

APPLICATION DE L'ANCIENNE ET DE LA NOUVELLE ORTHOGRAPHE DANS LA PRESSE FRANCOPHONE ET FRANÇAISE

Iva Dedková

Ostravská univerzita, Filozofická fakulta, Katedra romanistiky
Reální 5, 701 03 Ostrava, République tchèque
iva.dedkova@osu.cz

Application of old and new orthography in the Francophone and French press

Abstract: The present paper deals with the use of traditional and new French orthography in journalistic texts and is interested in finding out which orthography was more common during the period from 2014 to 2021. It is based on exploring a vast textual corpus composed of Francophone periodicals which is entitled "Timestamped JSI web corpus 2014-2021 French" and is integrated into Sketch Engine, a tool for studying how languages function. The paper aims to examine the competing graphic variants of 35 selected lexemes belonging to the grammatical category of nouns (anomalies and loanwords), first in the entire corpus, made up of Francophone periodicals, then in a sub-corpus, restricted to French periodicals, and finally in a sub-sub-corpus, made up of eleven selected French newspapers and magazines. The results of the study show not only the most frequent graphic variants in these corpora, but also a potential link between predominant spellings and other characteristics, such as the frequency of selected lexemes and the political or thematic orientation of the periodicals that are inspected. The paper concludes with the findings that some periodicals have a greater preference for new orthography than others, although old orthography remains widespread in this sector, which is not obliged to apply the new spellings. While there is no particular association between orthographic practices and the political orientation of the periodicals that are examined, there is a certain connection between orthographic practices and the thematic orientation of the publications that are analysed.

Keywords: French; old orthography; new orthography; press; frequency of occurrences

Résumé : Le présent article porte sur l'utilisation de l'ancienne et de la nouvelle orthographe française dans les textes journalistiques et s'intéresse au type d'orthographe le plus appliqué au cours de la période 2014–2021. Il se base sur l'étude d'un vaste corpus textuel constitué de périodiques francophones qui s'intitule « Timestamped JSI web corpus 2014–2021 French » et qui est intégré dans le Sketch Engine, un outil pour explorer le fonctionnement du langage.

L'article propose d'examiner les variantes graphiques concurrentes de 35 lexèmes choisis, appartenant à la classe grammaticale des noms (anomalies et emprunts), d'abord dans le corpus entier, constitué de périodiques francophones, ensuite dans un sous-corpus, restreint aux périodiques français, et puis dans un sous-sous-corpus, composé de onze journaux et magazines français sélectionnés. Il résulte de la présente étude non seulement les variantes graphiques les plus fréquentes dans les (sous-)corpus définis, mais également un lien possible ou en revanche manquant entre les graphies prédominantes et d'autres caractéristiques, telles que la fréquence des lexèmes sélectionnés et l'orientation politique ou thématique des périodiques dépouillés. L'article conclut par les constats que certains périodiques ont plus un penchant pour la nouvelle orthographe que d'autres, bien que l'ancienne orthographe demeure largement répandue dans ce secteur, qui n'est pas obligé d'appliquer l'orthographe rectifiée. Alors qu'on ne peut détecter de lien particulier entre les pratiques orthographiques et l'orientation politique des périodiques dépouillés, il existe un certain lien entre les pratiques orthographiques et l'orientation thématique des périodiques.

Mots-clés : Français ; ancienne orthographe ; nouvelle orthographe ; presse ; fréquence d'occurrences

1. Introduction

« La forme des mots est trompeuse : ce qui compte sur le plan graphique, c'est le rapport entre l'écrit et l'oral à une époque donnée, autrement dit la traduction du graphème en phonème et inversement » (Catach 1996 : 66).

Le français compte peu de mots dont l'orthographe reproduit la prononciation avec précision et économie (par ex. *il, sur, la, bel, ami*). D'autre part, il y a des mots qui s'écrivent de deux ou trois façons différentes selon les dictionnaires ou qui varient même au sein d'un même dictionnaire (*cuiller* ou *cuillère* ; *mot-clé*, *mot-clef* ou *mot clef*). Étant très complexe, l'orthographe française a fait l'objet de plusieurs réformes. Les dernières rectifications, proposées par un comité d'experts composé notamment de linguistes et lexicographes, approuvées à l'unanimité par l'Académie française et soutenues par d'autres instances francophones, ont été énoncées dans le rapport du Conseil supérieur de la langue française¹ publié le 6 décembre 1990 dans le *Journal officiel de la République française (Documents administratifs)*.² Elles avaient pour objectif de rendre l'orthographe plus facile à apprendre et à utiliser, en éliminant certaines anomalies, exceptions et irrégularités de la langue française. Ainsi, les rectifications essaient de simplifier son utilisation en francisant des mots étrangers (*un igloo* > *un iglou*), en unifiant la formation du pluriel des noms composés (*un gratte-ciel* > *des gratte-ciels*), en uniformisant l'usage des accents (*une boîte* > *une boite*, *un peso* > *un péso*), des traits d'union (*trois cents* > *trois-cents*) et des soudures (*un porte-monnaie* > *un portemonnaie*).³

¹ Le Conseil supérieur de la langue française a été créé par le premier ministre Michel Rocard en 1989 dans le but de veiller sur le destin et la qualité de la langue française, et également de préparer non pas une réforme mais des rectifications orthographiques portant sur cinq points : i) trait d'union, ii) pluriel des mots composés, iii) accents et notamment accent circonflexe, iv) corrections d'anomalies, v) participe passé avec l'auxiliaire *avoir* (cf. Catach 1996 : 86 ; Grevisse – Goosse 2008 : 83).

² Cf. Conseil supérieur de la langue française (1990) ; Lenoble-Pinson (2021 : 21).

³ « Les rectifications ne constituent pas une révolution, ni une grande réforme, ni une remise en cause de tout le système graphique du français. Les linguistes parlent d'une réformatte, affaiblie par des exceptions (parfois regrettables). [...] Elles touchent 2 400 mots, dont certains sont très rares dans l'usage [...] ».

Actuellement, les deux orthographe coexistent : l'emploi de la nouvelle orthographe est recommandé sans être imposé, et l'ancienne orthographe reste toujours valable. La nouvelle orthographe aurait dû devenir l'orthographe à enseigner prioritairement depuis la rentrée scolaire 2016–2017 en France. Cependant, la réalité ne le prouve pas, l'ancienne orthographe est toujours largement utilisée dans les écoles et universités françaises. Plusieurs dictionnaires ont été mis à jour pour attester les graphies rectifiées, citons à titre d'exemple la 9^e édition du *Dictionnaire de l'Académie française*, le *Dictionnaire de français Larousse*, le *Robert*, le *Wiktionnaire* et l'*Usito*.⁴ A contrario, par exemple, le *Trésor de la langue française informatisé* (TLFi) reste conservateur et continue à afficher les anciennes graphies (cf. Dedková – Lazar 2020 : 11), car sa rédaction est terminée depuis 1994 et le dictionnaire ne serait pas mis à jour, comme on peut le lire sur sa page Internet d'accueil. La nouvelle orthographe est appliquée dans certains manuels scolaires en France et en Belgique⁵, ceux-ci portent la mention « Nouvelle orthographe » (insistons sur l'adjectif *certain*, car en France, on continue à publier des manuels et des cahiers d'orthographe appliquant l'ancienne orthographe). Cependant, les rectifications sont peu répandues dans le grand public, qui y est peu confronté, ce qu'attestent même les enseignants chercheurs belges (cf. Dister – Naets 2020 : 12). La connaissance des nouvelles graphies demeure partielle non seulement au sein de la société française, mais également auprès de ceux qui sont le plus touchés par l'enseignement et l'apprentissage de l'orthographe, c.-à-d. les enseignants et les élèves et étudiants (cf. Ewing 2018 : 3). Nos discussions avec des étudiants et collègues français prouvent que la nouvelle orthographe est peu répandue dans le monde universitaire français.⁶ La réforme de l'orthographe de 1990 a suscité beaucoup de

Lorsqu'on les applique toutes, moins d'un mot par page est modifié. Et souvent, il ne s'agit que d'un accent » Lenoble-Pinson (2019 : 10). Ajoutons à cela que des linguistes continuent à proposer d'autres rationalisations de l'orthographe française. Entre autres, mentionnons l'ouvrage *Penser l'orthographe de demain* (Dister et al. 2009).

