
PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 1-345

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22392](http://hdl.handle.net/11143/22392)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22392](https://doi.org/10.17118/11143/22392)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Table des matières

Présentation	5
--------------------	---

Nicolas Sorba et Nadine Vincent

Ce que dictionnaires et plateformes en ligne (ne) nous disent (pas) du monde contemporain : le binôme langue-culture est-il encore actuel ?	10
--	----

Chiara Molinari

Les néologismes du Covid-19 dans les dictionnaires francophones en ligne	35
--	----

Kaja Dolar et Marie Steffens

Dictionnaires et corpus numériques pour l'enseignement du FLE : un parcours didactique sur le lexique	59
--	----

Francesco Faresin

Analyser des lexèmes du quotidien dans une correspondance privée corse (1881-1918) : ressources lexicographiques existantes et apports du corpus	84
---	----

Beatrice Dal Bo et Anna Giaufret

Les enjeux d'un cadre de référence en ligne pour le corse	104
---	-----

Nicolas Sorba et Alain di Meglio

Les dictionnaires de langue des signes française à l'ère du numérique : un premier état des lieux	122
--	-----

Amélie Josselin-Leray et Jérémie Segouat

Pour une analyse qualitative *et* quantitative, manuelle *et* computationnelle, synchronique *et* diachronique, des dictionnaires numériques154

Franck Sajous

La lexicographie numérique et la norme grammaticale :
le *Wiktionnaire* est-il aussi descriptif que voulu ?188

Sophie Piron

La rubrique *Dire, Ne pas dire* du dictionnaire numérique de l'Académie française.
Un changement de paradigme ?207

Caroline Salagnac

Études métalexigraphiques par analyse d'échantillons :
toute représentativité est-elle illusoire ?240

Franck Sajous

Les préjugés sexistes dans les dictionnaires de professionnels et de profanes :
un petit espoir de changement?274

Annick Farina

Féminicide : un mot ambigu pour une société divisée.....297

Laetitia Chicoine

Le Québec est-il une province?
Regard lexicographique sur un emploi polémique.....324

Nadine Vincent

TITRE: PRÉSENTATION

AUTEURS: NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT,
UNIVERSITÉ DE SHERBROOKE

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES
PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT,
UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 4-8

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22416](http://hdl.handle.net/11143/22416)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22416](https://doi.org/10.17118/11143/22416)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS
ATTRIBUTION 4.0 INTERNATIONALE.

Présentation

Collectif dirigé par Nicolas Sorba et Nadine Vincent

La lexicographie numérique s'impose aujourd'hui comme la manière usuelle de consulter les ouvrages de référence. C'est maintenant en ligne que se renouvellent les pratiques de production, d'archivage, de diffusion et de consultation. La présente publication s'intéresse à l'effet du numérique sur la conception et la réception des dictionnaires de langue dans l'espace francophone et réunit des spécialistes qui réfléchissent aux retombées de ces transformations sur la pratique lexicographique en particulier, mais aussi sur la description de la langue au sens large.

Il est établi que « le dictionnaire est une description de la culture et, en ce sens, [qu']il est un texte culturel » (Dubois et Dubois, 1971 : 99). Plusieurs chercheurs ont fait de cet ouvrage emblématique un corpus sociolinguistique (Rey, 2008) et y ont identifié les valeurs d'une société, y compris ses aspects moins glorieux : sexisme, racisme et autres discriminations et idéologies (Beaujot, 1989; Boulanger, 1999, 2001; Lehmann, 1981, 1989; Kottelat, 2009 et 2010, etc.). Les choses ont-elles changé depuis l'arrivée des dictionnaires en ligne ? L'accessibilité du numérique donne-t-elle une voix à des langues ou des variétés de français minoritaires peu traitées traditionnellement ? Les dictionnaires collaboratifs présentent-ils une image plus éclatée, moins homogène que les dictionnaires émis par une source unique ? Les agrégateurs font-ils remonter à la surface des préjugés que l'on croyait disparus ? Une nouvelle lexicographie émerge-t-elle sur le Web ou les modèles dominants sont-ils reproduits malgré le changement de format ? Les néologismes sont-ils attestés avec ouverture ou sont-ils soumis à des filtres conservateurs ? Peut-on toujours, aujourd'hui, parler du dictionnaire comme d'un texte culturel ?

Ce sont les questions que nous avons posées à des spécialistes de la langue et de la lexicographie. Les réponses obtenues sont riches en diversité et en perspectives, et témoignent du fait que la culture et les idéologies ne sont pas des compartiments étanches.

Dictionnaires et culture

- **Chiara MOLINARI**, Università degli Studi di Milano (Italie) : *Ce que dictionnaires et plateformes en ligne (ne) nous disent (pas) du monde contemporain : le binôme langue-culture est-il encore actuel ?*
- **Kaja DOLAR**, CREE – Inalco (France) et **Marie STEFFENS**, Université de Liège (Belgique) et Universiteit Utrecht (Pays-Bas) : *Les néologismes du Covid-19 dans les dictionnaires francophones en ligne*
- **Francesco FARESin**, Università di Verona (Italie) et Université Grenoble-Alpes (France) : *Dictionnaires et corpus numériques pour l'enseignement du FLE : un parcours didactique sur le lexique*
- **Beatrice DAL BO**, Université Sorbonne Nouvelle (France), et **Anna GIAUFRET**, Università di Genova (Italie) : *Analyser des lexèmes du quotidien dans une correspondance privée corse (1881-1918) : ressources lexicographiques existantes et apports du corpus*
- **Nicolas SORBA et Alain DI MEGLIO**, Università di Corsica Pasquale Paoli (Corse) : *Les enjeux d'un cadre de référence en ligne pour le corse*
- **Amélie JOSSELIN-LERAY et Jérémie SEGOUAT**, CLLE, CNRS et Université Toulouse – Jean Jaurès (France) : *Les dictionnaires de langue des signes française à l'ère du numérique : un premier état des lieux*
- **Franck SAJOUS**, CLLE, CNRS et Université Toulouse – Jean Jaurès (France) : *Pour une analyse qualitative et quantitative, manuelle et computationnelle, synchronique et diachronique des dictionnaires numériques*

Dictionnaires et idéologies

- **Sophie PIRON**, Université du Québec à Montréal, CRIFUQ (Québec) : *La lexicographie numérique et la norme grammaticale : le Wiktionnaire est-il aussi descriptif que voulu ?*
- **Caroline SALAGNAC**, Paul Valéry Université Montpellier 3 – Praxiling (France) : *La rubrique Dire, Ne pas dire du dictionnaire numérique de l'Académie française. Un changement de paradigme ?*
- **Franck SAJOUS**, CLLE - CNRS et Université de Toulouse 2 (France) : *Études métalexico-graphiques par analyse d'échantillons : toute représentativité est-elle illusoire ?*
- **Annick FARINA**, Università degli Studi di Firenze (Italie) : *Les préjugés sexistes dans les dictionnaires de professionnels et de profanes : un petit espoir de changement ?*
- **Laetitia CHICOINE**, Université de Sherbrooke (Québec) : *Féminicide : un mot ambigu pour une société divisée*

- **Nadine VINCENT**, Université de Sherbrooke , CRIFUQ, laboratoire LexiQcorpus (Québec) : *Le Québec est-il une province ? Regard lexicographique sur un emploi polémique*

Références

- BEAUJOT, Jean-Pierre (1989), « Dictionnaires et idéologies » dans Franz-Josef Hausmann (dir.), *Encyclopédie internationale de lexicographie*, Berlin : De Gruyter.
- BOULANGER, Jean-Claude (1998), « À propos de l'arrimage entre le dictionnaire et la néobienséance » dans Denise Deshaies et Conrad Ouellon (dir.), *Les linguistes et les questions de langue au Québec: points de vue*, Québec : Centre international de recherche en aménagement linguistique, p. 160-171.
- BOULANGER, Jean-Claude (1999), « Un épisode de la néobienséance dans les dictionnaires scolaires : le protocole de rédaction des exemples comprenant des prénoms », *Romanistik in Geschichte und Gegenwart*, vol. 5, n° 1, p. 67-95.
- DUBOIS, Jean et Claude Dubois (1971), *Introduction à la lexicographie : le dictionnaire*. Paris : Larousse.
- KOTTELAT, Patricia (2009), « Lexicographie et clichés doxiques : la permanence de stéréotypes sexistes dans un dictionnaire d'apprentissage contemporain, le *Robert Junior 99* », *Studi e Ricerche, Quaderni del Dipartimento di Scienze del Linguaggio e Letterature Moderne e Comparate dell'Università di Torino*, no 4, Alessandria : Edizioni dell'Orso, p. 165-185.
- KOTTELAT, Patricia (2010), « Définitions lexicographiques et idéologie : ambiguïtés discursives dans les définitions des races, traces de permanence de stéréotypes racistes? », *Publif@rum*, n° 11 *Autour de la définition*.
- LEHMANN, Alise (1981), *Analyse du discours lexicographique : le corps sexué dans le Petit Larousse illustré de 1906 à 1980*, thèse de doctorat, Université Paris VIII.
- LEHMANN, Alise (1989), « Les représentations idéologiques dans le discours du dictionnaire. Diversité des lieux et des procédures », *Lexiques*, p. 106-112.
- REY, Christophe (2008), *Dictionnaire et société*, Paris : Honoré Champion.

TITRE: CE QUE DICTIONNAIRES ET PLATEFORMES EN LIGNE (NE) NOUS DISENT (PAS) DU MONDE CONTEMPORAIN : LE BINÔME LANGUE-CULTURE EST-IL ENCORE ACTUEL ?

AUTEUR: CHIARA MOLINARI, UNIVERSITÀ DEGLI STUDI DI MILANO

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 9-33

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22393](http://hdl.handle.net/11143/22393)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22393](https://doi.org/10.17118/11143/22393)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONALE.

Ce que dictionnaires et plateformes en ligne (ne) nous disent (pas) du monde contemporain : le binôme langue-culture est-il encore actuel ?

Chiara Molinari, Università degli Studi di Milano

Résumé :

Dans cet article, nous nous proposons de réfléchir aux rapports entre nouvelles créations lexicales et supports lexicographiques. Autrement dit, nous nous intéressons à la façon dont les nouvelles plateformes lexicographiques (que ce soit des dictionnaires en ligne, des dictionnaires collaboratifs ou d'autres typologies d'outils lexicographiques) intègrent et/ou reproduisent les néologismes, ceux-ci étant normalement considérés comme un miroir des dynamiques socioculturelles. Pour ce faire, nous allons étudier un corpus hétérogène, composé d'articles de la presse consacrés à la langue et d'articles lexicographiques proposés par les dictionnaires collaboratifs et bases de données. La confrontation avec une ressource lexicographique traditionnelle, *Le Petit Robert*, permettra d'évaluer les apports des outils en ligne.

Mots-clés :

Lexicographie collaborative, dictionnaires en ligne, presse, néologismes, linguistique populaire

Abstract :

In this article, we propose to study the relationships between new lexical creations and lexicographic supports. In other words, we are interested in the way in which new lexicographic platforms (whether online dictionaries, collaborative dictionaries or other typologies of lexicographic tools) integrate and/or reproduce neologisms, these being normally considered as a mirror of socio-cultural dynamics. For this purpose, we will study a heterogeneous corpus, composed of press articles devoted to the language and lexicographic articles offered by collaborative dictionaries and databases. The comparison with a traditional lexicographical resource, *Le Petit Robert*, will allow us to evaluate the contributions of the tools online.

Keywords :

Collaborative lexicography, online dictionaries, press, neologisms, folk linguistics

1. Prémises

De nos jours, les liens entre « dictionnaires » et « culture(s) » ou entre « dictionnaires » et « idéologie(s) » ne sont plus à prouver. Déjà en 1979, avec Simone Delasalle, Alain Rey réfléchissait aux relations entre dictionnaires de langue et culture et affirmait :

Il [le dictionnaire] ne construit pas un modèle abstrait et simplifié des ensembles lexicaux essentiels, mais tente d'utiliser la richesse et la complexité des discours, ceux-ci véhiculant en bloc la totalité des problèmes linguistiques [...] et de très nombreuses questions socioculturelles, épistémologiques, etc. (Rey et Delasalle, 1979 : 7)

Le dictionnaire de langue est, pour les auteurs, un « dictionnaire culturel » (Rey et Delasalle, 1979 : 7) et ceci malgré les difficultés qu'implique une lexicographie qui se veut culturelle (Rey et Delasalle, 1979 : 11). Citons, à titre d'exemple, la dichotomie entre la « description d'une "compétence lexicale" unique et fictive » et la « description des variables sociales et des connotations » (Rey et Delasalle, 1979 : 11). Ces premières réflexions aboutiront, après plusieurs années, au célèbre ouvrage en quatre volumes, *Le dictionnaire culturel* où le lexicographe explore la dimension culturelle des mots conçus en tant que reflet d'une conception du monde (Rey, 2006). Elles seront aussi reprises par Pruvost, d'après qui les dictionnaires ne seraient pas de simples recueils de mots mais des « lieux privilégiés des savoirs sur la langue et sur le monde » (Pruvost, 2006 : 11).

Si l'on estime normalement que le but du dictionnaire est de décrire la langue telle qu'elle est parlée et écrite à une époque déterminée et dans un endroit précis, il n'en reste pas moins que la langue y est saisie dans son dynamisme :

« Tout dictionnaire, écrit-il [Alain Rey], puise dans la pluralité d'usages et prétend fournir une image de la "langue" ; en fait, il construit une proposition d'usage fondée sur une hiérarchie » ([Rey, 1983], p. 546). Il précise plus tard : le dictionnaire « est une construction, une tentative pour proposer une description normée, équilibrée d'une réalité dynamique et extrêmement complexe » (1985, p. 280) qu'est le lexique d'une langue à travers les usages multiples de ses nombreux utilisateurs. (Corbeil, 2003 : 273)

Dans le même sillage, Pruvost ajoute : « Alors même qu'ils [les dictionnaires] semblent destinés à fixer des savoirs, ils ne sont en vérité, à l'échelle du temps, que des instantanés » (2006 : 12). En d'autres termes, les dictionnaires sont les documents d'une période. Ils reflètent les préoccupations des individus ainsi que les sujets relayés par les médias. Et, c'est bien en tant que reflet d'une culture et d'une vision du monde que les dictionnaires sont censés être ouverts aux évolutions linguistiques et lexicales. Et pourtant, l'on connaît le scepticisme des lexicographes face à l'intégration de mots et/ou de sens nouveaux dans les dictionnaires (Pruvost, 2006 ; Sablayrolles, 2015 : 19). La méfiance à l'égard de tout ce qui tient à la néologie semble avoir été dépassée par le passage au numérique : il suffit d'ouvrir la version en ligne du *Grand Robert* ou du *Petit Robert* pour qu'un encadré invitant les usagers à découvrir les nouveautés intégrées frappe notre regard :

poké

Précédent

Suivant

Figure 1. Le *Petit Robert* en ligne, section « Découvrez une sélection des nouveautés du Petit Robert¹ »

Il sera donc question de réfléchir aux changements introduits grâce à l'irruption d'internet : est-ce que ce dernier accélère l'intégration de mots nouveaux, considérés comme représentatifs des tendances culturelles et des dynamiques sociales, dans la nomenclature ? Les dictionnaires numériques seraient-ils effectivement plus ouverts aux changements, comme le remarquaient des utilisateurs déjà en 2016 (Molinari, 2017) ? Et, par conséquent, seraient-ils plus à même par rapport aux versions papier de refléter les cultures des contextes sociaux contemporains, ceux-ci étant de plus en plus changeants ?

2. Corpus et méthodologie

Pour essayer de répondre à ces questions, nous développerons une réflexion articulée en trois parties. Dans une première partie, nous porterons le regard sur les articles tirés de la presse où les journalistes relèvent et commentent les mots nouveaux intégrés dans les millésimes des dictionnaires Robert et Larousse.

Dans une deuxième partie, nous examinerons le retentissement que les mots intégrés par les dictionnaires des maisons Robert et Larousse ont dans les dictionnaires collaboratifs (tels que le *Wiktionnaire* et le *Dictionnaire de la Zone*²) afin de mesurer l'écart entre ces outils plus récents et les dictionnaires traditionnels. Notre choix est justifié par les opinions relevées il y a quelques années dans les forums où les participants, tout en affirmant que les dictionnaires collaboratifs ne peuvent pas être considérés comme des références, en reconnaissent la modernité et la souplesse à l'égard des néologismes (Molinari, 2017).

Enfin, à l'époque actuelle, d'autres instruments reflètent la relation langue et réalités socioculturelles et participent, par conséquent, à l'élaboration progressive des représentations que les usagers élaborent de la langue et des contextes dans lesquels ils vivent. La troisième partie sera donc consacrée à des plateformes en ligne, telles que le site *Dico en ligne*, géré par la maison Robert³. Donc, d'une part le public est confronté à la mise en ligne de dictionnaires papier ou à de véritables dictionnaires conçus pour être mis directement en ligne. De l'autre, il existe tout un appareil qu'on pourrait appe-

1. [page consultée en février 2024]

2. https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d'accueil ; <https://www.dictionnairedelazone.fr>

3. <https://dictionnaire.lerobert.com>

ler, en paraphrasant Genette (1987), « paratexte lexicographique » et dont le but est éminemment vulgarisateur. Cependant ces plateformes, de par leurs contenus et de par la façon dont ceux-ci sont présentés, attirent le regard des lecteurs et captivent leur intérêt, stimulant de ce fait leur sensibilité à l'égard des dynamiques lexicales.

Sur le plan méthodologique, nous allons mobiliser des instruments différents selon le corpus. Pour ce qui concerne les articles tirés de la presse, notre but est de relever les représentations sur le lexique que la presse reflète et en même temps contribue à élaborer. Parmi les nombreuses définitions qui ont été proposées pour le concept de représentation (Jodelet, 1989 ; Calvet, 1999 ; Boyer, 2003), nous adoptons celle de Petitjean qui réfléchit, de façon spécifique, aux représentations linguistiques. Celles-ci correspondent à

un ensemble de connaissances non scientifiques, socialement élaborées et partagées, fondamentalement interactives et de nature discursive, disposant d'un degré plus ou moins élevé de jugement et de figement, permettant au(x) locuteur(s) d'élaborer une construction commune de la réalité linguistique et de gérer leurs activités langagières au sein de cette interprétation commune de la réalité (Petitjean, 2010 : 294).

Il sera donc question d'explorer les « contenus représentationnels mais aussi leurs processus de co-construction discursive » (Gajo, dans Petitjean, 2010 : 294). Dans la mesure où nous allons sélectionner des articles de la presse, nous nous situons dans le cadre de la sociolinguistique et de la linguistique populaire (Achard-Bayle et Paveau, 2008), à savoir ce courant de la linguistique qui s'intéresse aux savoirs spontanés et profanes et qui réunit, d'après Achard-Bayle et Paveau, « trois sortes de pratiques linguistiques : descriptives (on décrit l'activité de langage), normatives (on prescrit les comportements langagiers) et interventionnistes (on intervient sur les usages de la langue) » (2008 : 9-10). Dans ce cadre spécifique, nous nous situons dans l'approche descriptive de la linguistique populaire, notre but étant de relever les représentations qui circulent et se diffusent dans (et grâce à) la presse et leurs enjeux dans la construction des savoirs lexicographiques profanes. En général, les journalistes qui rédigent ces articles ne sont pas des linguistes, mais on peut les inscrire dans la catégorie des « amoureux de la langue », dont les attitudes peuvent être les plus diverses et inclure une inclination pour le purisme ou pour la norme ou encore se limiter à reproduire « les représentations spontanées et populaires sur la langue et sur son enseignement » (Paveau et Rosier, 2008 : 91).

Ensuite, les deux dernières sections prévoient une démarche de nature lexicologique et métalxicographique : en effet, il sera question d'explorer les termes traités par les articles de presse dans les dictionnaires collaboratifs et bases de données afin de vérifier si le lexique intégré renvoie aux phénomènes culturels de la période contemporaine et, dans le cas spécifique, quelles sont les stratégies mises en œuvre pour le décrire.

2.1. La presse

Les articles qui composent le corpus ont été sélectionnés à partir de la base de données *Europresse* en croisant une série de paramètres. La sélection a été opérée à partir de l'entrée *dictionnaires* dans la presse nationale française des deux dernières années, ce qui nous a permis de parvenir à 123 résultats. Parmi ceux-ci, nous avons choisi ceux qui présentent au grand public les mots intégrés dans les nouvelles éditions des dictionnaires. De façon plus spécifique, nous étudierons dans le détail les cinq articles proposés par le journaliste William Audureau dans *Le Monde* et dans lesquels il aborde, tour à tour, l'un des aspects des nouveautés lexicales des nouvelles éditions des dictionnaires *Robert* et *Larousse* en valorisant les points forts de la relation entre dictionnaires et phénomènes sociaux. Des renvois à d'autres articles portant sur les mêmes sujets ne seront pas négligés.

Dans le premier de ces articles, Audureau s'appuie à la fois sur le travail des décodeurs du *Monde*⁴ et sur les considérations de Carine Girac-Marinier, directrice des départements dictionnaires, encyclopédies et périscolaire de *Larousse* et de Géraldine Moinard, directrice de rédaction des éditions *Le Robert*. Audureau retrace de façon limpide le lien entre tendances lexicales apparues à partir de 2020 d'une part et phénomènes socioculturels de l'autre. Il parvient à identifier trois champs sémantiques dans lesquels s'inscrivent les mots acceptés dans les millésimes du *Robert* et du *Larousse* dans la période 2021-2023 (Audureau, 2022a). Plus précisément, trois domaines sont concernés : la médecine, les questions sociétales et l'environnement. Les versions en ligne des dictionnaires traditionnels semblent donc parfaitement à même de refléter les tendances qui traversent le monde contemporain. La relation entre la diffusion des termes médicaux liés à la pandémie et la société actuelle n'a pas besoin d'explication. Cependant les spécialistes remarquent trois phases lexicales, en parfait accord avec la dynamique pandémique : d'abord une perte de la nature spécialisée du lexique médical (*asymptomatique*) justifiée par la volonté des sujets de s'approprier ce lexique pour mieux dominer la peur, ensuite un pic de créativité (*déconfiner*, *vaccinodrome*) et, enfin, une banalisation lexicale où les sujets commencent en quelque sorte à jouer avec les mots à travers les procédés de dérivation (*covidé*, *vaccinateur*).

Au niveau sociétal, le journaliste remarque l'émergence de mots ayant suscité des débats : *woke* et *wokisme* atteignent un haut degré de diffusion, entourés de *glottophobie*, *dégenerer*, *iel*, *invisibilisation*.

Cependant, le choix des mots ne suffit pas pour pouvoir affirmer que les dictionnaires reflètent les opinions qui s'affrontent au niveau social. Il faut donc aller au-delà du choix des mots et vérifier la façon dont les dictionnaires en rendent compte. La difficulté de définir le substantif *wokisme* est révélatrice à cet égard. Dans l'article, en effet, le journaliste souligne l'écart existant entre les définitions du *Larousse* et du *Robert* :

4. Il s'agit d'une rubrique du site *Le Monde* dont le but consiste à vérifier des informations à propos de sujets variés concernant l'actualité (<https://www.lemonde.fr/les-decodeurs/> [page consultée le 23 février 2024]).

Wokisme *nom masculin*. [Anglicisme, souvent péjoratif] Courant de pensée d'origine américaine qui dénonce les injustices et discriminations ; mouvement, pensée woke. (Robert 2023)

[...]

Wokisme *nom masculin*. Idéologie d'inspiration woke, centrée sur les questions d'égalité, de justice et de défense des minorités, parfois perçue comme attentatoire à l'universalisme républicain. (Larousse 2023) (Audureau, 2022a)

La définition du *Larousse* est considérée par Carine Girac-Marinier comme étant plus neutre dans la mesure où elle évoque le sens propre qui est positif et l'utilisation négative de la part des médias. En revanche les lexicographes du *Robert*, non seulement introduisent une marque de registre diatopique et péjorative, mais renvoient dans la définition à sa nature polémique (Vincent, 2022 : 133-134).

La catégorie de l'environnement, d'après Audureau (2022a), serait particulièrement marquée par les néologismes dont la plupart ont été acceptés par le *Robert*, exception faite pour ceux dont l'usage est encore trop rare (c'est le cas de *climathypocrite*, création attribuée à Marine Le Pen, mais pas encore assez diffusée pour qu'elle puisse enrichir la nomenclature du *Robert*).

Dans un deuxième article, Audureau s'attache à la diversification du paysage linguistique. Autrement dit, il signale que la présence d'autres langues dans le lexique français est bien enregistrée par les dictionnaires (Audureau, 2022c). L'anglais domine surtout dans le domaine des nouvelles technologies ou des phénomènes sociaux. Cependant il n'est pas considéré comme un risque réel grâce aussi aux tentatives de créer des équivalents français (*verdissement* et *écoblanchiment* pour remplacer *green washing*⁵). Loin de s'en tenir aux considérations coutumières concernant l'influence de l'anglais, Audureau rend compte d'une diversification du paysage linguistique français, ce qui témoigne à notre sens de la volonté de restructurer – en les élargissant – les représentations au sujet des langues qui interagissent avec le français. Parmi les langues qui enrichissent le lexique français, il signale l'arabe et le japonais dans le domaine culinaire, suivis par le suédois. L'engouement pour la cuisine végétarienne entraîne aussi l'introduction de langues plus exotiques (sanskrit avec *dahl*, grec avec *halloumi*). Celles-ci n'offusquent pas les créations qui viennent d'autres pays francophones : le Québec en premier, suivi de la Belgique (*communautarisation*), de la Suisse (*mobard*, *natel*), de l'Afrique (*kiosque à essence*) et du Liban (*chalet*).

Dans le troisième article de sa rubrique, Audureau s'appuie sur les considérations de décodeurs du *Monde* au sujet des grands absents. Citons notamment celle que l'on connaît avec l'appellatif de « langue des jeunes » :

5. Les tentatives de produire des listes de mots en français pour éviter l'emploi d'anglicismes sont nombreuses. Citons à ce sujet la rubrique « Dire, ne pas dire » hébergée sur le site de l'Académie française (<https://www.academie-francaise.fr/dire-ne-pas-dire> [page consultée le 20 mars 2024]) ou encore la page du gouvernement français « Quelques bonnes pratiques rédactionnelles » où une section est consacrée aux anglicismes (<https://www.gouvernement.fr/marque-de-letat/quelques-bonnes-pratiques-redactionnelles>, [page consultée le 20 mars 2024]).

« La hchouma ! Les tèces se font hagra, après ils espèrent bicrave leurs dicos ? Ça deuh. » Si vous ne comprenez pas cette phrase, les dictionnaires généralistes ne vous seront d'aucune aide. Seule une poignée de termes issus des cités y ont fait leur entrée, comme *go* (« petite amie ») et *babtou* (ou *toubab* ; « blanc », en wolof), alors qu'ils circulent depuis trente ans dans les banlieues. (Audureau, 2022b)

Cette absence ne fait que souligner davantage le décalage qui existe entre la langue dans son usage quotidien et la façon dont elle est (sous-)représentée dans les dictionnaires. Considérés pendant longtemps comme la référence absolue de la norme, les dictionnaires ne constituent en effet qu'une représentation partielle. L'absence d'une partie du lexique, dont le rôle social n'est pas à remettre en question, n'est pas à notre sens un choix neutre. Doit-on supposer que l'exclusion de cette tranche de lexique dans des instruments officiels obéit à la volonté de cacher une portion, à vrai dire non négligeable, de la société française ? Ou, plus simplement, ces outils lexicographiques obéissent-ils encore, malgré tout, à une norme axée sur un français standard neutre et dont tout écart est proscrit ? Le journaliste parvient à croiser, voire à opposer, les deux points de vue. D'une part les responsables des deux maisons d'édition qui rappellent les critères lexicographiques de la stabilité des mots et de leur clarté ; critères qui ne se combinent pas avec la nature cryptique de la langue des jeunes :

En outre, pour les éditeurs de dictionnaires, une double difficulté méthodologique se pose, celle de s'entourer de lexicographes connaisseurs du sujet et de faire le tri entre un simple effet de mode et un terme appelé à rester, une gageure dans une langue aussi bouillonnante. « Tant que les termes restent un peu cryptiques, ils n'ont pas vocation à figurer dans le dictionnaire », résume Géraldine Moinard, directrice de la rédaction des éditions Le Robert. (Audureau, 2022b)

De l'autre, le point de vue du linguiste Jean-Pierre Goudaillier, qui n'adhère pas à l'idée de neutralité des dictionnaires. D'après lui, l'effacement de cette catégorie de mots relève de l'idéologie, à savoir d'une idéologie invisibilisant une tranche importante de la société :

« La reconnaissance par un dictionnaire relève de l'idéologie », juge-t-il. « Un dictionnaire sacralise. À partir du moment où un mot y entre, même s'il vient de l'étranger, il appartient au patrimoine linguistique français, ce qui explique leur frilosité. Le dictionnaire est aussi un lieu qui interdit et qui crée des tabous linguistiques » (Audureau, 2022b).

Soulignons que la créativité de la « langue de la rue » ne concerne pas les champs sémantiques décrits plus haut, mais porte plutôt sur la langue de tous les jours, transformée selon des procédés qui, sans être novateurs (verlan, apocopes, aphérèses), aboutissent à des résultats originaux :

Celles-ci [les inventions lexicales] portent principalement sur le vocabulaire de la vie quotidienne, des rapports de force, de la séduction et du succès matériel, à l'image des nombreux termes désignant une femme (*gadji, go, tchoin, zouz*) ou l'argent (*biff, lovs, roro, kichta, moulaga...*), popularisés par les principaux vecteurs de l'argot des cités, le rap et les réseaux sociaux. (Audureau, 2022b)

Une autre absence remarquable, mise en relief par le journaliste, concerne le français des pays africains. Un décalage important s'installe donc entre la réalité linguistique et la représentation que les dictionnaires contribuent à élaborer. En d'autres termes, si les spécialistes signalent que l'Afrique

inclut plus de la moitié des locuteurs de langue française, il n'en est pas moins vrai que les mots faisant partie du français d'Afrique sont sous-représentés non seulement par rapport à d'autres parties de la francophonie (Belgique et Canada, notamment) mais aussi par rapport à d'autres pays encore plus éloignés. Cette absence revient aussi dans des articles d'autres journalistes. Dans *Le Monde des idées*, Tégua Bogni rappelle que :

À titre comparatif, le Japon, pays asiatique qui n'est par ailleurs pas francophone, a vu ses mots comme « yuzu », « miso », « gomasio », « teriyaki », « wasabi », etc. entrer dans les dictionnaires français ces dernières années. Le moins que l'on puisse dire, c'est que ces mots, qui se rapportent à la cuisine, participent fortement au rayonnement du Japon. Alors, comment comprendre que malgré la gastrodiversité des pays africains, les gastronymes comme « ndolè », « maafé », « saka saka », « tièèbou dien », « maboké » et bien d'autres encore n'aient pas bonne presse auprès des lexicographes ? (Bogni, 2021)

Dans les deux derniers articles de la rubrique, Audureau aborde la question du lexique régional et du lexique concernant le domaine des jeux vidéo, ces deux domaines ayant désormais acquis droit d'accès aux dictionnaires.

