui du PS ([24], p. 47 ; [25], p. 9 ; [26], p. 375). Ainsi, dès le XV^e siècle, l'emploi du PC comme prétérit commença à se répandre graduellement. Foulet écrit qu'au XV^e siècle, le PC est employé avec la valeur du passé achevé dans un recueil de conversations en français pour les Anglais ainsi que dans les lettres de la même époque ([6], p. 285). Au XVI^e siècle, le PC n'aurait cessé de s'imposer. Schøsler note l'« utilisation prépondérante » du PC correspondant au prétérit latin dans la traduction de *L'Institution de la religion chrétienne* par Jean Calvin ([27], p. 331). Le PS pouvait toujours apparaître à l'oral au XVII^e siècle : Fournier signale sa présence dans le langage oral dans le *Journal* d'Héroard ou *Dom Juan* de Molière ([7], p. 399). Cependant, le PC n'en était pas moins rare : Dauzat constate que le PC est dominant dans *Dom Juan* ([5], p. 66) et Larthomas confirme la basse fréquence du PS chez Molière, qui utilisait plus l'imparfait du subjonctif que le PS ([28], p. 201). C'est dans ce contexte que Jean Pillot a reconnu que la distinction entre les deux temps était difficile et incertaine ([29], p. 21), et que Charles Maupas a affirmé qu'« [i]l sera souvent indifférent duquel nous usions défini [= le PS] ou indéfini [= le PC] » ([30], f. 138^o). Les auteurs des grammaires et des manuels de français distinguaient toutefois le PS du PC en évoquant des critères comme la distance temporelle par rapport au présent, la détermination par certains marqueurs de temps, ou l'inclusion ou non du moment d'énonciation à sa référence ([31], pp. 151-152). La fameuse règle des 24h peut être liée à une insécurité linguistique provenant de cette distinction délicate des deux temps verbaux.

Certains chercheurs forment des hypothèses concernant le moment de la disparition du PS. Dauzat considère que le PS était en péril dans le français populaire au XVII^e siècle et n'était plus usité dans le langage courant à Paris au XVIII^e siècle, même si l'on entend le PS dans les variétés régionales plus récente ([5], pp. 68, 77-79). Foulet écrit que le PS « a dû disparaître de la langue familière dans la première moitié du XVII^e siècle » ([6], p. 307). Pour Fournier, c'est durant la deuxième moitié du XVII^e siècle que le PS s'affaiblit dans le

langage naturel, avant que cette tendance ne devienne nette au XVIII^e siècle ([7], p. 399). Gadet, quant à elle, estime que le PS est rare vers la fin du XVIII^e siècle ([8], p. 608). Vu ces remarques, il semble raisonnable d'étudier la distribution du PS et du PC pendant ces deux siècles dans un style plus ou moins proche de l'oral. L'un des corpus prometteurs est celui des lettres informelles car ce genre reflète, partiellement certes, un aspect du langage informel à l'oral ([32, 33] ; également voir [34], p. 387). Il y a trois contributions à citer ici qui portent sur l'emploi du PS dans le genre épistolaire de l'époque.

En premier lieu, Galet donne la fréquence dans la correspondance de Madame de Sévigné. Ses données sont issues des lettres écrites entre les années 1648-1671, 1674 et 1675 ([9], pp. 34-35). Son étude recense la distribution du PS et du PC dans les phrases interrogatives et exclamatives, soit « la structure de la phrase qui représente le DISCOURS par excellence » ([9], p. 396). Selon les chiffres qu'elle donne, le PS est rare : une seule occurrence avec la deuxième personne du singulier *tu* a été observée, mais il y a 61 cas du PC avec *je* (16 occ.), *tu* (27 occ.), *on* (3 occ.), ou avec les pronoms de la troisième personne (15 occ.) ([9], pp. 419-420). En deuxième lieu, Caron et Liu ont examiné des lettres du XVII^e-XIX^e siècles dans la base textuelle *Frantext*³. Ils ont inventorié une vingtaine de scripteurs célèbres (Madame de Sévigné, Voltaire, etc.) du XVII^e-XVIII^e siècles pour déterminer la fréquence des deux formes verbales dans le contexte de la co-occurrence avec les adverbiaux de temps comme *hier*, *la veille*, *lundi* [10]. En troisième lieu, Jacobs a exploré des textes tirés de la base de données American and French Research on the Treasury of the French Language (ARTFL) [11]. Le corpus de Jacobs contient le genre « personal writings » dont des lettres par des auteurs comme Guez de Balzac, Madame de Sévigné, ou Voltaire. Ayant vérifié trois périodes différentes, à savoir 1620-1630, 1670-1680, et 1720-1735, Jacobs a trouvé que le PC était de 3 (Louise de Coligny) à 10 fois (comte de Bussy) plus fréquent dans toutes ces périodes d'investigation ([11], pp. 263-265). Ces trois études soulèvent deux questions. D'abord, il sera intéressant de se demander si la vitalité du PC dans le genre de la lettre était un fait établi parmi ceux qui n'appartenaient pas à la plus haute strate de la société. Cette question peut être abordée à l'aide du Corpus Macintosh [13] et la transcription des lettres d'une famille de marchands réfugiés d'origine rouennaise, la famille Caillouel. Ensuite, nous pouvons également nous intéresser à des facteurs qui ne sont pas examinés dans ces trois études. À part la personne grammaticale du sujet et la co-occurrence avec les marqueurs de temps, n'y a-t-il pas d'autres facteurs influant sur le choix entre le PS et le PC ? Il faudrait d'abord se procurer la liste des facteurs qui sont potentiellement à l'œuvre lorsqu'un scripteur choisit entre ces deux temps.