⁴ Concernant l'origine de nos sources, précisons que l'*Usito* est un dictionnaire québécois, région où les rectifications de l'orthographe sont attestées encore plus tôt qu'en Belgique. Les autres dictionnaires de références sont d'origine française.

⁵ Précisons qu'en Belgique, l'enseignement de la nouvelle orthographe est imposé depuis bien plus longtemps qu'en France. En effet, en Belgique, les enseignants de français de tous niveaux ont été invités à enseigner la nouvelle orthographe de manière prioritaire depuis la rentrée scolaire de 2008. De plus, son application a été recommandée par une circulaire datant de 1998.

⁶ En novembre 2023, lors d'un séminaire sur les activités ludiques destiné aux étudiants en Master 1 Sciences du langage, parcours Linguistique et didactique du Français langue étrangère et seconde, inscrits à l'Université Clermont Auvergne, UFR Lettres, Culture et Sciences Humaines, nous avons demandé à une trentaine d'étudiants quelle orthographe devrait être enseignée. Les réponses en faveur de l'ancienne orthographe soulignaient le fait que c'est l'orthographe qui est utilisée par la majorité des gens et dans les livres et qui est attendue lorsque l'on remplit ou rédige des documents administratifs et officiels. Il faut enseigner l'orthographe traditionnelle pour éviter les exclusions sociales que la nouvelle orthographe peut apporter, pour ne pas pénaliser les apprenants dans certains contextes sociaux (par ex. dans le cadre du travail), car la nouvelle orthographe n'est pas assez démocratisée et est assez dépréciée, elle n'est pas assez connue et acceptée dans le monde du travail et universitaire. Les opinions en faveur de la nouvelle orthographe accentuaient en revanche le fait qu'elle est plus simple pour les nouveaux apprenants et les personnes avec des difficultés, donc plus inclusive par rapport aux groupes sociaux, qu'il y a une meilleure cohérence entre l'orthographe et la phonologie, que les graphies sont plus simples et qu'il n'y a aucun impact sur l'oral. La nouvelle orthographe s'éloigne d'un

débats et a fait beaucoup de polémique en France. Il semble que les Français sont majoritairement contre la nouvelle orthographe. De nombreux scripteurs sont rétifs à réapprendre à écrire des mots qu'ils savent orthographier correctement : la maîtrise de la langue écrite étant une certaine marque sociale et de pouvoir, il y a une résistance des « sachants » contre la dilution de leur savoir. Nommons également à cette occasion la revue associative *Défense de la langue française*, partisane de l'orthographe traditionnelle.⁷ Cependant, il semble que le débat sur la nouvelle orthographe commence à être abandonné en France. Nous avons recherché l'occurrence du mot « orthographe » et des collocations « nouvelle orthographe » et « réforme de l'orthographe » dans notre sous-corpus de travail intitulé presse française (2014–2021) (cf. section 4) et avons obtenu des résultats qui le confirment (cf. Tableau 1). Les débats sur l'orthographe en France ont atteint leur pic en 2016, ce qui est sans doute lié à la recommandation de son enseignement prioritaire à partir de la rentrée scolaire 2016–2017, communiquée par le Ministère de l'éducation nationale.

	2014	2015	2016	2017	2018	2019	2020	2021
Orthographe	363	481	1 438	965	1 149	933	541	145
Nouvelle orthographe	0	0	48	4	0	6	1	0
Réforme de l'orthographe	1	3	193	77	16	4	4	3

Tableau 1. Occurrence du terme « orthographe » et de ses collocations dans le sous-corpus presse française (2014–2021)

À notre connaissance, la nouvelle orthographe n'est presque jamais appliquée dans les méthodes de FLE (à titre d'exemple de méthode utilisant l'ancienne orthographe, mentionnons *Imagine*, parue chez Didier en 2022). La méthode *Le français ENTRE NOUS plus 1–3*, parue en République tchèque chez Fraus entre les années 2018 et 2019, constitue un exemple rarissime.

La présente étude se focalise, dans cette optique, sur l'utilisation de l'ancienne et de la nouvelle orthographe dans la presse francophone et notamment française pendant la période 2014–2021 en analysant les variantes graphiques (graphies concurrentes) de 35 lexèmes choisis, constitués de noms comportant des anomalies diverses et des mots empruntés (parmi lesquels 12 anglicismes).

D'après le Conseil supérieur de la langue française (1990 : 12), les anomalies « sont des graphies non conformes aux règles générales de l'écriture du français (comme *ign* dans *oignon*) ou à la cohérence d'une série précise (comme *cz* dans *eczéma*) ». En ce qui concerne les mots empruntés à d'autres langues, le Conseil supérieur de la langue française (1990 : 11) indique : « Traditionnellement, les mots d'emprunt s'intègrent à la graphie du français après quelque temps. Certains, malgré

certain élitisme ou traditionalisme délétère pour la propagation de la langue, elle démystifie le français qui est une langue très difficile et qui peut paraître inaccessible. Il convient de rajouter que nous avons également discuté cette problématique avec François Trouilleux, Maître de conférences en linguistique à l'Université Clermont Auvergne, et tenons le remercier pour ses suggestions.

⁷ A contrario, l'association EROFA défend les recommandations orthographiques et travaille même sur celles à venir. Pareillement, le site internet <orthographe-recommandée.info> contribue à leur diffusion.

leur ancienneté en français, n'ont pas encore subi cette évolution ». Alors les mots empruntés sont dorénavant adaptés à l'alphabet et à la graphie du français. Ils sont désormais accentués conformément aux règles du français (*revolver* – *révoluer*) et forment leur pluriel en suivant les mots français (*un match, des matches* – *un match, des matchs*).⁸ Néanmoins, nous nous intéresserons à d'autres changements qu'ont subis les mots empruntés, par exemple la modification de la graphie *oo* en *ou* (cf. infra).

Cet article vise ainsi à observer et à déterminer :

- a) Quelle orthographe est la plus utilisée dans la presse, l'ancienne ou la nouvelle ;
- b) Dans quelle proportion les graphies rectifiées sont utilisées dans la presse ;
- c) Si l'emploi des anciennes et des nouvelles graphies est lié à la fréquence⁹ des lexèmes sélectionnés ;
- d) Si l'application de l'ancienne et de la nouvelle orthographe est en relation avec l'orientation politique ou le contenu des journaux et magazines dépouillés.

Plusieurs sources nous inspirent pour cette étude, parmi lesquelles l'article de Dister et de Naets (2020) portant, entre autres, sur les graphies rectifiées dans la presse belge francophone. En vue de notre analyse, nous avons utilisé le Sketch Engine, un outil permettant d'explorer d'énormes corpus textuels dans environ quatre-vingt-dix langues et ainsi d'examiner et d'observer le fonctionnement du langage. Le Sketch Engine rend possible la création de sous-corpus selon des critères prédéfinis tels que la langue source, le pays d'origine, le domaine, la plage temporelle, etc. Notre étude est basée sur les textes journalistiques rédigés en français et intégrés dans le corpus « Timestamped JSI web corpus 2014–2021 French » (cf. section 4).