Les considérations présentées par Audureau dans les rubriques analysées permettent deux réflexions principales. La première concerne l'ouverture des dictionnaires à la néologie. Si les dictionnaires jouissent d'une autorité qui leur vient de l'histoire, il n'en reste pas moins que pendant longtemps ils ont été considérés comme étant imperméables à tout ce qui tient de la néologie. Or, ces rubriques contribuent à modifier les représentations communes et cherchent à installer une nouvelle représentation des dictionnaires en tant qu'outils miroirs des dynamiques socioculturelles. La deuxième, en revanche, concerne le concept de représentativité : si, d'après Rey-Debove, le dictionnaire « ne peut sans doute pas tout inclure et tout dire, mais on peut exiger de lui qu'il donne une image représentative de la "totalité" du lexique » (Rey-Debove dans Béjoint, 2007 : 15), Audureau montre que la souplesse des dictionnaires en ligne à l'égard des nouvelles créations lexicales n'aboutit pas forcément à une représentativité accrue, ce qui nous permet d'adhérer au positionnement de Goudaillier, d'après qui l'absence d'une partie du lexique résulte de la volonté de passer sous silence un groupe social spécifique.

La relation entre nouveaux mots admis dans les dictionnaires et événements qui caractérisent l'époque contemporaine est aussi mise en valeur dans d'autres articles qui se situent dans le même sillage que les écrits d'Audureau, ce qui témoigne d'une part de la passion des Français pour leur langue et de l'autre de la circulation des discours. Le sujet est constamment relayé d'une voix à l'autre et cette circulation ne se déploie pas simplement selon une direction horizontale (d'un quotidien à l'autre), mais aussi selon une direction verticale, à savoir dans le temps. La question de l'intégration des mots nouveaux dans les millésimes des dictionnaires revient, en effet, dans d'autres articles publiés simultanément, mais aussi d'une année à l'autre. Dans la même période, signalons un extrait d'un article paru dans *Sud Ouest* où le journaliste souligne la capacité du Robert de reproduire le débat actuel concernant le pronom inclusif *iel* ainsi que l'ouverture du Robert à l'égard d'autres mots :

Dans la même proximité, le verbe « dégener » a été inclus dans le Robert 2023, qui comme toujours, accueille des mots liés à l'époque : « covidé », « covid long », « écouillonner », ou encore « écoanxiété » qui se passe de définition. Sans oublier « instagrameur » : « qui exerce son activité en publiant des photos et des vidéos sur Instagram ». « Influenceur » avait été ajouté dès 2018 par Larousse. (Monnier, 2022)

Plus centré sur les nouveautés introduites par le Larousse, *Le Figaro* reproduit à quelques différences près les mêmes considérations que celles relevées dans *Le Monde*. Les journalistes Mohammed Aïssaoui et Alice Devely, en effet, reviennent sur les champs sémantiques liés à la pandémie, sur les nouvelles tendances culinaires et gastronomiques en citant le lexique intégré par le Larousse mais ajoutent quelques réflexions sur la féminisation des noms de métier et jettent un regard critique et polémique au sujet de l'intégration du pronom *iel* dans le Robert :

Le « iel » est tombé sur la tête du Robert

En octobre 2021, Le Petit Robert intégrait le mot « iel » mais uniquement dans sa version numérique. Un curieux pronom, inconnu du grand public, qui répondait à cette définition : « *Pronom personnel sujet de la troisième personne du singulier (iel) et du pluriel (iels), employé pour évoquer une personne quel que soit son genre.* » [...]. Cette intégration n'a rien d'innocent en ce qu'elle relève d'une démarche militante. D'abord, parce que la vocation d'un dictionnaire, comme le note Le Petit Robert sur son site, est de se faire « *le reflet de la langue d'hier et d'aujourd'hui* », donc de l'usage. Or voilà une première incohérence. À peine intégré dans ses colonnes, le mot est dit « *rare* » ! Observons qu'un mot est dit rare lorsqu'il est ancien et non lorsqu'il est apparu depuis peu. Mais voilà un aveu intéressant : ce terme n'a jamais dépassé le stade de tracts, des associations ou des études idéologiquement orientées. (Aïssaoui et Devely, 2022)

2.2. Les dictionnaires collaboratifs

Après ce survol rapide des représentations que la presse fournit au sujet des relations entre dictionnaires et monde contemporain, nous allons vérifier la façon dont les dictionnaires collaboratifs, de par leur nature moins rigides et moins liés à une norme restrictive, reproduisent et renforcent ce lien. Il suffit de considérer les propos des intervenants au Forum *ABC de langue française* : si ceux-ci admettent que les dictionnaires collaboratifs ne peuvent pas être considérés comme la référence lexicographique, ils reconnaissent toutefois leur souplesse à l'égard des nouvelles créations lexicales (Molinari, 2017).

Étant donné ces considérations, nous avons choisi d'étudier deux dictionnaires en ligne : le *Wiktionnaire* et le *Dictionnaire de la Zone*. Le premier a été sélectionné pour sa richesse et sa diffusion ; le deuxième pour sa spécificité dans le domaine concerné. Des comparaisons avec *Le Petit Robert* seront aussi proposées pour mieux faire ressortir les écarts entre les ressources lexicographiques. Afin de réfléchir à la façon dont les ouvrages lexicographiques reflètent et/ou cachent les phénomènes culturels des différentes époques, notre choix a porté sur les néologismes.

La page du *Wiktionnaire* consacrée aux néologismes est articulée en six sous-sections⁶ : mots fantaisistes en français, néologismes dans la littérature en français, néologismes du médiéval fantastique en français, néologismes recommandés en français, néologismes de science-fiction en français, néologismes en sciences humaines en français. L'analyse de ces sections permet de remarquer que la section des néologismes recommandés en français réunit « les néologismes et équivalents en français des termes liés au développement de la technologie, proposés par les instances gouvernementales⁷ ». L'on précise ensuite que ces instances correspondent à la Délégation générale à la langue française et aux langues de France pour la France et à l'Office québécois de la langue française pour le Québec. Il s'agit d'une section riche qui inclut, à l'heure actuelle, 5659 articles. En général, ce sont des articles très techniques : c'est d'ailleurs dans le domaine des spécialités que les néologismes sont acceptés le plus aisément (*s'écraser*, *sélectionné*, *tablette*, pour n'en citer que quelques-uns).

Donc, au lieu de procéder en analysant ces néologismes, comme nous l'avions prévu au début, nous avons choisi de revenir sur quelques-unes parmi les catégories sémantiques et lexicales mises en relief par la presse. Pour ce qui concerne les mots qui reflètent les phénomènes sociaux, nous allons passer à côté de *woke*, déjà étudié par Vincent (2022), bien qu'il fasse l'objet d'une analyse de la part des responsables des maisons d'édition. Nous porterons davantage le regard sur *invisibilisation*, substantif qui renvoie aux phénomènes qui touchent plusieurs couches sociales (les femmes, les immigrés pour n'en citer que quelques-unes). D'après les considérations d'Audureau, la présence du substantif dans le PR est récente. Voici l'article consacré au verbe *invisibiliser* (*invisibilisation* constituant une entrée cachée) :

invisibiliser [ɛ̃vizibilize] verbe transitif (conjugaison 1)

ÉTYM. 1788 ◊ de *invisible*



■ Rendre invisible.

◦ PRONOM. « Avec la Carte Bleue l'argent s'invisibilisait » (Ernaux).

◆ FIG. Soustraire au regard social. → effacer, ignorer, masquer.

Invisibiliser les femmes, les précaires.

◦ Minorité invisibilisée.

◦ N. f. INVISIBILISATION.



■ CONTRAIRES : Exposer, montrer, souligner.

Figure 2. Article *invisibiliser* dans *Le Petit Robert 2024* en ligne

6. https://fr.wiktionary.org/wiki/Catégorie:Néologismes_en_français [page consultée le 20 mars 2024].

7. Ibid.

Le substantif fait aussi partie de la nomenclature du *Wiktionnaire* où l'historique des différentes versions montre qu'il est présent depuis 2010. Bien évidemment, les changements ont été importants. Ceux-ci n'ont pas porté sur la définition mais plutôt sur les exemples qui contribuent à contextualiser les emplois et permettent aux usagers de mieux saisir les enjeux socioculturels liés au substantif en question. Voici l'article en 2010 :

Ligne 1:

```

- == {{=fr=}} ==

- {{-étym-}}
- : {{cf|invisible|-ation}}

- {{-nom-|fr}}
  {{fr-rég|ĕ.vi.zi.bi.li.za.sjɔ̃}}
- '''invisibilisation''' {{f}}
  # Fait de rendre [[invisible]].
  #* ''Le destinataire doit accepter cette '''invisibilisation'''. Ne pas l'accepter, vouloir humilier, c'est rentrer dans un processus de violence.'' {{source|Jean Foucart, ''Sociologie de la souffrance'', 2004}}

```

Figure 3. Historique de l'article *invisibilisation* du *Wiktionnaire* en 2010

C'est en 2017 que l'article s'enrichit d'un exemple révélateur, dans la mesure où il permet de comprendre l'identité des sujets qui font l'objet d'*invisibilisation* :

Ligne 1 :

```
== {{langue|fr}} ==
=== {{S|étymologie}} ===
- : {{cf|invisible|-isation}}
=== {{S|nom|fr}} ===
```

Ligne 7 :

```
'''invisibilisation''' {{pron|ɛ̃.vi.zi.bi.li.za.sjɔ̃|fr}} {{f}}
# Fait de rendre [[invisible]].
- #* ''Le destinataire doit accepter cette '''invisibilisation''' . Ne pas l'accepter, vouloir humilier, c'est rentrer dans un processus de violence.'' {{source|Jean Foucart, ''Sociologie de la souffrance'', 2004}}
- #* ''De ce point de vue, l'''invisibilisation''' des femmes immigrées sur le marché du travail, dont il conviendra d'analyser les présupposés, participe à la fois de la manière dont s'est construite socialement la figure de la femme immigrée et, phénomène concomitant, des mécanismes de la construction sociale des situations subalternes.'' {{source|Sabah Chaib, ''Femmes immigrées et emploi : le bas de l'échelle pour propriété ?'', dans ''Le bas de l'échelle : La construction sociale des situations subalternes'', sous la direction de Pierre Cours-Salies & Stéphane Le Lay, Toulouse : Érès, 2006}}
```

Figure 4. Historique de l'article *invisibilisation* du *Wiktionnaire* en 2017

En revanche, avec le verbe *dégenerer*, verbe faisant partie des derniers admis par *Le Petit Robert*, la perspective est renversée. En effet, dans le *Wiktionnaire*, le verbe – qui porte la marque de néologisme – est défini de façon très neutre :

⇒ Verbe [modifier le wikicode]

dégenerer \de.ʒɑ̃.ʁe\ transitif direct 1^{er} groupe (VOIR LA CONJUGAISON)

1. (Néologisme) Ôter le genre ou toute notion de genre de (quelque chose).

- Dans le monde humain, le comportement de soin maternel ne peut être conçu hors de l'entrelacs des sphères vitale, morale et sociale dont nous réfléchissons la complexité spécifique en ramenant le comportement de soin à son fondement, à une ontologie de la relation dont on expliquera pourquoi elle contribue à « **dégenerer** » le soin maternel et à l'affranchir de la spécialisation sexuée des tâches où la société occidentale l'enferme depuis les Lumières. — (Frédérique Bisiaux, *Le soin maternel à la lumière de l'ontologie de la relation*, colloque de Cerisy : « Qu'est-ce que le vivant ? », aout 2012)
- Le système à trois genres comprenant le neutre, le féminin, le masculin a été réduit à deux dans la plupart des langues romanes, sauf en allemand ou dans les langues slaves. Cette évolution montre une tendance à « **dégenerer** » les langues en simplifiant leur système de genre. — (Philippe Clauzard, *Conversations sur le sexisme : éduquer pour l'égalité filles-garçons*)

Figure 5. Article *dégenerer* dans le *Wiktionnaire*

À l'inverse, *Le Petit Robert* introduit, dans la définition de *dégenrer*, une dimension polémique en évoquant des concepts universaux et au premier plan dans les débats contemporains, tels que la mixité et l'égalité. Les lexicographes laissent entendre que l'élimination de tout renvoi au « genre » est un pas à franchir pour atteindre des valeurs universelles et normalement considérées comme des idéaux positifs.

dégenrer [dezãre] verbe transitif (conjugaison 1)

ÉTYM. 1998 gramm. ◊ de *dé-* et *genre*

Famille étymologique ⇨ GENS.



■ DIDACT. Supprimer toute distinction en fonction du genre pour favoriser la mixité et l'égalité. *Dégenrer les métiers, l'urbanisme.*

= P. p. adj. *Éducation dégenrée.*

Figure 6. Article *dégenrer* dans *Le Petit Robert* 2024 en ligne

Ce qui peut frapper le lecteur averti est l'absence de renvoi, dans les deux cas, à la polémique sur le pronom *iel*, pourtant admis dans les deux dictionnaires.

Signalons aussi que parmi les nouveaux mots admis dans l'édition 2024 du *Petit Robert*, l'on retrouve le verbe *mégenrer*⁸. Le verbe est présent dans le *Wiktionnaire* depuis 2017 et, dans ce cas spécifique, la dimension polémique est au premier plan dans les exemples où l'action de mégenrer est comparée à une micro-agression :

mégenrer \me.ʒã.ʁe\ transitif direct 1^{er} groupe (VOIR LA CONJUGAISON)

1. (Transitide) (Emploi transitif et parfois absolu) Attribuer à une personne, volontairement ou non, un **genre** dans lequel la personne ne se reconnaît pas^[1].

- *La plupart des cissexuel·les sont inconscient·es de la nature subjective du genrement, essentiellement car elleux·mêmes ne vivent pas l'expérience d'être régulièrement mégenrées – i.e. assigné·es par erreur à un genre qui ne leur correspond pas.* – (Julia Serano, « Démanteler le privilège cissexuel », dans *Manifeste d'une femme trans*, traduit de l'anglais par Noémie Grunenwald, Tahin Party, 2007, page 54.)
- *Sachez que même si vous nous mégenrez involontairement, l'impact est présent. L'équation est simple : un mégenrage = une micro-agression.* – (UESG, « Mégenrer c'est violent » sur le blogue *Unique en son genre : Le blog trans-lucide*, 1^{er} janvier 2016.)
- *Je passe mon temps à me demander si je dois reprendre les gens ou laisser courir quand ils me mégenrent et en même temps je sais à quoi je ressemble. Ça me saoule, ça me rappelle la binarité malsaine de cette société, [...].* – (Mark, « J'ai 30 ans, je mue et je flotte : je suis adolescent », sur le site *Female2male*, 2 septembre 2015.)
- *Les personnes qui te genrent correctement après une journée entière à être mégenréx c'est vraiment les bests de la vie (et ça sauve accessoirement, genrez correctement les gens)* – (Aztruc. twitter. 2021 → lire en ligne)
- *Mégenrer peut se traduire par l'utilisation de pronoms, d'un vocabulaire, d'accords ou de salutations qui ne correspondent pas à l'identité de genre de la personne dont il est question.* – (Bureau de la traduction, *Lexique sur la diversité sexuelle et de genre* sur *Gouvernement du Canada*, février 2019)

Figure 7. Article *mégenrer* dans le *Wiktionnaire*

8. <https://www.radiofrance.fr/franceinter/complosphere-nasser-megenrer-les-nouveaux-mots-du-petit-robert-5102807> [page consultée le 20 mars 2024].

La dimension polémique est évidente aussi dans l'espace discussion où les intervenants réagissent à l'égard de plusieurs aspects : la nature des sources plus ou moins fiables (s'agissant de blogs considérés comme moins crédibles que d'autres genres textuels) ; l'idée de « mauvais » genre et, enfin, la création de néologismes à partir de théories considérées comme fumeuses.

Les anglicismes semblent ne pas poser problème dans les dictionnaires collaboratifs. *Instagram* et tous ses dérivés⁹ (*instagramer*, *instagrameur*, *instagramable*, *instagramiser*, *instagramabilité*) font partie de la nomenclature du *Wiktionnaire*, qui s'avère donc particulièrement accueillant à l'égard des emprunts à l'anglais et du champ lexical créé grâce aux procédés de dérivation morphologique. Nous remarquerons que le seul mot accompagné d'une remarque est *instagramabilité* : même s'il ne porte pas la marque de néologisme, le substantif est considéré comme n'appartenant pas au français standard.

instagramabilité \ɛ̃s.ta.ɡʁa.ma.bi.li.te\ féminin — ⚠ Il s'agit d'un terme utilisé qui n'est pas d'un usage standard. — **Note** : L'usage du pluriel n'est pas avéré.

Figure 8. Article *instagramabilité* dans le *Wiktionnaire*

Doit-on en déduire que les autres dérivés font désormais partie du standard ? Ou, au contraire, faut-il questionner la notion de « standard », dont les frontières semblent de moins en moins étanches ? Signalons aussi que seulement les substantifs *instagramateur/euse* (et leurs variantes graphiques *instagrammeur/euse*) font partie de la nomenclature du *Petit Robert* et ne sont accompagnés d'aucune marque de registre.

Les dictionnaires collaboratifs sont aussi plus accueillants à l'égard de la langue des jeunes. Le lexique exclu des maisons d'édition Robert et Larousse y est intégré avec des articles complets. Considérons les mots employés dans la langue des jeunes pour indiquer la femme : à côté du plus connu *meuf* et de *go*, la presse indique d'autres synonymes, tels que *gadjji*, *tchoin* et *zouz*. C'est dans le *Wiktionnaire* et dans le *Dictionnaire de la Zone* que l'on pourra non seulement retrouver ce lexique mais aussi en saisir les nuances sémantiques. Les deux signalent que *gadjji* est issu du romani et s'emploie pour indiquer une fille qui n'est pas tzigane, mais sans aucun jugement de valeur.

9. Sont aussi présentes certaines variantes avec double *m* : *instagrammer*, *instagrammeur* et *instagrammable*.

Étymologie [modifier le wikicode]

(Date à préciser) Mot emprunté au romani *gadji* (« non-tsigane »).

Nom commun [modifier le wikicode]

gadji \gad.ʒi\ féminin (pour un homme, on dit : gadjo)

1. (Argot) Fille ou femme qui n'est pas tsigane.

- *La première femme de Papa était une gadji – une Slovaque. Elle provenait d'une famille extrêmement pauvre qui comptait quatre sœurs et un frère.* — (Elena Lacková, *Je suis née sous une bonne étoile : Ma vie de femme tsigane en Slovaquie*, récit recueilli par Milena Hübschmannová, traduit du tchèque par Frédéric Bègue, Éditions Harmattan, 2000, p. 34)
- *Je suis comme toi, une gadji, devant mon ordinateur.* — (Thierry CROUZET, *La quatrième théorie*, 2013)
- — *La gadji lui menait la vie dure. Et les petites gadjis aussi. C'est ce qui arrive quand tu te laisses mener par le bout du nez, je le lui ai dit cent fois.* — (Carmen Mola, *La Fiancée gitane*, Éditions Actes Sud Littérature, 2019, chap. 41)

Singulier	Pluriel
gadji	gadjis
\gad.ʒi\	

Synonymes [modifier le wikicode]

- meuf

Figure 9. Article *gadji* dans le Wiktionnaire

Tchoin, en revanche, indique une fille facile et sans moralité. L'article du *Dictionnaire de la Zone*, grâce aux exemples, met en évidence les écarts sémantiques par rapport aux autres substantifs appartenant au même champ sémantique.

tchoin *nom féminin.*
Désigne une fille facile et sans moralité.

J'arrive en coulisses, taffe sur 3-4 joints | En un coup d'fil j'velé¹ 3-4 tchoins | Et ce dans 3-4 coins... Bigo (L'ENTOURAGE, « L'Introduction », 2012) 🎵

0:00 | 0:05 🔊 📺 >>

Alors les p'tits frères vont faire comme nous, tirer sur des joints comme nous | Tiser et baiser des tchoins comme nous, c'est juste une question d'exemple (NIRO, « Déçu », *Paraplégique (Réédition)*, 2012) 🎵

0:00 | 0:07 🔊 📺 >>

La go là c'est p'tête une fille bien | Mais on préfère les tchoins, tchoins, tchoins (KAARIS, « Tchoin », *Okou Gnakouri*, 2016) 🎵

0:00 | 0:08 🔊 📺 >>

VAR. tshoin.

ÉTYM. De l'argot ivoirien *tchoin* « putain ».

Figure 10. Article *tchoin* dans le Dictionnaire de la Zone

Zouz n'est présent qu'à la nomenclature du Wiktionnaire et désigne une « jolie fille », et par extension, une « amie de cœur ».

Pour ce qui concerne les emprunts à d'autres langues, c'est le même phénomène qui se produit. Certains mots, employés couramment et dont on déplore l'absence dans le Robert ou dans le Larousse, sont présents dans le *Wiktionnaire*.

L'article consacré à *saka-saka* est complet grâce à la présence de la marque de registre, des synonymes et d'une photo qui permet de mieux visualiser un plat peu connu.

Étymologie [modifier le wikicode]

Emprunt aux langues congolaises^[1].

Nom commun [modifier le wikicode]

saka-saka \sa.ka.sa.ka\ masculin

1. (*Cuisine*) (*Afrique*) Plat à base de feuilles de **manioc**.

- J'ai choisi le premier plat venu, du ndolé avec de la viande de bœuf. J'avais pas mangé ça avant, mais ça ressemblait au **saka-saka**, le plat de feuilles de manioc de chez nous. — (Alain Mabanckou, *Black Bazar*, Seuil, 2009, page 96)

Singulier	Pluriel
saka-saka	saka-sakas
\sa.ka.sa.ka\	

Variantes orthographiques [modifier le wikicode]

- saka saka
- sakasaka

Synonymes [modifier le wikicode]

- pondu
- madamba

Traductions [modifier le wikicode]

[Enrouler ▲]



Du saka-saka.

Figure 11. Article *saka-saka* dans le *Wiktionnaire*

Les dictionnaires collaboratifs seraient donc, encore une fois, plus perméables à l'égard d'un lexique qui circule mais qui ne peut pas être considéré comme standard. Les contours de la nomenclature des dictionnaires permettent donc de mieux saisir les dynamiques de la culture par rapport aux dictionnaires traditionnels, qu'ils soient en ligne ou pas.

2.3. La plateforme Le Robert. Dico en ligne

La plateforme en ligne que nous allons examiner est la page du *Robert. Dico en ligne*¹⁰. Celle-ci se présente dès le début comme « le site de référence de la langue française », ce qui justifie notre choix : de consultation immédiate et gratuite, les usagers lui accordent leur préférence par rapport au site payant du dictionnaire *Le Petit Robert*.

La page d'accueil du site présente une série de rubriques qui permettent, à notre sens, d'une part d'observer la façon dont les représentations linguistiques sont élaborées et de l'autre d'explorer la façon dont la langue reflète le monde contemporain.

10. <https://dictionnaire.lerobert.com> [page consultée le 20 mars 2024].

La première de ces rubriques concerne les *Tendances*. Il s'agit d'un encadré articulé en deux pages où sont signalés les dix mots les plus recherchés dans les 7 derniers jours et les dix conjugaisons les plus recherchées. Donc : lexique et grammaire. C'est le lexique que nous allons prendre en compte. Voici les mots indiqués dans la semaine du 22 avril 2023 :



Figure 12. Saisie d'écran des 10 mots les plus recherchés dans la semaine du 22 avril 2023 dans le *Dico en ligne*

Bien que sa consultation semble instable (la flèche rouge descendante le 22 avril est remplacée par une flèche montante le 11 mai), le substantif *wokisme*, déjà relevé plus haut, est toujours en première position¹¹. Une confrontation avec la base de données *Europresse* semble confirmer cette donnée : la recherche effectuée sur les 7 derniers jours dans la section « sources en français » signale 63 résultats (dont 28 le 20 avril) ; les attestations augmentent à 99 dans la première semaine de mai (41 documents le 10 mai).

11. Signalons que, au moment de la dernière relecture de cet article (semaine du 18 mars 2024), *wokisme* figure encore en deuxième position parmi les mots les plus cherchés et est accompagné d'une flèche verte signalant une hausse dans la consultation du mot.

Dans le *Dico en ligne*, il suffit de cliquer sur le mot pour obtenir la définition suivante :



wokisme

DÉFINITION

Définition de **wokisme** [nom masculin]

anglicisme (souvent péjoratif) Courant de pensée d'origine américaine qui dénonce les injustices et discriminations ; mouvement, pensée woke.

Figure 13. Article *wokisme* tiré du *Dico en ligne*¹²

La définition est assez synthétique ; aucun exemple, qui pourrait en illustrer l'usage, n'est intégré. Le dictionnaire signale que le mot revient fréquemment et donc amène les usagers à s'interroger à ce sujet mais ne les aide pas à mieux comprendre les enjeux qui se cachent derrière l'idéologie *woke*. On suggère cependant une dimension polémique à travers la marque d'usage.

En deuxième position, nous trouvons le verbe *promulguer*, verbe apparemment neutre et qui n'appartient pas à la catégorie des néologismes. La deuxième position et donc la fréquence d'emploi (confirmée, elle aussi par la base de données *Europresse* avec 162 documents dans la même période), est à mettre en relation avec les événements qui se produisaient en France dans la période considérée. C'est en effet dans cette semaine que la loi concernant la réforme des retraites, qui a soulevé autant de protestations, a été promulguée. Le lien lexique-culture/événements politiques et sociaux est donc confirmé. En revanche, l'on regrettera que les exemples cités ne permettent pas d'établir le lien avec les événements contemporains ce qui rend plus difficile pour un lecteur qui ne connaît pas la réalité française de comprendre les tendances lexicales, celles-ci pouvant aussi être liées aux recherches des étudiants, aux programmes scolaires, etc.

D'autres rubriques nourrissent la page du *Dico en ligne*. Citons, à ce titre, *Le mot du jour* où il est possible de retrouver tous les « mots du jour » ayant fait l'objet d'un approfondissement. L'on retrouvera donc *motion* publié au mois de mars 2023, *retraite* en janvier 2023 (les deux se rattachent au débat concernant la réforme de la loi sur les retraites), *iel*, *iels* en novembre 2021, *hommage* en octobre 2021, *même* en mai 2021 et les mots qui relèvent du champ sémantique des « vaccins » en 2020 pour n'en citer que quelques-uns.

L'on remarquera aussi une transition dans la façon de présenter ces mots : c'est à partir de juin 2022 que les mots sont décrits selon une structure qui se rapproche de l'architecture d'un article lexicographique : définition et, parfois, exemples. Avant, l'on avait souvent des billets de la part des spécialistes dont le but est d'expliquer le sens des mots en relation avec les réalités socioculturelles. C'est le cas de *glottophobie*, *iel/iels*, *hommage*, *même*. *Vaxxie*, par exemple, est présenté par une spécialiste qui se fait ici divulgatrice conformément à la nature « grand public » de la plateforme. Une portion im-

12. <https://dictionnaire.lerobert.com/definition/wokisme?source=tendances>

portante du lexique présenté relève de ces mots qui, tout en étant issus du phénomène de la covid 19, ne concernent pas seulement les dérivations morphologiques à partir du substantif *covid* mais toutes les conséquences qu'il a entraînées. Citons, à ce titre, *télétravail*, *déconfinement*, *crise*, *reprise*. Souvent le lien avec la période n'est pas évoqué de façon explicite mais la date des billets permet de remonter à l'époque de la covid, ce qui témoigne de la volonté de la part des spécialistes qui gèrent la base en ligne non seulement de montrer un lexique nouveau, mais aussi de mettre en valeur le lexique commun qui acquiert une nouvelle valeur à la suite des événements de l'époque. Il nous semble donc pouvoir affirmer que le choix des mots participe de la construction des représentations, dans la mesure où l'élargissement des champs lexicaux implique une prise en compte de l'ensemble des transformations sociales provoquées par la pandémie.

Signalons la rubrique au sujet des questions concernant l'environnement, au premier plan dans le monde contemporain. Il s'agit, plus précisément, d'une liste de dix mots qui portent sur le réchauffement climatique et dont la fréquence est aussi confirmée par la presse¹³ : *anthropocène*, *écocide*, *collapsologie*, *survivaliste*, *écoanxiété*, *climatosceptique*, *sobriété énergétique*, *fléxitarien*, *zéro déchet*, *greenwashing* (ou *écoblanchiment*). *Écoanxiété* a retenu notre attention car il est aussi signalé par Audureau parmi les mots ayant été admis dans *Le Petit Robert 2023*. Cependant, la définition du Robert et celle du *Wiktionnaire* sont extrêmement synthétiques :

écoanxiété [ekoãksjete] **nom féminin**

ÉTYM. 2007 au Canada ◊ de *éco-* et *anxiété*

■ Anxiété provoquée par les menaces environnementales qui pèsent sur notre planète.

FORMES
singulier : écoanxiété
pluriel : écoanxiétés

PRONONCIATION
❖ écoanxiété
[ekoãksjete]

Figure 14. Article *écoanxiété* tiré du *Petit Robert 2024* en ligne

13. « Top dix des nouveaux mots du réchauffement climatique », <https://dictionnaire.lerobert.com/dis-moi-robert/raconte-moi-robert/top-10-mots-plus-etonnants/top-10-des-nouveaux-mots-du-rechauffement-climatique.html> [page consultée le 20 mars 2024].

écoanxiété \e.ko.āk.sje.te\ féminin

1. Anxiété provoquée par l'anticipation de modifications écosystémiques provoquées par l'humanité.

- Marie-Ève Tremblay croit aussi que cette génération serait plus touchée par l'écoanxiété en raison de sa présence en ligne, beaucoup plus importante que celle des personnes de 55 ans et plus. — (L'écoanxiété, le mal d'une génération, radio-canada.ca, 20 juin 2019.)

- Or, pollution de l'air et réchauffement climatique sont imbriqués, puisqu'il y a des pics de pollution aux particules fines en période de forte chaleur. C'est là que l'on voit le plus de problèmes respiratoires, l'asthme notamment, en particulier chez les enfants. Depuis deux ou trois ans, les jeunes, eux, sont de plus en plus nombreux à présenter des signes d'écoanxiété. — (Thomas Baietto, Marie-Adélaïde Scigacz, Camille Adaoust, COP26 : viticulteur, monitrice de ski, médecin... Comment la crise climatique bouleverse leur métier, FRANCE TELEVISIONS. Mis en ligne le 12 novembre 2021)

Singulier	Pluriel
écoanxiété	écoanxiétés
\e.ko.āk.sje.te\	

Figure 15. Article écoanxiété tiré du Wiktionnaire

Les deux articles mettent en évidence deux aspects différents du phénomène : le Robert souligne les menaces environnementales, alors que dans le Wiktionnaire la responsabilité revient à l'homme, voire à l'humanité entière. Dans la rubrique *Top 10 des nouveaux mots du réchauffement climatique*, le Robert Dico en ligne, tout en renvoyant à l'article et à la définition du Petit Robert en ligne, se rapproche davantage de l'article d'un dictionnaire encyclopédique. Le renvoi à l'activiste Greta Thunberg, mondialement connue pour sa lutte en faveur de l'environnement, et donc au contexte social, ne peut que favoriser chez les usagers la compréhension et l'adoption du nouveau mot.

