



LUNDS
UNIVERSITET

FRAK11
Mémoire en linguistique (15 crédits)

Les rectifications orthographiques de 1990

L'usage dans le français écrit contemporain

Åsa Forsberg
Printemps 2022
Directrice : Malin Ågren

Table des matières

1. Introduction	3
2. But de l'étude	3
3. Questions de recherche.....	4
4. Le cadre théorique	4
4.1 Les caractéristiques du français écrit et de l'évolution de la langue	4
4.1.1 L'évolution de la langue française	4
4.1.2 Le système de l'écrit	4
4.1.3 L'orthographe française	5
4.2 Les rectifications de l'orthographe française de 1990	6
4.3 Les rectifications	7
4.3.1 Les attitudes des francophones vis-à-vis des rectifications.....	8
4.4 Recherches antérieures	9
5. Méthode.....	12
5.1 La méthode de corpus	12
5.2.1 Les corpus sélectionnés	12
5.2.2 Méthode de recherche dans les corpus	14
5.3 Catégories d'analyse et items sélectionnés	15
5.4 Analyse en trois étapes	16
6. Analyse des résultats	16
6.1 Analyse des items sélectionnés dans chaque corpus individuel.....	16
6.1.1 FW	16
6.1.2 Wiki	19
6.1.3 JSI1	21
6.1.4 JSI2.....	22
6.2 Comparaison des résultats pour chaque item dans les quatre corpus.....	24
6.2.1 Le trait d'union remplacé par la soudure	24
6.2.2 Le pluriel de mots composés réunis de trait d'union	24
6.2.3 Groupe accents A	25
6.2.4 Groupe accents B circonflexe	25
6.2.5 Les mots étrangers.....	26
6.2.6 La simplification des consonnes doubles	27
6.2.7 La suppression d'anomalies	28
6.3 Comparaison des résultats avec ceux de la littérature antérieure.....	28
7. Discussion	31
7.1 L'usage de la nouvelle orthographe dans le français écrit contemporain	31
7.2. La variation dans le temps de l'application de la nouvelle orthographe.....	32

7.3 Quelques facteurs qui peuvent influencer l'usage de la nouvelle orthographe.....	32
8. Conclusion.....	33
9. Références bibliographiques	35

1. Introduction

Le 7 février 1989 le journal *Le Monde* a publié un manifeste signé par dix linguistes éminents, « Moderniser l'écriture du français » (Catach et al, 1989). Environ 19 mois plus tard, en décembre 1990, les rectifications orthographiques ont été publiées, approuvées à l'unanimité par l'Académie française. Ces rectifications, cherchant à simplifier l'orthographe, portent d'entre autres sur la formation du singulier et du pluriel des mots composés, l'utilisation de l'accent circonflexe, la francisation des mots étrangers, la simplification des consonnes doubles et la suppression d'anomalies. La publication a immédiatement été suivie d'un débat intensif et parfois véhément, et au fil des années ce débat a ressuscité de temps en temps, comme en 2016 quand l'Éducation nationale 26 ans après la publication des rectifications, a pris la décision d'appliquer la nouvelle orthographe dans les manuels scolaires et le hashtag #JeSuisCirconflexe est devenu viral sur Twitter¹.

Mais est-ce que la nouvelle orthographe est réellement rentrée dans les pratiques ? Les bilans réalisés par les linguistes en 2005 (Petit), 2006 (Biedermann-Pasques), 2012 (Farid) et 2020 (Lazar & Mudrochová) indiquent d'une part que les rectifications sont peu connues et d'autre part que l'ancienne orthographe domine toujours.

Depuis 1990, ont eu lieu des études scientifiques sur les attitudes envers les rectifications, ainsi que sur les connaissances et les pratiques de la nouvelle orthographe. Grâce à la digitalisation et la constitution de corpus de web, la possibilité de rassembler des données a évolué, et dernièrement des études de corpus ont été faites sur l'usage des domaines spécifiques des rectifications. Nous n'avons pas pu repérer des études plus récentes sur les pratiques de l'ensemble de la nouvelle orthographe.

Dans cette étude, nous allons étudier les pratiques de plusieurs domaines rectifiés en 1990, en faisant des recherches dans des corpus de web. Quelle est l'usage contemporaine ? Peut-on discerner les variations entre les domaines ? Nous allons également comparer l'usage contemporain avec la recherche antérieure, pour repérer des variations au fil du temps.

2. But de l'étude

L'objectif de cette étude est d'examiner les usages contemporains des rectifications d'orthographe de 1990, pour voir dans quelle mesure les rectifications sont appliquées par les scripteurs francophones. Elles ont été controversées dès leur publication et mis longtemps

¹ <https://twitter.com/hashtag/JeSuisCirconflexe?src=hash>

avant d'être incorporées dans l'enseignement de l'orthographe et les œuvres de référence. Malgré ces faits, peut-on repérer des variations dans l'usage au fil du temps ?

3. Questions de recherche

Dans ce but nous avons formulé les questions de recherche suivantes :

1. Dans quelle mesure la nouvelle orthographe introduite par les rectifications en 1990 est-elle rentrée dans l'usage du français écrit contemporain ?
2. Des études faites en 2002-04 montrent une différence de l'application des règles incluses dans les rectifications. Quelle est la variation dans l'application de ces règles dans l'usage du français écrit contemporain ?
3. À partir de l'usage indiqué, peut-on identifier quelques facteurs qui soit promeuvent soit freinent l'usage de la nouvelle orthographe ?

4. Le cadre théorique

4.1 Les caractéristiques du français écrit et de l'évolution de la langue

4.1.1 L'évolution de la langue française

Henriette Walter donne un aperçu de l'évolution de la langue française. Le français appartient aux langues romanes, et la source de la langue est le latin vulgaire parlé par les romains qui ont conquis la Gaule en 1^{er} siècle avant J.-C. Mais l'influence des langues germaniques dans les siècles suivants a eu une très grande importance, surtout la langue des Francs, qui occupaient le nord de la France actuelle au V^e siècle. Au cours des siècles la langue qui va devenir le français a évolué, à cause de la diffusion du christianisme et des volontés politiques de Charlemagne de réintroduire le « bon » latin, des influences des langues germaniques des vikings et des dialectes régionaux. Au Moyen Âge, l'Île-de-France est devenu le centre politique où résidait la cour royale, et le dialecte parlé dans cette région a emporté sur les autres. En 1539 François Ier prend l'ordonnance de Villers-Cotterêts, qui déclare que désormais tous les textes officiels doivent être écrits en français. L'orthographe, le domaine linguistique pour notre étude, a été fixée en 1673, quand les membres de l'Académie française ont pris la décision d'adopter une orthographe unique (Walter, 1988 : 13-127).

4.1.2 Le système de l'écrit

Nina Catach (1986) décrit le système de l'écrit comme un système référant à l'ensemble des règles employées dans une langue particulière pour l'orthographe et la ponctuation. Le phonème est l'unité la plus petite dans la langue orale et le graphème l'unité la plus petite « de

la chaîne écrite ayant une référence phonique et/ou sémique dans la langue parlée » (Catach, 1986 : 27). Cook et Bassetti (2005) expliquent que dans une langue avec une graphie phonographique les graphèmes correspondent aux phonèmes, et on peut également la qualifier d'alphabétique. Cependant, une langue peut être plus ou moins alphabétique. En italien par exemple, chaque phonème correspond d'une manière relativement régulière à un graphème et l'inverse, donc l'italien peut être qualifié d'être phonologiquement transparent. Une langue où les phonèmes correspondent à plusieurs graphèmes ou l'inverse peut être qualifié d'être phonologiquement opaque. C'est le cas de l'anglais et également du français. Toutefois, la lecture de ce dernier est plus transparente que l'écriture. (Cook & Bassetti, 2005 : 7-9)

4.1.3 L'orthographe française

Nina Catach (1986) définit l'orthographe comme

...la manière d'écrire les sons ou les mots d'une langue, en conformité d'une part avec le système de transcription graphique propre à cette langue, d'autre part suivant certains rapports établis avec les autres sous-systèmes de langues (morphologie, syntaxe, lexicale). Plus ces rapports secondaires sont complexes, plus le rôle de l'orthographe grandit. De plus, un tissu d'antagonismes se crée entre les relations phonie/graphie, et les autres considérations qui entrent en ligne de compte. L'orthographe est un *choix* entre ces diverses considérations, plus ou moins réglé par des lois ou des conventions diverses. Pour des raisons faciles à comprendre, le choix est beaucoup plus impératif à l'écrit qu'à l'oral, en France tout particulièrement. (1986 : 26)

Selon Fayol et Jaffre (2016 : 23) il existe 130 graphèmes pour transcrire les 32 à 36 phonèmes dans le français. Par conséquent l'orthographe française peut être considérée comme une orthographe opaque, c'est à dire que chaque phonème ne correspond pas à un graphème unique et l'inverse. Donc, la maîtrise alphabétique ne suffit pas pour maîtriser l'orthographe. L'espagnol, cependant, a une orthographe transparente, avec des liens entre phonèmes et graphèmes relativement réguliers (Fayol et Jaffre, 2016 : 18).

Une autre caractéristique de l'orthographe française est le fait que l'on y trouve des configurations de lettres sans correspondants phonologiques. C'est le cas du premier « h » dans « orthographe » par exemple mais aussi des doubles consonnes. D'autres exemples sont des consonnes muettes à la fin des mots (lard, puit), le -s qui termine les noms au pluriel et le -nt chez les verbes à la troisième personne au pluriel.

Les accords posent une autre difficulté. Fayol et Jaffre (2016 : 26) constatent que le scripteur est obligé de mobiliser de l'attention pour pouvoir écrire les accords correctement.

[...] même les meilleurs auteurs sont victimes de telles erreurs, non qu'ils ignorent les règles, non qu'ils ne sachent pas les appliquer, mais plus simplement parce que la vigilance requise leur fait

défaut à un moment ou à un autre. En d'autres termes, la gestion des phénomènes d'accord soulève en français des problèmes tenant à l'homophonie et oblige les rédacteurs à faire preuve de vigilance pour prévenir les omissions ou substitutions de marques. Cela vaut pour les accords des noms, des adjectifs et des verbes. Cela vaut aussi pour les confusions entre infinitif et participe passé comme entre adverbe et participe présent (Pacton & Fayol, 2003). (Cité de Fayoul et Jaffre, 2016 : 26)

Catach soulève l'accent circonflexe, qui peut remplir plusieurs fonctions dans le français, mais d'une manière aléatoire, c'est-à-dire qu'il apparaît dans des certains cas mais pas dans d'autre cas, même que les conditions sont les mêmes. L'accent circonflexe a historiquement pu indiquer qu'une voyelle est longue et soit fermée soit ouverte (frêle/cône). Parfois il remplace un *s* étymologique (forêt). Dans certains cas il est utilisé pour distinguer des homophones (bailler/bâiller), ou encore il est utilisé dans certains formes des mots, comme des conjugaisons verbales (je connais, il connaît) ou des flexions nominales.