Une piste se trouve dans la littérature sur la concurrence entre le *past* et le *present perfect* en anglais, deux formes morpho-syntaxiquement parallèles au PS et au PC. L'analyse systématique la plus récente est celle de Yao, qui a étudié l'alternance entre le *past* et le *present perfect* dans la variété américaine et la variété anglaise en 1750-99 et 1950-99 en se basant sur le corpus ARCHER [36]. Dans ses analyses, Yao prend en compte onze facteurs : 1) genre (*Drama*, *Letters*, *News*, *Science*), 2) présence ou non d'une spécification temporelle, 3) télicité du syntagme verbal, 4) transitivité du verbe, 5) négation, 6) type de sujet, 7) type de proposition (principale/subordonnée), 8) temps du verbe précédant le verbe examiné, 9) interaction de 1 et 3 ci-dessus, 10) interaction de 1 et 4 ci-dessus, et 11) interaction de 1 et 7 ci-dessus. L'analyse de Yao est multivariée : elle consiste à intégrer tous ces facteurs dans un modèle statistique afin d'en déduire simultanément la signification statistique et la valeur de chaque facteur sur le choix entre le *past* et le *present perfect*. Cette approche est déjà adoptée par ses prédécesseurs [37-40], mais l'originalité de Yao est que sa modélisation est réalisée en tenant compte de l'effet aléatoire au niveau du scripteur (intercept aléatoire), c'est-à-dire, en tenant compte de l'effet individuel de chaque scripteur superposé au niveau des exemples. Cette modélisation a été effectuée avec un modèle linéaire généralisé à effets mixtes, ou le GLMM. Étant

donné que la variable de réponse (*past/present perfect*) est binaire, Yao a adopté une analyse de régression logistique, à la suite de laquelle Yao a réussi à identifier plusieurs facteurs favorisant le *past* ou le *present perfect* dans la variété anglaise en 1750-1799. Juste pour en citer deux, un pronom sujet de troisième personne (*he, she, it, they*) est un facteur plus favorable pour le *past* en comparaison avec un pronom de première/deuxième personne ; similairement, un verbe téléique est un facteur favorable pour le *present perfect* en comparaison avec un verbe atélique ([36], p. 120 sq.).

La contribution de Yao est intéressante pour deux raisons : cette étude liste plusieurs facteurs influant potentiellement sur notre distinction PS/PC en français d’un côté, et de l’autre sa méthode d’analyse permet d’intégrer à la fois plus d’un facteur ainsi que l’effet aléatoire au niveau du scripteur. Nous allons alors chercher à appliquer cette analyse multivariée à la façon de Yao, aux données récupérées depuis les deux sources, le Corpus Macintosh et les lettres de la famille Caillouel.

3 Méthode

3.1 Corpus et collecte des données

Les deux sources sur lesquelles cette étude est basée sont constituées de lettres produites par des scripteurs qui ne sont pas répertoriés dans *Frantext* et ne sont donc pas examinés par les études précédentes.

La première des deux sources, le Corpus Macintosh [13] offre une transcription de lettres françaises constituant le fonds des « Prize Papers » abritées aux Archives nationales du Royaume-Uni ([41], pp. 5-6 ; [42], p. 32). Ce corpus contient des lettres de scripteurs de divers degrés d’alphabétisation allant de ceux qui sont à peine alphabétisés jusqu’à ceux qui savent très bien écrire ([42], p. 36). Afin de constituer notre corpus d’enquête, nous avons utilisé les lettres du XVII^e-XVIII^e siècles qui étaient disponibles le 6 février 2025 sur le site du Corpus Macintosh. Le deuxième corpus est la transcription faite par le premier auteur de cet article des lettres de la famille Caillouel. Il s’agit d’une famille de réfugiés protestants français, qui a fui la politique antiprotestante de Louis XIV et est arrivée en Angleterre vers l’an 1686. Leurs lettres en français sont actuellement conservées sous la cote F/CA dans la *Huguenot Library* de la *Huguenot Society of the Great Britain and Ireland*. Comme dans le cas des « Prize Papers », cette bibliothèque est accessible dans la localité des Archives nationales du Royaume-Uni. Parmi les documents dans ce fonds, nous avons sélectionné des lettres reçues par Isaac Caillouel (la première génération des réfugiés) ainsi que ses ébauches et des copies de ses propres lettres (F/CA/18), et des lettres venant de Gédéon Vincent père et fils et reçus par un autre Isaac Caillouel, fils d’Isaac Caillouel ci-dessus (F/CA/21). Les transcriptions du corpus Caillouel ont été faites à partir des documents primaires originaux et revues à plusieurs reprises afin de minimiser les erreurs de transcription. Le tableau 1 décrit les lettres que nous avons incluses dans le corpus.

Tableau 1. Les deux sources examinées dans cette étude.

	Corpus Macintosh (récupéré le 06/02/2025)	Transcription des lettres de la famille Caillouel
Taille	110 lettres = 47 852 mots 104 scripteurs	82 lettres = approx. 23 800 mots 14 scripteurs
Période	1670-1756	1686-1719

À partir de nos sources, nous avons collecté manuellement toutes les occurrences du PS et du PC. À chaque fois qu’il y avait un doute sur la transcription, autant pour le Corpus

Macintosh que pour les archives de la famille Caillouel, nous avons toujours consulté le document original ou la photographie du document original pour en vérifier la lecture. Une lettre pour laquelle la photographie de l'original n'était pas disponible sur le Corpus Macintosh a été exclue de notre corpus (Lettre 116, Port-au-Prince, 1755, Jeanne Commeau, Corpus Macintosh). Nous avons également exclu les cas où la lecture n'était pas certaine, soit à cause de l'état de conservation du matériel, soit à cause de l'ambiguïté morpho-syntaxique. Un spécimen de ce dernier cas est donné dans la thèse de Mareschal :

- (1) a ma cher fille je te pris / bien da voir soint de notre anfant / je te diré que **je resté** a pin beuf /// quatre jour de tans (TSB123_14 Fort Royal→Pornic 1778, cité par Mareschal [42], p. 43. Souligné par nous)⁴

Ici, comme Mareschal l'a remarqué, il est difficile de trancher si la partie en gras <je resté> est un PS *je restai* ou un PC avec un auxiliaire *avoir* (c.-à-d. <j'e> = *j'ai*), combiné avec un verbe de mouvement *rester*. Ajoutons un autre cas cité depuis le Corpus Lettres Macintosh :

- (2) Cet pour la dus sieme que je vous / ecris a pres avoir resu deus votre que / **je resu[PS/PC]** qui ma fet bien dus plesir anaprant que tus et bons sancte dieus mesi (Chaterine Bernade, 26/02/1755, Corpus Macintosh 114. Souligné par nous)