5) écoanxiété

Avez-vous déjà ressenti de la peur, de la tristesse et de la colère face aux menaces environnementales qui pèsent sur notre planète ? Si oui, alors vous souffrez peut-être d'écoanxiété. Cette angoisse teintée d'impuissance et de culpabilité, particulièrement ressentie par les jeunes générations, peut être paralysante mais peut aussi pousser à agir, en modifiant son mode de vie et en participant à des actions militantes. La figure emblématique de cette jeunesse engagée dans la lutte contre le changement climatique est l'activiste suédoise Greta Thunberg, qui a fait de son écoanxiété le moteur de son engagement écologique.

Figure 16. Texte sur écoanxiété tiré de la rubrique *Top 10 des nouveaux mots consacrés au réchauffement climatique* dans le Dico en ligne

3. En guise de conclusion

Le parcours que nous avons décrit au cours de cette réflexion ne peut que confirmer le rôle de la lexicographie collaborative et de la paralexigraphie en ligne en tant qu'outils à même (encore plus que les dictionnaires traditionnels en ligne) de refléter les tendances et les courants socioculturels qui sillonnent le monde contemporain. Bien évidemment, tout instrument propose une sélection à partir de l'idéologie qui l'inspire. Il n'en reste pas moins que les événements qui font débat au niveau

social trouvent un espace lexicographique. Par ailleurs, cet accueil diffère d'une ressource à l'autre. C'est grâce à leur intégration que l'on parvient à ébaucher une représentation plus complète des relations langue-société et à faire émerger des potentialités souvent inimaginées du français. Il serait aussi important de sensibiliser le public à l'égard des possibilités lexicographiques offertes par les nouvelles technologies : il est vrai que les réseaux sociaux permettent aux usagers (souvent amateurs lexicographes) d'entrer en contact plus facilement avec ces outils. Cependant, cette sensibilisation devrait être accompagnée d'un usage plus responsable de ces instruments et en mesure de mieux comprendre la relation qu'ils entretiennent avec le monde contemporain. Si l'on reprend les mots de Dubois et Dubois, d'après qui le dictionnaire serait un « miroir dans lequel le lecteur doit se reconnaître à la fois comme locuteur natif et comme participant à une culture » (Dubois et Dubois, 1971 : 99), l'on pourra conclure que c'est d'abord dans les dictionnaires collaboratifs que certains traits socioculturels apparaissent, alors que les dictionnaires traditionnels (même dans leur version en ligne) font toujours état d'une certaine difficulté à intégrer le lexique désignant les nouveaux phénomènes.

Quant aux dictionnaires collaboratifs, il serait important de parvenir dans le futur à mieux cerner le poids des apports et des contributions. Si la section « historique » du dictionnaire permet de comparer les changements et les transformations qui sont intervenues dans le temps, il n'en reste pas moins que souvent les espaces de discussions sont vides alors qu'il serait intéressant de suivre les débats et les polémiques que l'intégration de certains mots soulèvent.

Enfin, la presse aussi joue un rôle important dans la diffusion et/ou dans la mise sous silence d'une partie du lexique : le retentissement ou l'absence d'un champ lexical témoigne de la volonté de rendre visible ou d'occulter une partie de la société pour des questions idéologiques.

Bibliographie

ACADÉMIE FRANÇAISE, « Dire, ne pas dire », <https://www.academie-francaise.fr/dire-ne-pas-dire>.

ACHARD-BAYLE, Guy et Marie-Anne Paveau (2008), « Présentation. La linguistique "hors du temple" », *Linguistique populaire ?*, *Pratiques*, n° 139-140, p. 3-16.

AÏSSAOUI, Mohammed et Alice Devely (2022), « Le Petit Larousse et Le Petit Robert dévoilent leur millésime 2023 », *Le Figaro littéraire*, 19 mai.

AUDUREAU, William (2022a), « Comment la langue française a évolué depuis la pandémie de Covid-19 : "écouvillonner", wokisme", "écoanxiété" », *Le Monde*, 25 juillet, https://www.lemonde.fr/les-decodeurs/article/2022/07/25/ecouvillonner-wokisme-ecoanxiete-comment-la-langue-francaise-a-evolue-depuis-la-pandemie_6135998_4355770.html

AUDUREAU, William (2022b), « "Deh", "hchouma", "tchop" : l'argot des cités, grand oublié des dictionnaires », *Le Monde*, 25 juillet, https://www.lemonde.fr/les-decodeurs/article/2022/07/27/deh-hchouma-tchop-l-argot-des-cites-grand-oublie-des-dictionnaires_6136318_4355770.html

AUDUREAU, William (2022c), « "Flow", "tataki", "coolitude" : d'où viennent les nouveaux mots de la langue française », *Le Monde*, 26 juillet, https://www.lemonde.fr/les-decodeurs/article/2022/07/26/flow-tataki-coolitude-d-ou-viennent-les-nouveaux-mots-de-la-langue-francaise_6136172_4355770.html

AUDUREAU, William (2022d), « "Faluche", "schmutz", "s'entrucher" : les mots des régions, des coquetteries très attendues », *Le Monde*, 26 juillet, https://www.lemonde.fr/les-decodeurs/article/2022/07/28/faluche-schmutz-s-entrucher-les-mots-des-regions-des-coquetteries-tres-attendues_6136410_4355770.html

AUDUREAU, William (2022e), « "Boss", "game designer", "vidéoludique" : le vocabulaire du jeu vidéo enfin reconnu dans les dictionnaires », *Le Monde*, 26 juillet, https://www.lemonde.fr/les-decodeurs/article/2022/07/29/boss-game-designer-vidéoludique-le-vocabulaire-du-jeu-vidéo-enfin-reconnu-dans-les-dictionnaires_6136570_4355770.html

BÉJOINT, Henri (2007), « Informatique et lexicographie de corpus : les nouveaux dictionnaires », *Revue française de linguistique appliquée*, 1(12), p.7-23.

BOGNI, Téguia (2021), « Francophonie : "Il y a un manque de connaissances de l'Afrique et de ses langues qui a une influence certaine sur le français parlé" », *Le Monde des idées*, 21 avril.

BOYER, Henri (2003), *De l'autre côté du discours. Recherches sur les représentations communautaires*, Paris, L'Harmattan.

CALVET, Louis-Jean (1999), *Pour une écologie des langues du monde*, Paris, Plon.

- CORBEIL, Jean-Claude (2003), « La contribution de Josette Rey-Debove et d'Alain Rey à l'aménagement de la langue au Québec », dans Monique Cormier, Aline Francœur et Jean-Claude Boulanger (dir.), *Les dictionnaires Le Robert*, Montréal, Les Presses de l'Université de Montréal, p. 263-279.
- DUBOIS, Claude et Jean Dubois (1971), *Introduction à la lexicographie : le dictionnaire*, Paris, Larousse.
- GENETTE, Gérard (1987), *Seuils*, Paris, éd. du Seuil.
- GOUVERNEMENT DE LA RÉPUBLIQUE FRANÇAISE, « Quelques bonnes pratiques rédactionnelles » où une section est <https://www.gouvernement.fr/marque-de-letat/quelques-bonnes-pratiques-redactionnelles>
- JODELET, Denise (éd.) (1989), *Les représentations sociales*, Paris, PUF.
- Le Dictionnaire de la Zone, tout l'argot des banlieues*, <https://www.dictionnairedelazone.fr>
- Le Robert. Dico en ligne*, <https://dictionnaire.lerobert.com>
- Les décodeurs*, <https://www.lemonde.fr/les-decodeurs/>
- MOLINARI, Chiara (2017) « Nouvelle lexicographie vs anciennes représentations », dans Chiara Molinari et Nadine Vincent (dir.), *Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, Repères Do.Ri.F., 14, http://dorif.it/ezone/ezone_articles.php?dorif_ezone=57d6c60563ef-6c21aaff0ff5d8426615&art_id=381
- MONNIER, Rémi (2022), « La chronique du médiateur : entre deux mots, le lecteur balance », *Sud Ouest*, 21 octobre.
- PAVEAU, Marie-Anne (2007), « Les normes perceptives de la linguistique populaire ». *Langage et société*, n° 119. *Les normes pratiques*. p. 93-109.
- PAVEAU Marie-Anne (2017), *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*, Paris, Hermann.
- PAVEAU Marie-Anne et Laurence Rosier (2008), *La Langue française. Passions et polémiques*, Paris, Vuibert.
- PETITJEAN, Cécile (2010), « La notion de représentation linguistique : définition, méthode d'observation, analyse », dans Henri Boyer (dir.), *Pour une épistémologie de la sociolinguistique*, Limoges, Lambert-Lucas, p. 293-300.
- PRUVOST, Jean (2006), *Les dictionnaires français : outils d'une langue et d'une culture*, coll. L'Essentiel français, Paris, Ophryx.
- RADIOFRANCE, <https://www.radiofrance.fr/franceinter/complosphere-nasser-megenrer-les-nouveaux-mots-du-petit-robert-5102807>
- REY, Alain et Simone Delasalle (1979), « Problèmes et conflits lexicographiques », *Langue française*, 43, p. 4-26.
- REY, Alain (dir.) (2006), *Le Dictionnaire culturel en langue française*, Paris, Le Robert.

SABLAYROLLES, Jean-François (2015), « Néologie et dictionnaire(s) comme corpus d'exclusion », dans Jean-François Sablayrolles (dir.), *Néologie et terminologie dans les dictionnaires*, Paris, Champion, p. 19-36.

VINCENT, Nadine (2022), « Faut-il adapter les dictionnaires à l'air du temps ? Proposition d'un traitement polyphonique du mot *woke* », *Circula*, n°15, p. 122-145, https://circula.recherche.usherbrooke.ca/wp-content/uploads/2023/01/2022_Circula_15_007_Vincent.pdf

Wiktionnaire, le dictionnaire libre, https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d'accueil

TITRE: LES NÉOLOGISMES DU COVID-19 DANS LES DICTIONNAIRES FRANCOPHONES EN LIGNE

AUTEURS: KAJA DOLAR, CREE – INALCO, ET MARIE STEFFENS, UNIVERSITÉ DE LIÈGE ET UNIVERSITEIT UTRECHT

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 34-57

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22394](http://hdl.handle.net/11143/22394)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22394](https://doi.org/10.17118/11143/22394)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Les néologismes du Covid-19 dans les dictionnaires francophones en ligne

Kaja Dolar, CREE – Inalco

Marie Steffens, Université de Liège et Universiteit Utrecht

Résumé :

La pandémie de Covid-19 a donné lieu à une explosion néologique mondiale sans précédent. Cet article propose un état des lieux de la lexicographie francophone actuelle en ligne à cet égard. Pour ce faire, notre contribution analyse le traitement d'une sélection de termes nouvellement créés ou chargés d'un sens nouveau pour rendre compte de la situation sanitaire dans quatre dictionnaires aux objectifs différents : le *Dictionnaire de français Larousse*, le *Dico en ligne Le Robert*, le *Wiktionnaire* et le *Dictionnaire des francophones*. Après avoir fait l'inventaire des unités sélectionnées dans chacun des dictionnaires de notre corpus, nous nous focaliserons sur des exemples plus précis pour comparer ces dictionnaires selon sept aspects macro- et microstructuraux pour répondre aux questions suivantes : les dictionnaires francophones actuels ont-ils réussi à suivre cette vague néologique ? Quelle couverture et quel traitement des néologismes liés au Covid-19 offrent-ils ? Quelle est la qualité des articles dictionnaires décrivant ces néologismes ? Le bilan se dessine moins positif que nous ne pourrions le penser.

Mots-clés :

Néologie, pandémie, variation, neutralité, lexicographie collaborative

Abstract :

The Covid-19 pandemic led to an unprecedented global explosion of neologisms. The present article offers an analysis of current French online lexicography and focuses on the treatment of a selection of either terms that were newly created or acquired a new meaning in the wake of the pandemic. Our corpus is comprised of four dictionaries that differ in terms of their structure and objective: *Dictionnaire de français Larousse*, *Dico en ligne Le Robert*, *Wiktionnaire* and *Dictionnaire des francophones*. First, we created an inventory of entries related to Covid-19 in each of the dictionaries of our corpus, and then focused on a smaller sample that allowed us to compare these dictionaries according to seven macro- and microstructural aspects. The article is addressing the following questions : did the French dictionaries manage to integrate the wave of neologisms related to Covid-19? What kind of coverage and treatment of these neologisms do they offer? What is the quality of the dictionary articles describing these neologisms? The results are not as positive than one might expect.

Keywords :

Neology, pandemic, variation, neutrality, collaborative lexicography

1. Introduction

La pandémie de Covid-19 a entraîné un processus sans précédent de création lexicale simultanée dans toutes les langues, produisant différents types de néologismes : la situation était nouvelle pour tout le monde, avec des conséquences sociétales majeures (Kolářiková, 2021 : 5), expliquant ainsi la « nécessité à nommer les choses et en parler » (Kolářiková, 2021 : 3). Cette « urgence de nommer ce phénomène quasiment inédit » (Altmanova, Murano et Preite, 2022 : 1) a produit un phénomène néologique hors pair, avec un début si net et une prolifération si intense qu'il offre une possibilité exceptionnelle aux lexicographes, celle de faire l'état des lieux et d'évaluer la lexicographie francophone en ligne actuelle ; ce sera précisément l'objectif du présent article.

Plusieurs questions se posent alors : les dictionnaires francophones actuels ont-ils réussi à suivre cette vague néologique ? Quelle couverture et quel traitement des néologismes liés au Covid-19 offrent-ils ? Quelle est la qualité d'articles dictionnaires décrivant ces néologismes ? Dans notre étude, nous consacrerons d'abord quelques mots au processus néologique lié au Covid-19. Ensuite nous présenterons nos objectifs, le contexte scientifique ainsi que le corpus. Nous discuterons enfin les résultats de notre analyse qui porte à la fois sur la micro- et la macrostructure des dictionnaires étudiés.

2. Processus néologique lié au Covid-19

Il semble que la crise sanitaire ait entraîné une vraie « épidémie lexicale » (Lardellier, 2022 : 87). Citons quelques exemples : *coronapéro* (apéro en visioconférence), *coronologue* (par ironie, « spécialiste » du coronavirus), *coronapiste* (piste cyclable servant à fluidifier le trafic durant la pandémie), *coronasceptique* (qui doute de l'impact du virus) ayant tous pour base *corona*, ou encore *covidiot* (personne au comportement irrationnel), *covidéprimer* (déprimer à cause du Covid-19), *mélanCovid* (morosité causée par la pandémie), *covid manager* (responsable de mesures au sein d'une entreprise), *c'est pas très covid* (qui ne tient pas compte des mesures) à partir de *covid*. Une autre matrice particulièrement féconde semble être *confinement* : ainsi nous trouvons *reconfinement*, *reconfinement*, *déconfinement*, *redéconfinement*, *après-confinement*, *semi-confinement*, *quasi-confinement*, *grand confinement*, *surconfinement*, *anticonfinement*, *antidéconfinement*. Nous trouvons également *balconner* (applaudir sur le balcon), *airgasmer* (prendre une bouffée d'air en enlevant le masque), *connardvirus* (virus qui pourrit la vie), porter le masque à *raz-le-nez*, *zoûter* (Zoom au moment du goûter), mais aussi des néologismes comme *pneumonie chinoise*, *pneumonie de Wuhan*, *kung flu* (exemples tirés du *Wiktionnaire* et du *Dicovid* du *Petit Robert*¹).

1. Si la démarche du *Dicovid* est collaborative, le projet est axé exclusivement sur le caractère ludique des créations néologiques : *Robert*, en partenariat avec l'OULIPO, a sélectionné les créations les plus drôles des internautes. Les mots listés ne sont pas considérés comme des lexèmes avec une réelle valeur d'échange mais uniquement comme le produit d'un processus de création gratuite.

Rien qu'en regardant ces exemples nous pouvons constater que la création néologique a donné lieu aussi bien aux unités monolexicales (comme *reconfinement*, *déconfinement*, *rereconfinement*, etc.) que polylexicales (comme *covid manager*, *pneumonie chinoise*, *c'est pas très covid*, etc.) qui s'inscrivent tant dans des domaines de spécialité (médical, militaire, physique, etc.) que dans la langue de la vie courante. Nous y trouvons aussi bien des néologismes sémantiques qui donnent des sens nouveaux aux unités lexicales existantes (Pruvost et Sablayrolles, 2003 : 10; notamment *confinement*) que formels qui proposent des unités nouvelles (fondées notamment sur les matrices *corona* et *covid*). Parmi ces derniers, nous constatons la présence de procédés simples – la dérivation (comme *déconfinement*, *rereconfiner*, etc.) et la composition (comme *covid manager*, *raz-le-nez*, etc.), mais aussi des mots-valises (comme *airgasmer*, *zoûter*, etc.) – ainsi que des mécanismes plus complexes. Les emprunts et calques de l'anglais (comme *kung flu*, *grand confinement*, etc.) semblent particulièrement féconds comme matrice néologique. Une dimension ludique (Tallarico, 2022) et humoristique (Guo-Gripay, Berbinski et Veleanu, 2022) est incontestable. Il est difficile de juger si nous pouvons parler ici d'un plaisir lié à la création néologique ou si cette dimension relève d'une volonté de surmonter l'angoisse par l'humour (Berbinski et Veleanu, 2022).

Quoi qu'il en soit, nous avons pu constater l'omniprésence de cette nouvelle terminologie dans les médias et les conversations dès le début de la pandémie. La nécessité de définir précisément ces termes a accéléré leur introduction dans les ouvrages de référence.

3. Contexte scientifique et objectifs

Cette floraison néologique a suscité l'intérêt de la communauté scientifique ; de nombreuses études et analyses pour différentes langues ont vu jour, comme l'anglais (Akut, 2020; Asif, Zhiyong, Iram et Nisar, 2021; Mweri, 2021; Al-Azzawi et Haleem, 2021; Sajous, 2022), l'allemand (Klosa-Kuckelhaus, 2022), le roumain (Dinca, 2022 ; Stoichitoiu Ichim, 2022), etc. Par ailleurs, cette création simultanée dans de nombreuses langues a permis des comparaisons interlinguistiques inédites (Altmanova *et al.*, 2022 pour un panorama), des études bilingues (Jacquet-Pfau et Kacprzak, 2022) et plurilingues (Guo-Gripay, Berbinski et Veleanu, 2022; Martí Solano, 2022).

Notre étude se situe en prolongement des travaux sur le lexique français qui ont porté sur les néologismes formels et sémantiques liés à la pandémie (Grimaldi, 2022; Maldussi, 2022), y compris la création métaphorique (Rollo, 2022), ainsi que sur les aspects humoristiques et expressifs du lexique du Covid-19 (Tallarico, 2022) ou sur la visée pragmatique et la dimension énonciative (Pennec, 2021). Elle se veut également complémentaire de l'inventaire macrostructural de Kolářiková (2021) et de la comparaison de Sajous et Humbley (2022) entre le traitement lexicographique dans le *Wiktionnaire* et le traitement encyclopédique dans *Wikipédia* de la terminologie de l'isolement sanitaire. Notre démarche s'inscrit dans la même dynamique que Sajous et Humbley (2022) dans la mesure où elle entend associer à l'étude d'un phénomène néologique particulier une réflexion épistémologique plus générale sur la néologie elle-même et son traitement lexicographique. Notre perspective est un peu

plus large parce qu'elle se fonde sur un éventail plus diversifié de termes, rejoignant ainsi l'approche méthodologique de Kolářiková (2021).

Notre objectif est de fournir un état des lieux de la lexicographie francophone en ligne et d'analyser la qualité aussi bien de la micro- que de la macrostructure à travers les néologismes liés au Covid-19. Avant de passer à l'analyse proprement dite, nous allons consacrer quelques mots à notre corpus et à la méthodologie utilisée.

4. Corpus et méthodologie

Dans le cadre de cette étude, nous nous basons uniquement sur des dictionnaires francophones gratuits et accessibles en ligne. Afin d'arriver à un corpus diversifié et équilibré, nous avons tenu compte de plusieurs critères, tels que la diversité institutionnelle ainsi que la diversité dans le mode de collecte de données dictionnaires. Ainsi, nous avons choisi :

- *Dictionnaire de français Larousse* (DFL) – dictionnaire professionnel gratuit² en ligne édité en France, <https://www.larousse.fr/dictionnaires/francais-monolingue> ;
- *Dico en ligne Le Robert* (DLR) – dictionnaire professionnel gratuit en ligne édité en France, <https://dictionnaire.lerobert.com/> ;
- *Le Dictionnaire des francophones* (DDF) – dictionnaire institutionnel hybride intégrant plusieurs dictionnaires existants tout en permettant l'enrichissement collaboratif. Accès gratuit en ligne, <https://www.dictionnairedesfrancophones.org/>;
- *Le Wiktionnaire* – dictionnaire général complètement collaboratif du français ; bien référencé par les moteurs de recherche, il présente une source d'information lexicale de plus en plus privilégiée. Accès gratuit en ligne, https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil.

Les quatre dictionnaires choisis diffèrent également quant à leur politique éditoriale : les deux dictionnaires professionnels se concentrent essentiellement sur les usages européens, même s'ils intègrent quelques variantes diatopiques, que nous examinerons sous 5.6. La nomenclature de ces deux dictionnaires en ligne reste tributaire des éditions papier, même si la mise à jour peut être plus rapide en ligne (Kolářiková, 2021). Les dates de mise à jour des articles en ligne n'est toutefois pas indiquée. La microstructure de ces deux dictionnaires est assez différente. Le DFL propose les définitions des entrées, les expressions contenant le mot-vedette et éventuellement des exemples, les homonymes, les difficultés, etc. Le DLR identifie pour chaque entrée des définitions, des citations générées automatiquement et parfois aussi des synonymes ainsi que des combinaisons triées en fonction de leurs éléments constitutifs. Il propose également un accès au *Dictionnaire universel de Furetière* (1690).

2. La politique de cookies mise en place par *Larousse* tend à réduire l'accessibilité des données dans la mesure où il faut les accepter tous ou s'abonner au dictionnaire pour pouvoir le consulter.

Le DDF, contrairement aux deux précédents, est centré sur la variation diatopique. Il intègre plusieurs sources dictionnaires, anciennes ou plus récentes, dans un même modèle de données. Les deux seules ressources régulièrement mises à jour sont le *Grand dictionnaire terminologique* québécois et le *Wiktionnaire*, versés dans le DDF sur base régulière. Le DDF permet aux lecteurs de contribuer à l'extension et à la mise à jour du contenu, sans possibilité de suppression. Même si les ressources intégrées jusqu'ici ne lui permettent pas tout à fait d'atteindre cet objectif, le DDF se veut un dictionnaire général du français qui vise à décrire tous les usages de la langue dans le monde, tous registres confondus. Ce parti pris est aussi celui du *Wiktionnaire*, qui est lui entièrement collaboratif. Sa mise à jour et le degré de généralité qu'il peut atteindre sont dépendants du dynamisme et de la composition de la communauté des contributeurs.

Le corpus de termes constitué par Kolářiková (2021), intégré par ailleurs dans son relevé des descriptions lexicographiques de la terminologie Covid-19 en français, nous a servi de point de départ. Nous avons écarté les lexèmes certes très présents pendant la pandémie mais dont le caractère n'était pas forcément néologique (par exemple *ventilateur*, *télétravail*, *taux de létalité*, etc.) même si ce critère n'était pas toujours évident à appliquer et de nombreux cas sont en réalité limites ; ainsi nous avons gardé environ la moitié (27 sur 59) de ses entrées. Nous avons ensuite enrichi cet inventaire des relevés dans les dictionnaires choisis. Nous avons procédé à une recherche par mots-clés et avons veillé à y inclure différentes bases lexicales ainsi que des unités mono- et polylexicales, associées au Covid-19 (y compris leurs variantes graphiques). Nous avons aussi inclus les entrées reliées aux entrées relevées.

La deuxième phase a notamment servi à nettoyer le corpus : nous avons d'abord regroupé les différentes variantes graphiques et éliminé les doublons (qui sont nombreux notamment entre le *Wiktionnaire* et le DDF ; cela s'explique par le fait que le *Wiktionnaire* est aussi inclus dans le DDF). Puis, nous avons écarté les expressions reliées trop éloignées (ce qui était notamment courant dans le DDF). Ainsi nous sommes parvenues à un corpus non exhaustif couvrant 27 entrées venant du corpus Kolářiková (2021) enrichi par 83 entrées ; soit au total 110 entrées différentes liées au Covid-19 venant de quatre dictionnaires différents (voir relevé en annexe). Dans la section suivante, nous allons présenter notre analyse du corpus faite à travers différents critères, touchant des aspects généraux mais aussi plus concrètement la micro- et la macrostructure, nous basant donc sur une approche qualitative non exhaustive.

5. Analyse

Afin de pouvoir fournir un bilan aussi complet et approfondi que possible, nous avons décidé d'analyser notre corpus à l'aide de sept critères, regroupés dans trois catégories. Parmi les aspects généraux, nous avons examiné la rapidité de la mise à jour et la réactivité des dictionnaires à décrire les néologismes ainsi que l'exhaustivité de la nomenclature et le traitement de la variation interne à la francophonie. Dans les aspects touchant à la macrostructure, nous avons analysé la présence

des variantes graphiques, le traitement d'unités polylexicales ainsi que les renvois structurés. Pour ce qui relève de la microstructure, nous nous sommes focalisées sur l'homogénéité du marquage métalexicographique et sur la neutralité des définitions et des exemples.

5.1. Mise à jour et réactivité

Dans son étude, Kolářiková note que c'est seulement dans les éditions 2022 (sorties en juin 2021) que les néologismes liés au Covid-19 trouvent leur place dans le *Petit Robert* et le *Petit Larousse illustré* (26 néologismes formels ou sémantiques dans le *Petit Robert* 2022 et 48 dans le *Petit Larousse illustré* 2022, Kolářiková, 2021 : 5). Ce retard semble considérable or

[...] les éditeurs ont réussi à intégrer quelques-uns de ces mots dans la version informatisée (à titre d'exemple, nous pouvons mentionner *immunité collective*, *patient zéro*, *cluster*, *confinement*, etc. qui ont été introduits dans la version en ligne du *Petit Robert* en 2020). Les versions informatisées prennent ainsi un peu d'avance sur celles en papier. Alors, les deux versions se désynchronisent, ce qui n'est pas habituel mais témoigne de la nécessité de réagir en temps réel (Kolářiková, 2021 : 5).

Même si cette « situation inhabituelle » voulait répondre à l'urgence, elle ne le faisait que partiellement ; de fait la réactivité des dictionnaires professionnels n'a pas été comparable aux dictionnaires collaboratifs qui ont intégré les néologismes massivement et systématiquement dès mars 2020 et constituaient ainsi la référence lexicographique principale en la matière.

En raison de la lenteur de la mise à jour des dictionnaires professionnels tenus à des cycles assez longs de maisons d'édition, les ressources collaboratives sont souvent les premières à décrire les nouvelles créations lexicales et font preuve d'une réactivité remarquable (Sablayrolles, 2018 : 239). Cette réactivité permet une introduction très rapide d'un grand nombre de néologismes, même ceux dont la pérennité dans la langue n'est pas garantie. Cette caractéristique les oppose aux dictionnaires professionnels, qui adoptent une approche plus prudente en attendant d'être certains que le mot en question s'implante durablement dans le langage courant. Cette terminologie nouvelle a donc été décrite d'abord, et parfois seulement, par des ressources collaboratives (par exemple, *gestes barrières* dans le *Wiktionnaire* en mars 2020 ou *covidiot* en mai 2020) ou des glossaires profanes dans la presse. Les dictionnaires professionnels n'ont été mis à jour que dans un deuxième temps, suivis plus tard par des dictionnaires spécialisés, ludiques (*Le Dicorona*, *Le Dicovid : D'asymptomatique à Zoonotique*) ou plus sérieux (*Les mots du coronavirus*).

Le *Dictionnaire des francophones* (DDF) occupe une position particulière dans cet écosystème lexicographique. En raison du fait que sa communauté de contributeurs est encore en phase de développement, sa réactivité n'est pas maximale. Cependant, l'intégration régulière du contenu du *Wiktionnaire* vient compenser cette lacune.

5.2. Exhaustivité de la nomenclature

Si l'étendue de la nomenclature des dictionnaires papier est nécessairement limitée par les contraintes matérielles, celle de leurs homologues numériques permet théoriquement l'exhaustivité. Cependant, même sous forme numérique, les dictionnaires présentent des différences notables. Notre recherche par mot-clé met en avant deux pôles : 34 (DLR) et 44 (DFL) résultats pour les dictionnaires professionnels versus 76 (DDF) et 99 (*Wiktionnaire*) résultats pour les dictionnaires à volet collaboratif (dont certains sont des épaves - entrées sans articles dictionnaires au moment où nous écrivons ces lignes, en août 2024) (voir relevé en annexe).

Ce chiffre élevé, qui illustre parfaitement deux tendances, est dû à plusieurs facteurs. D'une part le principe de prudence, déjà évoqué, limite l'introduction de néologismes dans les dictionnaires professionnels numériques ; d'autre part les dictionnaires collaboratifs aspirent à une inclusivité totale dans la nomenclature - non seulement des néologismes, mais aussi du lexique francophone dans toute sa variation, des unités polylexicales et des variantes graphiques - éléments sur lesquels nous reviendrons encore dans la suite.

Cet idéal d'exhaustivité est toutefois soumis aux aléas de la dynamique de contribution. Le processus d'enrichissement est souvent non structuré, impulsé au gré des envies des contributeurs. Par exemple, dans le *Wiktionnaire*, la liste des termes désignant les apéritifs virtuels organisés pendant le confinement est impressionnante : *coronapéro*, *apéro virtuel*, *e-apéro*, *visio-apéro*, *web-apéro*. Par contre, pour des raisons obscures, *covid-apéro*, pourtant attesté, n'y est pas repris. Ce simple exemple et les innombrables cas similaires soulignent l'impact de la politique éditoriale des ressources lexicographiques qui favorisent ou non la réactivité et l'inclusivité.