Catach constate que l'emploi n'est pas systématique et qu'il sera conseillé de supprimer l'accent circonflexe sur parmi d'autres *u, i, ai, ei, ou, eu*. (1986 : 69-73).

4.2 Les rectifications de l'orthographe française de 1990

Béchennec et Sprenger-Charolles (2014 : 3) soulèvent le fait que l'orthographe a été reformée plusieurs fois depuis la première édition du Dictionnaire de l'Académie française (1694). Dans l'édition de 1740, 5 000 mots sur 18 000 au total ont été modifiés, de même que pour certains, l'ancienne orthographe a été rétablie dans l'édition de 1835. En 1901 et 1935 des modifications de l'orthographe ont été proposées, mais pas appliquées.

En février 1989 dix linguistes, dont Nina Catach², ont publié un appel sur l'orthographe, pour une « modernisation de l'écriture du français ». Selon eux, une modernisation était nécessaire pour répondre à trois défis : défi technique, pédagogique et politique. Ils préconisaient que les propositions faites par une commission travaillant sur l'initiative du ministre de l'Éducation nationale fussent réexaminées. Celles-ci concernaient entre autres l'accent circonflexe, les consonnes doubles, les règles d'accord du participe passé et l'acceptation de doubles graphies dans certain cas (oignon et ognon). (Catach et al : 1989).

En juin 1989 le Premier ministre de l'époque, Michel Rocard, a fondé le Conseil supérieur de la langue française (désormais CSLF), chargé « de formuler des propositions claires et précises sur l'orthographe du français, d'y apporter des rectifications utiles et des ajustements afin de résoudre, autant qu'il se peut, les problèmes graphiques, d'éliminer les

² Le manifeste n'est pas signé mais selon Lazar et Mudrochova les linguistes étaient Nina Catach, Bernard Cerquiglini, Jean-Claude Chevalier, Pierre Encrevé, Maurice Gross, Claude Hagège, Robert Martin, Michel Masson, Jean-Claude Milner et Bernard Quémada (2020:1)

incertitudes ou contradictions, et de permettre aussi une formation correcte aux mots nouveaux que réclament les sciences et les techniques» (CSLF, 1990 : 3). Des ressortissants français, belges, québécois, suisses et marocains ont participé aux travaux du CSLF (Béchenec & Sprenger-Charolles, 2014 : 10). Le travail du CSLF a été approuvé par l'Académie française et en décembre 1990 les rectifications ont été publiées dans le *Journal officiel de la République française* (CSLF, 1990).

4.3 Les rectifications

Aussi bien Béchenec et Sprenger-Charolles (2014) comme l'organisation Renouvo (2014) donne une liste complète de points concernés par les rectifications, ainsi qu'un grand nombre de mots concernés. La liste ci-dessous, basée sur Béchenec et Sprenger-Charolles et *Renouvo*, inclue la plus grande partie de points rectifiés. Il est important de mentionner que ce sont des recommandations, aussi bien l'ancienne que la nouvelle orthographe est tolérée.

Le trait d'union

- Le trait d'union est remplacé par la soudure dans les composés de « contr(e)-, « entr(e) », « extra », « infra », « intra », « ultra » et « auto », ainsi que dans les composés de formation onomatopéique ou des mots d'origine étrangère (« blabla », « cowboy »).
- Le trait d'union est également remplacé par la soudure dans certains composés d'un verbe et un nom (« portemonnaie »), d'un verbe et de « -tout » (« mangetout ») et dans des mots composés avec « basse- », « mille- », « haut(e) », (« millepatte »)

Le pluriel

- Dans les mots composés unis par un trait d'union, dont le deuxième mot est un nom commun, celui-ci prend toujours la marque du pluriel (« un abat-jour, des abat-jours », « un sans-abri, des sans-abris »).
- Les mots étrangers suivent la règle du singulier et du pluriel des mots français (« les gentlemens »).

Les accents

- Devant une syllabe contenant un « e » muet on écrit « è » (évènement », « cèderait »).
- L'accent circonflexe est supprimé sur « i » et « u ».
- Le tréma est déplacé sur la lettre « u » (« aigüe »).

- Les mots empruntés suivent la règle des mots français (« à cappella »).

Le participe passé

- Le participe passé de « laisser » suivi d'un infinitif est invariable.

Simplification des consonnes doubles

- Des verbes en -eler et -eter s'écrivent avec un accent grave et une consonne simple devant une syllabe contenant un « e » muet (« il épèle »).
- Une consonne qui suit un « e » muet est simple (« lunette/lunetier »).
- Les mots qui se terminent en « -olle » et « -otter » se terminent en « -ole » et « -oter » (« girole »)

Suppression d'anomalies

- Des familles de mots sont réaccordées (« imbécilité » comme « imbécile »).
- Des anomalies sont supprimées (« oignon – ognon », « asseoir – assoir »)

(Béchenec & Sprenger-Charolles, 2014 : 14-19 ; *Renouvo*)

4.3.1 Les attitudes des francophones vis-à-vis des rectifications

Les rectifications proposées par le CSLF ont donc été approuvées à l'unanimité par l'Académie en mai 1990 et publiées dans le *Journal officiel de la République française* en décembre 1990. Elles ont tout de suite suscité beaucoup de critique, et provoqué un débat animé, et même parfois véhément, dans la presse quotidienne en décembre 1990 et janvier 1991. Dans ce débat, deux graphies rectifiées ont souvent été soulevées, la nouvelle orthographe de nénuphar, nénufar, et de l'oignon, ognon.

Ousselin (2004) résume le débat : « Au lieu d'un progrès, d'une évolution nécessaire des règles de l'écriture, les rectifications de l'orthographe représentaient pour certains la voie ouverte à tous les dérapages langagiers, voire à l'acceptation pure et simple du déclin, de la décadence de la langue française » (Ousselin, 2004 : 491).

Participaient à ce débat le CSLF et les linguistes voulant faciliter l'écriture, à l'opposé des intellectuels (dont des écrivains et des journalistes) et des politiciens qui voulaient préserver l'orthographe telle qu'elle était. « ...ce sont précisément ceux qui ont réussi, au prix de longs et patients efforts, à maîtriser les règles et les nombreuses exceptions de l'orthographe, qui estiment que les autres locuteurs du français, en particulier les jeunes générations, n'ont qu'à faire de même » (Ousselin, 2004 : 493). Parmi toutes les rectifications les protestes concernaient surtout la suppression de l'accent circonflexe.

En 2005 *Renouvo, Réseau pour la nouvelle orthographe du français*, fait un bilan sur l'usage de rectifications orthographiques dans la francophonie (Honvault (2005)). Pendant qu'en Belgique deux circulaires officielles avaient rappelé le caractère officiel des rectifications et qu'en Suisse une circulaire et une brochure avaient été diffusées aux enseignants, en France et au Québec aucune consigne n'avait encore été donnée aux enseignants. Malgré cela, les rectifications étaient prises en compte au Québec. En France, par contre, la majorité des enseignants attendait une information du Ministère de l'éducation. Honvault, l'auteur du bilan des rectifications en France, constate que la neuvième édition du dictionnaire de l'Académie française intègre toutes les rectifications qui concernent son corpus de mots, et d'autres dictionnaires les intègrent également. Honvault donne également un survol sur le débat dans la presse, en constatant que les rectifications sont toujours controversées. (Honvault, 2005 : 17-34).

En 2012 Farid fait encore un bilan, et constate que « Les rectifications orthographiques à pas de tortue » (2012 : 2060). Il soulève le fait qu'en France le Ministère de l'Éducation a précisé sa position envers les rectifications dans le Bulletin officiel en 2007, en proclamant que les rectifications orthographiques seront suivies. (2012 : 2061)

Pourtant, ce n'est qu'à la rentrée 2016 que les rectifications ont été inclus dans les manuels scolaires, ce qui provoque encore des réactions dans la presse.

4.4 Recherches antérieures

En 2002-04 Biedermann-Pasques (2006) a fait une enquête sur les attitudes et les pratiques envers des rectifications de l'orthographe, basé sur une autoévaluation et un test graphique. Deux questions avaient été inclus dans un questionnaire sur le langage ordinaire, soumis à environ 500 étudiants en France, Belgique, Suisse et Québec. La première était une question ouverte sur l'usage des rectifications, « Est-ce que vous appliquez les rectifications de l'orthographe lorsque vous écrivez ? » (Biedermann-Pasques : 11). Les réponses à cette question portent sur la connaissance, les attitudes et l'application des rectifications. La deuxième question était un test graphique, où les répondants allaient choisir entre la nouvelle et l'ancienne graphie pour un certain nombre de mots.

Les réponses à la première question montrent que 60,61% des étudiants belges, 53,57% des étudiants suisse, 37,68% des étudiants québécois et 10,25% des étudiants français connaissent les rectifications. Parmi les répondants qui ignoraient les rectifications, il y avait des cas de confusion entre rectification et corrections orthographiques.

Dans le test graphique les répondants devaient indiquer qu'elle graphie ils utilisent dans 6 paires de mots appartenant aux 5 domaines des rectifications (voir section 6.3.1 pour la forme exacte de la question et un résumé des résultats). Biedermann-Pasques trouve les pratiques relativement avancées pour trois domaines des rectifications. Premièrement, la régularisation du pluriel des mots composés (abat-jours) parce que grâce aux rectifications ces mots suivent les formes régulières. Deuxièmement, l'accentuation en *è* ouvert avant une syllabe contenant un *e* muet (cèderai), en conformité avec la prononciation. Troisièmement, la francisation des mots d'emprunt. Selon Biedermann-Pasques, l'usage serait normalisé si ces trois domaines étaient intégrés dans l'enseignement scolaire de l'orthographe. L'enquête montre peu d'usage de deux derniers domaines des rectifications, la suppression de l'accent circonflexe sur *i/u* et la régularisation d'anomalies (ognon).

Une observation fait par Biedermann-Pasques est que les étudiants qui préféraient la forme rectifiée dans l'accentuation (cèderai) utilisaient le plus grand nombre de rectifications dans l'enquête, pendant que ceux qui préféraient l'ancienne orthographe (céderai) utilisaient très peu de rectifications. Elle observe également une négative attitude envers les rectifications, même chez répondants qui ne les connaissaient pas.