Ce cas de figure est d'autant plus intéressant qu'il y a une scriptrice qui a écrit <jai> pour le pronom sujet *je* :

- (3) **jai ne / point resu[PC]** pas une detes nouvelle depuis le 3 avril 1744 [...] (Francoises[sic] Joyeux, 10/05/1745, Corpus Macintosh 106. Souligné par nous)
- (4) alor **je / resus[PS/PC]** le sacremant il me restait encore que que / petite operance e pour masurai du fait **jai demandé[PC]** / a monsieur fenolo (Francoises[sic] Joyeux, 10/05/1745, Corpus Macintosh 106. Souligné par nous)

Nous avons également mis à part les cas où deux lectures sont possibles entre le PS et le présent (également voir [34], p. 388):

- (5) appris auoir Receu La Vostre Jefus voir monsieur / Lemery a qui **Je dis[PS/PRÉSENT]** quayant obtenu sentence deRenuoy en possession de Vos / biens (F[r.] Lesauvage à "Monsieur Isaac Caillouel", 17/08/1690, Huguenot Library F/CA/18/38. Souligné par nous)
- (6) Hoche / Vint hier chez nous **il me dit[PS/PRÉSENT]** quils sestoient accommodes ensemble quon Leurs / Redonnoit vne petite mesure (F[r.] Lesauvage à "Monsieur Isaac Caillouel", 17/08/1690, Huguenot Library F/CA/18/38. Souligné par nous)

Bien que ces exemples ne soient pas pris en compte dans l'analyse quantitative, ils suscitent une hypothèse relative au déclin du PS. Nous reviendrons sur ce sujet dans la section 5.

Dans notre corpus, le nombre de lettres n'est pas équilibré. Afin d'éviter que notre observation soit biaisée par certains scripteurs occupant une large proportion du jeu de données, nous avons limité le nombre de lettres par scripteur à une seule en excluant tous les locuteurs pour lesquels le total des occurrences du PS et du PC était inférieur à 10 ou supérieur à 40. Cela nous a conduit à éliminer 86 scripteurs. Par conséquent, notre corpus est composé de 32 scripteurs provenant des deux sources, avec un nombre total d'occurrences du PS et du PC par lettre compris entre 10 et 40 (médiane = 12,5 ; IQR =

4,25). En définitive, nous avons considéré un total de 500 exemples, PS et PC confondus. Nous avons ensuite étiqueté les occurrences ainsi obtenues en indiquant pour chacun la présence ou non des facteurs affectant potentiellement le choix entre le PS et le PC. La section suivante présente ces facteurs ainsi que le protocole d'étiquetage.

3.2 Facteurs examinés avec le GLMM

Comme nous l'avons vu dans la section 2, il y a deux facteurs pris en compte dans la littérature sur l'alternance PS/PC : la co-occurrence de marqueurs temporels et la personne grammaticale du sujet. La liste peut être ensuite augmentée en adaptant les variables traitées dans l'étude de Yao [36]. Pourtant, tous les facteurs ne sont pas pertinents dans notre cas : ainsi en est-il par exemple de la question du genre (*Drama*, *News*, etc.) puisque notre corpus ne contient que des lettres. De plus, une catégorisation objective et reproductible n'était pas possible pour tous les facteurs ; c'est pourquoi la télicité du verbe est exclue. Déterminer si un syntagme verbal est télélique ou atélélique nécessite un test introspectif, alors qu'il semble dangereux pour un chercheur d'aujourd'hui non natif de se fier à sa propre catégorisation. Nous avons également exclu les facteurs comme le type de proposition (principale/subordonnée) et le temps du verbe précédant le verbe en question. Le premier facteur suppose que cette distinction principale/subordonnée soit applicable à l'écrit informel d'une lettre, mais la configuration des phrases n'est pas si nette surtout dans l'écrit des scripteurs moins alphabétisés. En ce qui concerne le temps du verbe précédant le verbe en question, l'introduction de cette catégorie rend difficile le traitement du verbe à la tête d'une unité discursive. En conséquence, nous avons choisi les quatre facteurs suivants :

- La co-occurrence ou non avec une spécification temporelle du passé (*hier*, *date*, *marqueur du temps passé*, etc.)
- La transitivité du verbe (*transitif / intransitif*)
- La présence ou non d'une négation
- Le type de sujet (*première ou deuxième personne du singulier / autres*)

Chaque facteur est codé selon deux modalités (présence ou absence, etc.), ce qui facilite l'interprétation d'une analyse multivariée. Nous avons distingué les première et deuxième personnes des autres types de sujet parce que les données de Martin ([4], pp. 358, 388) indiquent que le PS apparaît moins souvent avec les première et deuxième personnes, au singulier comme au pluriel. Limiter ainsi le nombre de facteurs est crucial dans notre cas. En effet, dans une analyse de régression logistique comme la nôtre, avec une variable de réponse à deux modalités (PS et PC), le nombre de variables explicatives considérées dans le modèle devrait idéalement ne pas être supérieur au nombre de cas appartenant à la modalité la moins fréquente de la variable de réponse divisé par dix [43]. Or, comme on le verra dans la section 4, les données dont nous disposons contiennent seulement 48 cas d'utilisation du PS. Il est donc raisonnable de nous limiter ici aux quatre facteurs mentionnés ci-dessus. Nous n'avons pas pris en compte non plus le genre des scripteurs et la période de production parce que la distribution du genre en fonction de la période est loin d'être équilibrée : les huit scripteurs du XVII^e siècle sont des hommes, tandis qu'il y a 21 femmes et un seul homme parmi les scripteurs du XVIII^e siècle.