2. Lexèmes analysés

Pour notre analyse, nous avons sélectionné 35 lexèmes. Il s'agit tantôt de noms qui présentent diverses anomalies, car leurs anciennes graphies ne sont pas conformes à la prononciation ou car elles font partie de séries désaccordées, tantôt de noms empruntés à des langues étrangères dont les formes anciennes, n'ayant pas été francisées, ont été nouvellement adaptées à la graphie du français. Afin de supprimer certaines irrégularités et de simplifier leur orthographe, le Conseil supérieur de la langue française y a apporté les modifications suivantes (cf. Dupriez 2018 : 187–222) :

- Alignement de mots isolés au sein d'une même famille ou série : *eczéma m* – *exéma m* (famille ou série : *examen m, exécution f*), *oignon m* – *ognon m* (famille ou série : *rognon m, trognon m*).
- La graphie *cch* peut s'écrire *cc* dans le mot suivant et ses dérivés : *saccharine f* – *saccarine f*.
- Francisation des mots anglais terminés par le morphème *-er* et prononcés *-eur* : *baby-sitter m/f* – *babysitteur m/f*, *leader m* – *leadeur m*, *thriller m* – *thrilleur m*.

⁸ Explicitons que pour les exemples, l'ancien est mentionné d'abord, le nouveau ensuite.

⁹ Cf. Dal – Namer (2015 : 48–49) pour l'explication de la notion « la fréquence d'un lexème X », qui correspond « à la somme des fréquences d'occurrences des mots-formes qui le réalisent dans un corpus donné ».

- La graphie *gh* peut s'écrire *g* devant les lettres autres que *e* et *i* dans quelques mots d'origine étrangère : *yoghourt*¹⁰ *m* – *yogourt m* (emprunt au turc).
- La graphie *-illier* peut s'écrire *-iller*, conformément à la prononciation : *serpillière f* – *serpillère f*.
- La graphie *k* ou *kh* peut s'écrire *c* : *diktat m* – *dictat m* (emprunt à l'allemand).
- La graphie *oo* d'origine étrangère peut s'écrire *ou*, conformément à la prononciation française : *boom m* – *boum m* (emprunt à l'anglais), *hooligan m* – *houligan m* (emprunt à l'anglais), *igloo m* – *iglou m* (mot inuit).
- La graphie *ph* peut s'écrire *f* (les deux graphies sont antérieures aux dernières rectifications orthographiques) : *képhir m* – *kéfir m* (mot caucasien), *phantasme m* – *fantasme m*, *téléphérique m* – *téléférique m*, *nénuphar m* – *nénufar m*.
- La graphie *un* de quelques mots d'origine étrangère peut s'écrire *on*, conformément à la prononciation : *acupuncture f* – *acuponcture f*.
- La graphie *y* à la finale de certains mots anglais peut s'écrire *i* ou *ie* : *garden-party f* – *garden-partie f*, *grizzly m* – *grizzli m*.
- Introduction d'un *e* à la finale de mots (étrangers) terminés par une consonne : *pogrom m* – *pogrome m* (emprunt au russe).
- Modifications diverses de certaines graphies de mots empruntés : *cacahuète f* – *cacahouète f* (emprunt à l'espagnol), *corned-beef m* – *cornedbif m* (emprunt à l'anglais), *freesia m* – *frésia m* (mot allemand latinisé), *muesli m* – *musli* ou *muesli m* (mot d'origine suisse),¹¹ *spéculoos m* – *spéculos m* (emprunt au néerlandais), *tagliatelles f* – *taliatelles f* (emprunt à l'italien).¹²
- Réduction de la consonne double qui suit un *e* muet : *prunellier m* – *prunelier m*.
- Réduction du *l* à la finale des mots terminés par *olle* : *corolle f* – *corole f*.
- Réduction des consonnes et voyelles doubles de quelques mots d'origine étrangère : *cannette f* – *canette f* (emprunt à l'anglais), *shopping m* – *shoping m* (emprunt à l'anglais), *freesia m* – *frésia m* (emprunt à l'allemand), *spéculoos m* – *spéculos m* (emprunt au néerlandais), *shampooing m* – *shampoing m* (emprunt à l'anglais).
- Uniformisation des consonnes doubles : *bonhomie f* – *bonhomme f* (famille ou série : *homme m*, *bonhomme m*), *cahute f* – *cahutte f* (famille ou série : *hutte f*, *lutte f*), *chariot m* – *charriot m* (famille ou série : *charrette f*, *charretier m*), *courbature f* – *courbature f* (famille ou série : *battre f*, *battue f*, *batture f*).
- Remplacement du trait d'union par une soudure dans les mots mentionnés supra : *baby-sitter m/f* – *babysitteur m/f*, *corned-beef m* – *cornedbif m*.

À propos de la graphie *nénuphar*, Duprier (2018 : 203) ajoute qu'elle a été indûment calquée sur le grec. Ce mot qui est venu de l'arabe (*nīnūfar*) fut orthographié

¹⁰ D'après nos recherches, la variante graphique *yaourt*, qui se maintient, est beaucoup plus fréquente que les variantes *yoghourt* et *yogourt* : *yaourt* bénéficie de 24 398 occurrences attestées dans le corpus entier et de 6 211 occurrences dans le sous-corpus presse française, contre *yoghourt* à hauteur de 309 occurrences dans le corpus entier et 28 occurrences dans le sous-corpus presse française, et *yogourt* à hauteur de 3 268 occurrences dans le corpus entier et 33 occurrences dans le sous-corpus presse française.

¹¹ Le Conseil supérieur de la langue française recommande la graphie *musli* (1990 : 18).

¹² Le phonème /g/ ne devrait pas être prononcé, selon *Le Robert*, quoiqu'il le souvent soit.

nénufar dans les dictionnaires, jusque dans la huitième édition du *Dictionnaire de l'Académie* de 1932.

Concernant la paire *boom* – *boum*, les occurrences de la variante rectifiée seront en réalité inférieures aux résultats obtenus, car elle dispose de deux homographes : *boum* fonctionnant comme interjection et *boum*, nom féminin, signifiant « réunion festive de jeunes chez l'un d'entre eux » (cf. *Dictionnaire de français Larousse*). Lorsque nous avons limité la requête à la classe des noms, seules certaines interjections ont été filtrées. Une analyse linguistique fine devrait être réalisée afin de discerner le nom masculin du nom féminin et de l'interjection. Vu le nombre de données à traiter (13 451 occurrences de *boum* dans le corpus entier et 3 205 occurrences dans le sous-corpus) et la difficulté à le faire, nous avons également recherché les occurrences de la collocation *baby boom* – *baby boum*, pour mieux pouvoir illustrer la disproportion entre les deux graphies.

Les variantes rectifiées *leadeur* (m, f) et *babysitteur* (m, f) entraînent les féminins *leadeuse* (cf. Usito ou Wiktionnaire) et *babysitteuse* (cf. Usito ou *Dictionnaire de français Larousse*). À côté de la graphie traditionnelle *baby-sitter*, le dictionnaire Usito établit également la graphie soudée *babysitter*. Cependant, les variantes *leadeur* et *leadeuse* ne figurent ni dans la 9^e édition du *Dictionnaire de l'Académie française*, ni dans le *Dictionnaire de français Larousse*.

Certains lexèmes peuvent présenter diverses variantes graphiques, par exemple *baby-sitter*, *babysitter*, *babysitteur/babysitteuse*, *baby-sitteur/baby-sitteuse*, le morphème français -eur s'appliquant au masculin aussi bien qu'au féminin.¹³

Les résultats globaux ne contiennent ni le pluriel *garden-parties*, car on ne sait pas s'il est relatif à la graphie ancienne *garden-party* ou à la graphie nouvelle *garden-partie*, ni le pluriel *grizzlies*, car de nombreuses occurrences de ce dernier se rapportent aux dénominations des équipes sportives, par exemple *les Memphis Grizzlies*. Remarquons que les noms sont souvent conservateurs, et que le dernier cas cité présente une structure anglaise donc écrite à l'anglaise.

3. Hypothèses

En nous basant sur nos observations préalables, nous émettons les hypothèses suivantes sur l'application des anciennes et des nouvelles graphies dans les textes journalistiques français :

Primo, nous nous attendons à ce que l'ancienne orthographe soit toujours largement répandue dans les articles journalistiques.

Secundo, nous prévoyons que certains journaux et magazines auront une préférence pour la nouvelle orthographe tandis que d'autres en auront une pour l'orthographe traditionnelle.