Pour promouvoir l'usage des rectifications, Biedermann-Pasques recommande quelques mesures. Elle soulève l'importance d'inclure les variantes rectifiées dans les dictionnaires, d'appliquer les rectifications pour les néologismes et d'intégrer la nouvelle orthographe pour les mots composés dans la terminologie médicale. En outre, les rectifications devraient être inclus dans l'enseignement de l'orthographe, aussi bien pour les francophones que pour les apprenants du français langue étrangère. Finalement, les variantes rectifiées devraient être intégrées dans les correcteurs orthographiques.

Dans un mémoire de maîtrise, Sicotte (2011) étudie la suppression de l'accent circonflexe sur *i* et *u* et dans quelle mesure les dictionnaires ont inclus la variante sans circonflexe pour les mots concernés. Elle a constitué un corpus de 21 dictionnaires, sur papier et en ligne. L'examen du corpus a été effectué en trois étapes. D'une part Sicotte a examiné si les dictionnaires incluent un échantillon de variantes sans accent circonflexe dans les articles et tableaux, d'autre part si une section sur les rectifications est incluse dans l'œuvre et si les pages liminaires contiennent des variantes sans accent circonflexe. L'échantillon a compris 21 mots, dont de mot de la famille du verbe *bruler/brûler* et des verbes en *-aitre/âtre*, en *-oitre/ôître* et en *-aire*.

Dans 10 de ces 21 dictionnaires la variante sans accent circonflexe n'est pas incluse dans les articles, et elle n'est jamais en vedette dans un article. En outre, aucune variante sans

accent circonflexe figure dans les pages liminaires. Seulement 8 dictionnaires donnent l'information sur les rectifications.

La conclusion de Sicotte est que la « résistance à la suppression de l'accent circonflexe se manifeste en particulier dans le traitement que font les dictionnaires, encore aujourd'hui, de la variante sans accent circonflexe introduite par les recommandations de 1990 » (2011 : 107). Elle identifie la cause de cette résistance au faible usage des variantes sans accent circonflexe.

En 2020 un numéro thématique de la revue *Cahier praxématique* (Lazar & Mudrochová) rassemble plusieurs études sur l'application des domaines différents des rectifications. Les études portent sur des secteurs différents des rectifications.

Dedková et Lazar ont étudié la formation selon les rectifications du singulier et pluriel des noms composés dans un corpus de cinq dictionnaires en ligne. D'une part ils ont examiné les mots composés réunis d'un trait d'union, où le dernier mot prend un *s* pluriel (un *sans-abri*/des *sans-abris*), d'autre part les mots composés où le trait d'union a été remplacé par la soudure (un *portemonnaie*/des *portemonnaies*). Ils ont également examiné leur échantillon de mots dans le corpus de presse *Timestamped JSI web corpus 2014-2020* (SI). À part *Le Trésor de la Langue Française* les dictionnaires ont relativement bien intégré la nouvelle orthographe. En revanche, les résultats des recherches effectuées dans JSI indiquent que dans la presse l'usage de l'ancienne orthographe domine pour ces domaines des rectifications. Pour la majorité des mots composés ils ont trouvé plus de 50% d'occurrences de l'ancienne orthographe. « Il semble donc que les journalistes ne suivent que partiellement les préconisations des dictionnaires sur l'orthographe de mots composés. » (Dedková & Lazar, 2020 : 43)

Parmi eux, dans une étude de corpus publiée en 2020, Puchovská a examiné l'orthographe des verbes en *-eter* et *-eler*. Dans l'ancienne orthographe ces verbes prennent des doubles consonnes dans quelques formes conjuguées (*il dételle, épelle, feuillette*) et dans la forme rectifiée ils s'écrivent avec un accent grave et une consonne simple devant une syllabe contenant un *e* muet (*il détèle, épèle, feillète*). Le corpus utilisé est *Araneum Francogallicum Maximum*. Puchovská a effectué des recherches sur plusieurs formes conjuguées de trois verbes en *-eter* et trois en *-eler*. Elle conclut que l'ancienne orthographe reste dominante pour ces verbes.

5.Méthode

5.1 La méthode de corpus

Puisque l'objet de l'étude est d'examiner l'usage actuel de la nouvelle orthographe, il convient d'étudier des corpus du français écrit, et donc utiliser la méthode de corpus. Étant donné qu'il est nécessaire de parcourir de grandes masses de textes contemporains, les techniques du traitement automatique des langues (ci-après TAL) seront utilisées.

Selon *Le dictionnaire de l'Académie française* un corpus dans le sens linguistique est un « Ensemble de documents servant à une analyse linguistique ». Mellet fait la distinction entre les corpus clos, comme une monographie, et les corpus échantillonnés. Aujourd'hui, ces derniers peuvent, grâce aux techniques de l'information, contenir des mégadonnées, et ils constituent de très grands corpus. Dans toute sorte de corpus, il est important de garantir la représentativité des données et éviter le biais par un choix trop spécifique de données. Mellet (2002 : 3) constate que « le corpus est un objet construit », et il faut en être conscient.

Dalbera distingue également deux manières d'utiliser un corpus dans une étude linguistique. Un corpus peut être utilisé pour observer et identifier des phénomènes linguistiques, ou pour confronter des hypothèses avancées à la réalité. Il souligne l'importance, dans le premier cas, d'assurer la validité du corpus, et en parlant des corpus d'échantillons il souligne l'importance d'être transparent avec toute délimitation des données dans un corpus. (Dalbera, 2002 : 7-17)

Le TAL inclut un grand nombre de techniques dans plusieurs domaines, dont la linguistique, pour traiter des textes avec un logiciel. Cori (2008) fait la distinction entre le TAL théorique et le TAL robuste, et ce dernier peut être un instrument dans la linguistique appliquée qui permet de quantifier la représentation d'un phénomène linguistique spécifique dans un corpus donné (Cori 2008 : 100). Dans cette étude nous allons utiliser la méthode de corpus pour observer et quantifier l'usage des règles de la nouvelle orthographe.

5.2.1 Les corpus sélectionnés

Parmi les corpus de la langue française, un certain nombre consiste en grands « dumps » ou extractions d'Internet, soit d'un site individuel comme le Wikipédia en français, soit d'un secteur spécifique comme la presse numérique française, soit d'une grande partie de l'Internet francophone extraite à une date spécifique.

Nous allons utiliser les quatre corpus suivants :

A. French Web 2017 (frTenTen17) (ci-après FW)

Un corpus du web français, extrait en décembre 2017. Il contient environ 14 millions de documents, qui étaient en ligne en 2017 mais dont on ne connaît pas la date de création. Le nombre de mots de ce corpus est 5 752 261 039.

La date de création de ces documents n'est pas connue, tout ce que nous savons est qu'ils étaient publiés au moment de l'extraction. Il faut supposer qu'ils consistent des textes d'une très grande diversité, des citations de textes antiques jusqu'au blogues et forums de discussions, ainsi que des documents venant des sites commerciaux. Malgré cela, grâce à la taille et à la diversité de ce corpus, il doit être possible de trouver des tendances concernant l'usage de la nouvelle orthographe. En effet, c'est le seul corpus où aussi bien la nouvelle comme l'ancienne orthographe des mots sélectionnés pour cette étude sont représentées (voir section 6.1.1).

B. Wikidiscussion 2018 (ci-après Wiki)

Ce corpus inclut des discussions sur Wikipédia en français dans sa version du 20/10/2018. Le corpus intègre 439 638 discussions, 1 243 676 sections, 3 145 943 messages et 213 286 460 mots.

Le corpus Wiki est constitué des discussions sur des articles sélectionnés de Wikipédia en français depuis son début jusqu'en 2018. Elles ont été extraites le 20 octobre 2018. Puisque chaque fichier est tagué avec la date de publication, nous avons pu limiter la recherche à l'époque la plus récente, de 2016 à 2018 (Ho-Dac, 2021). Le corpus contient 439 638 discussions, mais il n'est pas possible de savoir combien de fichiers qui datent de 2016-2018, aussi le nombre d'occurrences est relativement faible. Malgré cela, nous avons voulu l'inclure dans cette étude surtout à cause de la diversité des scripteurs. N'importe qui peut contribuer à Wikipédia et participer aux discussions sur les articles. Donc, l'usage de la nouvelle orthographe dans ces discussions peut indiquer des tendances générales et relativement actuelles.

C. Timestamped JSI web corpus 2014-2021 French (ci-après JSI1)

Ce corpus est encore plus grand, avec 22 510 360 documents et 6 998 186 326 mots. Il contient uniquement des documents publiés entre 2014 et 2021, et il inclut des articles de presse. Contrairement au corpus précédent, le contenu de ce corpus est donc plus homogène et les textes sont écrits par des professionnels qui peuvent avoir des recommandations linguistiques à suivre, selon les journaux qu'ils représentent.

D. Timestamped JSI web corpus 2021-04 French (ci-après JSI2)

Le corpus JSI2 est compilé chaque trimestre sur la base d'articles publiés dans la presse française (extraits de feeds rss) et par la suite rendu accessible sur Sketchengine. La plus

récente édition accessible (en avril-mai 2022) date du quatrième trimestre de 2021. Le contenu est environ 358 000 documents, et il fait également partie du corpus JSI1 (voir ci-dessus). Malgré cela, nous avons choisi d'effectuer une recherche dans ce corpus, pour avoir une idée de l'usage de la nouvelle orthographe dans les articles des actualités pendant une période très récente. Puisque le corpus est petit, le nombre d'occurrences est faible, mais nous avons choisi de l'inclure tout de même puisque ce sont les proportions entre l'usage de la nouvelle et l'ancienne orthographe qui nous intéressent.

5.2.2 Méthode de recherche dans les corpus

Les corpus utilisés dans cette étude sont structurés et les catégories des mots sont taguées avec des balises pour faciliter leur utilisation. C'est le lemme qui constitue la forme de base d'un mot et une recherche d'un lemme donne toutes les formes flexionnelles du mot.

Pour un lemme, le corpus donne des collocations, c'est-à-dire les mots qui sont associés fréquemment avec le lemme, ce qui peut être très utile pour comprendre leur utilisation.

Les lemmes sont spécifiques à chaque corpus, c'est le créateur du corpus qui en spécifie les critères. Dans les corpus que nous avons utilisés, c'est le plus souvent la graphie ancienne qui constitue le lemme, ainsi dans le FW *boîte* est un lemme mais pas *boite*. Alors, *boite* ne sera pas retrouvé dans une recherche de lemme.