Une fois l'étiquetage terminé, la variable de réponse a été recodée en variable numérique binaire : 1 pour « PS » et 0 pour « PC ». Les données ont été ensuite traitées avec le logiciel d'analyse statistique R (ver. 4.5.1) [44]. Dans un premier temps, nous avons réalisé un test unilatéral sur la proportion de PS (l'hypothèse nulle H_0 étant que cette proportion est égale ou supérieure à 0,5). Puis, nous avons estimé les paramètres d'un modèle de régression logistique multivariée à effets mixtes : les variables explicatives

introduites dans le modèle étaient les quatre facteurs identifiés ci-dessus ; un intercept aléatoire permettait en outre de prendre en compte une éventuelle variabilité inter-scripteurs. La présence d'éventuelles interactions entre les quatre facteurs retenus a ensuite été testée en estimant les modèles correspondants : aucune des interactions ne s'étant révélée statistiquement significative, le modèle finalement retenu a donc été celui incluant les quatre facteurs sans interactions. En raison du phénomène de séparation quasi-complète observé pour la négation (en effet, aucune forme verbale au PS n'a été observée après une négation dans notre jeu de données), l'estimation des paramètres des différents modèles testés a été effectuée selon l'approche de régression pénalisée décrite par Clark et al. [45] utilisant la fonction *bglmer()* de la librairie *blme* (Bayesian Linear Mixed-Effects Models) [46]. En particulier, il est apparu que le choix d'une distribution *a priori* normale avec écart-type fixé à 2 permettait d'obtenir des résultats stables en termes d'erreur-type du paramètre associé à la négation.

La section suivante présente les résultats de cette analyse multivariée ainsi que la distribution des deux formes verbales dans notre corpus.

4 Résultats

Dans ce qui suit, nous allons d'abord présenter la distribution du PS et du PC en fonction des facteurs examinés (section 4.1). Les résultats de l'analyse multivariée seront ensuite donnés (section 4.2).

4.1 Fréquence du PS et du PC

La figure 1 ci-dessous montre la fréquence absolue du PS et celle du PC. La dominance du PC par rapport au PS est claire : 48 occurrences du PS contre 452 occurrences du PC, soit une proportion de PS dans l'échantillon de 0,096 : un test sur la proportion de PS dans la population théorique dont est issu notre échantillon permet de rejeter l'hypothèse nulle H_0 selon laquelle cette proportion est égale à 0,5 ($p < 0,05$). La population théorique représente l'ensemble des contextes dans lesquels le PS alternait avec le PC dans des lettres écrites par des scripteurs n'appartenant pas à la plus haute strate de la société du XVII^e-XVIII^e siècles.

Une comparaison avec les données soumises par les études antérieures met en exergue la basse fréquence du PS dans notre corpus. Dans les données de Jacobs, le scripteur qui utilise le moins le PS est Roger de Rabutin, comte de Bussy, qui a employé le PS quatre fois contre 40 occurrences du PC ; le scripteur suivant le comte de Bussy est Rousseau avec sept cas du PS et 62 cas du PC [11]. La fréquence du PS dans notre corpus de lettres, très basse, est ainsi comparable à la fréquence des scripteurs lettrés qui utilisent le moins le PS (et qui utilisent donc le plus le PC) dans l'étude de Jacobs.

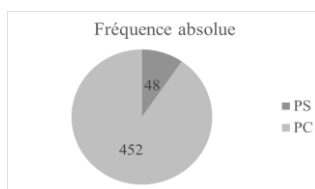


Fig. 1. Fréquence absolue du PS et du PC.

Cette tendance devient nette si l'on tient compte de la fréquence relative, soit la fréquence du PS par rapport à la fréquence totale du PS et du PC, par scripteur (Fig. 2). Parmi les 32

scripteurs examinés, la moitié des scripteurs n’ont utilisé que le PC et il y a une seule scriptrice qui a préféré le PS. La fréquence moyenne du PS par scripteur est de 0,082.

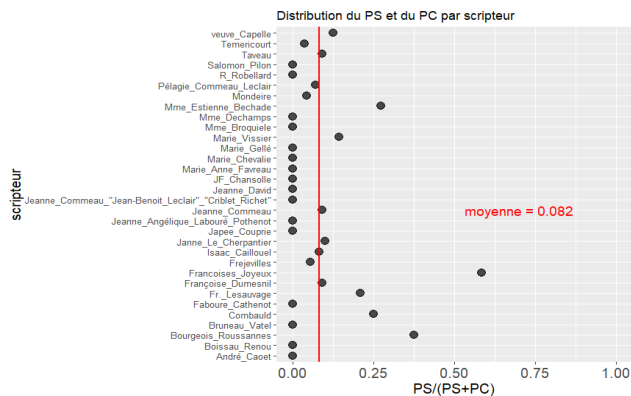


Fig. 2. Distribution du PS et du PC par scripteur.

En ce qui concerne la distribution des deux temps par facteurs, deux groupes peuvent être distingués. Le premier groupe est la transitivité et la personne grammaticale du sujet. Pour ces deux facteurs, la distribution du PS et du PC n’est pas si différente selon la valeur : c’est toujours le PS qui est de loin moins fréquent que le PC (Fig. 3 et 4).

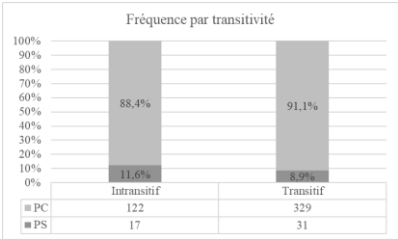


Fig. 3. Fréquence par transitivité.

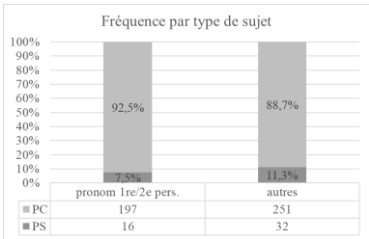


Fig. 4. Fréquence par type de sujet.

La deuxième catégorie englobe la spécification temporelle du passé et la négation. Le PS est plus fréquent si la phrase est accompagnée d’une spécification temporelle du passé (Fig. 5 : 31/435 avec spécification temporelle du passé, et 17/65 sans spécification temporelle). En revanche, le PS est absent lorsque la phrase contient une négation (Fig. 6).

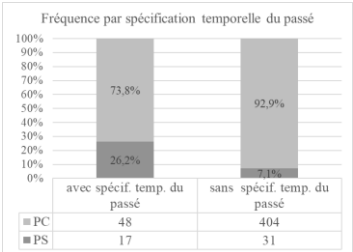


Fig. 5. Fréquence par spécification temporelle du passé.