Tertio, nous croyons que les pratiques orthographiques pourraient également être liées à la fréquence des lexèmes sélectionnés : vu le caractère simplificateur et

¹³ Les formes hybrides *babysitter* et *baby-sitteur/baby-sitteuse* n'ont pas été incluses dans notre analyse, car elles portent les marques des deux graphies et il n'est pas possible de les classer dans l'une ou l'autre catégorie.

unificateur de l'orthographe rectifiée, les lexèmes avec une faible fréquence d'occurrences lui seront-ils plus soumis ?

Enfin, les deux orthographes coexistant, nous croyons que les pratiques orthographiques pourraient également dépendre d'autres phénomènes, tels que l'orientation politique et/ou thématique des périodiques dépouillés, l'écriture propre à chaque journaliste ou le pays d'origine des textes (la France, la Belgique, le Canada, etc.). Néanmoins, cette hypothèse n'a pas vocation d'être vérifiée dans le cadre de cette étude, basée sur l'analyse d'un vaste corpus dont la source est tout l'espace francophone et son sous-corpus restreint à la France.

4. Analyse du nombre d'occurrences des variantes graphiques dans la presse francophone et française

Dans un premier temps, nous avons recherché les occurrences des lexèmes sélectionnés dans le corpus entier, intitulé « Timestamped JSI web corpus 2014–2021 French ». Ce corpus comprend 8 166 710 150 tokens, c.-à-d. unités lexicales, et 6 998 186 326 mots. C'est un vaste corpus francophone composé d'articles de journaux d'actualité portant sur des sujets variés et publiés entre les années 2014 et 2021. Le corpus contient 319 sources médiatiques, parmi les plus importantes figurent *Le Parisien*, *Le Figaro*, *L'Équipe*, *La Dernière Heure*, *Point Magazine*, *Le Monde*, *La Voix du Nord*, *20 Minutes* et *Montagne* ; nous pouvons y trouver également *Le Matin*, *La Libre Belgique*, *Courrier Picard*, *Tribune de Genève*, etc. Les domaines de premier niveau sont : fr, com, ca, be, net, ch, org, lu, ma et tn. Parmi les pays d'origine des médias d'information figurent la France, le Canada, la Belgique, la Suisse, l'Algérie, le Maroc, le Luxembourg, la Tunisie, la Réunion, le Congo, le Liban, la Polynésie française, le Mali, la Mauritanie, Maurice, le Burkina Faso, le Sénégal, etc. L'outil utilisé pour travailler avec le corpus était concordance (exemples en contexte).

Dans un deuxième temps, nous avons défini un sous-corpus composé uniquement de journaux et magazines de provenance française, que nous avons par suite nommé « presse française ». Ce sous-corpus contient 1 333 288 131 tokens et 1 142 516 214 mots. Il comporte 76 sources médiatiques variées (genre : journal & magazine), telles que la presse quotidienne nationale (PQN), la presse quotidienne régionale (PQR), la presse hebdomadaire nationale (PHN) ou la presse hebdomadaire régionale (PHR). Voici les titres les plus importants : *Le Figaro*, *Le Monde*, *Le Parisien*, *Point Magazine*, *La Voix du Nord*, *L'Équipe*, *Montagne*, *Libération*, *20 Minutes*, *Nord Éclair*, *La Provence*, *L'Express magazine*, *Courrier Picard* et *Paris Match magazine*.

Chaque lexème a d'abord été recherché dans sa graphie traditionnelle, ensuite avec sa graphie rectifiée, y compris ses formes au singulier et au pluriel. Prenons pour exemple le lexème *can(n)ette* et ses formes *cannette*, *cannettes*, *canette* et *canettes*.

Certains mots, par exemple les noms propres de personnes tels que Lars Boom (cycliste néerlandais), en lien avec le terme *boom*, Florian Fresia orthographié Frésia

(joueur de rugby), en lien avec le terme *frésia*, ou le toponyme Ognon, en rapport avec la nouvelle graphie *ognon*, ont été enlevés des résultats obtenus.

Tous les résultats ont été minutieusement listés dans les tableaux A1 et A2 qui se trouvent en annexe. Ils concernent la fréquence d'occurrences des lexèmes examinés, convertie de plus en pourcentages arrondis à une décimale. Pour chaque paire de variantes graphiques, la fréquence d'occurrences la plus élevée est surlignée afin de faciliter la lecture des données. D'ailleurs, les graphiques n° 1 et 2 infra permettent une comparaison globale en pourcentages des graphies dominantes.

L'analyse des données concernant le corpus et le sous-corpus permet de mettre en évidence les éléments suivants :

De prime abord, nous remarquons que la plupart des lexèmes apparaissent avant tout en ancienne orthographe. En effet, pour le corpus entier ainsi que le sous-corpus, l'ancienne orthographe prédomine dans 29 des 35 lexèmes examinés.

Quelques mots gardent presque exclusivement leur ancienne orthographe. Les lexèmes *courbature* – *courbature* et *corned-beef* – *cornedbeef* sont exclusivement écrits selon l'ancienne orthographe. Les lexèmes *garden-party* – *garden-partie* et *saccharine* – *saccharine* conservent une graphie traditionnelle dans le sous-corpus presse française.

Ensuite, certains lexèmes sont écrits à la fois avec l'ancienne et la nouvelle orthographe, mais c'est souvent l'ancienne orthographe qui prévaut. Les 5 lexèmes suivants sont écrits selon l'ancienne orthographe dans plus de 99 % des cas : *thriller* – *thrilleur*, *shopping* – *shoping*, *oignon* – *ognon*, *chariot* – *charriot*, *téléphérique* – *téléférique*, *hooligan* – *houligan*, *baby-sitter* – *babysitteur*, *eczéma* – *exéma* et *garden-party* – *garden-partie*. Les 7 lexèmes suivants sont écrits selon l'ancienne orthographe dans plus de 92 % des cas : *leader* – *leadeur*, *diktat* – *dictat*, *cacahuète* – *cacahouète*, *acupuncture* – *acuponcture*, *igloo* – *iglou*, *pogrom* – *pogrome*, *corolle* – *corole*, *cahute* – *cahutte* et *tagliatelles* – *taliatelles*. Les graphies traditionnelles sont également majoritaires, dans 58 % à 91 % des cas, dans les lexèmes *boom* – *boum*, *bonhomie* – *bonhomie*, *serpillière* – *serpillère*, *grizzly* – *grizzli*, *néphar* – *néphar*, *spéculoos* – *spéculos*, *freesia* – *frésia* et *prunellier* – *prunelier*. Ajoutons à cela que les formes féminines *leadeuse* et *babysitteuse* présentent une très faible occurrence dans le corpus et le sous-corpus, et que celle de *babysitteuse* n'apparaît même pas dans le sous-corpus.¹⁴

La nouvelle orthographe, en revanche, prédomine uniquement dans les six lexèmes suivants : *phantasme* – *fantasme*, *cannette* – *canette*, *shampooing* – *shampoing*, *muesli* – *musli* + *muesli*, *kéfir* – *kéfir* et *yoghourt* – *yogourt*. Les résultats ne sont pas étonnants, car les deux graphies de ces mots sont antérieures aux dernières rectifications orthographiques, ce n'est pas le cas pour les paires *oignon* – *ognon* ou *tagliatelles* – *taliatelles*. Aucun lexème ne se rencontre exclusivement en nouvelle orthographe.