Pour cette raison, nous avons trouvé la recherche de concordances plus utile. En outre, nous voulons le plus souvent trouver les occurrences pour une forme spécifique d'un mot, et pour cet objet le mieux est presque toujours d'effectuer une recherche de concordances. Une concordance est le contexte du mot, les mots ou phrases à gauche et à droite de celui-ci. Sketchengine, l'outil que nous avons utilisé pour accéder aux corpus et pour travailler avec leur contenu, donne plusieurs possibilités de filtrer le résultat d'une recherche, ainsi que la possibilité de télécharger la liste des concordances. Les concordances peuvent donner une idée de l'objet d'un texte. Dans notre cas, certains mots rectifiés ont fréquemment été utilisés dans la présentation des exemples des rectifications de l'orthographe en 1990, comme par exemple *nénufar* et *ognon*. (voir section 4.3). Il n'est pas certain que les scripteurs de ces textes ont intégré la graphie dans leur usage, nous pouvons seulement savoir qu'ils l'utilisent pour décrire les rectifications.

Image 1 Concordances pour ognon dans JSI1

on qui pourra s'écrire	ognon	.	</s><s>	Nénufar sera autorisé à la
questions...	</s><s>	"	Ognon	", "nénufar", "weekend" : la France
a, imbécilité, nénufar,	ognon	,	persiffler, relai, joailler, etc.	</s><s>
reuses nouveautés, "	ognon	"	(pour oignon), sera considéré com	
pourra ainsi s'écrire "	Ognon	"	et "nénuphar", "nénufar".	</s><s>
flexe et bienvenue à "	ognon	", "nénufar"...	à moindre "cout"	</s><s>
après-midi.	</s><s>	"	Ognon	", "nénufar", accent circonflexe... </s><s>
ignon" devient ainsi "	ognon	", "nénuphar"	s'écrira désormais "né	
être rédigé comme "	ognon	",	et nénuphar sera transformé en "	
seul fait de lire le mot "	ognon	"	suffira à nous faire pleurer mainter	

Pour la plus grande partie des mots, nous avons donc effectué une recherche simple de concordance dans les quatre corpus sélectionnés. Dans deux cas, cette méthode n'a pas pu être appliquée. Premièrement, il s'agit du groupe du pluriel des mots composés unis d'un trait d'union. Dans l'ancienne graphie, le deuxième mot ne prend pas la forme plurielle (des *sans-abri/après-midi*). Une recherche pour *sans-abri* va donc donner aussi bien des mots au pluriel qu'au singulier. Pour éviter ce problème, nous avons dans les deux grands corpus effectué deux recherches de phrase pour chaque variante orthographique, pour nous assurer que le résultat ne contient que la forme plurielle. Nous avons fait une recherche d'abord pour les phrases *des sans-abris* et *les sans-abris*, et ensuite pour *des sans-abri* et *les sans-abri*. Par conséquent, les concordances comme *deux personnes sans-abris* ne sont pas incluses dans les résultats, et le nombre d'occurrences présenté pour FW et JSI1 ne couvrent donc qu'une partie des occurrences dans ces corpus. Pour éviter le biais, nous avons suivi la même procédure pour les trois items sélectionnés. Dans les deux petits corpus, le nombre d'occurrences est plus petit et possible de parcourir manuellement. Deuxièmement, l'*ognon* est un homonyme, désignant à la fois le légume et un fleuve français, et dans ce cas spécifique il a fallu parcourir les concordances manuellement.

5.3 Catégories d'analyse et items sélectionnés

Nous avons d'abord fait une sélection de mots pour lesquels nous avons observé la fréquence de la nouvelle orthographe par rapport à l'ancienne. La base de cette sélection est constituée par les mots inclus dans l'étude de Biedermann-Pasques (2006), avec des ajouts dans chaque

catégorie. D'une part, certains des mots ne sont pas très fréquents dans les corpus de web, comme par exemple *abat-jours* et *à cappella*. D'autre part, il nous a semblé nécessaire d'avoir un échantillon plus large pour pouvoir identifier des tendances.

Les mots dans le test graphique de l'étude de Biedermann-Pasques (2006) sont les suivants :

- Le pluriel : *abat-jours/abat-jour*
- L'accent grave : *évènement/événement, cèderai/céderai*
- L'accent circonflexe : *connaître, dans notre étude nous avons choisi la forme connaît/connaît*
- Francisation des mots étrangers : *à cappella/a cappella*
- Régularisation d'anomalies : *ognon/oignon*

À cette liste relativement courte nous avons ajoutés un certain nombre de mots pour enrichir nos analyses. Ils ont été sélectionnés pour élucider des aspects spécifiques des rectifications (voir section 4.3).

- Le trait d'union : *portemonnaie/porte-monnaie, plateforme/plate-forme*
- Le pluriel de noms composés : *sans-abris/sans-abri, après-midis/après-midi*
- L'accent circonflexe : *boîte/boîte, bruler/brûler, brulure/brûlure, aout/août*
- Francisation des mots étrangers : *cacahouète/cacahuète, rockeur/rocker, bogue/bug*
- La simplification des consonnes doubles : *épèle/épelle, craquèle/craquelle, feuillète, feuillette, girole/girolle*
- La régularisation d'anomalies : *asseoir/asseoir, relai/relais, imbécilité/imbécillité*

5.4 Analyse en trois étapes

- A. Analyse des items sélectionnés dans chaque corpus individuel
- B. Comparaison des résultats pour chaque item dans les quatre corpus
- C. Comparaison des résultats avec ceux de la littérature antérieure

6. Analyse des résultats

6.1 Analyse des items sélectionnés dans chaque corpus individuel

6.1.1 FW

Ce corpus contient environ 14 millions de documents (5,7 milliards de mots) extraits de l'Internet francophone en décembre 2017. Le tableau 2 ci-dessous résume la recherche des mots rectifiés dans le corpus.

Tableau 1 FW 2017 : occurrences et proportions des items sélectionnés

Mots Nouvelle orthographe/ Ancienne orthographe	Nouvelle orthographe (nombre)	Nouvelle orthographe (pour cent)	Ancienne orthographe (nombre)	Ancienne orthographe (pour cent)
Trait d'union				
portemonnaie/porte-monnaie	400	3,5%	10 922	96,5%
plateforme/plate-forme	253 935	71,8%	99 756	28,2%
Le pluriel				
Des/les abat-jours/abat-jour	244	46,6%	280	53,4%
Des/les sans-abris/sans-abri	2 145	43,3%	2 805	56,7%
Des/les après-midis/après-midi	1 624	30,7%	3668	69,3%
Les accents				
évènement/événement	132 044	15,3%	729 059	84,7%
(je) céderai/céderai	169	31,5%	368	68,5%
... dont l'accent circonflexe				
(il) connaît/connait	114 076	27,5%	300 359	72,5%
boite/boîte	139 806	38,3%	224 747	61,7%
bruler/brûler	4 542	10%	40 888	90%
brulure/brûlure	1 393	4,4%	30 534	95,6%
aout/août	36 116	2,9%	1 192 825	97,1%
Mots étrangers				
à cappella/a cappella	125	4,3%	2 974	25,7%
cacahouète/cacahuète	1 166	10,1%	10 285	89,9%
roqueur/rocker	3 175	28,6%	7 938	71,4%
bogue/bug	4 290	8,1%	48 262	91,9%
Simplification des consonnes doubles				
(elle/il) épèle/épelle	73	15,6%	395	84,4%
(elle/il) craquète/craquelle	241	30,1%	559	69,9%
(elle/il) feuillète/feuillette	144	5,9%	2 304	94,1%
girole/girolle	38	1,2%	3 192	98,8%
Anomalies				
ognon/oignon	244 (1781)	0,3%	81 907	99,7%
assoir/asseoir	6 876	12,1%	49 892	87,9%
relai/relais	9 475	5,2%	172 614	94,8%
imbécilité/imbécillité	1 117	39,4%	1 720	60,6%

Selon les rectifications le trait d'union est remplacé par la soudure dans plusieurs cas, et les résultats de ce corpus indique un grand pourcentage de l'usage de la forme *plateforme* (71,8%) contrairement à l'autre item sélectionné, *portemonnaie* (3,5%). Quant au group Le pluriel, rappelons que les difficultés liées à la forme du singulier se terminant en –s ont limité

la recherche possible des items de ce groupe (voir la section 5.2.3). En conséquence, les résultats pour les items dans ce groupe ne montrent qu'une partie des occurrences réelles. En tenant ce fait en compte, le résultat indique un usage relativement élevé pour la nouvelle orthographe, allant entre environ 30 à 46 %. Les recherches pour explorer l'usage des accents montre un relativement faible usage de l'accent grave devant une syllabe contenant un *e* muet dans l'exemple *évènement* (15,3%), et plus élevé pour *cèderai* (31,5%). Il faut pourtant souligner que la présence du dernier est clairement plus faible que celle du premier. La suppression de l'accent circonflexe sur *i* est plus courante que sur *u* et nous constatons que ce sont les mots fréquemment utilisés dans le corpus, *boite* et *connait*, qui sont le plus souvent employés avec la nouvelle orthographe. L'usage de la forme francisée des quatre mots étranger sélectionnés est également faible, sauf pour *rockeur* (28,6%). Il en est de même pour les items dans le groupe Simplifications des consonnes doubles, sauf *craquèle*, mais encore une fois il faut noter sa faible fréquence.

Dans le corpus, il y a peu d'occurrences de la nouvelle orthographe *girole*, cette graphie apparait seulement 38 fois. Puisque *girole* et *girolle* sont tagués comme des lemmes dans le corpus, il est possible de faire une comparaison de ses collocations, les mots associés. Parmi les mots associés avec *girole* on trouve *fumerole*, *roussote* et *rousserole*, qui ont tous les trois été rectifiés, ainsi que *orthographier*. Les mots associés avec *girolle* sont des noms de champignons (*chanterelle*, *cèpe*) et de cuisine (*concasser*, *égoutter*). Alors, on peut supposer que les occurrences pour *girole* surtout viennent des textes sur les rectifications pendant que les occurrences pour *girolle* viennent des textes sur des champignon ou des recettes de cuisine.