5.3. Présence des variantes graphiques

Les variantes graphiques d'une même unité sont quasi-absentes des deux dictionnaires professionnels de notre corpus ; quand elles sont mentionnées, elles sont rassemblées dans un même article. Par contre, elles sont nombreuses dans le *Wiktionnaire* (et par voie de conséquences dans le DDF) où elles bénéficient d'articles distincts. Outre de nombreuses variantes orthographiques pour *Covid(-19)* (*Covid*, nom et adjectif ; *covid*, nom et adjectif ; *Covid-19*, nom et adjectif ; *covid-19*, nom ; *COVID-19*, nom), le *Wiktionnaire* mentionne également les appellations alternatives *maladie à coronavirus 2019*, *corona*, *pneumonie de Wuhan* - *pneumonie chinoise* qu'on trouve via les renvois lexicaux.

Le *Wiktionnaire* contient donc au total cinq entrées pour *Covid(-19)* en tant que nom ou adjectif. Comme il dépend de la vigilance de la communauté, le contenu de ces entrées n'est pas toujours totalement homogène : le sens métonymique « Qui intervient pendant le cours de la pandémie » apparaît par exemple sous *Covid* mais pas sous *covid*. Le tableau ci-dessous compare le nombre d'entrées pour une sélection de termes dans les différents dictionnaires de notre corpus (août 2024).

Parfois les chiffres du *Wiktionnaire* et du DDF ne coïncident pas car les articles-épaves du *Wiktionnaire* ne sont pas inclus dans le DDF.

	<i>covid</i>	<i>coronavirus</i>	<i>geste barrière</i>	<i>cas contact</i>
DFL	1	1	0	0
DLR	1	1	0	0
Wiktionnaire	5	2	2	2
DDF	4	2	1	1

5.4. Traitement des unités polylexicales

Dans les ouvrages lexicographiques généraux, il est fréquent que les unités polylexicales ne soient traitées que de manière incidente sous les entrées des lexèmes qui composent ces unités ; en d'autres mots elles n'apparaissent pas dans la nomenclature comme une entrée à part entière même si elles sont figées. C'est le cas dans les dictionnaires professionnels de notre corpus. Comme l'indique le tableau ci-dessous, les principales unités polylexicales liées à l'épidémie de Covid-19, lorsqu'elles sont traitées, apparaissent dans une sous-entrée, dans un champ « expression » sous une entrée ou simplement dans les exemples listés sous les définitions. Ces unités, dissimulées dans la microstructure, ne sont pas repérables par le moteur de recherche dans le DLR. Il faut en chercher chacun des termes et espérer que l'expression soit traitée dans l'entrée correspondant à l'un d'entre eux. Dans le DFL, par contre, le moteur de recherche permet de retrouver les expressions qui sont traitées, à condition d'encoder la forme attendue (par exemple *gestes barrières* et non *geste barrière*), ce qui constitue un atout important.

Dans le *Wiktionnaire* et dans le DDF, par contre, les unités polylexicales bénéficient d'entrées séparées, trouvables en tant que telles via le moteur de recherche et reprises dans les renvois comme unités liées. L'information est ainsi complètement accessible, sans contenu caché. Les mêmes expressions bénéficient toutefois d'autant d'entrées que de variantes possibles (*geste barrière*, *geste-barrière*, *gestes barrières* dans le *Wiktionnaire*), ce qui ne facilite pas toujours la recherche.

	<i>cas contact</i>	<i>geste(s) barrière(s)</i>	<i>passé sanitaire</i>	<i>crise sanitaire</i>
<i>DFL</i>	sous <i>contact</i> (expressions)	sous <i>barrière</i> (expressions, zootéchnie)	sous <i>passé</i>	non traité
<i>DLR</i>	sous <i>contact</i>	sous <i>barrière</i>	sous <i>passé</i> (exemples) et sous <i>sanitaire</i> (exemples)	sous <i>crise</i> (exemples)
<i>Wiktionnaire</i>	1 entrée, 1 épave	2 entrées	2 entrées	1 entrée
<i>DDF</i>	1 entrée	2 entrées	2 entrées	1 entrée

5.5. Renvois

Quasi inexistantes dans le DFL et le DLR, les renvois sont extrêmement nombreux dans le *Wiktionnaire* et surtout dans le DDF. Les renvois dans le DDF sont particulièrement intéressants à examiner. Leur nombre est impressionnant, comme le montre le tableau ci-dessous.

	<i>covid</i>	<i>coronavirus</i>	<i>confinement</i>	<i>geste barrière</i>
<i>DFL</i>	1	0	2	0
<i>DLR</i>	0	0	2	0
<i>Wiktionnaire</i>	22	12	22	2
<i>DDF</i>	15	48	79	21

Pour que des renvois lexicographiques puissent être complètement utiles et pertinents, l'enjeu principal est la clarté de leur structuration : les renvois sont-ils transparents, lisibles et explicites ? Le caractère explicite se manifeste dans le signalement des renvois en tant que tels, dans un champ séparé ou une mise en forme différente, idéalement cliquables pour mener directement à l'entrée qui fait l'objet d'un renvoi (Ligas, 2010). La lisibilité renvoie tant aux problématiques d'accessibilité aux différents profils d'utilisateur qu'aux choix de mise en page pour une distinction immédiatement visible entre les différents renvois. Enfin, la transparence désigne ici les principes d'organisation interne des renvois qui devraient idéalement être immédiatement apparents pour permettre à l'utilisateur de comprendre pourquoi et en vertu de quel critère un renvoi est proposé (synonymie, antonymie, variantes graphiques, dérivés, composés polylexicaux, etc.).

Pour des raisons de lisibilité et de qualité de l'information, plus les renvois sont nombreux, plus leur structuration doit être claire, ce qui n'est malheureusement pas le cas pour le DDF. Prenons deux exemples, *coronavirus* et *confinement*. Sous ces deux lexèmes, on ne trouve dans le DDF, contrairement au *Wiktionnaire*, que des renvois explicites, qui bénéficient eux-mêmes d'une entrée (Ligas, 2010).

Sous *coronavirus*, les renvois sont répartis en quatre catégories (« mot ou expression dérivée », « vocabulaire lié », « hyponyme », « hyperonyme »), qui ne sont pas clairement distinguées et définies. Dans la longue liste de dérivés, on retrouve des néologismes du Covid-19 formés à partir de *corona*, comme *coronapéro*, *coronapiste* ou *coronabond*, mais aussi une grande quantité de noms de virus, qui n'ont pas de rapport avec celui-ci. Le « vocabulaire lié » est particulièrement intéressant parce qu'il n'est pas fondé sur des ressemblances formelles mais analogiques (Zotti, 2014). Le problème ici c'est que la relation lexico-sémantique qui justifie ces renvois n'est pas indiquée : *premier de tranchée*, *monde d'après*, *Grand Confinement*, *corona*. Sauf pour *corona* qui est une abréviation, qu'on pourrait considérer comme synonymique, les autres lexèmes ont une relation plus métonymique avec *coronavirus*. La présence de ce type de relations, qui rattachent le lexème à définir à un cadre de référence, comparable aux « frames » de Fillmore (1982), donne un caractère onomasiologique au DDF. Ce caractère est un atout qui le distingue de la plupart des autres dictionnaires, mais malheureusement le lien entre l'entrée et le vocabulaire lié est parfois ténu et toujours non explicite. De plus, en ce qui concerne les renvois dont la nature est clairement définie, comme « hyponyme » et « hyperonyme », ils ne sont pas reliés au sens pertinent lorsqu'ils se rattachent à des polysèmes. Ainsi *SARS-CoV-2* est un hyponyme de *coronavirus* lorsqu'il désigne une vaste famille de virus, mais est relié plutôt par « métonymie³ », d'après la définition du *Wiktionnaire* reprise dans le DDF, à *coronavirus* pour désigner en particulier le virus qui provoque le Covid-19.

Sous *confinement*, un nouveau type de renvoi apparaît, qui n'est pas présent sous *coronavirus*, « terme apparenté », qui contient en fait des dérivés (*reconfinement*, *confinable*, *autoconfinement* ou *déconfinement*, par ailleurs aussi repris sous « antonyme »). Sous « mot ou expression dérivée », on

3. Par métonymie ou plutôt par synecdoque genre-espèce. La définition de cette relation peut être sujette à débat.

retrouve sensiblement les mêmes lexèmes, à quelques exceptions près (*autoconfiner* n'y est pas), auxquels vient s'ajouter *post-confinement*, ainsi qu'une série de dérivés et composés formés sur *fin* et qui n'ont aucun rapport avec *confinement* (*interrupteur de fin de course*, *fin de droits*, *toucher à sa fin*, etc.), listés juste avant *Grand Confinement*. Le renvoi à ces lexèmes est vraisemblablement dû à la génération semi-automatisée de ce type de liste, sur une base morphologique erronée dans ce cas. Sont également présents dans cette longue liste des dérivés et composés liés à un autre sens de *confinement*, défini dans le cadre de la production d'énergie nucléaire (*fusion par confinement inertielle*). Sous « vocabulaire lié », les expressions reprises renvoient quasi-exclusivement à la production d'énergie nucléaire, sauf *quarantaine* et *isolement*, en fin de liste.

Ces deux exemples montrent que la distinction est peu claire entre les types de renvois, tout y est mélangé, même lorsque les dérivés concernent des sens différents. Rien ne permet vraiment de s'y retrouver dans les dizaines de renvois proposés. Pour un utilisateur qui connaît les mots listés, il est encore possible de reconstruire une logique, mais pour des apprenants ces renvois, qui pourraient être vraiment intéressants, sont difficilement utilisables sans accompagnement.

5.6. Intégration de la variation et homogénéité du marquage

L'impact de la dynamique communautaire et de la politique éditoriale est aussi palpable en ce qui concerne l'intégration des variantes notamment géographiques dans les dictionnaires.

L'intégration de la variation diatopique dans les dictionnaires professionnels reste anecdotique. Elle est limitée à quelques cas clairement typés de « -ismes » (belgicismes, québécoismes, helvétismes) et n'intègre que rarement des acceptions géographiquement spécifiques (ex. *espadrille*, « (Canada) Chaussure de sport » dans DLR, mais aucune mention des sens québécois de *sucre*, en revanche) à côté des particularismes lexicaux (ex. *tuque* dans DLR). Les particularismes francophones liés au Covid-19 ne sont intégrés ni dans le DLR ni dans le DFL. Le *Wiktionnaire* tend à remettre la variation à sa juste place, au centre de la pratique linguistique. Les mots utilisés, le sens qu'on leur donne, les expressions dans lesquelles ils entrent changent plus ou moins d'une région francophone à l'autre ; c'est à la fois le fondement et la conséquence de la pratique sociale de la langue. Même si l'inclusivité est théoriquement totale dans un dictionnaire collaboratif comme le *Wiktionnaire*, le manque de systématisme, notamment dans le marquage, pose le problème de la qualité et de la complétude de l'information.

Ainsi, sous *passé sanitaire*, on trouve un renvoi vers *passéport vaccinal* avec la mention « (France), (Canada) », mais sous *passéport vaccinal* - où on attendrait « (Canada) » - aucun marquage n'est visible. Les variantes belge (*Covid safe ticket*) et suisse (*certificat Covid*) ne sont pas mentionnées. *Gestes barrières* (au singulier ou au pluriel) est étiqueté « (France) » alors que la locution est aussi largement utilisée avec le même sens dans le reste de la francophonie. Ces lacunes reflètent sans doute le profil de la communauté des contributeurs, essentiellement Français.

Alors que le marquage de variation est souvent aléatoire dans le *Wiktionnaire*, c'est la grande force des dictionnaires professionnels d'assurer la cohérence de l'ensemble - bien entendu dans la mesure où la variation est présente. À ce niveau, le DDF peut être considéré comme un dictionnaire professionnel parce qu'il rassemble et aligne des ressources lexicographiques scientifiques. À ces ressources s'ajoutent des contributions libres et le contenu du *Wiktionnaire*, régulièrement mis à jour. Les néologismes du Covid-19 n'y sont présents que par ces deux biais. Lorsqu'aucune balise géographique n'est présente dans le *Wiktionnaire*, le DDF étiquette l'entrée par défaut sous « Monde francophone », y compris lorsque des mentions géographiques, plus difficilement détectables automatiquement, sont contenues dans les définitions (voir, par exemple, *sucre de pomme*, « confiserie de Rouen »). Le manque de systématisme dans le marquage du *Wiktionnaire* donne ainsi lieu à une systématisme totale dans le DDF mais celle-ci est souvent trompeuse. *Passeport vaccinal*, par exemple, apparaît sous « Monde francophone », mais aussi sous « Québec », via le *Grand dictionnaire terminologique*. Si tous les francophones peuvent comprendre la locution, certains d'entre eux seulement (les Québécois, mais aussi les Tunisiens, par exemple) s'en sont activement servi. Pour les mêmes raisons de reprise du marquage non homogène du *Wiktionnaire*, *geste barrière* est marqué comme utilisé uniquement en France alors que *gestes barrières* est étiqueté « Monde francophone ».

5.7. Définitions, exemples et neutralité

Comme dans tous les domaines-clés d'une société humaine, la production néologique liée à la pandémie de Covid-19 a donné lieu à des polémiques, tant sur la forme que sur le fond. La question se pose de savoir dans quelle mesure la description lexicographique rend compte de ces polémiques. En ce qui concerne le débat sur la forme, tous les dictionnaires de notre corpus font état des hésitations de l'usage sur la question du genre de *covid* avec plus ou moins de précisions : si le DLR mentionne que le mot peut avoir les deux genres, le DFL indique que l'Académie française recommande le féminin et le *Wiktionnaire* ajoute à cette précision des informations sur les fondements de cette recommandation (le genre de *maladie*), sur la position de l'Office québécois de la langue française, sur l'usage qui tend à s'imposer au Québec et en France (voir Steffens, 2022; Dow et Drouin 2023, pour une analyse). Le DDF reprend le *Wiktionnaire* et mentionne que « l'usage hésite ». Il intègre également une contribution caractérisant *covid* comme un nom féminin, étiqueté « Canada ».

Un dictionnaire peut aussi dans une certaine mesure refléter les débats idéologiques sur les concepts dans les définitions des mots ou dans les exemples choisis. Se pose alors la question de sa neutralité. Cette question a été bien étudiée notamment en ce qui concerne le *Wiktionnaire* avant la pandémie (Sajous *et al.*, 2019). Qu'en est-il du traitement des néologismes de la pandémie dans les dictionnaires de notre corpus ? Prenons l'exemple d'*antivax*.

La définition donnée par le DLR à *antivaccin* (*antivax* étant mentionné comme abréviation familière) est minimale : « Qui est hostile à la vaccination⁴ ». Les citations reprises sous cette entrée ne sont pas neutres. Le DLR met en garde : les exemples sont extraits automatiquement de textes scientifiques, journalistiques, littéraires et institutionnels issus de « sites partenaires externes⁵ » et « ne font pas l'objet d'une relecture par les équipes du Robert⁶ ». Il est intéressant de noter que la moitié des citations (5/10) présentes dans cet article en août 2024 contiennent *antivaccin* en mention plutôt qu'en usage et sont issues du même article : « Le terme *antivaccin* et ses variantes ont pu être identifiés sur 1009 pages de notre corpus (sur un total de 27292) » [sic], dans *Déviance et Société* (2019) de Jeremy K. Ward, Paul Guille-Escuret et Clément Alapetite (Cairn.info). Plus intéressant encore, cet exemple : « La démarche *antivaccin*, y compris dans les pays en développement, est sous-tendue par la même idéologie : limiter la protection pour faire régresser la démographie.⁷ », issu d'un article militant de Pierre PAGESSE sur le « *bashing* » subi par les agriculteurs, paru dans *Paysans et société* en 2019. Ces exemples sont antérieurs à la pandémie, pendant laquelle le terme a connu une utilisation beaucoup plus importante et ciblée sur la contestation d'une campagne massive de vaccination contre le SARS-CoV-2 à partir de vaccins très récemment développés. Ces exemples pourraient induire le lecteur en erreur et mériteraient d'être mis à jour et vérifiés.

L'adjectif *antivax* est défini en tant que tel dans le DFL de manière assez neutre : « Se dit d'un mouvement d'opinion marqué par une opposition à certains vaccins ou à la vaccination en général, dont il remet en cause l'efficacité et l'innocuité. (On dit aussi *antivaccin*, *antivaccination* ou *antivaccinal*.)⁸ », et sans exemples. Par contre, sous la définition, se trouve le commentaire suivant qui exprime une position claire contre le mouvement *antivax* : « Les théories de ce mouvement, non fondées sur les données acquises de la science, exposent, dans la mesure où elles entraînent une baisse de la couverture vaccinale, à un risque de réapparition d'épidémies de maladies infectieuses⁹ ». Par ailleurs, le traitement par le DFL d'autres mots polémiques semble également partisan, par le biais du commentaire. C'est le cas pour *woke*, où la remarque « Ce terme, importé en France en 2020, est fréquemment employé par les tenants de l'universalisme républicain pour qualifier les excès relatifs au militantisme des défenseurs des minorités. » contextualise le terme en identifiant les deux camps antagonistes (Vincent, 2022), mais pas de façon tout à fait neutre, dans la mesure où *universalisme républicain* est connoté plutôt positivement.

4. <https://dictionnaire.lerobert.com/definition/antivaccin> [page consultée le 24 août 2024]

5. C'est la raison pour laquelle *passé sanitaire* et *crise sanitaire* apparaissent massivement dans les exemples repris sous *sanitaire*.

6. <https://dictionnaire.lerobert.com/definition/antivaccin> [page consultée le 24 août 2024]

7. <https://dictionnaire.lerobert.com/definition/antivaccin> [page consultée le 24 août 2024]

8. <https://www.larousse.fr/dictionnaires/francais/antivax/188487> [page consultée le 24 août 2024]

9. <https://www.larousse.fr/dictionnaires/francais/antivax/188487> [page consultée le 24 août 2024]

Pour *antivax*, le DDF reprend exclusivement les deux articles du *Wiktionnaire*, l'un pour le nom, l'autre pour l'adjectif, les définitions et exemples qu'ils contiennent. La présence d'exemples dans le DDF dépend ainsi de la source intégrée et des potentiels ajouts de contributeurs.

La définition du *Wiktionnaire* quant à elle s'inspire visiblement du DFL et insère le commentaire directement dans la définition du substantif *antivax* : *antivax* (nom) « Personne ou mouvement opposée [sic] à certains vaccins ou à la vaccination en général, et qui remet en cause, pour des raisons non fondées sur les données acquises par la science, son efficacité ou son innocuité.¹⁰ » ; la définition de l'adjectif par contre reste neutre : *antivax* (adjectif) « Qui s'oppose à la vaccination et remet en cause son efficacité ou son innocuité.¹¹ »

L'historique des modifications montre qu'à l'insertion de ces entrées dans le dictionnaire en 2020 les deux définitions étaient neutres. C'est en 2024 que la définition du substantif change radicalement pour passer de l'individu au groupe et indiquer une prise de position. Apparemment sans réel débat, juste ce coup de gueule le 27 juillet 2022 dans la section *Discussion* du *Wiktionnaire* : « Wikipedia est une encyclopedie ou bien un relai politique ? Tous les exemples de citation font passer les anti vax pour des idiots ou les coupables de l'épidémie. Aberrant » [sic], suivie d'une réponse, étonnante, le même jour : « Ici, cest le *Wiktionnaire*, pas l'encyclopédie partisane Wikipédia. Vous pouvez alimenter la page antivax en exemples de votre choix.¹² » [sic]. Cet échange traduit une confusion entre le dictionnaire et l'encyclopédie, mais aussi ce qui semble être des tensions internes au sein de Wikimedia France.

Dans l'historique des contributions pour *antivax*, un autre commentaire est intéressant : « Les références sont difficiles à trouver en dehors du wiki anglais. Je précise aussi que je ne suis pas membre du mouvement, j'améliore juste le Wiki de temps en temps¹³ ». Ce type de remarque soulève la question de la posture adoptée par les contributeurs pour justifier leur légitimité à leurs propres yeux et à ceux des autres membres de la communauté. Les facteurs intervenant dans la construction de cet ethos de légitimité des contributeurs mériteraient plus d'attention que nous ne pouvons y consacrer ici.

L'évolution de la définition d'*antivax* illustre donc très clairement les modalités de l'édition par consensus. Citons encore le cas de *passé sanitaire* dans lequel le mot *pas* dans le segment « certifiant qu'une personne n'est pas contagieuse » a progressivement été remplacé par *peu* puis par *moins* dans la version de la définition en août 2024. Le débat auquel on pourrait s'attendre au sein de cette communauté collaborative semble en réalité assez limité : quand le consensus n'est que mou parce

10. <https://fr.wiktionary.org/wiki/antivax> [page consultée le 24 août 2024]

11. <https://fr.wiktionary.org/wiki/antivax> [page consultée le 24 août 2024]

12. <https://fr.wiktionary.org/wiki/Discussion:antivax> [page consultée le 24 août 2024]

13. <https://fr.wiktionary.org/w/index.php?title=antivax&action=history> [page consultée le 24 août 2024]

que les contributeurs se taisent plutôt que de se concerter explicitement sur les modifications, le dernier qui parle l'emporte.

6. Conclusion

La pandémie de Covid-19 a entraîné une explosion néologique sans précédent. Aussi fascinant que ce phénomène ait été, décrire ces néologismes de manière systématique, adéquate et rapide n'a pas été une tâche aisée pour les lexicographes, qu'ils soient professionnels ou amateurs, si tant est que cette distinction puisse toujours être aussi claire. Nous en avons profité pour dresser un bilan de la lexicographie francophone actuelle en ligne à partir de deux dictionnaires professionnels hexagonaux intégrant la variation et de deux dictionnaires à portée explicitement mondiale. Il s'avère que la description reste circonscrite essentiellement aux usages européens, voire français, même si le *Wiktionnaire* et le DDF ont, comme attendu, une nomenclature plus large. Le bilan se dessine par ailleurs moins positif que nous ne pourrions le penser.

Notre parcours dans les entrées néologiques liées à la pandémie dans quatre dictionnaires nous amène aux conclusions suivantes : la réactivité des dictionnaires collaboratifs est incontestable. L'inclusivité théoriquement totale dont ils font preuve leur donne un atout clair en ce qui concerne la richesse de la macrostructure qui intègre par ailleurs de plein droit les unités polylexicales. En ce qui concerne les dictionnaires professionnels de notre corpus, il faut mettre en avant le moteur de recherche de *Larousse* qui – contrairement au *Robert* – permet de repérer les unités polylexicales même quand elles ne constituent pas d'entrée à part entière dans la nomenclature. La grande force du *Robert*, en revanche, est la présence de citations – certes, automatiquement générées et non relues mais souvent mieux mises à jour que les définitions. Nous regretterons l'absence de citations dans *Larousse*. L'homogénéité, la systématisme et la structuration des données peuvent être mises en avant dans les deux dictionnaires professionnels, tandis que ces qualités pourraient être plus développées dans le *Wiktionnaire*, mais surtout dans le DDF. L'intégration des variantes lexicales et sémantiques francophones semble dépendre fortement du dynamisme, de la composition et de l'ouverture de la communauté de contributeurs. Si le DDF compense en général l'imprécision de certains marquages diatopiques collaboratifs par l'intégration de sources professionnelles à côté du *Wiktionnaire*, c'est moins le cas pour les néologismes du Covid-19 qui ne sont que peu intégrés dans les autres sources du DDF, à l'exception du *Grand dictionnaire terminologique*. Par ailleurs, et de façon surprenante, le dictionnaire en ligne *Larousse* n'est pas plus exempt de prises de position idéologiques que le *Wiktionnaire*, dont le mode de fonctionnement devrait permettre des débats internes. Ces débats semblent malheureusement assez limités quant à la neutralité des définitions et leur dimension francophone.

Nous constatons d'importantes inégalités entre les descriptions retenues dans le corpus, avec chacune leurs forces et leurs faiblesses qui ne se situent pas nécessairement au même endroit (variantes graphiques, nomenclature, lexèmes reliés, etc.). Il semble malheureusement qu'à l'heure actuelle, aucune des descriptions étudiées ne soit tout à fait satisfaisante.

Annexe : Traitement lexicographique de néologismes du Covid

La liste ci-dessous contient 110 néologismes lexicaux ou sémantiques liés à la pandémie de Covid-19, et leur traitement dans les quatre dictionnaires de notre corpus. Si le lexème ou l'expression polylexicale n'est pas présente dans la nomenclature d'un des dictionnaires, cette absence est indiquée par une « / ». Si le lexème est inclus dans la nomenclature, avec le sens dans lequel il était employé pendant la pandémie, avec ou sans référence explicite au Covid-19 dans la définition, ce traitement est signalé par « ++ ». Il arrive qu'un sens nouveau lié à la pandémie ne soit pas décrit dans une entrée existante ou que la définition donnée ne renvoie pas précisément au contexte pandémique, le lexème est alors codé « +- ». Dans ce cas, il arrive que les exemples cités dans l'article situent plus clairement le lexème dans le contexte de la pandémie. Nous indiquons alors « +- EX. ». Quand une expression figée ne bénéficie pas d'une entrée séparée, mais est traitée sous l'un de ces constituants, nous indiquons lequel après « ss » pour *sous* (cette marque couvre également les renvois directs). Dans le *Wiktionnaire*, il arrive fréquemment que sous une entrée donnée soit indiqué un renvoi dans l'article dictionnaire mais le renvoi ne bénéficie pas encore d'un article. Dans ce cas, nous indiquons : « -* ss «entrée» ».

	<i>Wiktionnaire</i>	<i>DDF</i>	<i>DLR</i>	<i>DFL</i>
anticonfinement	+ - EX.	+ - EX.	/	/
anti-Covid	++	++	/	/
antidéconfinement	-* ss «déconfinement»	/	/	/
antivax	++	++	ss «antivaccin»	++
apéro virtuel	++	++	/	/
après-confinement	-* ss «confinement»	/	/	/
après-déconfinement	-* ss «déconfinement»	/	/	/
attestation de déplacement dérogatoire	/	/	EX. ss «déplacement» et ss «dérogatoire»	ss «dérogatoire»
autoconfinement	+ - EX.	+ - EX.	/	/
auto-confinement	+ - EX.	+ - EX.	/	/
autoconfiner	+ - EX.	+ - EX.	/	/
auto-confiner	+ - EX.	+ - EX.	/	/
c'est pas très Covid	++	++	/	/
cas contact	++	++	ss «contact»	ss «contact»
cas-contact	ss «cas contact»	ss «cas contact»	/	/

cluster	++	++	++	++
comportements barrières/barrière, comportement barrière	-* ss «mesure barrière»	/	/	/
confinable	+-	+-	/	/
confinement	++	++	+ EX.	++
confinement sanitaire	/	/	/	ss «confinement»
confiner	+-	+-	+ EX.	++
corona	++	++	ss «coronavirus»	/
Corona bond	/	/	/	/
coronabond	++	++	/	/
coronapéro	++	++	/	/
coronapiste	++	++	/	++
coronasceptique	++	++	/	/
coronaviral	++	++	/	/
coronavirus	++	++	++	++
coronologue	++	++	/	/
couvre-feu sanitaire	/	/	/	ss «couvre-feu»
covictature	-* ss «covid»	/	/	/
Covid	++	++	ss «covid»	ss «COVID-19»
covid	++	++	++	ss «COVID-19»
Covid 19	/	/	ss «covid»	ss «COVID-19»
covid long	++	++	ss «covid»	ss «COVID-19»
covid longue	++	++	/	/
covid manager	++	++	/	/
covid party	++	++	/	/
COVID-19	++	++	ss «covid»	++
Covid-19	++	++	ss «covid»	ss «COVID-19»
covid-19	++	++	ss «covid»	ss «COVID-19»
covidé	++	++	++	++
covidéprimé	++	++	/	/
covidéprimer	++	++	/	/
covidien	++	++	/	/
covidiot	++	++	/	/
covidote	++	++	/	/
covidiotie	++	++	/	/

covidisme	++	++	/	/
Covidistan	-* ss «covid»	/	/	/
covidiste	++	++	/	/
covid-positif	++	++	/	/
crise sanitaire	++	++	ss «crise»	/
déconfinable	-* ss «confinable»	/	/	/
déconfinement	+ - EX.	+ - EX.	+ - EX.	++
déconfiner	++	++	+ - EX.	++
décovider	++	++	/	/
distanciation physique	++	++	/	ss «distanciation»
distanciation sociale	++	++	ss «distanciation»	ss «distanciation»
état d'urgence sanitaire	++	++	EX. ss «urgence»	ss «urgence»
gestes barrières/ barrière, geste barrière	++	++	ss «barrière»	ss «barrière»
Grand Confinement	++	++	/	/
inconfinable	+ - EX.	+ - EX.	/	/
jauge	+ -	+ -	++	++
kung flu	++	++	/	/
kung-flu	-* ss «kung flu»	/	/	/
maladie à coronavirus 2019	++	++	/	ss «COVID-19»
masque sanitaire	++	++	/	/
mélancovid	++	++	/	/
mesures barrières, mesure barrière, mesures barrière	++	++	/	ss «barrière»
monde d'après	++	++	/	ss «monde»
monde d'avant	+ -	/	/	ss «monde», par opposition à monde d'après
nébulisation	+ -	+ -	+ -	++
nouveau coronavirus	-* ss «coronavirus»	/	/	/
pass sanitaire	++	++	ss «passe», EX. ss «sanitaire»	/
passé sanitaire	++	++	ss «passe»	ss «passe»
passé vaccinal	+ - EX.	+ - EX.	ss «passe»	ss «passe»

passport sanitaire	++	++	/	/
passport vaccinal	++	++	/	/
pic épidémique	/	/	ss «pic» Dér.	ss «pic»
PIMS	++	++	/	++
PIMS-Covid	++	++	/	/
plage dynamique	/	/	/	ss «dynamique»
plateau épidémique	/	/	/	ss «plateau»
pneumonie chinoise	++	++	/	/
pneumonie de Wuhan	++	++	/	/
politique zéro Covid	++	++	/	/
post-confinement	+ EX.	+ EX.	EX. ss «post»	/
pré-Covid	++	++	/	/
prélèvement nasopharyngé	++	++	/	ss «nasopharyngé»
premier de tranchée	++	++	/	/
préconfinement	-* ss «pré-reconfinement»	/	/	/
pré-reconfinement	+ EX.	+ EX.	/	/
quasi-confinement	-* ss «confinement»	/	/	/
quatorzaine	++	++	/	++
reconfinement	+ EX.	+ EX.	+ EX.	++
reconfiner	+ EX.	+ EX.	+ EX.	++
redéconfinement	+ EX.	+ EX.	/	/
rereconfinement	+ -	+ -	/	/
SARS-CoV-2	++	++	/	++
se confiner	+ -	+ -	++	+ EX.
semi-confinement	-* ss «confinement»	/	/	/
SRAS-CoV-2	++	++	/	/
sujet contact	-* ss «cas contact»	/	/	/
surconfinement	-* ss «confinement»	/	/	/
syndrome post-COVID	++	++	/	/
traçage	+ -	+ -	+ -	++
vaccinodrome	++	++	++	++
vie d'avant	/	/	/	/