Les résultats pour le corpus et le sous-corpus paraissent similaires, sans toutefois présenter les mêmes caractères. Une seule différence importante, liée aux lexèmes

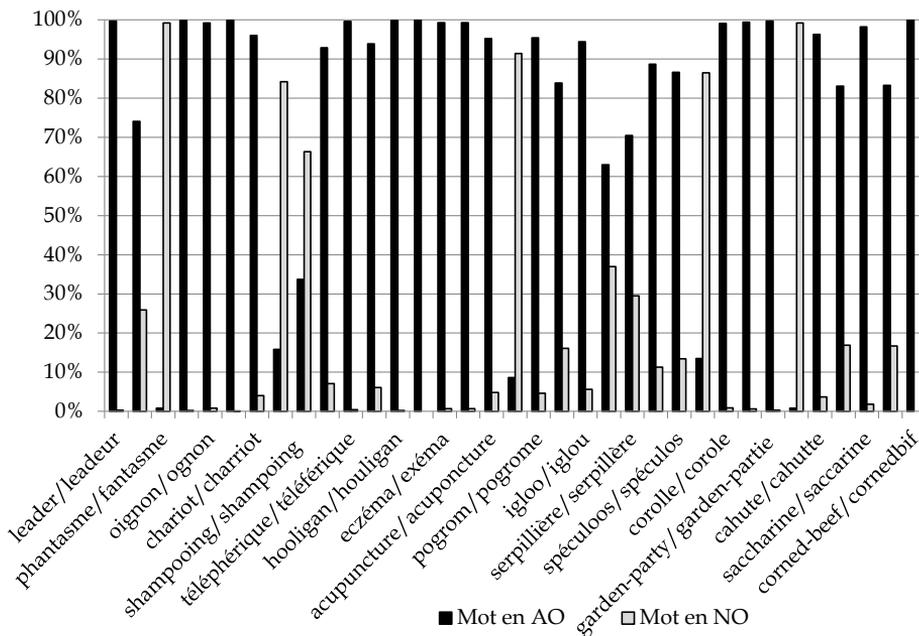
¹⁴ Voici le nombre d'occurrences par million d'unités lexicales : *leadeuse* 0,31 dans le corpus et 0,04 dans le sous-corpus, *babysitteuse* inférieure à 0,01 dans le corpus.

pour lesquels prédomine la nouvelle orthographe, concerne la paire *yoghourt* – *yogourt* : la variante rectifiée *yogourt* apparaît dans 91,4 % des cas dans le corpus entier mais uniquement dans 54,1 % des cas dans le sous-corpus, elle serait donc plus répandue dans d'autres pays francophones qu'en France.

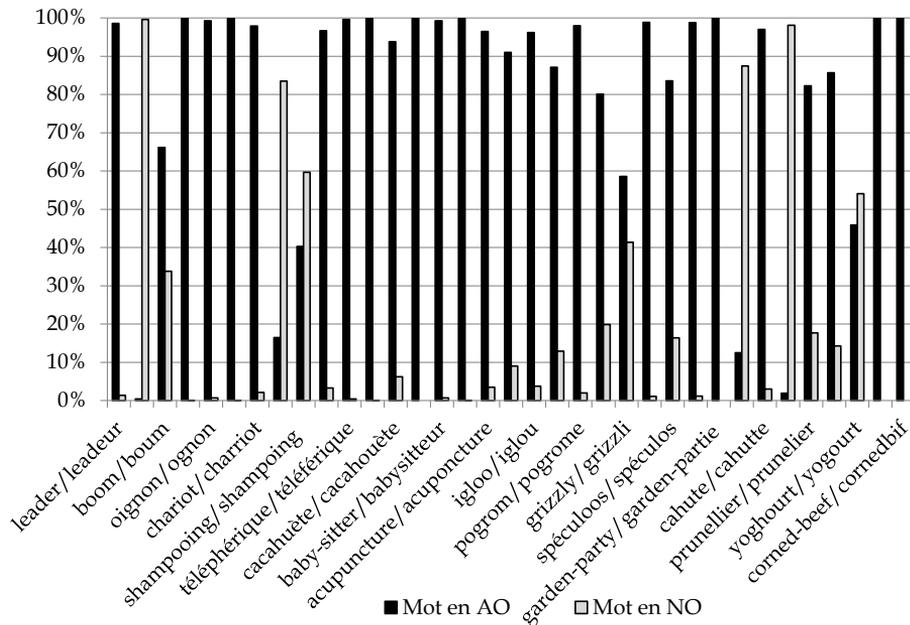
L'ancienne graphie la plus fréquente dans le corpus et le sous-corpus est *leader*, suivi de *shopping*, *oignon* et *thriller*, alors que les nouvelles graphies les plus fréquentes sont *fantasme*, *canette* et *boum*. Nous notons que l'ordre des lexèmes par fréquence d'occurrences dans le corpus et le sous-corpus n'est pas le même, notamment pour les anciennes graphies *freesia* (34^e rang contre 29^e rang), *yoghourt* (29^e rang contre 34^e rang) et *grizzly* (21^e rang contre 26^e rang) et les nouvelles graphies *yogourt* (5^e rang contre 18^e rang) et *exéma* (24^e rang contre 31^e rang). Somme toute, nous constatons que 91,51 % des mots du corpus entier et 88,65 % des mots du sous-corpus sont en ancienne orthographe, contre seulement 8,49 % des mots du corpus entier et 11,35 % des mots du sous-corpus en nouvelle orthographe.

Par ailleurs, concernant la collocation *baby boom* – *baby boum*, nous notons une forte prédominance de l'ancienne orthographe, plus élevée que pour la paire *boom* – *boum*.

Les graphiques n° 1 et 2 infra mettent en parallèle les pourcentages des graphies dominantes des 35 lexèmes analysés pour permettre un aperçu global : les anciennes graphies sont en noir et les nouvelles en gris.



Graphique 1. Corpus entier (2014–2021)



Graphique 2. Sous-corpus presse française (2014–2021)

5. Analyse du nombre d’occurrences des variantes graphiques dans les périodiques choisis

Pour la suite de notre analyse, nous avons restreint le sous-corpus presse française à onze périodiques, journaux et magazines, quotidiens ou hebdomadaires. Ce sous-sous-corpus (c.-à-d. sous-corpus du sous-corpus précédent) est basé sur les titres les plus importants et contient quatre quotidiens nationaux d’information politique et générale (Le Figaro – le quotidien le plus lu de France, Le Monde, Libération et La Croix ; L’Humanité a une diffusion plus petite que les précédentes), un quotidien national sportif (L’Équipe), un quotidien d’information générale (20 Minutes), deux quotidiens régionaux (Le Parisien et La Voix du Nord) et deux hebdomadaires nationaux (L’Express et Le Point). Dans le tableau 2, nous indiquons leurs titres, leur orientation politique ou thématique principale, leur date de création et leur site internet. Nous pouvons y trouver des périodiques de gauche à droite de l’échiquier politique. Notre recherche concerne également la période 2014–2021.

Tous les périodiques n’ont pas la même taille de corpus de textes intégrés dans le Sketch Engine, Le Figaro comprend le plus grand nombre d’unités lexicales et L’Humanité le moins (cf. Tableaux 3 et 4). Nous avons recherché les formes des lexèmes sélectionnés en ancienne et en nouvelle orthographe dans chaque titre couvrant la période des années 2014 à 2021 et incluant une diversité de thèmes traités.

Ensuite, pour chaque titre, nous avons additionné toutes les occurrences des formes analysées pour telle ou telle orthographe (cf. Tableaux 3 et 4) en vue d'obtenir un pourcentage des formes écrites selon l'ancienne et la nouvelle orthographe (cf. Tableau 5). Comme les pourcentages en nombres décimaux exprimant les occurrences des formes choisies sont très faibles, ils sont rapportés à l'ordre de dix puissance moins quatre et les résultats sont arrondis à deux décimales.

Titre	Orientation politique / thème	Quotidien ou hebdomadaire	Date de création	Site internet
Le Monde	Centre gauche	PQN	1944	lemonde.fr
Le Figaro	Droite	PQN	1826	lefigaro.fr
L'Équipe	Magazine sportif	PQN	1946	lequipe.fr
La Croix	Catholique	PQN	1880/3	la-croix.com
Libération	Gauche	PQN	1973	liberation.fr
20 Minutes	Centre	Quotidien	2002	20minutes.fr
Le Parisien	Centre	PQR	1944	leparisien.fr
La Voix du Nord	Gauche	PQR	1941	lavoixdunord.fr
L'Express	Droite	PHN	1953	lexpress.fr
Le Point	Centre-droit	PHN	1972	lepoint.fr
L'Humanité	Gauche	PQN	1904	humanite.fr

Tableau 2. Périodiques français choisis

Quoique la taille du corpus de l'Équipe ne représente que 43,82 % de celle du Figaro, elle contient le pourcentage le plus élevé des formes écrites suivant l'ancienne orthographe (348,39⁻⁴). L'Express et Le Parisien bénéficient en revanche des pourcentages les plus élevés des formes écrites selon la nouvelle orthographe (34,48⁻⁴ et 31,82⁻⁴).