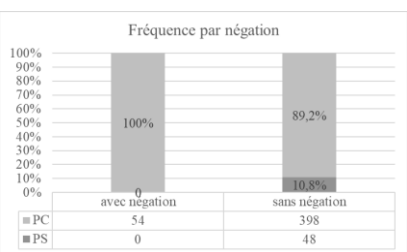


Fig. 6. Fréquence par négation.

4.2 Analyse avec le modèle de régression logistique à effets mixtes

Le tableau 2 ci-dessous récapitule les résultats de l’analyse avec le GLMM. En l’absence d’hypothèse sur la direction des effets des facteurs, tous les tests réalisés sont bilatéraux. Parmi les facteurs examinés (prédicteurs sur les tableaux 2 et 3), la présence ou non de la spécification temporelle du passé et la négation sont associées de façon statistiquement significatives à l’emploi du PS par rapport au PC.

Tableau 2. Résultats de l’analyse multivariée avec le GLMM. (arrondis au millième)

Résultats de l'analyse GLMM					
Prédicteurs	Effets fixes				Écart-type de l'Effet aléatoire (par scripteur)
	Estimation	Erreur-type	z	p	
(Intercept)	-2.773	0.493	-5.620	< 0.001	1.589
Type de sujet (autres -> pro. 1 ^{re} /2 ^e pers.)	-0.640	0.385	-1.662	0.096	—
Spécif. temp. du passé (non -> oui)	1.566	0.421	3.722	< 0.001	—
Négation (non -> oui)	-2.462	1.102	-2.234	0.025	—
Transitivité (intr. -> tr.)	-0.536	0.383	-1.398	0.162	—
Note : Nb d'exemples = 495*, Nb de scripteurs = 32. Formule du modèle : Temps ~ Sujet + Spéc_temp_passé + Négation + Transitivité + (1 scripteur). Valeur de la variable réponse : 1 = PS : 0 = PC.					
Marginal R2 = 0.159, Conditional R2 = 0.524, ICC = 0.434					
*Nb d'exemples est inférieur à 500 parce que nous avons écarté des cas inclassables					

Les rapports de cotes dans le tableau 3 (ci-après « OR », soit l’abréviation du terme anglais « odds ratio ») ainsi que leur intervalle de confiance à 95% ont été calculés avec la fonction *tab_model()* de la librairie *sjPlot*. L’OR de la spécification temporelle du passé est de 4,79 (intervalle de confiance à 95% : 2,10-10,92), indiquant une plus grande fréquence d’utilisation du PS en présence d’un marqueur temporel du passé⁵. À l’inverse, l’OR de la négation est de 0,09 (intervalle de confiance à 95% : 0,01-0,74), ce qui indique que le PS est moins fréquent en présence d’une négation.

Tableau 3. Estimation des rapports de cotes ou « OR ».

Estimation des rapports de cotes ou « OR »		
Prédicteurs	Odds ratio	Intervalle de confiance à 95%
Type de sujet (autre -> 1 ou 2 pronom)	0.53	[0.25, 1.12]
Spécif. temp. du passé (non -> oui)	4.79	[2.10, 10.92]
Négation (non -> oui)	0.09	[0.01, 0.74]
Transitivité (intr. -> tr.)	0.59	[0.28, 1.24]

5 Discussion

Le grand écart entre le PS et le PC en termes de fréquence (Fig. 1) suggère que le PS n’était pas fréquent dans les lettres écrites par nos scripteurs. Jacobs a déjà rapporté la diffusion du PC parmi les scripteurs nobles et intellectuels [11], mais cette tendance aurait été plus prononcée chez les scripteurs appartenant aux couches plus populaires. Comme la figure 2 l’indique, ceci n’est pas un phénomène limité à un certain individu : la plupart de nos scripteurs suivent en effet le même schéma. Toutes ces observations semblent soutenir l’hypothèse du déclin du PS dans le langage écrit informel aux XVII^e et XVIII^e siècles. Or, cette tendance ne relevait sans doute pas exclusivement de l’écrit. À ce propos, rappelons qu’au XV^e siècle, le PC était déjà employé dans un recueil de conversations pour les apprenants ([6]) et que le PS n’était pas fréquent chez Molière ([5, 28]). De plus, les auteurs des grammaires du XVII^e siècle reconnaissaient la difficulté de la distinction entre le PS et le PC [29-30]. À la lumière de toutes ces observations, nous reprenons, entre autres, pour la

langue orale comme pour la langue écrite, l'hypothèse de Dauzat, selon laquelle « au XVII^e siècle, le passé simple était en pleine crise dans la langue populaire » ([5], p. 68).

La basse fréquence du PS n'est probablement pas sans lien avec l'absence de cette forme verbale dans les phrases négatives. Nous avons vérifié d'abord avec la figure 6, puis avec les tableaux 2 et 3, que la négation entraînait rarement le PS. Ceci pourrait avoir un rapport avec le caractère sémantique du PS. Que ce soit dans une conversation ou dans une lettre, le locuteur rapporte la non-occurrence d'un événement parce qu'il lui convient de rappeler que cette situation reste inchangée au moment de l'énonciation. Ainsi dans notre corpus, on trouve de nombreux cas du type « je n'ai pas reçu votre lettre depuis ... », phrase souvent placée au début d'une lettre pour inviter le destinataire à continuer l'échange épistolaire, en lui rappelant que le destinataire n'a pas eu de nouvelles depuis un certain temps. La rareté du PS dans de telles formules ne signifie-t-elle pas que le PS n'était pas utilisé pour exprimer la non-occurrence d'un événement ? Vu la modeste taille de notre jeu de données, il serait trop hâtif d'exclure à ce stade toute compatibilité entre le PS et la négation, et le schéma pourra être examiné plus en détail soit en augmentant le nombre d'exemples, soit en entreprenant une étude plus détaillée d'approche qualitative sur chaque cas. Si cette incompatibilité s'avérait réelle, cela reviendrait à suggérer un caractère sémantique défectif du PS, qui expliquerait en partie sa rareté par rapport au PC.