Bibliographie

- AKUT, Katherine B. (2020), « Morphological Analysis of the Neologisms during the COVID-19 Pandemic », *International Journal of English Language Studies*, vol. 2, n° 3, p. 1–7. [en ligne : <https://al-kindipublisher.com/index.php/ijels/article/view/90/87>]
- AL-AZZAWI, Qasim Obayes et Haneen Ali Haleem (2021), « "Do you Speak Corona?": Hashtags and Neologisms since the COVID-19 Pandemic Outbreak », *International Journal of Linguistics, Literature and Translation*, vol. 4, n° 4, p. 113–122. [en ligne : <https://al-kindipublisher.com/index.php/ijllt/article/view/1544/1350>]
- ALTMANOVA Jana, Michela Murano et Chiara Preite (2022), « Le lexique de la pandémie et ses variantes », *Le lexique de la pandémie et ses variantes, Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/jana-altmanova-michela-murano-chiara-preite-le-lexique-de-la-pandemie-et-ses-variantes/>]
- ASIF, Muhammad, Deng Zhiyong, Anila Iram et Maria Nisar (2021), « Linguistic Analysis of Neologism Related to Coronavirus (COVID-19) », *Social Sciences & Humanities Open*, vol. 4, n° 1, p. 1-6. [en ligne : <https://doi.org/10.1016/j.ssaho.2021.100201>]
- AUROY, Olivier (2020), *Le Dicorona*, Paris, Éditions Intervalles.
- COPPENS, Bruno et Pierre Kroll (2021), *Le Dicovid : D'asymptomatique à Zoonotique*, Gerpinnes, Kennes.
- Dico en ligne Le Robert*. [en ligne : <https://www.lerobert.com/>]
- Dictionnaire de français Larousse*. [en ligne : <https://www.larousse.fr/dictionnaires/francais-monolingue>]
- DINCĂ, Daniela (2022), « Lexiqueroumain de la pandémie dans la communication institutionnelle », *Le lexique de la pandémie et ses variantes, Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/daniela-dinca-le-lexique-roumain-de-la-pandemie-de-covid-19-dans-la-communication-institutionnelle/>]
- DOW Michael et Patrick Drouin (2023), « Tracing the Evolution of the Gender of "COVID-19" in the French of Three Continents: A Traditional and Social Media Study », *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, vol. 68, n° 3, p. 486-513.
- DUHAMEL, Olivier et Laurent Bigorgne (2020), *Les mots du coronavirus*, Paris, Dalloz.
- FILLMORE, Charles (1982), « Frame Semantics », *Linguistics in the Morning Calm*, p. 111-137.
- GRIMALDI, Claudio (2022), « Les traits de la distance et de l'isolement dans le lexique autour de la pandémie », *Le lexique de la pandémie et ses variantes, Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/claudio-grimaldi-les-traits-de-la-distance-et-de-lisolement-dans-le-lexique-autour-de-la-pandemie/>]
- GUO, Weiwei, Sonia Berbinski et Corina Veleanu (2022), « La création lexicale de la pandémie, entre peur et humour », *Le lexique de la pandémie et ses variantes, Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/weiwei-guo-sonia-berbinski-corina-veleanu-la-creation-lexicale-de-la-pandemie-entre-peur-et-humour/>]

- JACQUET-PFAU, Christine et Alicja Kacprzak (2022), « De quelques mots-témoins d'une pandémie : les représentations du Covid 19 en français et en polonais », *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/christine-jacquet-pfau-alicja-kacprzak-de-quelques-mots-temoins-dune-pandemie-les-representations-du-covid-19-en-francais-et-en-polo-nais/>]
- KLOSA-KÜCKELHAUS, Anette (2022), « German Corona-Related Neologisms and their Lexicographic Representation », dans Anette Klosa-Kückelhaus et Ilan Kernerman (dir.), *Lexicography of Coronavirus-related Neologisms*, Berlin/Boston, De Gruyter, p. 27-42.
- KOLAŘIKOVA, Dagmar (2021), « Reflet de la pandémie de Covid-19 dans les dictionnaires de la langue française », *Studia Romanistica*, vol. 21, n° 2. [en ligne : https://dokumenty.osu.cz/ff/journals/studiaromanistica/21-2/SR_21_2_full.pdf#page=30]
- LARDELLIER, Pascal (2022), *La bonne distance ? Petite anthropologie d'une crise sanitaire*, Paris, MkF Éditions.
- Le Dicovid des mots inventés*. [en ligne : <https://dictionnaire.lerobert.com/dis-moi-robert/raconte-moi-robert/mot-annee/le-dicovid-des-mots-inventes.html>]
- Le Dictionnaire des francophones*. [en ligne : <https://www.dictionnairedesfrancophones.org/>]
- Le Wiktionnaire*. [en ligne : <https://fr.wiktionary.org>]
- LIGAS, Pierluigi (2010), « Renvois et circularité dans les définitions des dictionnaires spécialisés. Le cas du DAAFAPS », *Publifarum*, n° 11. [en ligne : <https://riviste.unige.it/index.php/publifarum/article/view/1550/1699>]
- MALDUSSI, Danio (2022), « De nouvelles dénominations pour un concept ancien : le rôle de l'adjectif qualificatif, de l'adjectif relationnel et du substantif épithète dans les processus d'innovation néologique en temps de pandémie », *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/danio-maldussi-de-nouvelles-denominations-pour-un-concept-ancien-le-role-de-ladjectif-qualificatif-de-ladjectif-relationnel-et-du-substantif-epithete-dans-les-processus-d/>]
- MARTÍ SOLANO, Ramón (2022), « Un regard contrastif sur des unités phraséologiques liées à la pandémie du coronavirus et sur leurs actualisations lexico-syntaxiques », *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/ramon-marti-solano-un-regard-contrastif-sur-des-unites-phraseologiques-liees-a-la-pandemie-du-coronavirus-et-sur-leurs-actualisations-lexico-syntaxiques/>]
- MWERI, Jefwa (2021), « Corona Virus Disease (COVID-19) Effects on Language Use: An Analysis of Neologisms », *Linguistics and Literature Studies*, vol. 9, n° 1, p. 36-47. [en ligne : https://www.researchgate.net/profile/Jefwa-Mweri/publication/349038339_Corona_Virus_Disease_COVID-19_Effects_on_Language_Use_An_Analysis_of_Neologisms/links/601bf84a92851c4ed54985cd/Corona-Virus-Disease-COVID-19-Effects-on-Language-Use-An-Analysis-of-Neologisms.pdf]
- PENNEC, Blandine (2021), *Les mots de la Covid-19 : Étude linguistique d'un corpus français et britannique*, Arras, Artois Presses Université.

- PRUVOST, Jean et Jean-François Sablayrolles (2003), *Les néologismes. Que sais-je ?*, Paris, PUF.
- ROLLO, Alessandra (2022), « Métaphores et "covidismes" aux temps de la Covid-19 ». *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/alessandra-rollo-metaphores-et-covidismes-dans-les-discours-du-president-emmanuel-macron-aux-temps-de-la-covid-19/>]
- Sablayrolles, Jean-François (2018), « Néologie, néonymie et dictionnaires », dans Agnieszka Konowska, Agnieszka Woch, Andrzej Napieralski et Anna Bobinska (dir.), *Le poids des mots. Hommage à Alicja Kacprzak*, Łódź, WUŁ, p. 239-251.
- SAJOUS, Franck, Nabil Hathout et Amélie Josselin-Leray (2019), « Du vin et devin dans le Wiktionnaire : neutralité de point de vue ou neutralité et point de vue ? », *Études de linguistique appliquée*, n°194(2), p. 147-164.
- SAJOUS, Franck (2022), « Using Wiktionary Revision History to Uncover Lexical Innovations Related to Topical Events: Application to Covid-19 Neologisms » dans Anette Klosa-Kückelhaus et Ilan Kernerman (dir.), *Lexicography of Coronavirus-related Neologisms*, Berlin/Boston, De Gruyter, p. 275-306.
- SAJOUS, Franck et John Humbley (2022), « Mesures d'isolement sanitaire dans *Wiktionnaire* et *Wikipédia* : néologie et lexicographie ou néonymie et terminographie ? », *Estudios Románicos*, n° 31. [en ligne : <https://revistas.um.es/estudiosromanicos/article/view/510631/324221>]
- STEFFENS, Marie (2022), « Pandémie, polémique et variation : le ou la Covid ? », *Circula*, n°15, p. 229-250. [en ligne : https://circula.recherche.usherbrooke.ca/wp-content/uploads/2023/01/2022_Circula_15_011_Steffens.pdf]
- STOICHITOIU ICHIM, Adriana (2022), « Hospitalité versus créativité dans le vocabulaire roumain de la pandémie », *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/adriana-stoichituiu-ichim-hospitalite-versus-creativite-dans-le-vocabulaire-roumain-de-la-pandemie/>]
- TALLARICO, Giovanni (2022), « Néologismes expressifs et ludiques dans le vocabulaire de la pandémie », *Le lexique de la pandémie et ses variantes*, *Repères DoRiF*, n° 25, juillet. [en ligne : <https://www.dorif.it/reperes/giovanni-tallarico-neologismes-expressifs-et-ludiques-dans-le-vocabulaire-de-la-pandemie/>]
- VINCENT, Nadine (2022), « Faut-il adapter les dictionnaires à l'air du temps? Proposition d'un traitement polyphonique du mot *woke* », *Circula*, n°15, p. 122-145. [en ligne : https://circula.recherche.usherbrooke.ca/wp-content/uploads/2023/01/2022_Circula_15_007_Vincent.pdf]
- ZOTTI, Valeria (2014), « Les renvois analogiques du Petit Robert : un système sémiotique complexe », dans Michaela Heinz (dir.), *Les sémiotiques du dictionnaire*, Berlin, Frank & Timme, p. 133-161.

TITRE: DICTIONNAIRES ET CORPUS NUMÉRIQUES POUR L'ENSEIGNEMENT DU FLE :
UN PARCOURS DIDACTIQUE SUR LE LEXIQUE

AUTEUR: FRANCESCO FARESin, UNIVERSITÀ DI VÉRONA ET UNIVERSITÉ GRENOBLE-ALPES

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES
PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT,
UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 58-82

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22395](http://hdl.handle.net/11143/22395)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22395](https://doi.org/10.17118/11143/22395)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS
ATTRIBUTION 4.0 INTERNATIONAL.

Dictionnaires et corpus numériques pour l'enseignement du FLE : un parcours didactique sur le lexique

Francesco Faresin, Università di Verona et Université Grenoble-Alpes

Résumé :

Cet article se concentre sur l'exploration du rôle des dictionnaires collaboratifs (*Wiktionnaire*, *Dictionnaire des francophones* et *Bob*) et des bases de traitement de corpus (*Lexicoscope* et *Compleat Lexical Tutor*) dans l'enseignement du FLE, en mettant particulièrement l'accent sur l'enseignement des unités phraséologiques. À travers une série d'exercices et d'activités didactiques, l'article propose une démarche pédagogique pour intégrer efficacement ces ressources numériques dans la salle de classe de FLE, tout en développant les compétences linguistiques des apprenants. Enfin, l'article présente les résultats d'une évaluation auprès d'un groupe d'enseignants, mettant en lumière leurs perspectives sur l'intégration des dictionnaires collaboratifs dans l'enseignement de FLE. Ces résultats soulignent la nécessité d'une formation continue des enseignants et d'une planification pédagogique attentive pour tirer pleinement profit du potentiel de ces outils.

Mots-clés :

Lexicographie collaborative, corpus numériques, enseignement du FLE, lexique, unités phraséologiques

Abstract :

This article explores the role of collaborative dictionaries (such as *Wiktionnaire*, *Dictionnaire des francophones*, and *Bob*) and digital corpora in teaching French as a foreign language (FLE), with particular emphasis on teaching phraseological units. Through a series of exercises and activities, the article proposes a pedagogical approach to effectively integrate these digital resources into the FLE classroom while developing learners' linguistic skills. Finally, the article presents the results of an evaluation among a group of teachers, shedding light on their perspectives on the integration of collaborative dictionaries and digital corpora in FSL teaching. These results highlight the need for ongoing teacher training and careful pedagogical planning to fully harness the potential of these tools.

Keywords :

Collaborative Lexicography, digital corpora, teaching FLE, vocabulary, phraseological units

1. Introduction

Tout au long des siècles, le statut du dictionnaire a évolué en raison des contextes sociaux et des modifications qui se sont produites au sein du public. Ces changements ne l'ont pourtant pas empêché de rester un point de repère pour les locuteurs et d'être le « reflet de l'histoire d'une langue » (Dubois et Dubois, 1971 : 99) et de sa culture.

Les ressources de la lexicographie collaborative se veulent une alternative capable de bien interpréter les évolutions du monde d'aujourd'hui. En effet, elles représentent de manière concrète le passage du web statique au web dynamique, qui bénéficie des interventions directes des utilisateurs (Dolar, 2018 : 34). De fait, comme le souligne Dolar, cette nouvelle frontière de la lexicographie, dite aussi *profane* (ou *populaire*), se configure comme un « observatoire de la langue qui se caractérise par le temps bref de mise à jour et par le nombre élevé de contributeurs » (Dolar, 2017 : 2). Du coup, plutôt que de s'appuyer exclusivement sur l'avis de linguistes experts, la lexicographie profane bénéficie également des contributions de nombreux utilisateurs qui enrichissent et diversifient le contenu. Ce processus collectif fait en sorte que le dictionnaire ne soit pas seulement une source d'autorité linguistique, mais aussi un reflet vivant et dynamique de la façon dont la langue est réellement employée dans les différentes communautés et couches sociales des locuteurs. Effectivement, les usagers, voire potentiels contributeurs – souvent des non-linguistes – par leur interaction « dans un espace virtuel [...] de construction des savoirs sur la langue » (Murano, 2014 : 149), apportent leurs propres expériences et perspectives sur la langue, aidant à créer un métalangage familier et accessible. Ce métalangage de nature plus informelle est basé sur les besoins quotidiens de communication et donc ressort d'une nécessité pratique de communication des utilisateurs.

En outre, par rapport à des ressources lexicographiques professionnelles en ligne (DAF, *Usito*, *Dico en ligne*...), les pages des dictionnaires collaboratifs sont souvent conçues pour être simples et directes : elles affichent souvent des expressions ou des locutions en entrée, contrairement aux ressources professionnelles. Cet aspect permet à l'utilisateur de trouver les informations clés en un coup d'œil, sans être submergé par des détails marginaux et périphériques. À la lumière de cela, l'un des aspects les plus saillants des dictionnaires collaboratifs concerne sans doute une certaine rapidité de consultation qui, d'un côté, semble attirer les apprenants mais, de l'autre, comporte une baisse de capacité d'évaluation critique. Cela notamment parce qu'ils s'accoutument seulement à « googler » les mots inconnus, s'arrêtant à la première définition fournie (Molinari, 2021 : 28). Il est essentiel de souligner, à ce propos, que ces ressources peuvent devenir des instruments très utiles surtout si elles sont utilisées de manière complémentaire avec des ouvrages traditionnels, qui restent d'importance capitale pour permettre l'accès à un riche patrimoine linguistique et culturel (Calvi et Dankova, 2023 : 17).

De plus, le passage au numérique, dans le contexte didactique, peut aussi amener à un changement de paradigme qui conduit les apprenants à acquérir une attitude plus active et directe face au processus d'apprentissage (Calvi et Dankova, 2023 : 31). Cela est vrai en particulier puisque les atouts

technologiques peuvent donner la possibilité à une classe de travailler sur plusieurs compétences à la fois, prenant en compte ainsi de manière plus concrète la « multicanalité » de la communication réelle et mettant en lumière la dimension socioculturelle de la conversation orale. Cet aspect est évoqué déjà par Cuq et Gruca (2002 : 423-425), par exemple, lorsqu'ils font référence à la notion d'*authenticité*. En effet, les outils de lexicographie collaborative interviennent de façon à favoriser une meilleure maîtrise des compétences de recherche, de révision et de communication, tout en permettant en même temps d'explorer les nuances socioculturelles qui peuvent émerger dans les définitions des mots et des expressions fournies.

En outre, dans un monde peuplé par plus de 300 millions de francophones, il est important que le numérique, y compris les dictionnaires collaboratifs, soit en mesure de valoriser les enjeux de la variation linguistique, évitant de présenter le français standard parisien comme seule source valable (Steffens, 2021 : 12). De manière générale, cet objectif est d'ailleurs rappelé aussi dans le *Cadre européen commun de référence pour les langues* lorsqu'on fait référence à la nécessité de la part des enseignants et des formateurs de créer un espace pluriculturel :

Il [le formateur/médiateur] cherche à améliorer et approfondir la compréhension interculturelle entre les participants afin de prévenir et/ou de surmonter d'éventuelles difficultés dues à des points de vue culturels contrastés. Le médiateur doit être constamment conscient des différences socioculturelles et sociolinguistiques qui affectent la communication interculturelle. (CECR, 2020 : 120)

L'objectif principal de cette contribution est de décrire quelques-uns de ces outils numériques (*Dictionnaire des francophones*¹, *Bob – l'autre trésor de la langue*², *Wiktionnaire – le dictionnaire libre*³), en s'interrogeant sur leur rôle en tant qu'instruments didactiques dans l'enseignement du FLE. Nous voulons montrer comment ces ressources, choisies en raison de leur accessibilité, de leur popularité et de leur capacité à refléter la diversité linguistique et culturelle du français, peuvent intervenir dans un contexte pédagogique, en emmenant des avantages tant du côté de l'enseignant que de celui de l'apprenant. À titre d'exemple, nous pouvons mentionner le fait que les plateformes collaboratives offrent des possibilités d'interaction, telles que des échanges et des commentaires dans des forums. Ces espaces peuvent être exploités par les enseignants, incitant les apprenants à participer activement à des discussions linguistiques, renforçant ainsi leur motivation et leur engagement.

Dans une première phase, nous allons répertorier, par le biais de ces outils, des exemples d'unités phraséologiques tirées du champ sémantique des animaux (à partir des lexies *chien*, *vache* et *cheval*).

1. Il faut préciser que par rapport au *Wiktionnaire* et à *Bob*, le *Dictionnaire des francophones* n'est pas entièrement une ressource collaborative : il présente un volet participatif qui permet aux utilisateurs de contribuer, mais les définitions fournies y sont majoritairement retravaillées par des linguistes ; les définitions de la majeure partie du DDF sont reprises d'autres ressources lexicographiques existantes (lien : <https://www.dictionnairedesfrancophones.org/>, page consultée le 26 août 2024).

2. <https://www.languefrancaise.net/Bob/Introduction> (page consultée le 18 mars 2024).

3. https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil (page consultée le 19 mars 2024).

Ce dernier, en effet, est parmi les plus productifs en ce qui concerne la création d'unités phraséologiques en langue française, en raison de sa prégnance dans la vie quotidienne et dans les représentations mentales et culturelles associées, partagées parmi les locuteurs. À ce propos, comme on le sait à partir d'études précédentes, le manque d'une méthode explicite d'enseignement du lexique et des unités phraséologiques, plus spécifiquement, demeure une faille importante (Frassi, 2018). Voilà pourquoi, lors des séances d'enseignement dédiées aux unités phraséologiques, l'enseignant devrait considérer « des mots concrets répondant à des objets que l'élève a sous les yeux, ou du moins qu'il ait vus et qu'il puisse aisément remplacer devant son imagination » (Puren, 2005 : 43). C'est pourquoi le choix des lexies tirées du domaine des animaux s'avère particulièrement judicieux, car il permet aux élèves de se connecter facilement avec les mots et les expressions, facilitant ainsi leur mémorisation et leur compréhension.

Pour les besoins de cet article, nous allons considérer la notion de phraséologie, ainsi que définie par Gonzalez-Rey en 2008 et par Bolly en 2011 :

Placée à mi-chemin entre le lexique et la syntaxe, la phraséologie s'occupe des signes polylexicaux composés d'au moins deux mots, stables, répétés et souvent figurés. (Gonzalez-Rey, 2008 : 5)

La phraséologie au sens strict a pour objet d'étude des *unités phraséologiques* (UP), i.e. toute unité polylexicale constituée de deux ou plusieurs mots graphiques catégoriellement liés [...] qui co-apparaissent de manière préférentielle dans l'usage (Bolly, 2011 : 23).

Ces définitions évoquent la notion de *figement* qui a été largement débattue pendant le vingtième siècle à partir de Charles Bally (1909) jusqu'à Ivan Fonagy, en passant par Louis Guilbert. Puisque le figement n'est pas le thème central de cette réflexion, nous allons nous contenter d'une définition assez générale qui considère une *expression figée* comme « une unité phraséologique constituée de plusieurs mots, contigus ou non, qui présentent un certain degré de figement sémantique, un certain degré de figement lexical et un certain degré de fixité morphosyntaxique » (Lamiroy, 2008 : 13). En d'autres mots, lorsque nous parlons d'*expression figée*, nous nous référons à une structure linguistique souvent immuable sur le plan syntaxique et dont la signification n'est pas déductible uniquement à partir du sens des mots qui la composent.

Par la suite, nous nous proposons de confronter les résultats de la recherche préliminaire à la consultation des corpus numériques, qui peuvent être exploités de manière efficace en synergie avec d'autres instruments, y compris les dictionnaires, dans un contexte didactique autour du lexique (Bonadonna, 2023). En particulier, nous allons nous appuyer sur les corpus utilisés au sein du projet *Compleat Lexical Tutor*⁴, développés par Tom Cobb de l'Université de Montréal (Québec), ou ceux consultables à partir de la ressource *Lexicoscope*⁵ (Université Grenoble-Alpes). Cela va être utile pour réfléchir sur les contextes où les unités phraséologiques recherchées sont attestées. La dernière

4. <https://www.lextutor.ca/conc/fr/> (page consultée le 19 mars 2024).

5. <http://phraseotext.univ-grenoble-alpes.fr/lexicoscope/> (page consultée le 19 mars 2024).

partie de cette contribution va être consacrée à la présentation d'un parcours didactique autour de l'apprentissage des unités phraséologiques préalablement analysées. Ce parcours, pensé pour une classe d'apprenants italophones de 12-14 ans ayant un niveau moyen de français (B1), va être soumis à l'évaluation d'un échantillon de cinq enseignantes de FLE de l'école secondaire. En ciblant précisément ce genre de public, ce module se veut un outil flexible et adapté, capable d'un côté d'inspirer les enseignants à la création de matériaux didactiques authentiques et, de l'autre, de susciter l'intérêt des élèves tout en favorisant un apprentissage efficace et engageant. Le parcours pédagogique qui sera présenté dans cet article a été élaboré en tenant en compte que des élèves de français de niveau B1 sont en mesure de « suivre relativement bien les conversations et les discussions de tous les jours » et « exprimer, à l'aide des mots simples, leur opinion sur des sujets de la vie courante » (CECR, 2001 : 51 ; 68).

2. Les ressources de lexicographie numérique populaire

2.1. Le Wiktionnaire : le dictionnaire libre

Né en 2004, le *Wiktionnaire* constitue le « compagnon lexical » (Sajous, Hathout et Calderone, 2014 : 663) de *Wikipédia*. Dans la page de présentation du projet, il nous est présenté comme ayant un objectif purement descriptif et ne jugeant pas « la valeur des mots » (*Wiktionnaire : à propos*⁶). Il s'agit donc d'un « produit d'un apport communautaire non règlementé » où la volonté centrale est celle de « légitimer les compétences personnelles au détriment des représentations légitimées par l'histoire » (Molinari et Vicari, 2023 : 6) et où la véridicité est confiée à la négociation entre contributeurs. Il est structuré de manière plutôt similaire aux dictionnaires traditionnels, au moins pour les informations qu'il fournit dans ses articles. Aux données définitives, étymologiques et phonétiques, s'ajoute une large section consacrée aux dérivés morphologiques et à une liste choisie de locutions et collocations où le mot recherché est utilisé. Cette partie spécifique est particulièrement intéressante dans un contexte didactique puisque les apprenants seraient en mesure de prendre conscience de l'extraordinaire productivité qui caractérise la langue française dans son ensemble. Par exemple, en recherchant la lexie *vache*, qui va nous servir de base pour l'élaboration de la proposition didactique dont il est question dans cet article, nous avons accès à une large variété de locutions nominales et d'expressions phraséologiques. Le choix de *vache* comme l'une des unités lexicales de base pour cette unité didactique est doublement motivé : d'un côté, elle se trouve à la source de nombreuses expressions idiomatiques et phraséologiques couramment utilisées en français comme *période de vaches maigres* ou *pleuvoir comme vache qui pisse*; de l'autre, il s'agit d'un animal familier pour la majorité des élèves, qu'ils vivent en milieu urbain ou rural.

6. *Wiktionnaire*, page de présentation du projet (lien : https://fr.wiktionary.org/wiki/Wiktionnaire:%C3%80_propos, (page consultée le 19 mars 2024).

En naviguant sur le site, une autre section qui illustre très bien la richesse linguistique du français est le *Portail de la langue française et de la francophonie*⁷ qui donne aux apprenants un aperçu sur plusieurs aspects relevant de la langue française dans le monde. En particulier, la section « Régionalismes » permet une réflexion approfondie en matière de variation diatopique dans l'espace francophone (Molinari, 2021 : 39-40) et explore les particularités du français québécois, belge, suisse ou africain. Un exemple pourrait être l'expression phraséologique québécoise *le diable est aux vaches*, utilisée pour décrire une situation complètement chaotique. Dans la section consacrée aux renvois étymologiques de cette expression, l'on précise, en outre, qu'elle était autrefois utilisée « lorsqu'un fermier constatait que ses bêtes s'agitaient lorsque le temps devenait orageux⁸ ». En raison de ce que nous venons de montrer, il est possible d'affirmer que le *Wiktionnaire* s'avère un instrument largement exploitable en classe de FLE, surtout pour ce qui concerne l'apprentissage du lexique et la mise en valeur de la variation linguistique.



Figure 1 Page d'accueil du « Portail de la langue française et de la francophonie » dans le *Wiktionnaire*⁹

7. <https://fr.wiktionary.org/wiki/Portail:Fran%C3%A7ais> (page consultée le 20 mars 2024)

8. https://fr.wiktionary.org/wiki/le_diable_est_aux_vaches (page consultée le 23 juillet 2024).

9. <https://fr.wiktionary.org/wiki/Portail:Fran%C3%A7ais> (page consultée le 20 mars 2024).

2.2. Le Dictionnaire des francophones

Le projet lexicographique numérique du *Dictionnaire des francophones* est né sous l'exigence soutenue par le ministère français de la Culture de faire rayonner la langue française et de faire de sorte qu'elle s'ouvre au monde, devenant une véritable « langue monde » (Vincent, 2019). Plus particulièrement, cette ressource se configure comme une plateforme élaborée à partir de l'intégration de plusieurs sources parmi lesquelles, par exemple, le *Wiktionnaire*, la *Base de données lexicographiques panfrancophones*¹⁰ et le *Dictionnaire des synonymes, des mots et des expressions des français parlés dans le monde*¹¹. Toutefois, il ne s'agit pas exclusivement d'un assemblage de ressources à consulter, mais plutôt d'un outil gratuit, pensé pour qu'il « ne soit pas [exclusivement] franco-français » (Boutard, 2021), qui offre à chacun la possibilité de « donner sa vision de sa langue quotidienne » (Vincent, 2019). Il s'enracine « dans le monde francophone pluricentrique, où toutes les variétés sont incluses et traitées comme égales » (Dolar, Gasparini et Steffens, 2023 : 8) et il se compose d'un contenu varié, bénéficiant des « apports érudits comme populaires » (Gasparini, Sefiane et Gathier, 2021 : 13). Dans une démarche de partage collaboratif des connaissances, l'intégration des contributions d'utilisateurs non experts est reconnue comme l'une des tâches essentielles qui anime le projet. Pour faire en sorte que cela soit possible et qu'il permette une réelle représentation des pratiques linguistiques populaires, des forums et des espaces de discussions sont disponibles (*Ibid.* : 10). Enfin, pour répondre aux divers soucis de fiabilité et de cohérence des contributions, un système de contrôle est prévu : un Conseil scientifique, dont les membres sont recrutés au sein de réseaux de spécialistes, et un Comité de relecture le chapeautent, en assurant que l'ajout de mots soit fait de la manière la plus transparente possible et répondant à la réalité le mieux possible (Sefiane, 2021).

Actuellement le *Dictionnaire des francophones* compte un total de 2 476 189 entrées, mais il est toujours en expansion et, selon les prévisions des professionnels qui sont en train d'y travailler, l'ajout de cartes des aires d'usage et des fichiers de prononciation audio sont envisagés (*Ibid.*). Ces avancements permettraient de répondre de manière adéquate aux soucis liés à la possibilité d'accès « à une visualisation claire [...] qui intègre la dimension dynamique des réseaux lexicaux » (Steffens, 2021 : 17). À ce propos, une fonction qui est déjà présente dans le DDF concerne la possibilité de consulter des sections dénommées « mot ou expression dérivée » et « proverbe ou phrase toute faite », où il est possible de retrouver des expressions linguistiques qui impliquent l'usage du mot préalablement recherché.

Sur la base de ce que nous venons d'affirmer, nous pouvons considérer le DDF comme un instrument numérique extrêmement valable pour des exploitations didactiques : en effet, il confronte les apprenants à la richesse lexicale de la langue française en valorisant la dimension variationnelle et socioculturelle qui caractérise l'ensemble des pays francophones. Ce dernier aspect sera d'ailleurs

10. <https://www.bdlp.org> (page consultée le 28 août 2024).

11. <http://www.dictionnaire-synonymes-francophones.fr> (page consultée le 28 août 2024).

partiellement présent aussi dans notre proposition, vu que nous allons travailler également sur des expressions phraséologiques d'origine québécoise, comme « avoir la chienne ». À titre d'exemple, à partir de la recherche de la lexie *chien* dans le DDF, il est possible de consulter 28 définitions différentes et des proverbes comme « bon chien n'aboie pas pour rien », « chien qui aboie ne mord pas » ou « bon chien, chasse de race ».