Périodiques AO	Nombre d'unités lexicales	Occurrences des formes choisies	Pourcentage .10 ⁻⁴
Le Figaro	125 955 854	17 596	139,70
Le Monde	121 557 860	13 348	109,81
Le Parisien	87 346 331	9 097	104,15
Le Point	78 675 882	12 043	153,07
20 Minutes	69 856 508	8 175	117,03
Libération	64 147 711	9 186	143,20
L'Équipe	55 196 785	19 230	348,39
La Voix du Nord	50 009 201	5 700	113,98
L'Express	44 314 178	6 381	105,90
La Croix	34 910 284	3 697	98,97
L'Humanité	136 169	8	58,75

Tableau 3. Ancienne orthographe dans les périodiques français choisis

Périodiques NO	Nombre d'unités lexicales	Occurrences des formes choisies	Pourcentage .10 ⁻⁴
Le Figaro	125 955 854	1 934	15,35
Le Monde	121 557 860	2 668	21,95
Le Parisien	87 346 331	2 779	31,82
Le Point	78 675 882	1 467	18,65
20 Minutes	69 856 508	1 320	18,90
Libération	64 147 711	1 468	22,88
L'Équipe	55 196 785	197	3,60
La Voix du Nord	50 009 201	698	13,96
L'Express	44 314 178	1 528	34,48
La Croix	34 910 284	327	7,37
L'Humanité	136 169	1	7,34

Tableau 4. Nouvelle orthographe dans les périodiques français choisis

Il ressort de la comparaison en pourcentages des lexèmes écrits selon les deux orthographe (cf. Tableau 5) que tous les journaux et magazines dépouillés ont une préférence pour l'ancienne orthographe. L'Équipe utilise presque exclusivement l'ancienne orthographe : en effet, 98,99 % des formes analysées sont écrites selon l'ancienne orthographe. C'est le lexème *leader* dont le nombre d'occurrences s'élève à 18 612 du total des 19 427 occurrences des lexèmes analysés (c.-à-d. 95,80 %), écrit exclusivement en ancienne orthographe, qui influence les résultats (sa présence étant indissociablement liée à la thématique sportive). Suivent la Croix avec 91,87 % des formes en ancienne orthographe, Le Figaro avec 90,10 % des formes en ancienne orthographe, Le Point avec 89,11 % des formes en ancienne orthographe et enfin La Voix du Nord avec 89,09 % des formes en ancienne orthographe. L'Équipe est, nous l'avons déjà souligné, un magazine sportif. Pour le reste, il s'agit de périodiques de gauche ou de droite sur l'échiquier politique.

Périodiques	Pourcentage AO	Pourcentage NO
L'Équipe	98,99 %	1,01 %
La Croix	91,87 %	8,13 %
Le Figaro	90,10 %	9,90 %
Le Point	89,11 %	10,89 %
La Voix du Nord	89,09 %	10,91 %
L'Humanité	88,89 %	11,11 %
Libération	86,22 %	13,78 %
20 Minutes	86,10 %	13,90 %
Le Monde	84,32 %	15,68 %
L'Express	80,68 %	19,32 %
Le Parisien	76,60 %	23,40 %

Tableau 5. Comparaison de la fréquence d'occurrences des graphies concurrentes

Parmi les périodiques qui appliquent le plus la nouvelle orthographe figurent d'abord Le Parisien avec 23,40 % des formes écrites selon la nouvelle orthographe, puis L'Express avec 19,32 % des formes en nouvelle orthographe, Le Monde avec 15,68 % des formes en nouvelle orthographe, 20 Minutes avec 13,90 % des formes en nouvelle orthographe et enfin Libération avec 13,78 % des formes en nouvelle orthographe. Il s'agit de titres classés à gauche, au centre et à droite de l'échiquier politique. Nous constatons que l'orientation politique n'a donc pas d'incidence sur le type d'orthographe appliquée. Comme l'Humanité n'a qu'une petite taille de corpus intégré dans le Sketch Engine, nous ne considérons pas ses résultats comme pertinents.

Par ailleurs, nous notons que certains lexèmes sont typiques de certains périodiques, ce que prouve le nombre d'occurrences par million d'unités lexicales (cf. Tableau 6). Mentionnons à cette occasion les anciennes graphies *cannette* (La Voix du Nord), *courbature* (L'Équipe), *diktat* (Le Monde), *hooligan* (L'Équipe, 20 Minutes), *pogrom* (Libération), *spéculoos* (La Voix du Nord), *boom* (Le Monde, Le Figaro, Le Point), *leader* (L'Équipe), *shopping* (Le Figaro, Le Parisien) et *thriller* (Le Point).

Terme AO	Périodique	Sous-corpus fréquence/ million ¹	Périodique fréquence/ million
Cannette	La Voix du Nord	0,56	1,5
Courbature	L'Équipe	0,96	1,76
Diktat	Le Monde	1,73	2,34
Hooligan	L'Équipe	1,45	3,42
Hooligan	20 Minutes	1,45	2,65
Pogrom	Libération	0,41	1,43
Spéculoos	La Voix du Nord	0,25	0,88
Boom	Le Monde	4,8	5,84
Boom	Le Figaro	4,8	5,93
Boom	Le Point	4,8	6
Leader	L'Équipe	95,77	337,19
Shopping	Le Figaro	5,62	6,11
Shopping	Le Parisien	5,62	7,08
Thriller	Le Point	6,07	13,89

Tableau 6. Comparaison du nombre d'occurrences des lexèmes examinés par million d'unités lexicales

Terme NO	Périodique	Sous-corpus fréquence/million	Périodique fréquence/ million
Fantasma	Le Monde	8,38	14,56
Canette	Le Parisien	2,83	4,43
Boum	20 Minutes	2,25	3,45
Leadeur	Le Parisien	1,4	16,92
Shampoing	L'Express	1,4	7,04

Tableau 7. Comparaison du nombre d'occurrences des lexèmes examinés par million d'unités lexicales

En ce qui concerne les nouvelles graphies (cf. Tableau 7), nommons *fantasme* (Le Monde), *canette* (Le Parisien), *boum* (20 Minutes), *leader* (Le Parisien) et *shampooing* (L'Express).

6. En guise de conclusion

Lors de l'étude des variantes graphiques des 35 lexèmes sélectionnés dans de vastes corpus textuels intégrés dans le Sketch Engine, nous avons noté une disproportionnalité dans l'application de l'ancienne et de la nouvelle orthographe dans la presse francophone et française.

Comme prévu, l'ancienne orthographe est toujours largement répandue dans les articles journalistiques. Nous constatons que seules 11,35 % des formes concernées dans le sous-corpus presse française sont écrites suivant la nouvelle orthographe. Les formes rectifiées sont majoritaires dans seulement 6 des 35 mots (*phantasme* – *fantasme*, *cannette* – *canette*, *shampooing* – *shampooing*, *müesli* – *musli* + *muesli*, *képhir* – *kéfir* et *yoghourt* – *yogourt*), dont les secondes graphies existent de plus longue date et sont depuis longtemps prônées par les dictionnaires (cf. par exemple le TLFi). Dans les autres cas, ce sont les graphies traditionnelles qui sont les plus utilisées (29 des 35 lexèmes examinés). D'ailleurs, les lexèmes *courbature* – *courbature* et *corned-beef* – *cornedbeef* se rencontrent exclusivement en orthographe traditionnelle dans le corpus entier, à quoi s'ajoutent les lexèmes *garden-party* – *garden-partie* et *saccharine* – *saccharine* qui ne conservent qu'une graphie traditionnelle dans le sous-corpus presse française. Nous doutons donc de l'utilité de leurs graphies nouvelles, qui n'ont pas été acceptées par les scripteurs. Cependant, nous devons ajouter que la presse représente tout de même un secteur qui n'est pas obligé d'appliquer la nouvelle orthographe. De plus, le choix des termes a certainement un impact. D'autre part, la presse est attachée à l'ancienne orthographe, car elle redoute de perdre des lecteurs, sachant que tout changement peut perturber sa clientèle.