Dans l'analyse multivariée, nous avons identifié la présence d'une spécification temporelle du passé comme étant statistiquement significative. Il y a deux points à discuter concernant ce facteur. Premièrement, nous avons montré avec les tableaux 2 et 3 que la présence d'une spécification temporelle du passé entraînait le PS. Parmi les 17 exemples du PS avec une spécification temporelle, 11 exemples étaient accompagnés d'un marqueur non déictique ou d'un marqueur exprimant l'ordre chronologique (ex. *lorsque*, *après* [+ un événement au passé]), tandis que les six autres cas étaient des cas avec des spécifications déictiques du passé (ex. *hier au soir*, ou *il y a 3 ans*). Localiser ainsi le procès dans un cadre temporel spécifique au passé est donc un contexte favorable pour le PS. Cette tendance n'est sans doute pas nouvelle dans le langage informel, parce que selon Wilmet, dans les farces, sotties et moralités en moyen français, le PS était plus fréquent lorsqu'il y avait une spécification temporelle ([15], p. 217). Pourtant, ces marqueurs temporels ne constituent pas une condition suffisante pour utiliser le PS puisque dans notre corpus, il y a des cas où le PC était employé en la présence d'un marqueur temporel : six cas du PS contre 14 cas du PC avec une spécification déictique comme *hier*, 11 cas du PS contre 34 cas du PC avec une spécification non déictique comme *lorsque*, *après* [+ un événement au passé].

Deuxièmement, il faut noter que les cas sans spécification temporelle du passé sont majoritaires (435 cas sur 500, Fig. 5). Certes, le PS est plus fréquent lorsqu'il y a une spécification temporelle. Cependant, les cas avec une spécification temporelle n'ont probablement pas eu de répercussion significative sur la distribution générale du PS et du PC dans le genre épistolaire parce qu'ils sont nettement moins fréquents que les contextes sans spécification, où le PC est dominant (404 cas du PC sur 435, Fig. 5). Une question intéressante soulevée ici — ou resuscitée car Lindschouw a abordé la question en se basant sur *Frantext* [21] — est le statut de la présence d'un marqueur de temps : est-elle un embrayeur de la diffusion du PC ou un catalyseur du changement déjà en œuvre ?

La fréquence élevée des contextes sans marqueur de temps prend un autre sens. Ne pas spécifier le cadre temporel peut avoir pour effet de laisser le lecteur à reconstruire soi-même ce cadre temporel. C'est partiellement pour cette raison que les séquences telles que <je dis> ou <il dit> peuvent être ambiguës dans certains cas, entre le présent et le PS (voir section 3.1). Ces cas sont encore plus complexes à cause de l'existence du *présent historique*. Ajoutons également ici la question de l'orthographe irrégulière. Nous avons déjà vu dans la section 3.1 que notre corpus contenait des exemples dont la lecture n'était pas certaine à cause de l'ambiguïté morpho-syntaxique (ex. <je resté> [PS/PC]; <je resu>

[PS/PC (=j'e resu)]; voir exemples 1-4 ci-dessus). Il est à noter que ces alternatives au PS sont à la première personne du singulier et peuvent alors être interprétées comme ayant un lien sémantique avec le moment de l'énonciation. Aux yeux du lecteur qui adopte cette interprétation, une telle séquence exprimerait un procès ayant sa pertinence dans le présent et pourrait alors déclencher la lecture comme un PC, soit résultatif « qui met en saillance la phase post-processive » (c.-à-d. la lettre étant reçue et donc ouverte/lue), soit processif qui « met en saillance la phase processive » (c.-à-d. le fait que la lettre a été reçue par le scripteur) ([2], p. 109). Autrement dit, l'ambiguïté morpho-syntaxique en question n'est pas un simple jeu d'orthographe mais une véritable possibilité d'interprétations différentes. Cette ambiguïté d'ordre graphique, qu'elle provienne d'une graphie irrégulière ou non, peut trahir l'intention du scripteur car c'est le lecteur qui décide la lecture. Pour les raisons que nous venons d'expliquer, le lecteur d'une lettre adressée à lui avait raison de choisir le présent ou le PC au lieu du PS, au moins dans certains contextes. De ce point de vue, même les formes « correctes » du PS ne seraient pas exemptes de malentendus ; par exemple, la graphie <je reçus> (PS) peut être lu comme *j'e reçu(s)* par ceux qui écrivaient la lettre <e> dans *je* pour le son /e/ (cf. [4], p. 399; également voir [1], p. 762). Notre spéculation ne semble pas complètement absurde puisque les scripteurs moins alphabétisés n'étaient pas ceux qui préféraient le PS ([42], p. 43). En d'autres mots, ceux qui écrivaient par exemple /e/ d'une façon irrégulière au moins de prime abord, n'étaient probablement pas ceux qui utilisaient souvent le PS comme *je reçus*. La direction inverse (ex. <j'ai reçu> lu comme un PS) ne devrait pas être courante parce que la graphie <jai/jay> pour *je* n'est clairement pas fréquente dans notre corpus et que le PC était d'ailleurs beaucoup plus fréquent que le PS. Ici encore, le déclin du PS est irrémédiable. Il nous convient d'ajouter rapidement que cette association du PS avec l'ambiguïté graphique reste une hypothèse. Pour la consolider, l'ensemble des graphies possibles et leur valeur phonétique seront à étudier du point de vue sociolinguistique et géolinguistique.

Pour finir, mentionnons les deux autres facteurs (le type de sujet et la transitivité du verbe) pour lesquels notre analyse multivariée n'a pas pu confirmer d'association statistiquement significative. Si leur signification statistique n'a pas été confirmée, cela ne prouve pas en soi que leur statut est secondaire dans l'usage du PS. Une observation plus fine de la variation auprès d'un petit groupe de scripteurs pourrait éclaircir la raison pour laquelle ces facteurs n'étaient pas, au moins statistiquement, très prononcés.

6 Conclusion

Cet article a analysé l'emploi du PS et du PC dans des lettres du XVII^e-XVIII^e siècles qui sont issues des deux corpus récemment construits, avec l'outil statistique qu'est le GLMM. Nous avons pu ainsi identifier deux facteurs influant sur le choix entre les deux formes verbales, à savoir la spécification temporelle du passé et la négation. Un des éléments majeurs de notre analyse multivariée a été son potentiel à traiter plusieurs facteurs tout en intégrant la variabilité inter-scripteur.