Terme recherché : chien

28 Définitions

<p>Aube, Champagne Source : ABC d...</p> <p>nom, masculin</p> <p>Repas célébrant la fin des vendanges.</p> <p>[...]</p>	<p>Champagne, Lorraine Source : Wikitia...</p> <p>nom, masculin</p> <p>Repas commun célébrant la fin de la période de travail (saison de martelage, vendanges, moissons...).</p> <p>[...]</p>
<p>Algérie Source : BDLP</p> <p>nom, masculin</p> <p>Terme insultant pour désigner un homme. (fiche originale)</p> <p>[...]</p>	<p>Québec Source : Wikitia...</p> <p>adjectif</p> <p>Injuste, méchant.</p> <p>[...]</p>
<p>Drôme Source : DROME</p> <p>nom, masculin</p> <p>Capitule (du grateron ou de la bardane).</p> <p>[...]</p>	<p>Drôme Source : DROME</p> <p>nom, masculin</p> <p>Cliquet (pièce métallique dont l'extrémité en sifflet vient se loger dans les trous aménagés sur le pignon d'entraînement dans un mouvement de va-et-vient vertical).</p> <p>[...]</p>

Figure 2 Définitions à partir de la recherche du mot *chien* dans le *Dictionnaire des francophones*

2.3. Bob – l'autre trésor de la langue

Par rapport aux autres ressources de lexicographie collaborative présentées jusqu'ici, *Bob* est un dictionnaire numérique consacré exclusivement à l'argot, c'est-à-dire à la langue populaire issue d'un registre familier. Cet élément est déjà présent dans le nom de la ressource qui marque une distinction nette par rapport aux ouvrages lexicographiques traditionnels. L'enseignement du français argotique et populaire, en contexte didactique, a été presque toujours négligé en faveur du français standard. Pourtant, la langue étant un fait social (Saussure, 2005 [1915] : 13), il est important de comprendre ses dimensions socioculturelles pour pouvoir s'adapter à toutes les situations de communication réelles. L'apprentissage de ce qui a été défini comme étant « un melting-pot de mots de tous les horizons » (Sun et Pu, 2018 : 134) apparaît donc essentiel pour maîtriser la langue de manière aussi complète que possible. Cela est indéniable si l'on considère surtout que, à la fonction expressive et identitaire

qui est attribuée à l'argot, s'ajoute le fait qu'il est « de plus en plus présent dans l'espace médiatique [...] francophone » (Goudaillier, 2015 : 208).

Le site qui héberge *Bob* peut être interrogé de plusieurs façons : par synonyme, par catégorie morphologique ou par domaine d'usage. La section « Usage » est à son tour organisée selon une série de thèmes comme « argot et armée », « argot de la police », « argot de la drogue », « argot et sexualité », etc. La recherche d'un mot donne une liste de lexèmes ou d'expressions phraséologiques où l'unité lexicale en examen est utilisée. Les entrées sont construites de façon assez intuitive : au-dessus de la définition choisie, il y a souvent des chiffres, allant de 1 (académique) à 10 (argotique), indiquant le degré de formalité/informalité de l'expression en question. La fréquence, en revanche, est signalée par une valeur numérique qui indique dans combien de sources l'élément défini a été relevé (Dolar, 2017 : 8). En outre, une grande importance est consacrée aux citations qui sont jugées essentielles « autant pour l'archéologie du mot (documenter ses origines) que pour en donner le mode d'emploi illustré¹² ». Dans la figure 3, la liste qui est proposée à partir de la lexie *cheval* nous fournit un exemple clair de la structuration de la nomenclature : nous y retrouvons des expressions telles que *changer son cheval borgne pour un aveugle* ou *c'est la mort du petit cheval*, qui signifient respectivement « un changement de situation de mal en pis » et « la fin de quelque chose ».

- - [c'est la mort du petit cheval](#) ^{grave} Pour exprimer la fin de qqch, d'une situation profitable
- - [c'est un cheval à l'écurie](#) C'est une chose qui coûte beaucoup d'entretien sans être d'aucune utilité
- - [c'est un vrai cheval](#) ^{travailleur} Se dit de qqun qui travaille beaucoup
- - [chameau](#) ^{HIPPISME} ^{cheval} Mauvais cheval, mot péj. contre cheval
- - [changer son cheval borgne pour un aveugle](#) ¹⁶³⁰ Changement de situation de mal en pis
- - [chargé à cul](#) ¹⁸⁸⁶ ^{ALCOOL} ^{ivre} Ivre, être très ivre ; ivre à ne plus pouvoir marcher
- - [chassebi](#) ¹⁸⁹⁷ ^{MILITAIRE} ^{soldat} (mil.) chasseur à pied, chasseur alpin ; (chasseur alpin ; chasseur à cheval ; chasseur d'Afrique)
- - [chasse-mar](#) ^{MILITAIRE} Chasseur d'Afrique ; chasseur à cheval
- - [chasse-marée](#) ^{MILITAIRE} Chasseurs d'Afrique
- - [cheval](#) ^{DROGUE} ^{héroïne} Héroïne ; (cocaïne)
- - [cheval](#) ^{grossier} Expression de mépris, individu ; individu grossier ; insulte
- - [cheval](#) ^{MILITAIRE} ^{mandat} Mandat, mandat-poste, bon de poste
- - [cheval](#) ^{femme laide} Grande femme
- - [cheval](#) ^{JOURNALISME} Se dit de l'article qui est à cheval sur la page 1 et la page 2
- - [cheval à double semelle](#) ¹⁸⁴⁴ ^{CORPS} ^{jambe} Jambe
- - [cheval à toute selle](#) ^{polyvalent} Homme à tout faire, qui n'est pas limité à un emploi, polyvalent
- - [cheval blanc](#) ^{montre} Montre d'argent (ou chose de valeur?)
- - [cheval blanc](#) Chose recherchée? situation désirée?
- - [cheval d'avoine, cheval de peine](#) Formule : un cheval bien nourri fait beaucoup de besogne
- - [cheval de bataille](#) Sujet auquel on tient, principe, sujet de prédilection, argument principal ; argumenter sur un sujet que l'on défend souvent
- - [cheval de bois](#) ^{HIPPISME} Cheval très docile?
- - [cheval de box](#) ^{MILITAIRE} ^{gradé} Sous-officier
- - [cheval de luxe](#) ¹⁹¹⁶ ^{MILITAIRE} ^{gradé} Officier ; militaires faisant un service agréable (bien nourris, bien habillés)
- - [cheval de retour](#) ¹⁸²⁸ ^{PRISON, POLICE, BAGNE} ^{forçat, récidiviste} ■ (crim.) Récidiviste ; forçat qui est à nouveau condamné, forçat recondamné ; forçat rattrapé, forçat évadé et repris ; criminel d'habitude plusieurs fois condamné, voleur, criminel, délinquant expérimenté, qui a de l'expérience ; > terme de mépris ; ■ qqun d'expérimenté, d'aguerri
- - [cheval de trompette](#) ¹⁶¹¹ ^{courage} Homme qui ne s'effraie pas aisément, ne se laisse pas intimider ; aguerri ; ne s'étonne, ne s'effrayer de rien

Figure 3 Résultats à partir de la lexie *cheval* dans *Bob*

12. <https://www.linguefrancaise.net/Aide/citations> (page consultée le 27 août 2024)

Dans un contexte pédagogique, l'utilisation de *Bob* permet aux apprenants d'explorer les variétés de la langue française au-delà de ses formes plus soutenues ou standardisées. *Bob*, donc, se présente comme une ressource que les enseignants pourraient exploiter pour faire en sorte que les apprenants entrent plus aisément en contact avec une langue familière et argotique qui est souvent présente dans les interactions orales.

Dans la figure 4, l'entrée correspondant à l'unité phraséologique *c'est un vrai cheval* est présentée : au-delà de la définition, nous pouvons trouver l'indication d'un synonyme (« travailleur ») et de la seule source où l'expression en question est attestée (« 1926 – *Locutions usuelles de la langue parlée, dans le français de tous les jours, deuxième partie* »). Pour ce qui est, en revanche, du registre d'usage, le numéro 5 indique un usage plutôt neutre, exactement à mi-chemin entre le registre soutenu et l'argotique.

c'est un vrai cheval #nom masc.	
Se dit de qqun qui travaille beaucoup	
synonyme	travailleur
fréquence	001
registre actuel	5
liens	GL Gallica Hathi Archive ULB
historique	dernière modification le 2024-03-20 23:46 (diff)

Chronologie et sources

Amis visiteurs, si vous connaissez une date plus ancienne, vous êtes chaleureusement invités à la partager en [discussion](#) avec indication de la source exacte. Elle sera vérifiée et intégrée à la notice sous votre signature. Merci.

Entrée relevée dans ces sources (sauf erreur) :

[1926](#)

- 1926. IV - Locutions usuelles de la langue parlée, dans *Le français de tous les jours, deuxième partie* ([fréquence](#) | [citations](#) | [datation](#) | [regis](#)

Figure 4 Définition de la locution *c'est un vrai cheval* dans *Bob*

3. L'élaboration d'un parcours didactique

Pour développer un parcours didactique sur l'enseignement du lexique qui soit utile pour des apprenants d'un niveau intermédiaire de français (B1), il faut avant tout bien définir les éléments théoriques avec lesquels nous œuvrons. Nous faisons référence à la distinction entre *leçon* et *unité didactique*, introduite par Balboni (2012 : 101). La notion d'*unité didactique* a pris le devant à partir des années 1960 sur la base du constat que tous les matériaux didactiques s'organisent par *unités* qui constituent un bloc autonome. Plus récemment, dans le domaine de la didactique, nous avons vu l'affirmation du *module* qui identifie des groupes différents de compétences linguistiques (Balboni, 2012 : 101).

Cette opération de découpage « en parties dont les contenus présentent une certaine cohésion » est celle que Puren appelle *séquentialisation* (Puren, 2005 : 41). Sur la base de cette précision terminologique, il est préférable de se référer à la fiche didactique qui sera proposée dans cet article comme un *module*, conçu pour l'apprentissage d'unités phraséologiques créées à partir du champ sémantique des animaux. D'ailleurs, ce module pourrait bien s'insérer à l'intérieur de l'unité didactique dédiée au vocabulaire thématique des animaux et, plus généralement, ayant l'objectif transversal de relativiser la conception de *standard linguistique*. Dans ce processus d'apprentissage, l'enseignant est appelé à jouer un rôle de *facilitateur* entre les élèves et les connaissances linguistiques à apprendre (Balboni, 2014 : 4), créant ainsi les conditions favorables pour un *apprentissage autodirigé*, dans lequel l'apprenant serait en mesure de « prendre les décisions constitutives de son apprentissage » (Cuq, 2003 : 83).

À ces réflexions, dans la tentative d'encadrer l'action didactique de l'enseignant de langues étrangères, Balboni ajoute qu'il est bien de partir toujours d'un constat sociolinguistique général pour après l'appliquer à une approche et à une méthode bien établies (Balboni, 2012 : 100). Par exemple, dans notre cas, ce constat pourrait être le suivant : « la langue se réalise à travers plusieurs variétés ». À partir de là, conformément à l'approche interculturelle de l'enseignement, la relativisation du standard, évoquée auparavant, apparaît nécessaire en vue de montrer aux élèves comment « dépasser les différences culturelles » en intériorisant une aptitude « à vivre dans des milieux et des situations pluriculturels » (Defays et Deltour, 2003 : 78). La relativisation de nature linguistique va de pair avec la relativisation culturelle et, en vue d'atteindre ce but, l'apport des outils lexicographiques collaboratifs peut s'avérer essentiel : en effet, leur objectif principal étant de préserver le patrimoine linguistique de toute la francophonie, ils permettent « d'obtenir un aperçu détaillé sur les expressions utilisées dans différents pays et régions francophones » (Calvi et Dankova, 2023 : 8). Ainsi faisant, la lexicographie collaborative contribue à rendre la communication interculturelle plus aisée, en emmenant à une compréhension de plus en plus approfondie des filtres culturels qui se posent lors d'une interaction entre locuteurs.

D'un point de vue méthodologique, nous traduisons ces perspectives en essayant de préciser et détailler davantage, par exemple, les modalités et le degré d'approfondissement nécessaires pour transmettre des nouvelles notions linguistiques aux apprenants. Pour atteindre ce but, dans une activité qui implique l'usage de ressources numériques collaboratives, l'enseignant se pose plusieurs questions : didactiquement, à ce stade, nous pouvons choisir entre une approche directe (fournir aux élèves les résultats déjà affichés par la machine) ou indirecte (laisser les élèves faire eux-mêmes la recherche). Une autre distinction méthodologique possible est celle entre perspective onomasiologique, où l'apprentissage s'organise par thème (« de quels mots a-t-on besoin pour dire ce thème ? »), et perspective sémasiologique, où par contre nous prenons chacune des unités lexicales singulièrement, en nous demandant dans quelle situation il est possible de l'employer (Cavalla, 2022). Dans notre cas, nous allons préférer une approche majoritairement indirecte et textuelle, en partant de l'analyse des résultats affichés par les corpus, pour après tester la capacité d'intuition des apprenants et les impliquer dans des activités pratiques.

3.1. Chien, vache et cheval : des unités phraséologiques francophones

Le module didactique que nous allons proposer dans cet article (voir Annexe 1) sera composé d'une partie de recherche et d'analyse individuelle de la part des apprenants, suivie de quelques activités et exercices pratiques, pour vérifier leur compréhension et les impliquer de façon plus active. En ce qui concerne l'organisation des cours, nous pouvons prévoir deux séances d'une heure chacune pour la phase introductive et deux ultérieures séances pour la partie consacrée aux exercices de compréhension et production écrite. Après avoir présenté aux apprenants les ressources de lexicographie collaborative dont nous avons parlé auparavant (voir section 2) et avoir brièvement introduit la notion d'expression phraséologique (voir section 1), nous leur demanderons de démarrer quelques recherches en autonomie pour se familiariser avec ces outils. En introduisant les élèves à ces ressources et à ces notions, il sera important d'insister sur les fonctions qui permettent d'avoir un aperçu général des variations diatopique et diastratique de la langue française. En effet, une unité phraséologique peut avoir des sens différents, selon la variété diatopique qui est prise en compte, sens qui souvent échappent à la compréhension des locuteurs non natifs. Par exemple, si l'expression « avoir mangé de la vache enragée » au Québec est traduite par le fait d'« être pris d'une folie violente », il semble qu'en français hexagonal le sens qu'on lui attribue le plus souvent est plutôt celui d'« endurer des fortes privations ». En tout cas, les deux sens semblent partager une racine commune et être reliés par une évolution sémantique : « créée au XVIII^e siècle, l'expression vient du fait que lors des périodes de privation, on est prêt à tout pour trouver de la nourriture, y compris à manger des animaux enragés¹³ », laissant entendre qu'on peut devenir enragé soi-même.

Au-delà du recours aux dictionnaires collaboratifs, pour ce module didactique, nous avons choisi de travailler principalement sur des corpus de presse et des corpus littéraires. Cela en raison du fait que le discours journalistique et la littérature sont des sources textuelles qui font un usage souvent important d'unités phraséologiques. Nous allons construire le premier exercice de notre fiche didactique en nous basant sur les occurrences d'unités phraséologiques construites à partir de *chien*, *vache* et *cheval*, attestées dans quelques-uns des corpus interrogeables à partir des plateformes *Compleat Lexical Tutor* et *Lexicoscope*. En particulier, nous allons nous référer à *Corpatext*, à la section *Fr_Journaux* et au corpus de presse de *Lexicoscope* qui inclue un échantillon d'articles du *Monde*, du *Figaro* et de *Libération* publiés entre 2010 et 2015. *Corpatext* est un corpus composé d'un total de 37 millions de mots, tirés principalement de textes littéraires français et francophones. *Fr_Journaux* est une section mineure (450 000 mots) du *Chambers-Rostand corpus*, constituée sur la base de trois sources de la presse nationale française : *Le Monde*, *L'Humanité* et *La Dépêche du Midi*¹⁴.

13. https://fr.wiktionary.org/wiki/manger_de_la_vache_enragée (page consultée le 27 août 2024)

14. Description des corpus disponibles (lien : https://www.lextutor.ca/conc/fr/corpus_descriptions_french.html, page consultée le 20 mars 2024) : « Le corpus est composé par 1723 articles tirés de trois quotidiens de la presse française : Le Monde (576 articles/355,046 mots), L'Humanité (576 articles/367,486 mots), La Dépêche du Midi (570 articles/257,299 mots) ».

Ce travail préliminaire sur les unités phraséologiques vise à évaluer la capacité des apprenants à déchiffrer et à interpréter les unités phraséologiques, qui valident leurs intuitions à l'aide de sources variées et authentiques. Cela aurait aussi le bénéfice d'introduire les élèves à l'usage structuré de ressources didactiques numériques, avec lesquelles ils commenceraient à se familiariser, en apprenant à les exploiter efficacement, ce qui s'avère de plus en plus une compétence précieuse.

À la suite d'une recherche dans les corpus ici décrits, nous allons soumettre aux apprenants des phrases choisies avec des unités phraséologiques à partir des unités lexicales *chien*, *vache* et *cheval* (voir Annexe 1). L'exercice qui va en dériver est pensé comme une activité introductive pour tester l'intuition des élèves : à partir des contextes d'emploi, les élèves tenteront de deviner la signification des unités phraséologiques repérées. Nous allons demander également aux apprenants de vérifier, par une recherche sur les ressources de lexicographie collaborative, si leur intuition était bonne ou non. Voici un exemple de recherche à partir de *chien battu* :

001. me dirigeai vers le lit, avec des hésitations de CHIEN battu mais fidèle. Souvent, j'ai tenté d'éc
 002. e rencontre de temps à autre, par une humilité de CHIEN battu On assurait qu'il allait entrer au Lo
 003. la persécution aigrie des anciennes. Cette vie de CHIEN battu rendait mauvaises les meilleures ; et
 004. stin qui arrive d'en bas et paraît avec un air de CHIEN battu Allant à lui, le prenant au collet et

Figure 5 Saisie d'écran des résultats à partir de la recherche de *chien battu* dans *Compleat Lexical Tutor*

831939 L'autre est un peu casse-pieds, une sorte de **chien battu** qui s'attache à sa personne avant tout.
 897851 A bord, les soldats ont des airs de **chien battu**.
 1382017 Pas un qui évolue depuis le début - le plus insupportable étant la mine de **chien battu** dont s'affuble, sans jamais s'en départir, Patrick Chesnais.
 175679 Le matin, elle arrive seule, avec ses grosses sandales noires, son imper noir, son air de **chien battu**.
 202114), sans oublier Pascal Elbé - faisant à peine son **chien battu** rituel, mais essentiellement au profit de sa casquette neuve de cinéaste-scénariste, maître d'oeuvre du projet, ethnopolar ambitieux et simple à la fois.

Figure 6 Saisie d'écran des résultats à partir de la recherche de *chien battu* dans *Lexicoscope*

La deuxième activité est, en revanche, apte à la consolidation des connaissances. Pour prouver la connaissance effective de ce lexique, les élèves devront relier les expressions aux définitions fournies par l'une des ressources de lexicographie collaborative connues. Cette activité peut avoir des effets très bénéfiques pour les élèves : en effet, non seulement ils sont emmenés à développer des associations mentales fortes, aidant à la mémorisation à long terme des expressions, mais aussi le fait de devoir distinguer des significations, qui peuvent paraître parfois proches, leur permettra de réfléchir de manière logique et d'affiner leur capacité de discernement linguistique.

Le troisième et dernier exercice du module didactique ici proposé constitue une activité ludique : les élèves sont invités à retrouver la lexie se référant à l'animal qui relie le mieux un petit ensemble d'autres mots, avec lesquelles des unités phraséologiques sont formées. À partir de l'adjectif *maigre* ou du verbe *manger*, nous voulons vérifier si les apprenants sont en mesure de les rattacher au ré-

férent *vache* et de reconstruire les unités phraséologiques *période de vache maigre* et *manger de la vache enragée*, qui signifient respectivement « période de temps avec un manque de ressources financières » et « endurer des privations ». De la même façon, à partir du substantif *sabots* et du verbe *miser*, les élèves devraient être capables de les relier au référent *cheval* et reconstruire les expressions *ne pas se trouver sous les sabots d'un cheval* et *miser sur le mauvais cheval*. En conclusion, en vue de stimuler aussi la compétence de production écrite, nous allons demander aux élèves d'inventer des phrases en employant les unités phraséologiques retrouvées et étudiées.

Somme toute, en combinant l'apport didactique des dictionnaires collaboratifs et des corpus numériques, les apprenants bénéficient d'une approche pédagogique plus nuancée, en ce qui concerne l'étude des expressions phraséologiques. Les définitions et explications détaillées des ressources collaboratives fournissent une base théorique solide, tandis que les exemples contextuels des corpus numériques permettent de voir comment ces expressions sont réellement utilisées dans la langue courante. Cette complémentarité renforce non seulement la compréhension des expressions mais aussi l'ensemble des compétences linguistiques des apprenants, y compris leur mémorisation et leur utilisation pratique en conversation.

4. L'avis des enseignants : un questionnaire d'évaluation

Par le biais d'un questionnaire d'évaluation que nous avons soumis à un petit groupe de cinq enseignantes travaillant dans l'école secondaire italienne (voir Annexe 2), il a été possible d'évaluer le module didactique ici présenté. Le retour des enseignantes a été précieux pour ajuster le contenu et répondre de la manière la plus adéquate possible aux besoins des élèves. En outre, nous avons pu vérifier la familiarité des enseignantes-mêmes, dans le contexte italien, avec les ressources de la lexicographie collaborative et avec les corpus numériques. Notre questionnaire, distribué en ligne par le biais de la plateforme de sondage Google Forms, a été construit suivant une perspective mixte : en effet, il comporte un mélange de six questions fermées, permettant de recueillir des réponses standardisées et faciles à quantifier, et quatre questions ouvertes, où nous avons laissé plus de liberté aux enseignantes de donner leur avis et des informations qualitatives plus détaillées (voir Annexe 2).

Les enquêtes précédemment menées sur ce sujet avaient mis en évidence « un écart significatif [...] entre les recherches, très avancées, sur les applications des corpus pour l'enseignement des langues et leur emploi réel, encore très limité, en classe » (Bonadonna, 2023 : 82). Cet aspect est tout à fait confirmé par les résultats du travail présenté dans cet article. En effet, bien que la majorité des interviewées déclarent connaître ces outils (60 % pour ce qui concerne les ressources de lexicographie collaborative et 80 % relativement aux corpus numériques), seulement 40 % ont eu l'occasion d'y travailler en classe au collège, avec des apprenants ayant un niveau B1.

Par contre, 80 % d'entre elles ont mis en évidence que, pour bien profiter de la fiche didactique proposée dans le cadre de cette étude, il faudrait une période de réflexion plus approfondie pour que

les élèves et les enseignants eux-mêmes puissent mieux se familiariser avec les instruments numériques. Le problème majeur qui émerge incontestablement est donc le manque de temps et parfois sa mauvaise gestion tout au long de l'année scolaire. De plus, Bonadonna (2023 : 91) constate qu'un autre problème qui demeure important est l'absence d'une formation *ad hoc* pour les enseignants, qui ont encore une conception très large et vague de ce qu'est vraiment un corpus numérique. Malgré cela, la totalité des enseignantes qui ont répondu au questionnaire se déclarent vivement intéressées par une insertion progressive mais constante du numérique dans leurs approches et méthodes d'enseignement de la langue française.

En outre, interrogées sur les possibles avantages des dictionnaires collaboratifs en contexte didactique, les enseignantes soulignent que ces instruments permettraient une consultation plus facile et immédiate, donnant en même temps aux élèves l'occasion de « participer activement à l'observation de la langue [...] et à leur propre apprentissage ». En revanche, en ce qui concerne les inconvénients, elles estiment que le principal défaut de la lexicographie collaborative réside dans le risque de faire croire aux élèves qu'il est facile d'accéder à toutes les informations possibles sans efforts. Toujours selon les personnes interviewées, ce risque s'accroît davantage avec les plus récents atouts de l'intelligence artificielle. Ces réponses démontrent qu'un premier travail à faire est sans doute celui de changer la perception de ces outils chez les apprenants, qui en font un usage « pragmatique » (lié exclusivement à la recherche des définitions), et chez les enseignants qui se montrent encore trop méfiants (Calvi et Danokova, 2023 : 4). En effet, il est nécessaire de rappeler à ce propos que « googler » un mot et consulter des dictionnaires collaboratifs n'est pas exactement la même chose : bien que les deux méthodes impliquent une recherche en ligne, elles diffèrent dans leurs approches et leurs résultats. Nous avons vu, en fait, que les dictionnaires collaboratifs offrent une richesse de perspectives et une mise à jour continue grâce aux contributions variées des utilisateurs. En outre, ces ressources encouragent une approche plus interactive et engageante, où les apprenants peuvent s'impliquer directement dans la documentation et l'évolution de la langue, contrairement à une simple recherche sur un moteur de recherche comme Google, qui peut parfois manquer de profondeur. À la lumière de cela, il est essentiel de former les apprenants à utiliser de manière critique et réfléchie les ressources en ligne, en leur enseignant non seulement à rechercher des informations, mais aussi à évaluer la qualité et la fiabilité des sources qu'ils consultent. Ainsi, une utilisation consciente et judicieuse des dictionnaires collaboratifs, en complément des méthodes traditionnelles, pourrait enrichir l'apprentissage en langues étrangères.

En conclusion, en proposant des suggestions pour intégrer le numérique avec la didactique traditionnelle, les enseignantes ont réfléchi, par exemple, à la possibilité de consacrer périodiquement des séances à l'usage des corpus et des dictionnaires numériques pour que les élèves arrivent à bien les maîtriser. En particulier, il serait intéressant, d'après elles, d'avoir recours aux corpus à chaque fois que du nouveau vocabulaire est introduit pour rechercher des exemples d'usage en contexte. Cela aurait un impact sans doute positif sur l'aptitude des apprenants envers l'apprentissage qui serait plus proche de l'usage quotidien réel de la langue.

5. Conclusions

Dans le cadre de ce travail, nous nous sommes interrogé sur l'évolution du statut du dictionnaire et sur l'émergence des ressources de lexicographie collaborative, telles que le *Wiktionnaire*, le *Dictionnaire des francophones* et *Bob*, notamment pour ce qui concerne l'apprentissage de la langue française. En outre, nous avons aussi brièvement considéré leur usage en relation avec les corpus numériques dans le contexte didactique italien actuel, en consultant les bases de traitement de corpus *Lexicoscope* et *Compleat Lexical Tutor*. Un exemple a été fourni grâce à la création d'une fiche didactique sur les unités phraséologiques à partir du champ sémantique des animaux, qui a été soumise à l'évaluation d'un groupe de cinq enseignantes italophones, travaillant à l'école secondaire.

L'intégration de ces instruments dans la pratique didactique soulève des questions quant à leur utilisation consciente et appropriée et à leur valeur ajoutée par rapport aux outils lexicographiques traditionnels : ce qui émerge, en effet, à partir des réponses des enseignantes au questionnaire, c'est que, bien que ces outils offrent une consultation rapide et une immersion dans la langue en contexte réel, ils peuvent aussi entraîner une dépendance à une recherche superficielle et une perte de capacité d'évaluation critique chez les apprenants. Par conséquent, il est essentiel pour les enseignants de guider les apprenants dans une utilisation réfléchie de ces ressources, tout en développant leur capacité à contextualiser et analyser les informations linguistiques. Pour ce faire, de la part des enseignantes, il serait essentiel de consacrer des séances à la familiarisation des élèves avec les corpus numériques et les dictionnaires collaboratifs.

Les réponses des enseignantes montrent un intérêt croissant pour l'intégration du numérique dans leurs pratiques pédagogiques, en soulignant, en même temps, la nécessité d'une gestion efficace du temps, ce qui demeure l'un des problèmes majeurs. Somme toute, en combinant les avantages du numérique avec une approche pédagogique réfléchie et une formation adéquate, les enseignants seraient en mesure d'aider les apprenants à développer une compréhension plus profonde et plus contextuelle de la langue française, tout en les préparant à naviguer dans un monde numérique en constante évolution.

Annexe 1

Fiche didactique pour l'apprentissage du lexique en FLE :

Les unités phraséologiques à partir du champ sémantique des animaux

1a) Lisez les phrases suivantes (tirées des outils de traitement de corpus numériques *Lexicoscope* et *Compleat Lexical Tutor*). Identifiez les unités phraséologiques relevant du champ sémantique des animaux, puis essayez de deviner leur signification et, si possible, la zone géographique d'usage.

- a) *Ses souffrances du début la poignaient encore, une pitié lui remuait le cœur, à chaque nouvelle venue qu'elle rencontrait dans les rayons, les pieds meurtris, les yeux gros de larmes [...]. Cette **vie de chien battu** rendait mauvaises les meilleures* (*Compleat Lexical Tutor*).
- b) *Dans une interview révélée par un remarquable documentaire qui lui est consacré, Michel Magne explique : « Je me suis tourné vers le cinéma car j'en avais marre de **manger de la vache enragée** »* (*Lexicoscope*).
- c) *Avant de devenir célèbre dans sa profession, Maurice Maman **connut des années de vaches maigres** durant ses études à Bordeaux* (*Lexicoscope*).
- d) *Frères **comme chien et chat**, Jean-Michel et Alexandre se partagent le restaurant familial* (*Lexicoscope*).
- e) *Interrogé vendredi au micro d'I-Télé et Radio Classique sur la déclaration de Guénant, la veille, préconisant l'interdiction des signes religieux pour les usagers des services publics, Guaino est **monté sur ses grands chevaux** et s'est désolé de voir le pays « dériver dans l'hystérie et la caricature »* (*Lexicoscope*).
- f) *Le Chinois avait **misé sur le mauvais cheval** en choisissant Bear Stearns, le numéro cinq à Wall Street, qui fut la première grande victime de la crise financière dès mars 2008* (*Lexicoscope*).
- g) *Erin m'explique que ce voyage sera hard, parce que l'attachée de presse qui les attend à la destination idyllique est une bitch (une prostituée, c'est-à-dire une **peau de vache**), mais qu'elle a une bonne maîtrise du print (la presse écrite) et du Web locaux* (*Lexicoscope*).
- h) *Envoyé en mission à Detroit pour encadrer un projet automobile en pleine crise financière, un jeune ingénieur français découvre une ville au bord du chaos, **livrée aux chiens** de défense errants* (*Compleat Lexical Tutor*).
- i) *Aujourd'hui, de « Scènes de ménages », sur M6 (qui en est tout de même à sa quatrième saison), à « Nos chers voisins », sa copie sur TF1 depuis septembre 2012, les exemples de ce type **pleuvent comme vache qui pisse*** (*Lexicoscope*).
- j) *Il avait appris au fil des années à être plus coulant avec les joueurs qui lui faisaient gagner les matchs, « ceux qu'on **ne trouve pas sous les sabots d'un cheval** »* (*Lexicoscope*).
- k) *Elle **a tellement la chienne** de sortir pour aller aux toilettes qu'elle a pissé dans le lavabo de la chambre* (*Compleat Lexical Tutor*).

1b) Vérifiez si votre intuition initiale était correcte ou non, en démarrant une recherche sur les dictionnaires collaboratifs que vous connaissez (*Wiktionnaire*, *Dictionnaire des francophones*, *Bob*).

2) Reliez les unités phraséologiques de la colonne de gauche aux définitions de la colonne de droite.