Image 2 Concordances et collocations pour *girole* et *girolle*

"girole/girolle" and/or ...			verbs with "girole/girolle" as object		
gimiés	1	0	orthographier	1	0
fumerole	1	0	concasser	0	1
roussote	1	0	braiser	0	1
grole	1	0	persillées	0	1
rousserole	1	0	fricasser	0	1
coulemelle	2	3	égoutter	0	9
bolet	1	44	défraichies	0	1
chanterelle	2	92	sechez	0	1
morille	1	63	osciètre	0	1
cèpe	1	299	légumesfrottez	0	1
mousseron	0	17	réhydrater	0	4
pleurote	0	28	poêler	0	28

Le dernier groupe concerne les anomalies supprimées. *Ognon* peut désigner soit le végétal, soit une rivière dans la Franche-Comté. La recherche de concordance dans ce corpus a donné une liste de 1 781 occurrences. En parcourant la liste nous avons sélectionné les passages contenant *ognon* (avec *o* minuscule), selon l'hypothèse que le nom propre de la rivière doit normalement être écrit avec un *O* majuscule. Le nombre 244 est donc une estimation pour le légume. Aussi bien *ognon* que *oignon* sont des lemmes, ce qui donne la possibilité de comparer les collocations. En analogie avec *girole/girolle* la collocation la plus fréquente pour *ognon* est *nénufar*, pendant que les collocations pour *oignon* sont des autres légumes (*ail, poireau...*) ou termes de cuisine (*hacher, revenir...*). Intéressant à noter est également les occurrences d'*imbécilité*, 39 %.

6.1.2 Wiki

Le tableau 2 ci-dessous résume la recherche des mots rectifiés dans le corpus Wiki (2016-2018).

Tableau 2 Wiki (2016-18) : occurrences et proportion des items sélectionnés

Mots Nouvelle orthographe/ Ancienne orthographe	Nouvelle orthographe (nombre)	Nouvelle orthographe (pour cent)	Ancienne orthographe (nombre)	Ancienne orthographe (pour cent)
Trait d'union				
portemonnaie/porte-monnaie	0	0	0	0
plateforme/plate-forme	286	68,9%	129	31,1%
Le pluriel				
(des) abat-jours/abat-jour	0	0	0	0
(des) sans-abris/sans-abri	4	36,4%	7	63,6%
(des) après-midis/après-midi)	5	55,5%	4	44,4%
Les accents				
événement/événement	437	15,7%	2 343	84,3%
(je) céderai/céderai	2	33,3%	4	66,7%
...dont l'accent circonflexe				
(il) connaît/connaît	596	41,7%	831	58,3%
boite/boîte	263	29%	641	71%
bruler/brûler	10	8,4%	109	91,6%
brulure/brûlure	0	0	7	100%
aout/août	0	0	0	0
Mots étrangers				
à cappella/a cappella	0	0	4	100%
cacahouète/cacahuète	5	23,8%	16	76,2%
rockeur/rocker	4	12,1%	29	87,9%

bogue/bug	10	25,6%	29	74,4%
Simplification des consonnes double				
(elle/il) épèle/épelle	1	50%	1	50%
(elle/il) craquèle/craquelle	1	100%	0	0
(elle/il) feuillète/feuillette	0	0	1	100%
girole/girolle	0	0	0	0
Anomalies				
ognon/oignon	0	0	43	100%
assoir/asseoir	22	21,5%	81	78,5%
relai/relais	41	15,1%	230	84,9%
imbécilité/imbécillité	4	22,2%	14	77,8%

Wiki est un corpus avec des discussions sur des articles dans Wikipédia en français. Seulement les contributions faites pendant la période 2016-2018 ont été cherchées, et de ce fait le nombre d'occurrences est petit. La petite taille du corpus utilisé, ainsi que le fait que la terminologie utilisée dans les discussions dépend de sujets d'articles discutés, peuvent expliquer le fait que tous les mots sélectionnés n'ont pas été retrouvés.

Plateforme, dans le groupe Trait d'union, fait partie des mots qui ont un nombre d'occurrences plus élevé, et l'usage de la nouvelle graphie est haut aussi dans ce corpus (68,9%). La fréquence de mots du groupe Le pluriel est petite, et même que la proportion de l'usage de la nouvelle graphie est élevée pour *sans-abris* (36,4%) et *après-midis* (55,5), il faut traiter le résultat avec précaution. Dans le group Les accents, *évènement/événement* sont les items les plus fréquentes, mais le pourcentage pour la nouvelle graphie est relativement faible (15,7%). Par contre, le circonflexe sur *î* est supprimé dans 41,7 % des occurrences pour *connait* et dans 29% des occurrences pour *boite*. Le circonflexe sur *u* dans *bruler* est seulement supprime dans 8,4% des occurrences, et pas du tout dans *brulure*. Dans ce groupe les items *aout/août* ont été exclus, puisque chaque fichier est tagué selon le modèle « jour/mois/année/l'heure : 20 août 2018 à 13 : 35 ». Cette balise est générée automatiquement par Wikipédia. Alors, on peut noter que la balise *août* suit l'ancienne orthographe. Les occurrences des mots dans le groupe Mots étranger sont relativement faibles, mais toutefois ils sont de 23,8% pour *cacahouète* et 245,4% pour *bogue* 25,4%, et seulement 12,1% pour *rockeur*. Les mots dans le groupe Simplification des consonnes double occurrent seulement une ou deux fois dans ce corpus. Dans le dernier groupe, les occurrences pour *assoir* (21,5%) et *imbécilité* (22,2%) sont relativement élevées.

6.1.3 JSI1

JSI1 comprend des articles de presse 2014-2021. Le tableau 3 ci-dessous résume la recherche des mots rectifiés.

Tableau 3 JSI web corpus 2014-2021 : occurrences et proportion des items sélectionnés

Mots Nouvelle orthographe/ Ancienne orthographe	Nouvelle orthographe (nombre)	Nouvelle orthographe (pour cent)	Ancienne orthographe (nombre)	Ancienne orthographe (pour cent)
Trait d'union				
portemonnaie/porte-monnaie	468	2,8%	16 541	97,2%
plateforme/plate-forme	655 058	82%	144 065	18%
Le pluriel				
Des/les abat-jours/abat-jour	134	35,4%	229	64,6%
Des/les sans-abris/sans-abri	6 014	30,6%	13 648	69,4%
Des/les après-midis/après-midi	2 335	25,3%	6 871	74,7%
Les accents				
évènement/événement	138 072	8,3%	1 532 969	91,7%
(je) céderai/céderai	156	15,1%	874	84,9%
L'accent circonflexe				
(il) connaît/connaît	111 693	18,8%	482 677	81,2%
boite/boîte	52 470	13,5%	337 345	85,5%
bruler/brûler	1 607	0,8%	208 9930	99,2%
brulure/brûlure	222	0,8%	26 525	99,2%
aout/août	25 673	1,8%	1 364 204	98,2%
Mots étrangers				
à cappella/a cappella	61	1,7%	3 533	98,3%
cacahouète/cacahuète	196	2,3%	8 336	97,7%
rockeur/rocker	19 544	75,6%	6 306	24,4%
bogue/bug	2 424	6,6%	34 268	93,4%
Simplification des consonnes double				
(elle/il) épèle/épelle	32	14,6%	187	85,4 %
(elle/il) craquèle/craquelle	104	18%	473	82 %
(elle/il) feuillète/feuillette	38	2%	1 858	98 %
girole/girolle	15	0,9%	1589	99,1 %
Anomalies				
ognon/oignon	411 (934)	0,9%	18 802	99,1 %
assoir/asseoir	6 448	7,9%	75 485	92,1 %
relai/relais	8 320	3,7%	214 954	96,3 %
imbécilité/imbécillité	485	31,9%	1 036	68,1 %

À la différence de deux corpus précédents le contenu de JSI1 est relativement homogène, avec des articles de presse. Certaines maisons de presse imposent probablement

des recommandations linguistiques, ce qui pourrait expliquer les résultats légèrement diversifiés avec des occurrences modestes de formes rectifiées, sauf pour quelques mots spécifiques. Dans le premier groupe la forme rectifiée pour *plateforme* domine (82%), pendant que les occurrences pour *portemonnaie* est faible (2,8%). Le groupe Le pluriel est traité de la même manière que dans FW, voir la section 5.2.3. En conséquence, nous pouvons conclure que le nombre d'occurrences au total pour ces items est plus grand que les chiffres présentés dans le tableau. Les pourcentages de la nouvelle graphie chez les mots dans ce groupe se trouvent entre 25 et 35%. Pour tout le groupe Les accents, accent circonflexe inclus, le nombre d'occurrences pour les formes rectifiées est faible, par exemple *événement* (8,3%), *connait* (18,8%), *bruler* (0,8%). Le même peut être observé pour les Mots étrangers, sauf pour *roqueur* (75,6%), et pour les anomalies, sauf *imbécilité* (31,9).

Il est intéressant à noter le cas d'*ognon*. Nous avons parcouru la liste des concordances pour estimer le nombre d'occurrences qui concernent la légume *ognon* et non le fleuve. En effet, ce n'est pas loin de la moitié, 411 de 934 occurrences, et il semble que parmi eux, une partie relativement grande vient des articles sur les rectifications. Sketchengine donne la possibilité de filtrer la liste des concordances par combinaison avec d'autres termes. Ainsi, *ognon* et *nénufar* apparaissent ensemble dans 34 des concordances. En plus, c'est possible de limiter la recherche par an, et des 278 des concordances qui viennent des articles publiés en 2016, l'année où les rectifications ont été intégrées dans les manuels scolaires, 214 concernent le légume (77%). Nous pouvons supposer ici que les occurrences pour l'*ognon* ne témoignent pas d'un usage de la nouvelle orthographe, plutôt ils viennent des articles qui décrivent les rectifications.

6.1.4 JSI2

Ce corpus est un sous-corpus du JSI1, et couvre le dernier trimestre en 2021.