Il faut toutefois être conscient des limites de notre approche quantitative. Tout d'abord, nous n'avons pas pleinement considéré les autres formes verbales qui pouvait alterner avec le PS, notamment l'imparfait et le présent à valeur de passé. Une étude plus extensive requerrait également l'examen de ces formes. Ensuite, pour des raisons méthodologiques relatives à la difficulté de quantification objective, nous avons exclu certains facteurs tels que la télicité du verbe, le genre du scripteur, ou la période de production. Ces facteurs ne sont pas moins importants que les facteurs examinés dans ce travail. Une analyse qualitative sera alors nécessaire afin de compléter notre analyse quantitative en intégrant ces facteurs. Un autre sujet important que nous avons soulevé est celui de l'orthographe, en particulier l'ambiguïté des séquences telles que <je reçu> ou <je resté> et leur valeur dans

les variétés du français. Comme nous l'avons vu dans la section 5, cette question n'est pas anodine dans le contexte du déclin du PS ; pour comprendre mieux l'usage du PS chez les scripteurs populaires, une étude de l'orthographe irrégulière sera indispensable.

Références

1. GGLF = C. Marchello-Nizia, B. Combettes, S. Prévost et T. Scheer, Eds., Grande grammaire historique du français (De Gruyter Mouton, Berlin, 2020)
2. D. Apothéloz, Les temps verbaux : (I) temps simples, (II) temps composés, in Encyclopédie Grammaticale du Français (2021).
<https://nakala.fr/10.34847/nkl.0dbdpyf5>, <https://nakala.fr/10.34847/nkl.c93em866>.
3. C. Touratier, Le système verbal français (Description morphologique et morphématique) (Masson & Armand Colin, Paris, 1996)
4. R. Martin, Temps et aspect. Essai sur l'emploi des temps narratifs en moyen français (Klincksieck, Paris, 1971)
5. A. Dauzat, Études de linguistique française, 2^e éd. (Éditions d'Artrey, Paris, 1946)
6. L. Foulet, La disparition du prétérit. Romania **46**, 271-313 (1920)
7. N. Fournier, Grammaire du français classique (Belin, Paris, 1998)
8. F. Gadet, La langue française au XX^e siècle : L'émergence de l'oral, in Nouvelle histoire de la langue française, J. Chaurand, Ed. (Seuil, Paris, 1999), pp. 581-671
9. Y. Galet, Les corrélations verbo-adverbiales, fonction du passé simple et du passé composé, et la théorie des niveaux d'énonciation dans la phrase française du XVII^{ème} siècle, Vols 1-2, thèse de doctorat, Université de Paris X, Atelier reproduction des thèses, Université de Lille III/Honoré Champion, Lille/Paris (1977)
10. P. Caron et Y.-C. Liu, Nouvelles données sur la concurrence du passé simple et du passé composé dans la littérature épistolaire. Inf. Gramm. **82**(1), 38-50 (1999)
11. M. Jacobs, Problems and procedures in the construction of diachronic corpora: a case study of the *passé composé* and *passé simple* in Classical French, in Congrès Mondial de Linguistique Française - CMLF 2010, La Nouvelle-Orléans, États-Unis, 12-15 juillet 2010 (2010), 018
12. E. Labeau, The Decline of the French Passé Simple (Brill, Leiden, 2022)
13. M. Bergeron-Maguire, Corpus Macintosh, CLESTHIA/Sorbonne Nouvelle (2024-aujourd'hui). <https://lettres-outre-mer.huma-num.fr/>
14. H. Yvon, Emploi dans la *Vie de saint Alexis* (XI^e siècle) de l'imparfait, du passé simple et du passé composé de l'indicatif. Romania **81**, 244-250 (1960)
15. M. Wilmet, Le système de l'indicatif en moyen français (Droz, Genève, 1970)
16. S. Fleischman, Tense and Narrativity: From medieval performance to modern fiction (University of Texas Press, Austin, 1990)
17. G. Moignet, Grammaire de l'ancien français. Morphologie – syntaxe, 2^e éd. revue et corrigée (Klincksieck, Paris, 1976)
18. P. Ménard, Syntaxe de l'ancien français, Nouv. éd. entièrement ref. (Sobodi, Bordeaux, 1973)
19. R. Posner, Linguistic Change in French (Clarendon Press, Oxford, 1997)
20. P. Caudal et C. Vetters, Passé composé et passé simple : Sémantique diachronique et formelle. Cahiers Chronos **16**, 121-151 (2007)