	Unités phraséologiques
a)	<i>(Avoir des) yeux de chien battu</i> 🇫🇷/🇺🇸
b)	<i>Ne pas se trouver sous les sabots d'un cheval</i> 🇫🇷
c)	<i>Être une peau de vache</i> 🇫🇷
d)	<i>Pleuvoir comme vache qui pisse</i> 🇫🇷/🇺🇸
e)	<i>(Avoir) manger de la vache enragée</i> 🇫🇷/🇺🇸
f)	<i>Livrer aux chiens</i> 🇫🇷
g)	<i>(Période de) vaches maigres</i> 🇫🇷
h)	<i>Miser sur le mauvais cheval</i> 🇫🇷
i)	<i>S'entendre comme chien et chat</i> 🇫🇷/🇺🇸
l)	<i>Monter sur ses grands chevaux</i> 🇫🇷/🇺🇸
m)	<i>Avoir la chienne</i> 🇺🇸

Définitions
1) Endurer des privations / Être pris d'une folie violente. (<i>Wiktionnaire</i>)
2) Se dit de deux personnes qui sont en conflit perpétuel, qui se cherchent querelle pour tout motif. (tirée du <i>Wiktionnaire</i> – présente dans le <i>Dictionnaire des francophones</i>)
3) (Être) une personne méchante. (<i>Wiktionnaire</i>)
4) Jeter quelque chose en pâture aux journalistes et au public. (tirée du <i>Wiktionnaire</i> – présente dans le <i>Dictionnaire des francophones</i>)
5) Regard apitoyant, attendrissant ; regard de tristesse, de désolation. (tirée du <i>Wiktionnaire</i> – présente dans le <i>Dictionnaire des francophones</i>)
6) Symbole de disette ou de manque de ressources financières ou autres, en parlant d'une période de temps. (<i>Wiktionnaire</i>)
7) Pleuvoir abondamment. (<i>Wiktionnaire</i>)
8) Avoir de la difficulté à trouver ou se procurer quelque chose. (<i>Bob</i>)
9) S'emporter, prendre la situation avec excès de gravité. (<i>Bob</i>)
10) Avoir peur. (<i>Wiktionnaire</i>)
11) Choisir le mauvais camp, ne pas faire le bon choix. (<i>Bob</i>)

3) Trouvez l'animal qui relie les mots suivants et après créer des petits dialogues pour chaque expression repérée.

battu
livrer
chat
avoir



CHIEN

sabots
miser
monter
vrai



CHEVAL

pleuvoir
manger
peau
maigre



VACHE

Annexe 2

Questionnaire d'évaluation pour tester l'efficacité du numérique en contexte didactique

1. Vous connaissiez déjà (ou vous aviez déjà travaillé avec) quelques corpus numériques ?
 - Oui, j'en connaissais déjà et j'avais déjà travaillé avec.
 - Oui, j'en connaissais déjà mais je n'avais jamais travaillé avec.
 - J'en avais seulement entendu parler.
 - Non, je ne les connaissais pas.
2. En vous basant sur ce que vous avez pu observer dans le module didactique, pour ce qui concerne l'apprentissage et l'enseignement du lexique, quel serait d'après vous le potentiel des corpus numériques en classe de FLE ?
3. Vous connaissiez déjà (ou vous aviez déjà travaillé avec) quelques ressources de lexicographie collaborative ?
 - Oui, j'en connaissais déjà et j'avais déjà travaillé avec.
 - Oui, j'en connaissais déjà mais je n'avais jamais travaillé avec.
 - J'en avais seulement entendu parler.
 - Non, je ne les connaissais pas.
4. Est-ce que notre module didactique pourrait être utilisé aisément en contexte didactique ?
 - Oui, ce serait une extraordinaire opportunité d'apprentissage pour les élèves.
 - Oui, mais il faudrait une période de réflexion plus approfondie pour apprendre à utiliser ces instruments.
 - Non, ce sont des instruments trop avancés pour des élèves à l'école secondaire.
 - Non, il nous faudrait trop de temps.
5. Quels seraient d'après vous les avantages majeurs de la lexicographie collaborative par rapport à la lexicographie traditionnelle ?
 - Elle met plus en évidence l'aspect de la variation diatopique (variation liée à la zone géographique) par rapport aux dictionnaires traditionnels.
 - Ses ressources sont consultables plus facilement et plus rapidement par rapport aux dictionnaires traditionnels.
 - Leur exploitation permettrait de mettre en cause réellement la présumée centralité du standard franco-parisien et franco-français.
 - Il n'y a pas d'avantages.
 - Autre :

-
6. Quels seraient d'après vous les défauts majeurs de la lexicographie collaborative par rapport à la lexicographie traditionnelle ?
 - Elle cause une baisse de concentration et d'attention.
 - Les informations repérables dans les ouvrages de lexicographie traditionnelle sont plus complètes et fiables.
 - Elle mène les élèves à penser qu'il est suffisant de taper sur un clavier pour avoir accès à toutes les informations possibles.
 - Autre :
 7. Quelle est votre attitude actuelle envers les ressources de lexicographie collaborative ?
 - Je suis prêt/prête à utiliser ces nouveaux instruments numériques pour faire progresser la didactique des langues en ce sens.
 - Il est bien de savoir en profiter, en ayant conscience des limites.
 - Elle risque de mettre trop en cause l'autorité et la centralité de l'enseignant/enseignante dans le processus d'apprentissage.
 - Autre :
 8. En vous basant sur ce que vous avez pu observer dans le module didactique présenté, pour ce qui concerne l'apprentissage et l'enseignement du lexique, quel serait d'après vous le potentiel des dictionnaires collaboratifs en classe de FLE ?
 9. Pensez-vous que les matériaux didactiques traditionnels (manuels, livres...) sont suffisants pour acquérir une connaissance adéquate de la langue ? Motivez votre réponse.
 10. Avez-vous des suggestions pour mieux intégrer le numérique avec les matériaux traditionnels d'apprentissage (manuels, livres...) ?

Bibliographie

- BALBONI, Paolo (2012), « Modulo, Unità Didattica, Unità d'Apprendimento: dalla conoscenza alla progettazione in glottodidattica », dans Antonella Bennuci et Giuseppe Caruso (dir), *Competenze d'uso e integrazione*, Napoli, Edizioni Scientifiche Italiane, p. 99-110 [en ligne : <https://iris.unive.it/handle/10278/34467?mode=complete>].
- BALBONI, Paolo (2014), « Etica nell'approccio, nel metodo, nelle azioni dell'educazione linguistica », *EuroAmerican Journal of Applied Linguistics and Languages*, n°1, p. 1-14 [en ligne : <https://iris.unive.it/handle/10278/3659951>].
- BOLLY, Catherine (2011), *Phraséologie et collocations – approches sur corpus en français L1 et L2*, Peter Lang.
- BONADONNA, Maria Francesca (2023), *Didactique du lexique et corpus numériques pour le Français L2 – des applications pour le commerce et le marketing digital*, Peter Lang.
- BOUTARD, Anne-Sophie (2021), « Dictionnaire des francophones : le français par tous », *CNRS, Le journal*. [en ligne : <https://lejournel.cnrs.fr/articles/dictionnaire-des-francophones-le-francais-par-tous>]
- CALVI, Silvia et Klara Dankova (2023), « Lexicographie traditionnelle-collaborative et FLE: compétences lexicales et lexicographiques », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [en ligne : <https://journals.openedition.org/linx/10108>]
- CECR (2020 [2001]), *Cadre Européen Commun de Référence pour les langues : apprendre, enseigner, évaluer* [en ligne : <https://rm.coe.int/cadre-europeen-commun-de-reference-pour-les-langues-apprendre-enseigne/1680a4e270>].
- Compleat Lexical Tutor*. [en ligne : <https://www.lextutor.ca/conc/fr/>]
- CUQ, Jean-Pierre et Isabelle Gruca (2002), *Cours de didactique du français langue étrangère et seconde*, Presses Universitaires de Grenoble.
- CUQ, Jean-Pierre (2003), *Dictionnaire de didactique du français langue étrangère et seconde*, Paris, CLE International.
- DOLAR, Kaja (2017), « Les dictionnaires collaboratifs non institutionnels dans l'espace francophone : éléments de typologie et bilan », *Repères DoRif, Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, n°14, p. 1-11 [en ligne : <https://www.dorif.it/reperes/kaja-dolar-les-dictionnaires-collaboratifs-non-institutionnels-dans-lespace-francophone-elements-de-typologie-et-bilan/>].
- DOLAR, Kaja (2018), « Les dictionnaires collaboratifs en ligne, des objets métalinguistiques profanes », *Les Carnets du Cediscor*, p. 34-50 [en ligne : <http://journals.openedition.org/cediscor/1161>].
- DOLAR, Kaja, Noé Gasparini et Marie Steffens (2023), « L'utilisateur au centre de l'objet lexicographique : l'exemple du *Dictionnaire des francophones* », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [en ligne : <https://journals.openedition.org/linx/9854>]

- DUBOIS, Jean et Claude Dubois (1971), *Introduction à la lexicographie : le dictionnaire*, Paris : Larousse.
- FRASSI, Paolo (2018), « L'enseignement/apprentissage de la collocation entre contraintes grammaticales et contenu sémantique », *Éla – études de linguistique appliquée*, n°189, p. 63-84.
- GASPARINI, Noé, Nadia Sefiane et Sébastien Gathier (2021), *Compendium du DDF – Dictionnaire des francophones*, Institut international pour la Francophonie - Université Jean Moulin Lyon 3 [en ligne : https://www.academieoutremer.fr/wp-content/uploads/2021/03/Compendium_du_DDF.pdf]
- GONZÁLEZ-REY, Isabel (2008) *La didactique du français idiomatique*, Fernelmont (Belgique), E.M.E.
- GOUDAILLIER, Jean-Pierre (2015), « Culture *banlieues*, langue de cités et Internet », *Hermès – La Revue*, n°71, p. 208-213, [en ligne : https://www.cairn.info/revue-hermes-la-revue-2015-1-page-208.htm&wt_src=pdf].
- LAMIROY, Béatrice (2008), « Les expressions figées : à la recherche d'une définition », p. 1-14 [en ligne : <https://www.ling.arts.kuleuven.be/franitalco/papers/Lamiroy2008.pdf>]
- MOLINARI, Chiara (2017), « Nouvelle lexicographie vs anciennes représentations », *Repères DoRif, Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, n°14, p. 1-13 [en ligne : <http://www.dorif.it/reperes/chiara-molinari-nouvelle-lexicographie-vs-anciennes-representations/>].
- MOLINARI, Chiara (2021), « La lexicographie numérique : un nouvel outil pour l'enseignement du FLE ? », *Éla – études de linguistique appliquée*, n° 201, p. 27-47.
- MOLINARI, Chiara et Stefano Vicari (2023), « Quelle(s) collaboration(s) dans les dictionnaires en ligne? Les cas du *Wiktionnaire* et du *Dictionnaire des francophones* », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [en ligne : <https://journals.openedition.org/linx/10393>]
- MURANO, Michela (2014), « La lexicographie 2.0 : nous sommes tous lexicographes ? », dans Ruggero Druetta et Caterina Falbo (dir.), *Cahiers de recherche de l'École doctorale en linguistique française*, n° 8. *Docteurs et Recherche... une aventure qui continue*, EUT Edizioni Università di Trieste, p. 147-162.
- PUREN, Christian (2005), « L'évolution historique des approches en didactique des langues-cultures, ou comment faire l'unité des "unités didactiques" », *Nouveau bulletin de l'ADEAF*, n°89, p. 40-51 [en ligne : https://www.tesol-france.org/uploaded_files/files/OJ-Puren05.pdf].
- SAJOURS, Frank, Nabil Hathout et Basilio Calderone (2014), « Ne jetons pas le Wiktionnaire avec l'oripeau du Web ! Études et réalisations fondées sur le dictionnaire collaboratif », 4^e Congrès mondial de linguistique française, Berlin (Allemagne), p. 663-680 [en ligne : <https://shs.hal.science/halshs-00969260>].
- SAUSSURE, Ferdinand (2005 [1915]), *Cours de linguistique générale*, Genève, Arbre d'Or.
- SEFIANE, Nadia (2022), « Webinaire IFprofs – DDF : Le *Dictionnaire des francophones* et son utilisation en classe », [en ligne : <https://www.youtube.com/watch?v=d5UvN8UZGjA>].

STEFFENS, Marie et Esther Baiwir (2021), « Intégrer la variation diatopique à l'enseignement du français : le rôle des outils numériques », *Éla – études de linguistiques appliquées*, n°201, p. 11-22.

SUN, Juan et Zhihong Pu (2018), « Pour une approche sociolinguistique en didactique du français langue étrangère – l'argot français contemporain en classe », *Synergies Chine*, n°13, p. 131-142
[en ligne : https://gerflint.fr/Base/Chine13/sun_pu.pdf].

VINCENT, Nadine (2019), « Qu'est-ce qu'un dictionnaire ? Perspectives de la lexicographie au 21^e siècle », Anne Dister et Sophie Piron (dir.), *Les discours de référence sur la langue française*, Bruxelles, Presse de l'Université Saint-Louis, p. 127-148 [en ligne : <https://books.openedition.org/pusi/26417#text>].

TITRE: ANALYSER DES LEXÈMES DU QUOTIDIEN DANS UNE CORRESPONDANCE PRIVÉE CORSE (1881-1918) : RESSOURCES LEXICOGRAPHIQUES EXISTANTES ET APPORTS DU CORPUS

AUTEURS : BEATRICE DAL BO, UNIVERSITÉ SORBONNE NOUVELLE, ET ANNA GIAUFRET, UNIVERSITÀ DI GENOVA

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 83-102

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22396](http://hdl.handle.net/11143/22396)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22396](https://doi.org/10.17118/11143/22396)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONALE.

Analyser des lexèmes du quotidien dans une correspondance privée corse (1881-1918) : ressources lexicographiques existantes et apports du corpus

Beatrice Dal Bo, Université Sorbonne Nouvelle

Anna Giaufret, Università di Genova

Résumé :

Le projet Corr<si>Ca, qui porte sur la numérisation de la correspondance privée de la famille Canioni d'Olmi-Cappella (Corse), nous a menés à explorer quelques ressources lexicographiques qui se situent à la confluence des aires linguistiques italienne, française et corse. Cet article présente notre projet numérique, approfondit la dimension culturelle et lexicale de notre corpus et illustre les ressources lexicographiques, numériques et numérisées, utilisées pour mener des analyses lexicales. En conclusion, nous fournissons quelques pistes de recherche et suggérons la création de nouveaux outils pour le décodage de ces éléments culturels, en utilisant une approche « bottom-up ».

Mots-clés :

Correspondance, Corse, peu lettrés, ressources lexicographiques, édition numérique

Abstract :

The Corr<si>Ca project, which concerns the digitisation of the private correspondence of the Canioni family of Olmi-Cappella (Corsica), has led us to explore several lexicographical resources at the confluence of the Italian, French and Corsican linguistic areas. This article presents our digital project, explores the cultural and lexical dimensions of our corpus, and illustrates the digital and digitised lexicographical resources used to conduct lexical analyses. In conclusion, we provide some avenues of research and suggest the creation of new tools for decoding these cultural elements, using a 'bottom-up' approach.

Keywords :

Correspondence, Corsica, less-literate writers, lexicographical resources, text digitisation

1. Introduction

Le projet Corr<si>Ca, qui porte sur la numérisation de la correspondance privée de la famille Canioni d’Olimi-Cappella (Corse), nous a menés à nous confronter avec un certain nombre de ressources lexicographiques qui se situent à la confluence des aires linguistiques italienne, française et corse. Dans cet article, nous allons premièrement présenter notre projet numérique, approfondir dans un deuxième temps la dimension culturelle et lexicale de notre corpus, pour passer troisièmement à l’illustration des ressources lexicographiques, numériques et « numériquées », que nous avons utilisées pour mener des analyses lexicales. En conclusion, nous allons fournir quelques pistes de recherche et suggérer la création de nouveaux outils pour le décodage de ces éléments culturels, en utilisant une approche « bottom-up » : nous partons du corpus pour identifier les outils existants, mais aussi les besoins lexicographiques.

2. Le projet Corr<si>Ca

Ce projet d’édition numérique Corr<si>Ca prend la relève d’un projet précédent dirigé par Anna Giaufret, *MaPio (Histoire de la communauté de Pioggiola dans les années de la Révolution, 1787-1797¹)*, au sein duquel nous avons réalisé la publication numérique des procès-verbaux de la commune de Pioggiola à la fin du XVIII^e siècle. Le texte principal est également accompagné d’analyses historiques, linguistiques et sur le patrimoine matériel. De son côté, Beatrice Dal Bo fait partie du projet de recherche *Corpus 14*, qui a pour but de collecter, numériser et analyser des correspondances privées de la Première Guerre mondiale, échangées par des scripteurs et des scriptrices *peu lettrés* (voir Dal Bo, 2019 ; Steuckardt, 2015 *inter alia*). Ces expériences ont constitué le point de départ d’une ligne de recherche sur l’écriture privée de cette catégorie de scripteurs (dont nous traiterons infra, § 2) qui nous a poussées à nous intéresser aux correspondances locales. La générosité de la famille Canioni nous a permis d’avoir accès à la correspondance familiale écrite entre 1881 et 1918, correspondance constituée de lettres échangées entre le village d’Olimi-Cappella (qui se trouve, ainsi que Pioggiola, dans la vallée du Ghjunsani en Haute-Corse), d’autres villages corses et le continent. Nombreuses sont les raisons de s’intéresser à cette correspondance : du point de vue linguistique, on peut y voir l’évolution de la francisation et celle du passage de l’italien au français comme langue de l’écriture, ainsi que la transformation du répertoire des locuteurs corses² ; du point de vue historique, la correspondance contient de nombreuses références à la vie matérielle, mais aussi à la vie des soldats pendant la Première Guerre mondiale ; du point de vue culturel, on y retrouve, entre autres, la description des us et coutumes, des musiques et des données importantes sur l’échange de biens entre la Corse et le continent. Enfin, la présence de femmes parmi les scripteurs, qui appartiennent d’ailleurs à des

1. <https://mapio.unige.it/fr>. Le projet a donné lieu à une publication (Ferrando, 2022).

2. Pour des informations sur la langue corse et la situation sociolinguistique, voir Arrighi, 2008.

génération différentes et à des milieux socioculturels différents, rend cette correspondance particulièrement précieuse.

La correspondance Canioni regroupe les lettres de trois générations : celle de Xavier, le patriarche, né en 1835, celle de Jacques Antoine, son fils, né en 1855, et celle de son petit-fils Léon, né sur le continent en 1890 et mort au front en 1915. À ces scripteurs, il faut ajouter de nombreux autres membres de la famille et de leur entourage, en particulier Pierre François, le fils cadet de Xavier (né en 1858), qui passera toute sa vie à Olmi-Cappella avec sa famille, contrairement à Jacques Antoine, qui poursuit à Marseille sa carrière militaire, et à sa sœur Herminie, qui épouse un homme de Belgodère et fonde dans ce bourg de la Balagne une nouvelle famille. La correspondance se compose d'environ 270 lettres, rédigées en français pour la plupart, mais comprenant aussi huit lettres en italien de Xavier, écrites entre 1881 et 1893. À partir de 1895, Xavier fait rédiger ses lettres en français par son fils Pierre François, mais celui-ci maîtrise également l'italien, puisqu'en 1887 une de ses lettres contient un post-scriptum dans cette langue. Cette correspondance témoigne ainsi du processus de francisation de la Corse : le recul de la langue italienne comme langue d'écriture sur l'île aurait commencé autour de 1850, selon Thiers (1978 ; voir aussi Branca et Sorba, 2023) et aurait atteint seulement la deuxième génération Canioni. À l'italien et au français s'ajoute bien évidemment le corse, langue parlée par les scripteurs et scriptrices vivant sur l'île. Des difficultés de communication sont évoquées par Pierre François entre ses filles et leur cousin Léon : pour communiquer avec lui, ses enfants expriment le souhait d'apprendre le français. Le substrat corse émerge donc par moments dans l'écriture par le biais d'éléments lexicaux, de certaines tournures syntaxiques et de la graphie de certains mots ou expressions (par exemple, la sonorisation/insonorisation des consonnes occlusives intervocaliques, comme dans *sabere* ou *pacare*).

Le projet Corr<si>Ca a pour objectif la réalisation d'une ressource numérique, dans le cadre du doctorat en Humanités numériques³ des Universités de Gênes et Turin, et va se développer en plusieurs phases :

- Phase 1 : édition numérique de la correspondance Canioni (numérisation des lettres en haute définition, transcription diplomatique selon un protocole préétabli et encodage en XML-TEI⁴).
- Phase 2 : publication du corpus (possibilité d'interrogations complexes, de géolocalisation, etc.) en accès libre et création d'un blogue de vulgarisation scientifique.
- Phase 3 : analyse des données.

À l'état actuel du projet, la phase 1 est presque terminée.

3. [Benvenuti | Digital Humanities PhD \(unige.it\)](#)

4. Voir Pellino *et al.*, 2022 ; Stadler, 2016 et Walter, 2018 *inter alia* pour plus d'information sur l'édition numérique de correspondances.

3. La famille Canioni : une correspondance *peu lettrée* ?

Le terme *peu lettré* et la définition de cette notion sont empruntés à Sonia Branca-Rosoff et Nathalie Schneider. Les scripteurs *peu lettrés* sont en effet définis par les deux chercheuses, dans le cadre d'une étude sur des écrits publics de la période de la Révolution, comme le « groupe de ceux qui emploient une langue non conforme » (1994 : 9). Ils se situent ainsi entre les groupes des *lettrés*, qui maîtrisent la langue de façon conforme à la norme linguistique, et des *illettrés*, qui, au contraire, ne savent pas écrire. En se fondant sur cette première définition, on utilisera ici le terme de *peu lettrés* pour désigner des individus qui n'ont pas l'habitude de l'écrit et dont certains usages s'éloignent des normes linguistiques prescriptives. Soulignons que cette catégorie doit être envisagée en tant que *continuum* : les écrits des scripteurs *peu lettrés* montrent des degrés de compétences linguistiques à l'écrit différents d'un scripteur à l'autre, certains s'éloignant davantage de la norme linguistique que d'autres. Le corpus Canioni témoigne bien de ce caractère graduel : si les écrits de Xavier Canioni présentent des traits linguistiques qui permettent de le situer pleinement dans cette catégorie, cela est moins évident pour son fils cadet Pierre François et encore moins pour son fils aîné Jacques Antoine et son petit-fils Léon, dont les écrits répondent davantage aux attentes des normes (linguistique et épistolaire). Les lettres de cette correspondance familiale donnent accès à des usages linguistiques qui ont été souvent négligés dans les études scientifiques à cause de leur caractère « non standard » et, par conséquent, elles constituent une ressource précieuse pour l'étude de ces usages à différents niveaux d'analyse. En effet, l'étude de la graphie et des dimensions morphosyntaxique, lexicale et syntaxique permet de mieux décrire le contact entre l'italien, le corse et le français à un moment charnière de l'histoire linguistique de la Corse (cf. *supra*). Le travail sur ce corpus s'inscrit ainsi dans le cadre des études menées en sociolinguistique historique (Bergeron-Maguire (dir.), 2022 ; Branca-Rosoff, 1989 *inter alia*) et vise à apporter de nouvelles données authentiques à l'étude du contact entre plusieurs langues, ainsi qu'à celle du répertoire linguistique des Corses de l'époque.

4. Les éléments culturels dans la correspondance Canioni

Dans les études sur les éléments culturels répertoriés dans les dictionnaires, les auteurs se sont essentiellement concentrés sur l'ubiquité des éléments doxiques (Lehmann, 1989), sur la représentation des femmes et des minorités (Lehmann, 1981 ; Kottelat, 2009 et 2010) ou sur les dictionnaires d'apprentissage (Boulanger, 1999).

Pour l'analyse ici présentée, nous nous sommes penchées sur les éléments lexicaux qui sont porteurs d'information sur la culture matérielle, notamment les poids et les mesures, la nourriture et les maladies. Ces trois domaines nous permettent d'aborder à la fois l'évolution de la vie matérielle, les outils lexicaux pour la raconter, le passage de l'italien au français et de vérifier quels outils lexicographiques les répertorient. Cette étude est complétée par les données que la correspondance

Canioni fournit quant à la fréquence et au contexte d'emploi de plusieurs termes de ces domaines : ces informations permettront ainsi d'améliorer leur traitement lexicographique.

5. Quels outils lexicographiques ?

Les propos de Kevers (2021) cités ci-dessous sont tout à fait pertinents pour notre corpus et pour notre réflexion :

[Le corse est] une langue peu dotée, pour laquelle des variations dialectales sont enregistrées, et qui peut également souffrir d'une certaine proximité linguistique avec l'italien, ainsi que de la situation de diglossie avec le français, celle-ci pouvant se matérialiser par l'alternance de ces deux langues dans certains textes (Kevers, 2021 : 13).

On peut ajouter à ce constat le souhait, exprimé par Retali-Medori et Kevers, que les outils numériques puissent fournir « des ressources utiles à sa survie en donnant aussi toute sa place à la dimension variationnelle d'une langue non normée » (Retali-Medori et Kevers, 2022 : 9). Voilà pourquoi notre démarche s'est attachée en premier lieu à identifier les ressources numériques qui existent déjà, à constater la présence ou l'absence des unités lexicales identifiées, à analyser leur traitement et suggérer la création d'outils permettant de faire le pont entre les ressources existantes, voire de créer de nouveaux glossaires. Il faut toutefois préciser que les ressources existantes, à quelques exceptions près, sont des ressources numérisées plutôt que numériques (Paveau, 2017), c'est-à-dire des outils numérisés après avoir été conçus pour la version papier. Ceci est évident si l'on considère que nous avons utilisé des dictionnaires de la langue du XIX^e siècle et du début du XX^e ; néanmoins, leur numérisation modifie profondément leur statut, puisqu'ils peuvent maintenant participer à un réseau informationnel et entrer en résonance avec d'autres outils numériques. Il reste qu'une grande asymétrie existe en termes de disponibilité d'outils fiables entre le corse d'une part, l'italien et le français de l'autre. Si pour ces dernières langues nous avons consulté des dictionnaires institutionnels et des lexiques de spécialité, pour le corse, nous nous sommes également tournés vers des pages personnelles et des sites associatifs qui relèvent davantage d'une lexicographie spontanée ou collaborative (voir à ce sujet le numéro 14 de la revue *Repères Dorif*, coordonné par Molinari et Vincent, 2017). Le recours aux ressources lexicographiques dans les trois langues s'avère fondamental, puisque la correspondance Canioni se trouve précisément à l'intersection des trois, les scripteurs et scriptrices ayant un répertoire linguistique trilingue.

Une introduction à la langue corse et à ses spécificités de langue polynomique (Marcellesi, 1984) dépasse largement le cadre de cet article : nous renvoyons à Arrighi, 2002 ; Arrighi, 2008 et Corvetto et Nesi, 1993. La vallée du Ghjunsani, entourée par des chaînes de montagnes dépassant les mille mètres d'altitude, se trouve à la frontière sud-orientale de la Balagne et est reliée à celle-ci, ainsi qu'à la région de Caccia, par de nombreux sentiers. On y parle la variété du « cismonte » (Dalbera-Stefanaggi, 2002) dans sa forme « balanine », répertoriée dans un dictionnaire, celui du corse P. Tommaso Alfonsi, publié à Livourne en 1932, œuvre imprégnée d'irréductibilisme, qui vise à rapprocher le plus possible

le corse de l'italien et à condamner l'emploi des dangereux « francesismi, che sciupano il nostro dialetto, bello come il sole ! » (Alfonsi, 1932 :178, « gallicismes, qui gâchent notre dialecte, beau comme le soleil! »). Dans la préface, Alfonsi explique son objectif : tracer un portrait du « dialetto corso » (« dialecte corse ») tel qu'il est parlé en Balagne. L'entreprise originale, à laquelle Alfonsi admet avoir dû renoncer, était toutefois un dictionnaire des différentes variétés du corse. Il s'agit donc finalement d'une sorte de dictionnaire différentiel, qui ne contient pas les unités communes aux locuteurs de l'île. Toutefois, le caractère idéologique et artisanal de l'œuvre laisse pour compte de nombreuses particularités lexicales.

Voici une liste des ressources utilisées :

a) Pour le corse :

- Alfonsi, P. Tommaso, *Il Dialetto còrso nella parlata Balanina* (1932) : https://archive.org/stream/AlfonsiDialettoCorsoNellaParlataBalanina/AlfonsiDialettoCorsoNellaParlata-Balanina_djvu.txt
- Monti, Anton Domenico, « Essai sur les anciennes unités de mesure utilisées en Corse avant l'adoption du système métrique », ADECEC, 1982 : <https://adecec.net/parutions/pdf/anciennesmesures.pdf>
- Base de données Langue Corse : <https://bdlc.univ-corse.fr/bdlc/corse.php>
- Base de données Infcor ACEDEC : [infcor \(adecec.net\)](http://infcor.adecec.net)
- Dictionnaire Glosbe corse-français (dictionnaire communautaire multilingue) : <https://fr.glosbe.com/>
- Lexiques de l'ACEDEC : [lessichi \(adecec.net\)](http://lessichi.adecec.net)
- *Parlami Corsu* (ville d'Ajaccio) : Parlami Corsu - Le site de référence pour apprendre le corse

b) Pour l'italien :

- Tommaseo, Nicolò, *Dizionario della lingua italiana* : www.tommaseobellini.it
- Accademia della Crusca, *Lessicografia della Crusca in rete* : www.lessicografia.it
- Accademia della Crusca, *Il grande dizionario della lingua italiana*, Torino, Utet : www.gdli.it
- Giagnacovo, Maria, *Appunti di metrologia mercantile genovese*, 2014 : https://media.fu-press.com/files/pdf/24/2833/2833_21514

c) Pour le français :

- *Le Dictionnaire de l'Académie française* (7^e (1878) et 8^e éditions (1935)) : <https://www.dictionnaire-academie.fr/>
- Larousse, Pierre, *Grand Dictionnaire universel du 19^e siècle* (1866-1876), Classiques Garnier Numérique (édition abonnés)

- Dictionnaire *Le Grand Robert* (version numérique, 2023)
- *Le Trésor de la langue française informatisé* (1971-1994) : <http://atilf.atilf.fr/>

6. Les poids et les mesures

Le champ lexical des poids et des mesures constitue un domaine spécialement riche en ressources, puisqu'il est décrit dans de nombreux textes, notamment le Muntese, 2014 et Monti, 1982. On peut distinguer dans ce champ lexical trois grands ensembles : (i) les mesures anciennes, notamment celles de poids ; (ii) les mesures métriques, surtout de capacité, et (iii) les récipients employés pour mesurer ou exprimer des quantités.