Tableau 4 JSI2 : occurrences et proportion des items sélectionnés

Mots Nouvelle orthographe/ Ancienne orthographe	Nouvelle orthographe (nombre)	Nouvelle orthographe (pour cent)	Ancienne orthographe (nombre)	Ancienne orthographe (pour cent)
Trait d'union				
portemonnaie/portemonnaie	5	2,1%	229	97,9%
plateforme/plate-forme	15 349	80,2%	2 035	11,8%
Le pluriel				
(des) abat-jours/abat-jour	0	0	2	100%
(des) sans-abris/sans-abri	36	36%	64	64%
(des) après-midis/après-midi)	9	29%	22	71%

Les accent				
évènement/événement	2 361	18%	10 738	82%
(je) cèderai/céderai	1	20%	4	80%
L'accent circonflexe				
(il) connaît/connait	1 751	18,8%	7 537	81,2%
boîte/boite	461	11,3%	3 594	88,7%
bruler/brûler	18	2,4%	708	97,6%
brulure/brûlure	2	2,4%	79	97,6%
aout/août	174	1,8%	9 556	98,2%
Mots étrangers				
à cappella/a cappella	0	0	36	100%
cacahouète/cacahuète	1	0,7%	139	99,3%
rockeur/rocker	88	69,8%	38	30,2%
bogue/bug	29	6,1%	444	93,9%
Simplification des consonnes double				
(elle/il) épèle/épelle	0	0	1	100%
(elle/il) craquète/craquelle	1	25%	3	75%
(elle/il) feuillète/feuillette	1	6,6%	14	93,4%
girole/girolle	0	0	0	0
Anomalies				
ognon/oignon	3	0,7%	384	99,3%
assoir/asseoir	119	8,5%	1 282	91,5%
relai/relais	143	4,3%	3 156	95,7%
imbécilité/imbécillité	4	23%	13	77%

Ce corpus montre généralement les mêmes tendances que le précédent, avec quelques exceptions. D'une part, le corpus est beaucoup plus petit et pour certains items il y a très peu d'occurrences. D'autre part, la proportion d'occurrences pour *évènement* est nettement plus élevée, 18% comparé aux 8,3% pour le JSI1. Les items *évènement/événement* sont parmi les plus fréquemment utilisés dans le corpus, et c'est possible qu'ici on trouve une tendance, mais il faudrait plus d'étude pour savoir si elle est statistiquement fiable. Pour certains d'autres items le nombre d'occurrences est très petit, comme c'est le cas pour *cèderai/céderai*, *épèle/épelle*, *craquète/craquelle* etc, et dans ces cas il faut faire attention.

Dans le groupe « Le pluriel » nous avons fait une simple recherche sur les mots et lu les concordances pour sélectionner les occurrences qui correspondent aux critères de l'étude. Nous avons également fait une évaluation des concordances pour *ognon*, pour pouvoir sélectionner celles qui sont pertinentes.

6.2 Comparaison des résultats pour chaque item dans les quatre corpus

Dans cette section nous allons comparer les résultats pour chaque item dans les quatre corpus, en groupant les items par domaine des rectifications. Dans les tableaux la nouvelle orthographe sera abrégée NO et l'ancienne AO. Toutes les occurrences sont données en pourcentage.

6.2.1 Le trait d'union remplacé par la soudure

Dans cette section nous faisons une comparaison entre l'usage de la nouvelle et de l'ancienne orthographe pour les mots composés où le trait d'union a été remplacé par la soudure.

Tableau 5 Le trait d'union : proportion (%) d'utilisation de la nouvelle (NO) et de l'ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
portemonnaie /porte-monnaie	3,5	96,5	0	0	2,8	97,9	2,1	97,9
plateforme /plate-forme	71,8	28,2	68,9	31,1	82	18	68,9	31,1

On peut remarquer que le nombre d'occurrences de *portemonnaie/porte-monnaie* est petit dans tous les corpus sauf dans le Wiki, où aucune forme apparaît. Ce fait peut affecter la fiabilité des résultats, néanmoins, les pourcentages dans les trois autres corpus sont relativement conformes : c'est l'ancienne orthographe qui domine largement. Encore plus clairement, on peut distinguer une consistance entre corpus concernant les pourcentages observés pour les items *plateforme/plate-forme* mais dans le sens opposé. La proportion élevée de la nouvelle orthographe de ce mot (autour de 70 pourcents) indique qu'elle est beaucoup plus intégrée dans l'usage courante que les autres mots examinés dans cette étude.

6.2.2 Le pluriel de mots composés réunis de trait d'union

Dans cette section nous faisons une comparaison entre l'usage de la nouvelle et de l'ancienne orthographe pour le pluriel des mots composés réunis de trait d'union.

Tableau 6 Le pluriel de mots composés réunis de trait d'union : proportion (%) d'utilisation de la nouvelle (NO) et de l'ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
abat-jours/abat-jour	46,6	53,4	0	0	35,4	64,6	0	100
sans-abris/sans-abri	43,3	56,7	36,4	63,6	30,6	69,4	36	64
après-midis/après-midi	30,7	69,3	55,5	44,5	25,3	74,7	29	71

Les résultats indiquent un usage relativement élevé de la nouvelle orthographe, surtout dans FW et Wiki, les deux corpus plus généraux. L’usage des formes rectifiées est légèrement plus faible dans les deux corpus d’articles de presse.

6.2.3 Groupe accents A

En premier lieu nous allons comparer entre l’usage de la nouvelle et de l’ancienne orthographe pour l’accent grave devant une syllabe contenant un *e* muet.

Tableau 7 Accent grave devant syllabe contenant *e* muet : proportion (%) d’utilisation de la nouvelle (NO) et de l’ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
évènement/événement	15,3	84,7	15,7	84,3	8,3	91,7	18	82
(je) cèderai/céderai	31,5	68,5	33,3	66,7	15,1	84,9	20	80

Les items *évènement/événement* sont relativement fréquents dans les corpus. À partir des proportions d’utilisation, on peut distinguer une certaine conformité entre les deux corpus plus généraux, FW et Wiki, comparé au grand corpus d’articles de presse JSI1, où l’occurrence de la nouvelle orthographe est nettement plus faible. Toutefois, le résultat de JSI2, un petit corpus mais avec un contenu très récent, suggère un changement dans l’usage dans la presse. Pourtant, on peut avoir des doutes sur la validité de ces données, autant plus qu’une recherche dans JSI1 donne le même résultat que pour le corpus entier, 8,3 % d’occurrences pour *évènement*.

Les occurrences pour *cèderai/céderai* sont faibles dans les quatre corpus. Par conséquent, il faut traiter les données avec précaution, même si elles suivent les mêmes tendances que pour *évènement*.

6.2.4 Groupe accents B circonflexe

En deuxième lieu nous regardons l’usage de la nouvelle et de l’ancienne orthographe pour l’accent circonflexe.

Tableau 8 Accent circonflexe : proportion (%) d’utilisation de la nouvelle (NO) et de l’ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
(il) connaît/connâit	27,5	72,5	41,7	58,3	18,8	81,2	18,8	81,2

boite/boîte	38,4	61,6	29	71	13,5	85,5	11,3	88,7
bruler/brûler	10	90	8,4	91,6	0,8	99,2	2,4	97,6
brulure/brûlure	4,4	95,6	0	100	0,8	99,2	2,4	97,6
aout/août	2,9	97,1	0	0	1,8	98,2	1,8	98,2

Premièrement les résultats indiquent une différence entre l'usage de l'accent circonflexe sur *i* et sur *u*. Dans les quatre corpus le pourcentage pour la nouvelle orthographe de *connait* et *boite* est plus élevé que pour *bruler*, *brulure* et *aout*. Mais il est probable que la grande différence au moins en partie est dû à la sélection de mots pour examiner l'usage de circonflexe sur *u*. Une recherche supplémentaire dans FW 2017 donne 19 % d'occurrences pour *coute* contra 81 % pour *coûte*, 8 % pour *gout* contra 92 % pour *goût* et 14 % pour *croute* contra 86 % pour *croûte*. Il faudra donc plus d'études pour pouvoir tirer des conclusions concernant cette comparaison.

Deuxièmement on peut distinguer une différence entre FW et Wiki, quant aux *connait/connaît* et *boite/boîte*. Nous avons examiné un échantillon de concordances pour *boite* et *boîte* dans les deux corpus, pour voir si l'orthographe peut varier selon la définition du mot. Pourtant, nous n'avons pas trouvé des indications qui confirment une telle différence. Par exemple, *boite/boîte déroulante* est un terme informatique fréquent, utilisé aussi bien dans le code, voir (1) et (2) que dans le texte courant, comme dans les deux dernières citations (3) et (4) :

- (1) « </s><s> Boîte déroulante/fin </s><s> »
- (2) « </date></s><s> boite déroulante|titre=non pertinent|contenu===== Suggestion =====</s><s> »
- (3) « Peut-être suffirait-il de l'ajouter sans *boîte déroulante*, quoique [...] »
- (4) « [...] qui pourrait donner ça (en enlevant le modèle de la *boite déroulante* of course) [...] »
(italiques ajoutées)

6.2.5 Les mots étrangers

Dans cette section nous faisons une comparaison entre l'usage de la nouvelle et de l'ancienne orthographe pour les mots étrangers.

Tableau 9 Francisation des mots étrangers : proportion (%) d'utilisation de la nouvelle (NO) et de l'ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
à cappella / a cappella	4	96	0	100	1,7	98,3	0	100
cacahouète / cacahuète	10,1	89,9	23,8	76,2	13,5	85,5	0,7	99,3

roqueur / rocker	28,6	71,4	12,1	87,9	75,6	24,4	69,8	30,2
bogue / bug	8,2	91,8	7,7	92,3	6,6	93,4	6,1	93,9

Les mots étrangers concernés par les rectifications orthographiques sont diversifiés, et les items dans ce groupe reflètent cette diversité. Les premiers, à *cappella/a cappella*, sont très peu nombreux dans les corpus, et la plupart d’entre eux sont utilisés avec l’ancienne orthographe. Les occurrences pour *cacahouète* varient entre les corpus, mais il faut traiter avec caution les résultats des corpus petits (Wiki et JSI2), étant donné que le nombre d’occurrences est faible. Dans les grands corpus (FW et JSI1) les résultats pour *cacahouète* sont plus conformes et nous montrent que l’usage de la francisation de ce mot reste faible.

Pour *roqueur/rocker* on peut distinguer une très grande variation entre les corpus généraux et les corpus d’articles de presse. Dans FW l’usage de la nouvelle orthographe est relativement élevé, avec 28 %. Dans les corpus JSI1 et JSI2, nous trouvons l’usage de la nouvelle orthographe le plus élevé dans cette étude. En tout, *roqueur/rocker* apparaissent 25 850 fois dans JSI1, et dans 75,6 % de ces occurrences la forme *roqueur* est utilisée. Les occurrences dans JSI2 sont beaucoup moins nombreuses, mais suivent la même tendance. Une hypothèse, qu’il reste à confirmer, est que la presse a adopté cette orthographe d’après le modèle établi de *footballeur*.

Pour la francisation du terme informatique *bug, bogue*, les corpus montrent tous un faible usage où l’ancienne orthographe domine encore.

6.2.6 La simplification des consonnes doubles

Dans cette section nous faisons une comparaison entre l’usage de la simplification des consonnes doubles.

Tableau 10 La simplification des consonnes doubles : proportion (%) d’utilisation de la nouvelle (NO) et de l’ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
(il) épèle/épelle	15,6	84,4	50	50	14,6	85,4	0	100
(il)craquèle/craquelle	30,1	69,9	100	0	18	82	25	75
(il) feuillète/feuillette	5,9	94,1	0	100	2	98	6,6	93,4
girole/girolle	1,2	98,8	0	0	0,9	99,1	0	0

Les items sélectionnés pour ce groupe sont peu fréquents dans les corpus, surtout les premiers, *épèle/épelle* et *craquèle/craquelle*. Pour *feuillète/feuillette* et *girole/girolle* nous

trouvons plus d'occurrences dans les grands corpus FW et JSI1, et la grande majorité sont pour l'ancienne orthographe.