Si l'on inspecte de plus près les nouvelles formes les plus utilisées, nous remarquons que les modifications effectuées portent sur les simplifications (réductions) et non sur les séries désaccordées (notamment les ajouts). Côté anglicismes, si les scripteurs préfèrent la forme rectifiée *shampooing* à la forme traditionnelle *shampooing*, le contraire est vrai pour la paire *shopping* – *shopping*.

Dans le cadre de notre étude, nous n'avons pas trouvé de lien particulier entre les pratiques orthographiques et l'orientation politique des périodiques dépouillés. Quoiqu'ils aient tous une préférence pour l'orthographe traditionnelle, certains se servent davantage des graphies rectifiées (notamment Le Parisien et L'Express) et d'autres moins (La Croix ou Le Figaro). A contrario, il existe un certain lien entre les pratiques orthographiques et l'orientation thématique des périodiques, qui se manifeste par l'occurrence élevée de l'ancienne graphie *leader* dans le magazine sportif l'Équipe, ce qui en fait le périodique conservant le plus les graphies traditionnelles. En tout cas, il faut prendre en compte le fait que les graphies rectifiées peuvent être employées soit consciemment, parce que les scripteurs les ont apprises, soit spontanément, parce que les formes rectifiées sont plus simples et

plus rationnelles (cf. Dister – Naets 2020 : 12). Néanmoins, nous ne supposons pas que les journalistes, en tant qu'agents glottopolitiques, soient entièrement libres de leurs choix et puissent décider eux-mêmes de l'orthographe à utiliser lors de la rédaction d'articles. Chaque organe de presse possède son protocole avec des choix internes. Pour illustration¹⁵, le journal *La Croix* a une charte orthotypographique répertoriant les choix orthographiques et typographiques propres au journal, ainsi qu'une charte de transcription des noms étrangers (transcription des caractères arabes, cyrilliques, chinois...) et une politique visant à éviter autant que possible les anglicismes. Le journal *Le Monde* dispose également d'une liste de mots avec une orthographe recommandée pour chacun. Étant donné que la réforme des années 1990 est peu appliquée, les deux journaux cités ne la suivent que pour certaines règles et certains termes (ils ont par exemple recours à la soudure et écrivent *plateforme* ou *minijupe* sans trait d'union). Concernant les modes de révision de textes au sein des rédactions, les correcteurs ou journalistes réviseurs professionnels utilisent un logiciel de correction et les dictionnaires usuels (*Le Monde* a une préférence pour le *Petit Larousse*), travaillant sur écran et intervenant directement dans les textes (à *La Croix*, une relecture de l'intégralité du journal est effectuée, tandis que pour *Le Monde*, environ 90 % des articles du journal papier et 70 % des articles en ligne sont corrigés). En fait, les journalistes et les correcteurs suivent certaines conventions pour que les écrits du journal soient unifiés et harmonisés au mieux.

La présente problématique soulève d'autres interrogations qui n'ont pas été abordées dans le cadre de cette étude, telles que l'application de la nouvelle orthographe dans la presse d'autres pays francophones, les directives de rédaction adoptées par les journalistes et les correcteurs en matière d'orthographe, ou encore les effets de la scolarisation de la jeune génération selon la nouvelle orthographe. Pour cette raison, nous envisageons, dans nos recherches ultérieures, d'analyser l'emploi de l'ancienne et de la nouvelle orthographe dans la presse d'autres pays francophones tels que la Belgique ou le Canada et de comparer leurs pratiques orthographiques avec celles obtenues pour la presse française.

Références bibliographiques

- CATACH, Nina (1996), *L'orthographe*, Paris : PUF (6^e éd.).
- COCTON, Marie-Noëlle (dir.) (2022), *Imagine 1 – 2 (méthode de français)*, Paris : Didier FLE.
- DAL, Georgette – NAMER, Fiametta (2015), « La fréquence en morphologie : pour quels usages ? », *Langages* 2015/1 (N^o 197), 47–68.
- DEDKOVÁ, Iva – LAZAR, Jan (2020), « La formation du singulier et du pluriel orthographiques des noms composés faisant l'objet de la dernière réforme orthographique – entre recommandation et application », *Cahiers de praxématique* 74, 1–16.
- DISTER, Anne et al. (2009), *Penser l'orthographe de demain*, Paris : Conseil international de la langue française.

¹⁵ Nos remerciements vont à Geneviève Peillon, responsable du service de révision de *La Croix*, et à Muriel Gilbert, membre de l'équipe de correction du journal *Le Monde*, pour avoir partagé avec nous ces informations importantes.

- DISTER, Anne – NAETS, Hubert (2020), « Les Rectifications de l'orthographe en Belgique francophone : de la politique linguistique aux pratiques des écoliers et de la presse », *Cahiers de praxématique* 74, 1-16.
- DUPRIEZ, Dominique (2018), *La nouvelle orthographe en pratique*, Bruxelles : De Boeck/Éditions Duculot (3^e éd.).
- GEFFROY-KONŠTACKÝ, Danièle et al. (2018-2019), *Le français ENTRE NOUS plus 1-3*, Plzeň : Fraus.
- EWIG, Anna (2018), « Le rôle de la presse dans la discussion de la réforme de l'orthographe française », *Circula* 7, Les Éditions de l'Université de Sherbrooke, 71-89.
- GREVISE, Maurice – GOOSSE, André (2008), *Le Bon usage*, Bruxelles : De Boeck & Larcier s.a. (14^e éd.).
- LENOBLE-PINSON, Michèle (2019), « L'Académie française et la nouvelle orthographe », *XLinguae* 12 (1XL-2019), 1-10.
- LENOBLE-PINSON, Michèle (2021), « La guerre du nénufar n'a pas eu lieu en Belgique », *Studia Romanistica* 21/1, 19-31.

Sitographie

- ATILF, *Trésor de la langue française informatisé*, <<http://atilf.atilf.fr/>> [1/11/2022].
- CONSEIL SUPÉRIEUR DE LA LANGUE FRANÇAISE (1990), « Les Rectifications de l'orthographe », *Journal officiel de la République française n° 100*, <https://www.academie-francaise.fr/sites/academie-francaise.fr/files/rectifications_1990.pdf> [22/6/2022].
- Défense de la langue française*, <<http://www.langue-francaise.org/index.php>> [1/11/2022].
- Dictionnaire de l'Académie française*, <<https://www.dictionnaire-academie.fr/>> [1/11/2022].
- EROFA, *Études pour une Rationalisation de l'Orthographe Française d'Aujourd'hui*, <<http://erofa.free.fr/>> [4/5/2023].
- LAROUSSE, *Dictionnaire de français*, <<http://www.larousse.fr/dictionnaires/francais/>> [1/11/2022].
- LE ROBERT, *Dico en ligne*, <dictionnaire.lerobert.com> [1/11/2022].
- La nouvelle orthographe, parlons-en !* <<https://www.orthographe-recommandee.info/>> [4/5/2023].
- SKETCH ENGINE, <<https://www.sketchengine.eu/>> [3/12/2022].
- UDS (Université de Sherbrooke), *Usito*, <<https://usito.usherbrooke.ca/>> [1/11/2022].
- WIKTIONNAIRE, *Le dictionnaire libre*, <https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil> [1/11/2022].