21. J. Lindschouw, Passé simple et passé composé dans l'histoire du français : changement paradigmatique, réorganisation et régrammation. *RLiR*, **77**, 87-119 (2013)
22. J. Lindschouw et L. Schøsler, Parfait ou aoriste ? Problèmes liés à l'identification des valeurs. *Cahiers Chronos* **28**, 175-198 (2016)
23. M. Wilmet, *Études de morpho-syntaxe verbale* (Klincksieck, Paris, 1976)
24. M. Harris, The past simple and present perfect in Romance, in *Studies in the Romance Verb: Essays Offered to Joe Cremona on the Occasion of his 60th Birthday*, N. Vincent et M. Harris, Eds. (Croom Helm, London, 1982), pp. 42-70
25. D. Apothéloz et M. Nowakowska, La résultativité et la valeur de parfait en français et en polonais. *Cahiers Chronos* **21**, 1-23 (2010)
26. M. Nowakowska, Combien y a-t-il de passés composés en français ?, in *Problemy współczesnej neofilologii. Wybrane zagadnienia*, A. Grabowska, J. Graca et L. Smutek, Eds. (Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Tarnów, 2013), pp. 371-381
27. L. Schøsler, Sur l'emploi du passé composé et du passé simple. « ... ayant reçu de voz nouvelles, ie communicquay avec luy, et la conclusion fust telle que je vous ay mande ... », in *Le changement en français : études de linguistique diachronique* (Vol. 101 de Sciences pour la communication), C. Guillot, B. Combettes, A. Lavrentiev, E. Oppermann-Marsaux et S. Prévost, Eds. (Peter Lang, Bern, 2012), pp. 321-339
28. P. Larthomas, *Le langage dramatique. Sa nature, ses procédés* (A. Colin, Paris, 1972)
29. J. Pillot, *Gallicae linguae institutio, latino sermone conscripta...* (Slatkine, Genève, 1972[1550]), BnF, Gallica. <https://gallica.bnf.fr/ark:/12148/bpt6k4584f>
30. C. Maupas, *Grammaire et syntaxe française*, 2^e éd. (Chez Oliuier Boynard, Et Iean Nyon, au Cloistre Ste. Croix, Orléans, 1618 [1607]), disponible dans le *Corpus des grammaires françaises du XVII^e siècle*, B. Colombat et J.-M. Fournier, Eds. (Classiques Garnier Numérique)
31. N. Fournier, Introduction. II. Le contenu grammatical de l'ouvrage, in *Grammaire et syntaxe française*, C. Maupas (N. Fournier, Ed.) (Classiques Garnier, Paris, 2020), pp. 50-174
32. S. Elspass, 9 The use of private letters and diaries in sociolinguistic investigation, in *The Handbook of Historical Sociolinguistics*, J. M. Hernández-Campoy et J. C. Conde-Silvestre, Eds. (Wiley-Blackwell, Malden, MA, 2012), pp. 156-169
33. P. Koch et W. Oesterreicher, 62. Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit, in *Lexikon der Romanistischen Linguistik*, vol. I/2, *Methodologie*, G. Holtus, M. Metzeltin et C. Schmitt, Eds. (Max Niemeyer Verlag, Tübingue, 2001), pp. 584-627
34. P. Larthomas, Note sur l'emploi des temps dans *Le Neveu de Rameau*. *TraLiLi* **18**(1), 387-394 (1980)
35. Y.-C. Liu, *Du français classique au français contemporain. Permanence et évolution dans la systématique des temps verbaux de l'indicatif : le cas de la littérature épistolaire*, Vols 1-2, thèse de doctorat, Université de Limoges (1999)
36. X. Yao, *The Present Perfect and the Preterite in Late Modern and Contemporary English. A corpus-based study of grammatical change* (John Benhamins, Amsterdam/Philadelphia, 2024)
37. J. Roy, *The Perfect Approach to Adverbs. Applying variation theory to competing models*, Ph.D. thesis, University of Ottawa (2014)
38. G. Van Herk, Letter perfect: The present perfect in early African American correspondence. *Engl. World-Wide* **29**(1), 45-69 (2008)

39. G. Van Herk, Aspect and the English present perfect. What can be coded?, in *Aspect in Grammatical Variation*, J. A. Walker, Ed. (John Benjamins, Amsterdam/Philadelphia, 2010), pp. 49-64
40. X. Yao, Developments in the use of the English present perfect: 1750-present. *J. Engl. Linguist.* **42**(4), 307-329 (2014)
41. M. Bergeron-Maguire, '[D]es navir qui ne vienderons poien ta la pointe' : quelques témoignages remarquables pour l'histoire de la liaison en français. *RLiR* **85**, 75-99 (2021)
42. C. de Mareschal, Français de France et français des Antilles à l'époque coloniale : étude de particularismes phonétiques, grammaticaux et lexicaux relevés dans les *Prize Papers* (1665-1793), thèse de doctorat, Sorbonne Université (2024)
43. P. Peduzzi, J. Concato, E. Kemper, T.R. Holford et A.R. Feinstein, A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49**(12), 1373-1379 (1996)
44. R Core Team, R: A Language and Environment for Statistical Computing (ver. 4.5.1). R Foundation for Statistical Computing, Vienna (2025). <https://www.R-project.org/>
45. R.G. Clark, W. Blanchard, F. K.C. Hui, R. Tian et H. Woods, Dealing with complete separation and quasi-complete separation in logistic regression for linguistic data. *Res. Methods Appl. Linguist.* **2**, 100044 (2023)
46. Y. Chung, S. Rabe-Hesketh, V. Dorie, A. Gelman et J. Liu, A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* **78**(4), 685-709 (2013)

¹ Cet article est partiellement basé sur la présentation en japonais faite par le premier auteur lors de la 352^e rencontre régulière de la Société Japonaise de Linguistique Française (20 sept. 2025) et sur une autre en japonais faite par le même auteur lors du *Workshop. Learner Language and Discourse Markers* (7 déc. 2025). Aucun contenu de notre article n'a fait l'objet d'une publication sauf les supports de ces communications orales. Notre projet est subventionné par le JSPS KAKENHI Grant N° JP25K22986.

² Yvon examine également l'imparfait dans la *Vie de saint Alexis* [14]. Comme il est hors de la portée de cette étude, nous ne discutons pas l'emploi du PS équivalent à l'imparfait. Pour cet emploi dit « atypique » du PS, voir [1, section 38.1.2 b; 16-19].

³ Nous n'avons pas eu l'occasion de consulter la thèse de Liu [35], qui contiendrait des résultats plus complets.

⁴ Mareschal emploie une barre oblique lorsqu'il y a un changement de ligne, et trois barres obliques pour indiquer qu'il y a un changement de page ([42], p. 41). Dans notre propre transcription, une barre oblique est insérée pour signaler un changement de ligne mais la triple barre oblique n'est pas utilisée. Nous avons respecté l'orthographe telle qu'elle apparaît dans les sources primaires.

⁵ Il convient de remarquer qu'un OR a qualitativement la même interprétation qu'un risque relatif, à savoir qu'un OR supérieur à 1 indique que l'événement étudié (ici, l'utilisation du PS) a une plus grande probabilité d'occurrence en présence du facteur considéré (par exemple, une spécification temporelle du passé), et inversement pour un OR inférieur à 1. Cependant, l'OR a une définition mathématiquement différente de celle du risque relatif, et sa valeur numérique ne peut donc pas être directement interprétée comme celle d'un risque relatif : elle en constitue toutefois une valeur approchée lorsque la prévalence de l'événement étudié est faible, ce qui est le cas pour l'utilisation du PS dans ce travail.