6.2.7 La suppression d'anomalies

En dernier lieu nous faisons une comparaison entre l'usage de la nouvelle et de l'ancienne orthographe pour les anomalies.

Tableau 11 La suppression des anomalies : proportion (%) d'utilisation de la nouvelle (NO) et de l'ancienne (AO) orthographe dans les quatre corpus analysés

Mots	FW NO	FW AO	Wiki NO	Wiki AO	JSI1 NO	JSI1 AO	JSI2 NO	JSI2 AO
ognon/oignon	0,3	99,7	0	100	0,9	99,1	0,7	99,3
assoir/asseoir	12,1	87,9	21,5	79,5	7,9	92,1	8,5	91,5
relai/relais	5,2	94,8	15,1	84,9	3,7	96,3	4,3	95,7
imbécilité/imbécillité	39,4	60,6	22,2	77,8	31,9	68,1	23	77

Pour les items de ce groupe, les résultats de tous les corpus sauf Wiki indiquent les mêmes tendances, c'est-à-dire que la graphie *ognon* est le moins utilisé et *imbécilité* le plus, et que l'utilisation de la graphie nouvelle des autres mots dans cette catégorie se trouve au milieu. Aussi bien dans les débats animés sur Internet que dans les articles de presse plus nuancés sur les rectifications orthographiques, *ognon* fait souvent objet d'exemple de la nouvelle orthographe, et une partie des occurrences traitent les rectifications plutôt que le légume, voir section 6.1.1 et 6.1.3. En tout cas, on trouve très peu d'occurrences dans les corpus, comme le montre le tableau 11. La nouvelle orthographe est employée dans 0,3 à 0,9% pourcent. À l'exception du Wiki, les corpus indiquent un faible usage également pour l'orthographe nouvelle *relai*, et un plus grand usage pour *assoir*. Les occurrences pour *imbécilité/imbécillité* suggèrent que la nouvelle orthographe est plus intégrée dans son usage. Une explication pourrait être le fait qu'*imbécilité* harmonise avec *imbécile*.

6.3 Comparaison des résultats avec ceux de la littérature antérieure

Dans l'étude réalisée en 2002-2004, Biedermann-Pasques (2006) et ses collègues examinent l'attitude envers les rectifications orthographiques chez des étudiants francophones en France, au Québec, en Belgique et en Suisse. L'enquête avait la forme d'une autoévaluation intégrée dans un questionnaire sur le « langage ordinaire ». La question « Est-ce que vous appliquez les Rectifications de l'orthographe lorsque vous écrivez ? » était conçue comme une question ouverte, avec un champ de texte, et elle était suivie par un test graphique :

Écrivez -vous :

1. *oignon* ou *ognon* ?

2-3. *événement* ou *évènement* ? Je *céderai* ou je *cèderai* ?

4. des *abat-jour* invariable ou des *abat-jours* ?

5. chanter *a cappella* ou *à cappella* ?

6. *connaître* ou *connaitre* ?

(2006 : 11)

Biedermann-Pasques fait une analyse très détaillée des résultats de l'enquête, pour examiner les attitudes et les pratiques par rapport aux facteurs comme nationalité, sexe, âge et sujet/forme des études sur ces attitudes. L'objectif ici est de faire une comparaison entre le résultat du test graphique avec les résultats trouvés dans les corpus, et par conséquent nous allons utiliser le tableau résumant les résultats du test graphique de Biedermann-Pasques ci-dessous. (2006 : 29)

Tableau 12 Un résumé des résultats du test graphique de Biedermann-Pasques

Domaine rectifié et mots étudiés	Pourcentage pour l'ensemble des répondants pratiquant les rectifications concernés
Pluriel de mots composés réunis de trait d'union: des abat-jours	49,83%
Accent grave devant une syllabe contenant <i>e</i> muet : cèderai évènement	40,26%
La francisation de mots étranger : à cappella	18,93%
La suppression de l'accent circonflexe sur <i>i</i> (et <i>u</i>) : connaitre	3,30%
La suppression d'anomalies : ognon	0%

Pour faire une comparaison dans le tableau 13 nous avons utilisé les résultats des recherches dans le corpus FW, parce que nous supposons qu'il grâce à sa taille et la grande diversité de son contenu est le corpus qui le mieux peut représenter les usages. Nous avons analysé les occurrences de la nouvelle orthographe (NO) pour les mêmes mots que Biedermann-Pasques (colonne 2), mais étant donné le faible nombre d'occurrences pour certains entre eux, nous avons également pris les pourcentages pour tous les mots rectifiés dans chaque domaine (colonne 3) et calculé la moyenne (colonne 4).

Tableau 13 Proportion de la nouvelle orthographe utilisée dans le corpus FW 2017

1.Domaine rectifié et mots étudiés	2.NO pour le mêmes mots que chez Biedermann-Pasques (%)	3.NO pour les mots dans le même domaine (%)	4.Moyenne (%)
Pluriel de mots composés réunis de trait d'union: des abat-jours	46,6%	<i>abat-jours</i> 46,6% <i>sans-abris</i> 43,3% <i>après-midis</i> 30,7%	40,2%
Accent grave devant une syllabe contenant e muet : <i>cèderai</i> <i>évènement</i>	31,5% 15,3%	<i>cèderai</i> 31,5% <i>évènement</i> 15,3%	23,4%
La francisation de mots étranger : à cappella	4%	<i>à cappella</i> 4,0% <i>cacahouète</i> 10,1% <i>rockeur</i> 28,6% <i>bogue</i> 8,2%	12,7%
La suppression de l'accent circonflexe sur <i>i</i> (et <i>u</i>) : <i>connaitre</i>	27,5%	<i>connait</i> 27,5% <i>boite</i> 38,4% <i>bruler</i> 10,0% <i>brulure</i> 4,4% <i>aout</i> 2,9%	16,6%
La suppression d'anomalies : <i>ognon</i>	0,3%	<i>ognon</i> 0,3% <i>assoir</i> 12,1% <i>relai</i> 5,2% <i>imbécilité</i> 39,4%	14,25%

Il faut d'abord constater qu'il n'est pas possible de faire une comparaison directe, étant donné que les méthodes sont différentes. L'enquête de Biedermann-Pasques a étudié les pratiques des rectifications chez les répondants, en les demandant quelle variante qu'ils utilisent. En ce faisant, l'enquête a informé les répondants qu'il a eu lieu des rectifications. Notre étude compte tout simplement les occurrences et le pourcentage d'utilisation des deux formes (nouvelle et ancienne) dans les corpus, sans savoir qui sont les scripteurs et quelle est leur connaissance sur les rectifications.

Néanmoins, nous pouvons constater un certain nombre de correspondances en comparant uniquement les mots qui font partie des deux études, sauf pour la suppression de l'accent circonflexe. Dans l'enquête de Biedermann-Pasques seulement 3,3% des répondants écrivait *connaitre* au lieu de *connâître*. La recherche dans FW a donné une fréquence de 27,5% pour la graphie *connait*.

Si nous traitons les mots dans l'étude de Biedermann-Pasques comme des exemples des catégories dans les rectifications, et calculons la moyenne pour les mots dans chaque groupe, nous pouvons toujours identifier des correspondances pour la régularisation du pluriel de mots

composés et pour l'utilisation de l'accent grave devant une syllabe contenant un *e* muet. La différence par rapport à la suppression de l'accent circonflexe subsiste mais est moins évidente, parce que la suppression de l'accent grave sur *u* est moins fréquente dans notre étude. Enfin, on peut constater une plus grande tendance à supprimer les anomalies outre-ognon dans l'étude dans FW sans pour autant tirer des conclusions.

7. Discussion

Dans cette étude nous avons voulu examiner dans quelle mesure les rectifications orthographiques publiées en 1990 sont rentrées dans l'usage des scripteurs francophones. En vue de cet objectif nous avons établi un échantillon de 25 mots concernés par les rectifications. Ensuite nous avons effectué des recherches dans quatre corpus pour trouver des occurrences pour aussi bien la graphie rectifiée des mots sélectionnés que la graphie non rectifiée.

Tout d'abord, il y a lieu de faire quelques remarques. D'une part, il faut rappeler que les corpus ne sont pas homogènes pour ce qui est de la taille et du contenu. Notamment, il n'est pas possible de savoir quand les textes présents dans FW ont été écrits. D'autre part, il y a peu d'occurrences pour certains mots dans l'échantillon. Malgré cela nous avons pu dégager quelques tendances.

7.1 L'usage de la nouvelle orthographe dans le français écrit contemporain

À partir de données rassemblées dans les corpus, on peut voir que l'usage de la nouvelle orthographe pour l'échantillon reste toujours relativement modeste. L'ancienne orthographe domine, à quelques exceptions près. Ceci est valable pour tous les corpus, néanmoins on peut discerner des variations régulières entre d'une part les deux corpus « libres » et d'autre part les deux corpus de presse. Pour la plus grande part des mots de l'échantillon, la proportion de l'utilisation de l'ancienne orthographe est plus grande dans JSI1 et JSI2 que dans FW et Wiki, sauf pour deux mots, *plateforme* et *rockeur*. La graphie *plateforme* domine dans tous les corpus, et peut être considérée comme étant rentrée dans l'usage général. Le cas de *rockeur* est plus étonnant, dans FW et Wiki l'ancienne orthographe domine. Cela dit, la presse paraît relativement conservatrice par rapport à l'orthographe. Dedkova et Lazar (2020) ont également fait cette observation dans leur étude sur les occurrences des formes rectifiées de noms composés dans des dictionnaires et dans JSI1³. Il est probable que beaucoup de

³ *Timestamped JSI web corpus 2014-2020*

maisons de presse imposent des recommandations linguistiques aux journalistes, et que l'ancienne orthographe domine dans ces recommandations.

Si on essaie de repérer les tendances générales, on peut voir que l'usage est assez élevé pour quelques domaines de rectifications. C'est surtout le cas pour le pluriel de mots composés réunis par un trait d'union, mais également pour la suppression de l'accent circonflexe sur *i* et l'accent grave devant une syllabe contenant *e* muet. L'usage est légèrement moins fréquent pour les autres domaines : la simplification de consonnes doubles, la suppression de l'accent circonflexe sur *u*, la francisation des mots étrangers et la suppression d'anomalies. Pourtant, il faut prendre en considération que ces derniers sont des domaines disparates, et que l'acceptation est plus grande pour *rockeur* et *imbécilité* que pour *à cappella* et *ognon*. Il convient également de remarquer que le choix d'items dans l'échantillon a pu influencer les résultats, ce que nous avons pu constater dans le cas de la suppression de l'accent circonflexe sur *u*.