Annexe

Mot en AO	Occurrences	Mot en NO	Occurrences	Total (100 %)
Leader <i>m/f</i>	899355 (99,7%)	Leadeur <i>m/f</i>	2845 (0,3%)	902200
Boom <i>m</i>	38573 (74,1%)	Boum <i>m</i>	13451 (25,9%)	52024
Phantasme <i>m</i>	395 (0,8%)	Fantasme <i>m</i>	49645 (99,2%)	50040
Shopping <i>m</i>	45654 (99,8%)	Shoping <i>m</i>	81 (0,2%)	45735
Oignon <i>m</i>	42156 (99,2%)	Ognon <i>m</i>	334 (0,8%)	42490
Thriller <i>m</i>	40135 (99,99%)	Thrilleur <i>m</i>	5 (0,01%)	40140
Chariot <i>m</i>	23320 (96%)	Charriot <i>m</i>	965 (4%)	24285
Cannette <i>f</i>	3752 (15,8%)	Canette <i>f</i>	19976 (84,2%)	23728
Shampooing <i>m</i>	5904 (33,7%)	Shampoing <i>m</i>	11636 (66,3%)	17540
Diktat <i>m</i>	15822 (92,9%)	Dictat <i>m</i>	1211 (7,1%)	17033
Téléphérique <i>m</i>	12101 (99,6%)	Téléférique <i>m</i>	50 (0,4%)	12151
Cacahuète <i>f</i>	8336 (93,9%)	Cacahouète <i>f</i>	538 (6,1%)	8874
Hooligan <i>m</i>	8359 (99,8%)	Houligan <i>m</i>	18 (0,2%)	8377
Courbature <i>f</i>	5912 (100%)	Courbature <i>f</i>	0 (0%)	5912
Eczéma <i>m</i>	5110 (99,3%)	Exéma <i>m</i>	35 (0,7%)	5145
Baby-sitter <i>m/f</i>	4811 (99,3%)	Babysitteur <i>m/f</i>	35 (0,7%)	4846
Acupuncture <i>f</i>	3752 (95,2%)	Acuponcture <i>f</i>	191 (4,8%)	3943
Yoghourt <i>m</i>	309 (8,6%)	Yogourt <i>m</i>	3268 (91,4%)	3577
Pogrom <i>m</i>	3391 (95,4%)	Pogrome <i>m</i>	162 (4,6%)	3553
Bonhomie <i>f</i>	2965 (83,9%)	Bonhommie <i>f</i>	570 (16,1%)	3535
Igloo <i>m</i>	3246 (94,4%)	Iglou <i>m</i>	192 (5,6%)	3438
Grizzly <i>m</i>	2035 (63%)	Grizzli <i>m</i>	1195 (37%)	3230
Serpillère <i>f</i>	2211 (70,5%)	Serpillère <i>f</i>	924 (29,5%)	3135
Nénuphar <i>m</i>	2017 (88,7%)	Nénufar <i>m</i>	257 (11,3%)	2274
Spéculoos <i>m</i>	1288 (86,6%)	Spéculos <i>m</i>	200 (13,4%)	1488
Müesli <i>m</i>	193 (13,5%)	Musli+muesli <i>m</i>	21+1220 (86,5%)	1434
Corolle <i>f</i>	1278 (99,1%)	Corole <i>f</i>	11 (0,9%)	1289
Tagliatelles <i>f</i>	1239 (99,4%)	Taliatelles <i>f</i>	7 (0,6%)	1246
Garden-party <i>f</i>	987 (99,7%)	Garden-partie <i>f</i>	3 (0,3%)	990
Képhir <i>m</i>	7 (0,8%)	Kéfir <i>m</i>	872 (99,2%)	879
Cahute <i>f</i>	783 (96,3%)	Cahutte <i>f</i>	30 (3,7%)	813
Prunellier <i>m</i>	246 (83,1%)	Prunelier <i>m</i>	50 (16,9%)	296
Saccharine <i>f</i>	276 (98,2%)	Saccarine <i>f</i>	5 (1,8%)	281
Freesia <i>m</i>	174 (83,3%)	Frésia <i>m</i>	35 (16,7%)	209
Corned-beef <i>m</i>	206 (100%)	Cornedbif <i>m</i>	0 (0%)	206

Tableau A1. Corpus entier (2014–2021)

Mot en AO	Occurrences	Mot en NO	Occurrences	Total (100 %)
Leader <i>m/f</i>	127683 (98,6%)	Leadeur <i>m/f</i>	1873 (1,4%)	129556
Phantasme <i>m</i>	44 (0,4%)	Fantasme <i>m</i>	11171 (99,6%)	11215
Boom <i>m</i>	6282 (66,2%)	Boum <i>m</i>	3205 (33,8%)	9487
Thriller <i>m</i>	8098 (99,96%)	Thrilleur <i>m</i>	3 (0,04%)	8101
Oignon	7787 (99,3%)	Ognon	52 (0,7%)	7839
Shopping <i>m</i>	7499 (99,9%)	Shoping <i>m</i>	9 (0,1%)	7508
Chariot	5724 (97,9%)	Charriot	121 (2,1%)	5845
Cannette <i>f</i>	743 (16,5%)	Canette <i>f</i>	3772 (83,5%)	4515
Shampooing <i>m</i>	1261 (40,3%)	Shampoing <i>m</i>	1871 (59,7%)	3132
Diktat <i>m</i>	2301 (96,7%)	Dictat <i>m</i>	79 (3,3%)	2380
Téléphérique <i>m</i>	2241 (99,6%)	Téléférique <i>m</i>	8 (0,4%)	2249
Hooligan <i>m</i>	1927 (99,9%)	Houligan <i>m</i>	2 (0,1%)	1929
Cacahuète <i>f</i>	1837 (93,8%)	Cacahouète <i>f</i>	122 (6,2%)	1959
Courbature <i>f</i>	1286 (100%)	Courbature <i>f</i>	0 (0%)	1286
Baby-sitter <i>m/f</i>	1240 (99,3%)	Babysitteur <i>m/f</i>	9 (0,7%)	1249
Eczéma <i>m</i>	1110 (99,9%)	Exéma <i>m</i>	1 (0,1%)	1111
Acupuncture <i>f</i>	976 (96,5%)	Acuponcture <i>f</i>	35 (3,5%)	1011
Bonhomie	772 (91%)	Bonhommie	73 (9%)	845
Igloo <i>m</i>	718 (96,2%)	Iglou <i>m</i>	28 (3,8%)	746
Néuphar	549 (87,1%)	Nénufar	81 (12,9%)	630
Pogrom <i>m</i>	546 (98%)	Pogrome <i>m</i>	11 (2%)	557
Serpillère <i>f</i>	451 (80,1%)	Serpillère <i>f</i>	112 (19,9%)	563
Grizzly <i>m</i>	267 (58,6%)	Grizzli <i>m</i>	189 (41,4%)	456
Corolle	444 (98,9%)	Corole	5 (1,1%)	449
Spéculoos <i>m</i>	327 (83,6%)	Spéculos <i>m</i>	64 (16,4%)	391
Tagliatelles <i>f</i>	399 (98,8%)	Taliatelles <i>f</i>	5 (1,2%)	404
Garden-party <i>f</i>	327 (100%)	Garden-partie <i>f</i>	0 (0%)	327
Müesli <i>m</i>	42 (12,5%)	Musli+muesli <i>m</i>	2+292 (87,5%)	336
Cahute <i>f</i>	262 (97%)	Cahutte <i>f</i>	8 (3%)	270
Képhir <i>m</i>	4 (1,9%)	Kéfir <i>m</i>	210 (98,1%)	214
Prunellier	102 (82,3%)	Prunelier	22 (17,7%)	124
Freesia <i>m</i>	54 (85,7%)	Frésia <i>m</i>	9 (14,3%)	63
Yoghourt <i>m</i>	28 (45,9%)	Yogourt <i>m</i>	33 (54,1%)	61
Saccharine <i>f</i>	45 (100%)	Saccarine <i>f</i>	0 (0%)	45
Corned-beef <i>m</i>	43 (100%)	Cornedbif <i>m</i>	0 (0%)	43

Tableau A2. Sous-corpus presse française (2014–2021)