7.2. La variation dans le temps de l'application de la nouvelle orthographe

Dans un questionnaire sur « langage ordinaire » distribué aux étudiants en Belgique, France, Québec et Suisse, Biedermann-Pasques (2006) a inclus deux questions sur les rectifications orthographiques, dont un test graphique. Nous avons comparé les résultats obtenus dans les recherches du corpus FW avec les résultats de ce test graphique, dans l'objectif de voir si on peut trouver des variations. En effet, nous avons pu identifier une variation en ce qui concerne la suppression de l'accent circonflexe sur *i*. Dans l'enquête de Biedermann-Pasques, seulement 3,30% des répondants déclaraient qu'ils utilisent la graphie rectifiée (*connaitre*). Dans notre étude, l'usage de la forme conjuguée du même verbe *connait* dans FW⁴ est de 27,5%. De même, l'usage du mot *boite* dans FW est de 38,4%. Ces résultats pourraient indiquer une variation au fil de temps pour l'acceptation de la suppression de circonflexe sur *i*.

7.3 Quelques facteurs qui peuvent influencer l'usage de la nouvelle orthographe

Nous pouvons constater que l'usage de la nouvelle orthographe est toujours relativement faible, même plus de trente ans après la publication des rectifications. Au fil des années, des bilans ont été faits dans l'objectif d'analyser les pratiques et les attitudes envers les rectifications dans la francophonie (Petit : 2005, Biedermann-Pasques : 2006, Farid : 2012, Lazar & Mudrochová : 2020).

⁴ Dans la comparaison nous avons uniquement utilisé les résultat du FW.

Petit (2005) identifie quelques facteurs pour expliquer la lenteur d'implémentation des rectifications, dont deux peuvent avoir un impact direct sur les pratiques. Selon Petit, « les reformes de l'orthographe sont d'autant plus déstabilisantes qu'elles offrent souvent un contenu incohérent ou inachevé, plongeant là aussi l'utilisateur dans une grande perplexité » (Petit : 95). En effet, les rectifications de 1990 présentent des *variantes* de l'ancienne orthographe, les scripteurs peuvent choisir de les utiliser ou non. De plus, les rectifications contiennent des exceptions, par exemple l'accent circonflexe est supprimé sur *u* sauf pour *dû*, *mûr* et encore d'autres mots. Des recommandations pour les auteurs de dictionnaires sont incluses dans les rectifications, entre d'autres ils doivent privilégier la graphie la plus simple. Pourtant, il a fallu du temps pour que les dictionnaires adoptent les graphies rectifiées. Dans l'étude de Sicotte (2011) sur le traitement de la suppression de l'accent circonflexe sur *i* et *u*, seulement 11 des 21 dictionnaires examinés incluaient la variante rectifiée dans les articles. Ajoutons que les rectifications ne sont pas très bien connues (Biedermann-Pasques : 2006), les scripteurs peuvent craindre « que les graphies rectifiées apparaissent comme des fautes d'orthographe » (Petit : 96)

Petit soulève également l'importance d'une position prise sur les rectifications par les institutions d'enseignement et d'autres ; l'enseignement de l'orthographe nouvelle n'était pas imposé d'une manière cohérente dans les programmes scolaires en France, en Belgique et au Québec.

Depuis 2005, la situation s'est améliorée, et les conditions pour l'usage de la nouvelle orthographe sont plus favorables. D'une part, les dictionnaires incluent au moins une partie des graphies rectifiées d'une manière ou de l'autre (Béchenec & Sprenger-Charolles : 2014, Dedkova & Lazar : 2020). D'autre part, l'enseignement de l'orthographe nouvelle est plus cohérent. En France, le ministère de l'Éducation s'est prononcé en faveur de la nouvelle orthographe en 2009, et en 2016 les rectifications ont été incorporées dans les manuels scolaires.

Sans pour autant pouvoir tirer des conclusions, on peut se demander si la normalisation en quelque sorte de la nouvelle orthographe dans les œuvres de référence et dans l'enseignement scolaire sont deux facteurs qui promeuvent son usage.

8. Conclusion

Les rectifications orthographiques de 1990 ont été controversées dès leur publication et il a fallu du temps avant qu'elles ne soient intégrées dans les programmes scolaires et dans les

œuvres de référence. Nous avons vu qu'elles sont toujours relativement peu appliquées par la presse quotidienne.

De même, l'usage des rectifications reste modeste dans le français écrit contemporain. Cependant nous avons pu repérer des variations entre les domaines rectifiés. Nous avons vu un usage plus élevé de la régularisation du pluriel des mots composés et la suppression de l'accent circonflexe ainsi que l'accent grave devant une syllabe contenant un *e* muet.

Pour la suppression de l'accent circonflexe sur *i*, nous remarquons une évolution pertinente dans le temps. Cette pointe de rectification a provoqué des réactions depuis la publication des rectifications et la recherche antérieure a trouvé une très faible acceptation. Pourtant, dans les corpus nous avons vu que la nouvelle orthographe est souvent employée pour les mots sélectionnés pour étudier ce domaine, *boite* et *connait*.

Nous avons également vu une variation dans l'usage de la nouvelle orthographe pour les mots spécifiques, comme *rockeur* à l'instar de *footballeur* dans les articles de presse.

Peut-être la nouvelle orthographe d'*ognon* et de *girole* ne sera-t-elle jamais courante dans l'usage du français écrit, mais nous avons pu repérer un usage légèrement croissant des variantes rectifiées dans les autres domaines.

9. Références bibliographiques

Béchenec, Danielle & Sprenger-Charolles, Liliane (2014). Guide pratique de l'orthographe rectifiée. *Pratiques : linguistique, littérature, didactique* 163-164.

<https://journals.openedition.org/pratiques/2227>

Biedermann-Pasques, Liselotte (2006) Un bilan des rectifications d'après une enquête en France et dans la francophonie (2002-2004). Dans Biedermann-Pasques, Liselotte & Jecic, Fabrice (2006) Les rectifications orthographiques de 1990 : Analyses des pratiques réelles (Belgique, France, Québec, Suisse, 2002-2004). *Cahiers de l'Observatoire des pratiques linguistiques*, N. 1. Presses universitaires d'Orléans, pp 9-52.

http://www.raynaldethien.fr/textes/Rapport_officiel_15ans_apres_rectifications_ortho.pdf

Catach, Nina (1986). *L'orthographe française : traité théorique et pratique*. Paris : Nathan.

Catach, Nina et al (1989). Modernisation de l'écriture du français. *Le Monde*, le 7 février.

https://www.lemonde.fr/archives/article/1989/02/07/un-appel-de-linguistes-sur-l-orthographe-moderniser-l-ecriture-du-francais_4112476_1819218.html

Conseil supérieur de la langue française (1990) : Rectifications de l'orthographe, Documents administratifs, *Journal Officiel de la République française*, 100, 6 déc.

http://www.academie-francaise.fr/sites/academie-francaise.fr/files/rectifications_1990.pdf

Cook, Vivian & Bassetti, Benedetta (2005). An introduction to researching second language writing systems. Dans Cook & Bassetti (eds) *Second language writing systems*. Clevedon : Multilingual matters, pp 1-70.

Cori, M. (2008). Des méthodes de traitement automatique aux linguistiques fondées sur les corpus. *Langages*, 171, 95-110. <https://www.cairn.info/revue-langages-2008-3-page-95.htm>

Dalbera, Jean-Philippe (2002). Le corpus entre données, analyse et théorie. *Corpus*, 1.

<http://journals.openedition.org/corpus/10>

Dedková, Iva & Lazar, Jan (2020). La formation du singulier et du pluriel orthographiques des noms composés faisant l'objet de la dernière réforme de l'orthographe — entre recommandation et application. *Cahiers de praxématique*, 74.

<https://journals.openedition.org/praxematique/6668>

Farid, Georges (2012). La « nouvelle orthographe », 21 ans plus tard. *SHS Web of Conferences*, pp 2055-2069. https://www.shs-conferences.org/articles/shsconf/pdf/2012/01/shsconf_cmlf12_000018.pdf

Fayol, Michel & Jaffre, Jean-Pierre (2016). L'orthographe : des systèmes aux usages. *Pratiques : linguistique, littérature, didactique*, pp 169-170.

<https://journals.openedition.org/pratiques/2984>

Lydia Mai Ho-Dac (2021). *WikiDisc* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, v1.1, <https://hdl.handle.net/11403/wikidisc/v1.1>.

Honvault, Rénée (2005). Bilan des rectifications en France. Dans *Le point sur les rectifications de l'orthographe en 2005*, le Renouvo (Réseau pour la nouvelle orthographe du français), p. 17-34. <https://www.renouvo.org/bilan2005.pdf>

Lazar, Jan & Mudrochová, Radka (eds) (2020). Les rectifications orthographiques de 1990 et leur application dans l'espace francophone, 30 ans plus tard. *Cahiers de praxématique*, numéro 74. <https://journals.openedition.org/praxematique/6098>

Mellet, Sylvie (2002). Corpus et recherches linguistiques : introduction. *Corpus*, 1 | 2002. <http://journals.openedition.org/corpus/7>

Mudrochová, Radka (2020). La francisation des emprunts à l'anglais d'après l'orthographe rectifiée : son application en français de France et en français québécois. *Cahiers de praxématique*, 74. <https://journals.openedition.org/praxematique/6367>

Ousselin, Edward. (2004). Aux accents, citoyens! la résistance à la réforme de l'orthographe. *The French Review*, 77(3), 490–499. <http://www.jstor.org/stable/25479388>

Petit, Christine. (2005). « Réforme de l'orthographe : les tribulations d'un siècle » dans *Le point sur les rectifications de l'orthographe en 2005*, le Renouvo (Réseau pour la nouvelle orthographe du français), p. 53-55. <https://www.renouvo.org/bilan2005.pdf>

Puchovska, Zuzana (2020). L'attitude et la pratique des locuteurs francophones face à la nouvelle orthographe : la résistance de la double consonne dans la conjugaison des verbes en -eler et -eter. *Cahiers de praxématique*, 74. <https://journals.openedition.org/praxematique/6251>

Renouvo (2014). Règles d'application des recommandations orthographiques. <http://www.renouvo.org/regles.php>

Walter, Henriette (1988). *Le français dans tous les sens*. Paris : Robert Laffont.