6.1. Les unités anciennes

En ce qui concerne les mesures anciennes, l'extrait suivant fournit la première attestation du corpus :

(1) Mi dimandi se le uve sono bone. Si, sono bone, e molto a marchato Ecco il prezzo, 30 soldi il **rubo**. Cioe 6, franchi il **cantaro** dunque sapia che ci vole, 13 **rubbi** duva per fare, 10020 litri di mosto. Viene dunque a, 3 soldi il litro. Il prezzo e bono. (30/01/1881, Xavier)⁵

Les termes *cantaro*⁶ et *rubo* désignent des mesures de poids anciennes, utilisées avant la généralisation du système métrique, et donc du kilo. Nous les retrouvons dans une des premières lettres du corpus rédigées en italien par Xavier. Ces mesures sont décrites dans de nombreux dictionnaires et répertoires, la description la plus complète à notre connaissance étant celle de Monti (1982 ; voir aussi Arrighi, 1970 : 186 et suiv.) :

Le *rubu* = 25 livres de Gênes = 8,162kg. [...] Le *cantaru* = 6 *rubbi* = 150 livres de Gênes = 48,972 kg. Il est sensiblement égal à 100 livres poids de marc (48,951 kg). Avec la vulgarisation du kilogramme, le *cantaru* fut compté 50 kg. C'est ainsi que dans un jugement de 1839 on lit : deux quintaux de fer, c.à.d. 100 kg (JUS). Plus tard le mot *cantaru* a été utilisé pour traduire le mot quintal. (Monti, 1982 : 4, nous reproduisons les choix typographiques de l'auteur)

Ces termes sont aussi répertoriés dans le dictionnaire de Tommaseo (en ligne), qui ne les attribue pas à la Corse, mais identifie le *cantaro* comme étant une mesure utilisée en Toscane, à Naples et à Gênes, avec des valeurs différentes :

Cantaro: S. m. (Metron.) Misura di diverse sorte di cose, di peso di libbre cinquanta circa di Toscana, e di maggiore o minore, secondo la diversità de' paesi e delle robe. [...] Detto anche di solo peso. [Camp.]

5. Nous reproduisons la graphie originale. La référence fournit la date de la lettre et l'auteur.

6. Nous utilisons ici la graphie employée par les scripteurs et scriptrices du corpus. Nous n'allons pas aborder ici la question de la graphie de la langue corse, sujet qui dépasse largement les limites de cet article.

Diz. mar. mil. *Cantaro è un peso di dugentocinquanta libre a Napoli, ed a Genova di centocinquanta.*
(Tommaseo en ligne, article *cantaro*)⁷

Le *rubbo* est défini comme une mesure de poids étrangère : « Rubbo: 2. [Cont.] † Nome di peso straniero. Cit. Tipocosm. 343. *Secondo diversi paesi stranieri è (un peso) il battimano, la sporta, il rubbo.* » (Tommaseo en ligne, article *rubbo*)

Le *Dialetto còrso nella parlata Balanina* de P. Tommaso Alfonsi, qui fournit des équivalents italiens pour chaque entrée, définit le *rubu* comme une mesure de poids : « Rubu — m. rubbio : misura di peso ». En revanche, *cantaru* n'est pas une unité de mesure, mais un outil pour peser, c'est-à-dire un type de balance : « Cantàru - m. stadèra (romaine) ». Cet exemple montre les incohérences du dictionnaire d'Alfonsi : pourquoi signaler une unité et pas l'autre ? Est-ce que *rubu* est une spécificité de la Balagne, contrairement à *cantaru* ? Il semblerait bien que l'emploi des deux soit généralisé dans toute la Corse et au-delà.

De son côté, *U Muntese* fournit au lecteur la valeur actuelle du *cantaru* et souligne l'instabilité du *rubbu* : « Cantaru : quintal actuel, 100 kg. À l'origine, une unité de poids valant 8 rubbi, soit près de 100 kg. » (*U Muntese*, article *cantaru*) et « Rubbu : ancienne mesure de poids de diverses valeurs suivant les régions. » (*U Muntese*, article *rubbu*)

Pour une représentation complète de ces deux termes, le chercheur se heurte donc à deux problèmes : la diversité des formes graphiques et la diversité du contenu référentiel. Une manière de les traiter du point de vue lexicographique et terminologique impliquerait une description plus complète, puisant aux différents répertoires, ainsi qu'une mise en réseau des documents déjà disponibles, ce que permet aujourd'hui la numérisation des dictionnaires historiques. De plus, les attestations dans les documents des XIX^e et XX^e siècles de ces termes peuvent contribuer à établir à quel moment ils ont été définitivement remplacés par les unités du système métrique.

6.2. Les unités métriques

Si le kilo fait son apparition dans le corpus en 1892, l'étude de cas la plus intéressante est celle de *décalitre*. C'est en effet dès 1887 (dans les lettres en italien) et à partir de 1905 (dans les lettres en français) que le *décalitre* apparaît, pour mesurer à la fois la capacité d'un récipient – et par conséquent la quantité d'une matière, non seulement liquide, mais également solide, comme la farine, les châtaignes et les noix – mais aussi la surface d'un terrain. Voici des occurrences de ces unités issues du corpus :

7. Nous ne fournissons pas la traduction des articles de dictionnaire puisqu'il s'agit de textes métalinguistiques.

(2) Avendoti rinviato il sacco delle castagne, cioè 10 **decalitri**.⁸ di noci, venerdì passato. (25/08/1887, Xavier)

(3) Comme nous avons décidé, de vous d'envoyer 4 ou 5 **décalitres** de châtaignes (20/11/1905, Pierre François)

(4) Or, la superficie du jardin doit être à peu près une trantaine de **cannes**, et le petit enclos, il n'est pas même 2 **décalitres**. Comme on a toujours dit que le plus grand prix qu'on peut donner au meilleurs terrains est de 100 fr. le **décalitre**, donc, ce lieu ne pourrait pas dépasser de plus de 600 fr. (16/02/1918, Pierre François)

Cette unité est représentée aussi par l'abréviation *dal* :

(5) Chez nous, Suzanne et Ruffine sont allées aux châtaignes à Pioggiola et en ont ramassé une centaine de **dal**. Nous avons fait une vingtaine de **dal** de farine et les autres pour les bêtes. (16/02/1918, Pierre François)

Si, du point de vue formel, l'unité est toujours représentée sous sa forme française, nous savons qu'en corse elle prend la graphie *decalitru* et que sa prononciation en Balagne suit l'accent tonique français et non pas italien (du moins selon Alfonsi⁹) : « Decalìtru — m. decàlitro (dé-calitre). I corsi mettono, alla francese, l'accento tonico su la penultima sillaba. — Cappèllu a de* calitru. cappello a stαιο, a cilindro (chapeau haute forme) ».

Du point de vue de l'usage, le corpus Canioni semble contredire Monti, qui affirme que : « Aux XIX^e et XX^e s. le mot *bacinu* fut conservé pour désigner le décalitre » (Monti, 1982 : 6) et que « Le *bacinu* est donc l'unité principale de mesure des capacités pour les matières sèches » (*ibid.*). Alors que le terme est largement employé dans le manuscrit des délibérations de Pioggiola (Colombani Giaufret et Giaufret, 2022 : 48-49) qui remontent à la fin du XVIII^e siècle, il est totalement absent du corpus Canioni.

L'emploi du décalitre comme mesure de surface, ainsi qu'illustré par l'extrait (4), n'est répertorié dans aucun outil parmi ceux que nous avons consultés : l'entrée apparaît dans la 5^e édition du *Vocabolario dell'Accademia della Crusca* (1863-1923) comme mesure de capacité. L'autre grand dictionnaire italien du XIX^e siècle, le Tommaseo (1865-1879), le définit ainsi : « Sorta di Misura de' Greci pe' liquidi. (Mat.) Misura multipla del litro che consta di dieci litri. V. LITRO. (Mt.) ».

8. Dans les lettres en italien, le point suit souvent les noms d'unités, même lorsque ceux-ci ne sont pas des abréviations.

9. Le dictionnaire d'Alfonsi, publié en 1932, est profondément marqué par l'irrédentisme et par le parti pris de démontrer que le corse, du moins celui qui est parlé en Balagne, est un dialecte de l'italien. Il consacre en effet un appendice de son dictionnaire à lutter contre les emprunts au français (Alfonsi, 1932).

En (4), le *décalitre* apparaît dans le même contexte que *canna*, que nous trouvons chez Monti :

Le *palmu* et la *canna* en carrés sont communément employés, et cela jusqu'au XX^e s. La *canna* (de 6,25 m² aux XIX^e et XX^e s.) sert pour mesurer les planches, les crépis des maisons, les murs, l'estimation faisant intervenir, en plus, l'épaisseur.

Une *canna* de lauze permet la construction de 3 m² de toiture. Dans un devis pour une toiture les *teghje* sont évaluées en *canne*, *palmi* et *palmetti*.

En 1851, à Cervioni, une *canna* de pierres correspond à 36 some. (Monti, 1982 : 5)

Il est toutefois impossible d'établir un rapport entre la *canna* et le *décalitre* et donc de définir quelle est la surface équivalant à ce dernier. D'après la lettre, il existe même un tarif standard au décalitre pour les terrains, ce qui implique un emploi fréquent de ce type de mesure. Une enquête plus approfondie sur le sujet pourra résoudre cette énigme.

Un dernier exemple intéressant concernant les mesures métriques est le suivant :

(6) Sa forte charrette monterait, à elle seule, de 20 à 24 **kintaux métriques** mais comme tes meubles sont volumineux, il descendra ses deux charrettes. (28/05/1919, Pierre François)

Outre la graphie phonétisante, qui semble rapprocher *quintal* de *kilo* avec l'emploi du *k* initial, le texte de Pierre François affiche une précision particulière dans l'emploi du terme *quintal métrique*, qualifié par le TLFi de « rare » (TLFi en ligne, s. v. *métrique*).

6.3. Les récipients

La dernière section que nous souhaitons consacrer aux mesures est celle qui concerne les récipients.

La dame-jeanne est souvent mentionnée comme récipient permettant de transporter du vin ou de l'huile. Le terme se présente sous l'aspect de nombreux allomorphes (*damigeane*, *damijane*) :

(7) Si tu veux un peu de vin, tu n'as qu'à nous envoyé l'argent pour acheter un bari ou une **damigeane** de 50 ou 60 litres, nous te l'enverons. (03/01/1893, Pierre François)

(8) Quant au reste, comme Rose se trouve là, tu nous enverra la **damijane** et même le petit barri. (04/06/1895, Xavier)

Dans ces réalisations écrites, on perçoit une influence de la graphie italienne (qui n'apparaît jamais ni dans les lettres en français ni dans celles en italien), tout en essayant de se rapprocher de la prononciation française.

Un cas complètement différent est celui de *pipe* :

(9) Comme j'ai l'intention d'acheter deux tonneaux, (pour te faire bien comprendre, ici on l'appel «**pipe**»), pour transporter le raisin et en même temps pour faire cuvé le vin. (27/05/1898, Xavier)

Xavier utilise pour ce petit tonneau un terme commun à l'aire génoise, que l'on retrouve chez Giagnacovo : « Essa [la grana spagnola, la *cochenille*] arrivava anche in pipe o botti di peso netto pari a lib. 302-418 ciascuna » (2014 : 233). Cependant, l'emploi de ce terme n'est répertorié dans aucun des dictionnaires ou documents consultés. Il est intéressant d'observer que Xavier (ou Pierre François qui est le rédacteur de la lettre) ressent le besoin d'en expliquer le sens à son fils Jacques Antoine, qui vit déjà sur le continent depuis une vingtaine d'années.

Malgré la relative abondance de descriptions des termes qui réfèrent aux unités de mesure, anciennes et métriques, la création de nouveaux outils de description permettant de mettre en réseaux ceux qui existent déjà serait d'une grande utilité pour le décryptage de la correspondance Canioni, mais aussi de nombreux autres textes écrits en Corse à différentes époques.

7. Maladies

Le deuxième champ lexical que nous allons aborder est celui des maladies, ponctuelles ou épidémiques. La correspondance Canioni nous informe sur les maladies qui sévissent dans le Ghjunsani, sur l'absence de médicaments, surtout pendant la guerre, mais aussi, et c'est ce qui nous intéresse le plus ici, sur les dénominations utilisées pour les désigner.

7.1. Maladies ponctuelles

Dans le corpus, le sujet des maladies est fréquemment abordé, comme il arrive souvent dans les correspondances familiales. Parmi les maladies ponctuelles on trouve *risibula* :

(10) Ma par varie circostanze che mi anno intratenuto prima, e che sono stato malato 15 giorni, da una **risibula** sula fronte, che mi a fatto molto incomoto. (09/02/1884, Xavier)

plobésie :

(11) L'autre jour, nous avons assister a l'enterrement du Pidarella, a succombé par suite de **plobésie**. (10/08/1916, Pierre François)

et *cathare* :

(12) Le reste des parents se porte assez bien, accepté notre mère, il y a quelques jours qu'elle a pris un peu de **cathare**, mélé de constipation qu'il lui fait gardé le lit, mais s'il n'y a rien de contraire nous esperons qu'elle reprendra bientôt sa santé. (20/11/1905, Pierre François).

Parmi ceux-ci, seulement le troisième terme est répertorié : on le trouve dans la base BDLC et dans Inforcor sous la forme *catarru* et dans le dictionnaire Glosbe sous la forme *cataru*. L'équivalent en français est toujours *rhume*, ce qui montre une stabilité sémantique.

Les deux premiers mots, en revanche, ne sont pas présents dans les dictionnaires consultés. Toutefois, on peut faire remonter *plobésie* au français *pleurésie* (il s'agissait d'une maladie mortelle à l'époque), alors que *risibula* semble désigner une excroissance sur la peau. Ce terme est apparemment connu et employé encore aujourd'hui en Corse (communication personnelle de locuteurs). On le retrouve sous cette forme dans un site italien sur la culture traditionnelle de Ferentino (village près de Frosinone), qui l'utilise dans une comptine contre les méfaits des sorcières (Ferentino, en ligne¹⁰). La nomenclature de la 4^e édition du *Dizionario della Crusca* (1729-1738) indique *risipola* (« Spezie di tumore infiammativo con macchia distesa di color rosso vivamente acceso »). Sous son allomorphe *erisipela*, il est présent plus tard dans la 5^e édition (1863-1923), qui le définit comme : « Inflammation della pelle, caratterizzata da colorito rosso carico con enfiamento della parte, e accompagnato per lo più da febbre. Comunemente *Resipola* o *Risipola* ». Dans le Tommaseo, l'article *erisipela* renvoie à *risipola* ou *resipola* : « Inflammation superficiale e parziale della cute, accompagnata quasi sempre da febbre. Nel pop. fior. più com. *Resipola*. Nel ling. scritto e in più dial. *Risipola* ». Les dictionnaires italiens semblent donc hésiter entre une forme plus courante (*risipola* ou *resipola*) pour la 4^e édition du dictionnaire de la Crusca et Tommaseo et une plus étymologique pour la 5^e édition du dictionnaire de la Crusca.

C'est cette dernière que l'on trouve dans les dictionnaires français. Le TLFi présente l'article *érésipèle*, *érésypèle*, *érysipèle* (subst. masc.), contenant la définition suivante : « Dermite streptococcique aiguë caractérisée par la présence d'une plaque rouge et œdématiée limitée par un bourrelet périphérique qui en constitue la zone de croissance. ». L'article est complété par la note sur la prononciation et l'orthographe ainsi que par une note historique et étymologique :

Ac. 1694-1740 donne uniquement *éré-* (écrit *érésipelle* ds *Ac.* 1718 et 1740). *Ac.* 1762-1932 donne en outre *éry-* qu'elle considère néanmoins, dans les éd. de 1835 et de 1878, comme hors d'usage bien que conforme à l'étymol. L'ensemble des dict. enregistrent les 2 formes, avec une préférence pour *éry-* ds LAND. 1834, LITTRÉ, DG, ROB. DUPRÉ 1972, p. 872, rappelle que Racine, parfait helléniste, admettait les 2. La forme *éré-* peut s'orthographier *érésypèle* (cf. SOULIÉ et POURRAT, *loc. cit.*). **Étymol. et Hist.** 1300 *herisipiles* (*La Chirurgie de l'abbé Poutrel*, ms. Reg. lat. 1211, Bibl. du Vatican d'apr. O. SÖDERGÅRD, *Mél. Lecoy*, 1973, p. 544); 1538 *érysipèle* (J. CANAPPE, 14^e *Livre de la méthode thérapeutique de Galien* ds *Fr. mod.* t. 18, p. 271). Empr. au lat. méd. *erysipelas*, « érysipèle, inflammation de la peau » (TLFi, en ligne).

La référence à la fièvre, qui apparaît en italien au XIX^e siècle, est absente de la définition française, mais les autres traits sémantiques indiquent qu'il s'agit bien de la même affection cutanée. La *risibula* du corse est proche de la forme courante italienne, correspondant selon Tommaseo à la fois à la forme employée dans la langue écrite et à celle employée dans plusieurs dialectes, *risipola*, qui

10. <https://www.ferentino.org/antiche-arti-e-mestieri/la-fattucchiera/> [page consultée le 28 mars 2024]

aurait subi les transformations phonétiques qui caractérisent le corse : sonorisation des consonnes intervocaliques ou lénition consonantique (Lai et Dalbera-Stefanaggi, 2005 : 50) et mutation de la voyelle [o] en [u].

7.2. Épidémies

Les maladies épidémiques sont également souvent mentionnées dans le corpus, en particulier la grippe. Si l'on observe les attestations en diachronie, on peut remarquer que la dénomination corse *influenza* (Infor, BDLC), son allomorphe *influenzza* et l'adjectif *enffluenzé* se distribuent entre 1893 et 1912, selon ce schéma : *influenzza* (30/05/1893, Xavier ; 10/05/1909, Pierre François), *influenza* (27/05/1898, Xavier), *enffluenzé* (21/04/1912, Pierre François). D'autre part, la forme *crippe* (30/06/1893, Xavier) ou *grippe* (lettre non signée et non datée attribuée à Pierre François prob. 1918) apparaissent aussi, la forme *grippa* étant également attestée dans la BDLC et dans le dictionnaire Glosbe :

(13) Le climat est assez sain, pas de maladie dans le village axcepté les petits enfants qui ont un peu d'**influenzza** qui les fait beaucoup tousser, l'est une espèce de **crippe**, mais ils s'en tirent bien. (30/05/1893, Xavier)

Ce qui est intéressant ici, c'est l'introduction par Xavier d'une différence sémantique par laquelle l'influenza serait « une espèce de grippe », mais pas exactement la même chose.

8. Nourriture

Le dernier champ lexical que nous allons aborder est celui de la nourriture, qui constitue sans surprise un des domaines les plus traités par les lettres Canioni et un des plus présents dans les répertoires lexicographiques.

L'échange de biens entre Olmi-Cappella et Marseille, où se trouvent Jacques Antoine et sa famille, consiste essentiellement en un échange de produits alimentaires. Il s'agit probablement du domaine dans lequel la surveillance du scripteur s'affaiblit et où le corse apparaît à côté du français, comme dans l'exemple suivant, dans lequel le corse ne se limite pas aux noms propres, mais émerge aussi dans les chiffres (*dui*) :

(14) Nous vous avons expédier un petit colis postal de cinque kilos contenant una **salticca**, deux **fiadelli**, una **pulpetta**, una fetta di **misina** e dui **furmagli**. Voilà le tout, nous avons coupé un petit morceau de **pulpetta** qui passait le poid. (02/02/1914 et 31/01/1915, Pierre François ; les caractères gras sont les nôtres)

Une autre occurrence de *misina* se trouve dans une lettre de Pierre François :

(15) Plus tard, comme la **misina** sera sèche, nous t'en ferons un petit colis. (16/02/1918, Pierre François)

Nous présentons ci-dessous l'analyse lexicale de chaque mot mis en gras en (14) et (15).

Fiadelli : probablement une graphie phonétique de *figatelli*, pluriel de *figatellu*, « saucisse à base de foie » (Infcor). Ce terme est répertorié dans la BDLC, le lexique « La cuisine » de l'ADECEC¹¹ et le site *Parlami Corsu*. La forme graphique représente la transcription de la sonorisation du « t » et de la chute du son [g] tous les deux intervocaliques, ce qui démontrerait que le corse parlé à Olmi-Cappella se rapproche davantage de celui de la Balagne (prononciation [vijad'elu] à Aregno, voir carte BDLC) que de celle de Moltifao (région de Caccia), pourtant plus proche en termes de kilomètres (prononciation [vigad'elu], voir carte BDLC). La consonne initiale « f », en revanche, qui subirait aussi cette mutation, n'imite pas la prononciation.

Furmagli : accompagné du chiffre *dui*, écrit lui aussi en corse, il s'agit du pluriel de *furmagliu*, *formaglui*, *furmaghju* (« fromage »), ainsi que nous en informe Infcor. Également répertoriée dans la BDLC, la variante *furmagliu*, comme l'indiquent ses cartes, est employée à Calvi et à Moltifao, mais, uniquement en Balagne (Sant'Antoninu et Felicetu), la forme toscanisante *casgiu* est plus attestée. Ce lexème est présent également dans ADECEC et *Parlami Corsu*.

Mizina / misina : ce terme est consigné dans Infcor sous les formes *mezina* et *mizina*, qui le définit comme « petit salé, poitrine de porc ». Le lexique français-corse « La cuisine » de l'ADECEC le répertorie avec le sens de « lard maigre ou petit salé », synonyme de *panzetta*, la BDLC comme « petit salé »¹². La variation dans la graphie chez les Canioni semble refléter pleinement celles qui y sont indiquées.

Pulpetta : mot qui possède trois allomorphes pour Infcor (*pulpetta*, *polpetta*, *porpetta*) et plusieurs équivalents en français (« boulette, croquette, godiveau ; quenelle »). La BDLC signale le sens de « boulette de pain » et la forme graphique *pulpetta* à Bisinchi (Castagniccia). Le lexique « La cuisine » de l'ACEDEC lui attribue le seul sens de « quenelle ». Dans *Parlami Corsu*, on ne mentionne que « *pulpetta di salmonu* » (c'est-à-dire, « paupiette de saumon »).

Salticca : terme absent de nombreux répertoires. La chute du « i » semblerait être accidentelle, puisque la forme répertoriée est *salticcia*. La carte de la BDLC indique l'emploi de cette forme (à côté de variantes de *salamu*) sur une ligne qui va de Santa Maria di Lota à Guagno, en passant par Moltifao. La BDLC relève deux sens : « un anneau de figatellu » et « saucisson ». Ici, c'est le deuxième référent que ce mot désigne.

Il est donc possible de conclure que le lexique de l'alimentation se révèle être le plus utile, du moins à l'état actuel de nos recherches, pour étudier le corse : d'une part, c'est dans les passages qui abordent l'échange d'aliments que les scripteurs et scriptrices s'abandonnent le plus facilement à la langue corse et, d'autre part, il s'agit de l'un des domaines les mieux étudiés du point de vue lexical.

11. Il faut toutefois effectuer une recherche plein texte, car le lexique est organisé uniquement par entrées en français. Ceci vaut pour tous les lexèmes suivants.

12. La BDLC signale toutefois que cette forme recouvre cinq sens différents (polysémie et homonymie).

Il est donc possible d'utiliser la correspondance Canioni pour confirmer ou infirmer les hypothèses existantes sur l'extension régionale de formes concurrentes, ou même d'ajouter de nouvelles informations sur leur prononciation.

9. Conclusion

En conclusion, la correspondance Canioni nous a permis de commencer l'analyse des rapports entre italien, français et corse chez des locuteurs et locutrices possédant un répertoire linguistique complexe, composé de trois codes linguistiques en contexte diglossique, le corse étant la langue quotidienne et orale, l'italien et/ou le français la langue de l'écriture. Le corpus se situe dans une période charnière de l'histoire linguistique de la Corse et se compose de scripteurs et scriptrices dont le répertoire varie selon plusieurs facteurs : la génération à laquelle ils appartiennent, leur lieu de naissance, leur genre, leur émigration éventuelle sur le continent, ainsi que leur degré de scolarisation. Le corpus a fait émerger des champs lexicaux qui nous ont semblé particulièrement pertinents pour étudier le contact de ces langues au niveau lexical. Pour ce faire, nous avons consulté les ressources lexicographiques existantes. Parmi les trois domaines analysés ici – les unités de poids et mesure, les maladies et la nourriture – il s'est avéré que le dernier était le mieux documenté : les occurrences présentes dans la correspondance Canioni sont également répertoriées dans la BDLC, Infcor, le lexique de l'ADECEC et *Parlami Corsu*. Ce corpus peut ainsi contribuer à l'amélioration des descriptions lexicographiques de ces termes : les attestations issues du corpus peuvent par exemple confirmer ou infirmer certaines observations qui ont été faites concernant l'aire d'emploi de ces formes ainsi que fournir de nouvelles données sur leur prononciation. L'étude du champ lexical des unités de poids et mesure a mis en lumière la nécessité d'une mise en réseau des ressources existantes pour avoir une description complète de certains items (*rubo, cantaro*, etc.), notamment pour les unités anciennes. Cette étude pourrait donc aboutir à la valorisation des documents disponibles, afin d'unifier les différentes descriptions lexicographiques et de les compléter éventuellement par les informations concernant l'usage de ces termes dont le corpus Canioni témoigne. Enfin, l'analyse du domaine des maladies permet d'attester des termes qui ne sont pas décrits dans les ressources existantes. Tous les termes relevés concourent donc à une description lexicographique plus fine du répertoire linguistique des locuteurs et locutrices corses entre la fin du XIX^e siècle et le début du XX^e à différents niveaux : dans les rapports phonie-graphie et sur le plan lexical, mais aussi en termes d'extension géographique de l'emploi de certaines formes et d'évolution des usages linguistiques en Corse.

Bibliographie

- ARRIGHI, Jean-Marie (2002), *Histoire de la langue corse*, Éditions Jean-Paul Gissenet.
- ARRIGHI, Jean-Marie (2008), « Langue corse : situation et débats », *Ethnologie française*, n° 38, p. 507-516. [En ligne : <https://www.cairn.info/revue-ethnologie-francaise-2008-3-page-507.htm>]
- ARRIGHI, Paul (1970), *La Vie quotidienne en Corse au XVIII^e siècle*, Paris, Hachette.
- BERGERON-MAGUIRE, Myriam (dir.) (2022), « Pour une histoire de la langue par "en bas" », *Linx*, n° 85 [En ligne : <https://doi.org/10.4000/linx.9304>]
- BOULANGER, Jean-Claude (1999) « Un épisode de la néobienséance dans les dictionnaires scolaires : le protocole de rédaction des exemples comprenant des prénoms », *Romanistik in Geschichte und Gegenwart*, n° 1, p. 67-95.
- BRANCA, Marina et Nicolas Sorba (2023), « Un siècle d'évolution de la transmission intergénérationnelle du corse », *Glottopol*, n° 38. [En ligne : <https://doi.org/10.4000/glottopol.3179>]
- BRANCA-ROSOFF, Sonia (1989) « Vue d'en bas : des écrits "malhabiles" pendant la période révolutionnaire », *Langage et société*, n° 47, Maison des sciences de l'homme, Paris, p. 9-27.
- BRANCA-ROSOFF, Sonia et Nathalie Schneider (1994), *L'écriture des citoyens. Une analyse linguistique de l'écriture des peu lettrés pendant la période révolutionnaire*, Paris, Éditions Klincksieck.
- COLOMBANI GIAUFRET, Hélène et Anna Giaufret (2022), « Il manoscritto dei verbali del comune di Pioggiola (Corsica), 1788-1797: analisi testuale e linguistica », dans F. Ferrando (dir.), *Una piccola comunità corsa negli anni della Rivoluzione. Pioggiola attraverso il manoscritto delle delibere 1787-1797*, Palermo, Palermo University Press, p. 35-59.
- COLLECTIF (2014) *U Muntese; Dizzionariu corsu-francese*, Ajaccio, Albiana.
- DAL BO, Beatrice (2019) *Aux frontières de la norme : usages linguistiques de scripteurs peu lettrés dans des correspondances de la Grande Guerre*, Thèse de doctorat, Université Paul-Valéry Montpellier 3 et Università degli studi di Genova.
- DALBERA-STEFANAGGI, Marie-Josée (2002) « Fondements de la partition dialectale », dans Marie-Josée Dalbera-Stefanaggi (dir.), *La langue corse*, Paris, Presses Universitaires de France. « Que sais-je ? », p. 54-68.
- FERRANDO, Francesca (dir.) (2022), *Una piccola comunità corsa negli anni della Rivoluzione. Pioggiola attraverso il manoscritto delle delibere 1787-1797*, Palermo, Palermo University Press.
- KEVERS, Laurent (2021), « L'identification de langue, un outil au service du corse et de l'évaluation des ressources linguistiques », dans Aarne Ranta et Cyril Goutte, *Diversité linguistique. Linguistic Diversity in Natural Language Processing*, p. 13-38. [En ligne : https://www.atala.org/sites/default/files/TAL_62_3_v2.pdf]

- KOTTELAT, Patricia (2009), « Lexicographie et clichés doxiques : la permanence de stéréotypes sexistes dans un dictionnaire d'apprentissage contemporain, le Robert Junior 99 », *Studi e Ricerche, Quaderni del Dipartimento di Scienze del Linguaggio e Letterature Moderne e Comparate dell'Università di Torino*, n° 4, Alessandria, Edizioni dell'Orso, p. 165-185.
- KOTTELAT, Patricia (2010), « Définitions lexicographiques et idéologie : ambiguïtés discursives dans les définitions des races, traces de permanence de stéréotypes racistes ? », *Publif@rum*, n° 11. [En ligne : <https://riviste.unige.it/index.php/publiforum/article/view/1572/1742>]
- LAI, Jean-Pierre et Marie-Josée Dalbera-Stefanaggi (2005), « Continuum dialectal et ruptures linguistiques : "prendre les langues au vol" », *Langage et société*, n° 112, p. 33-56. DOI : 10.3917/lis.112.0033.
- LEHMANN, Alise (1981), *Analyse du discours lexicographique : le corps sexué dans le Petit Larousse illustré de 1906 à 1980*, Thèse de doctorat, Université Paris VIII.
- LEHMANN, Alise (1989), « Les représentations idéologiques dans le discours du dictionnaire. Diversité des lieux et des procédures », *Le Français dans le monde*, n° spécial « Lexiques », p. 106-112.
- LOI CORVETTO, Ines et Annalisa Nesi (1993), *La Sardegna e la Corsica*, Milano, Utet.
- MARCELLESI, Jean-Baptiste (1984), « La définition des langues dans le domaine roman : les enseignements à tirer de la situation corse », *Actes du XVII^e congrès de Linguistique et Philologie Romane 5*, Sociolinguistique des langues romanes, p. 307-314.
- MOLINARI, Chiara et Nadine Vincent (dir.) (2017), *Dictionnaires culture numérique et décentralisation de la norme dans l'espace francophone*, Repères Dorif, n° 14. [En ligne : <https://www.dorif.it/reperes/category/14-dictionnaires-culture-numerique-et-decentralisation-de-la-norme-dans-lespace-franco-phone/>]
- PAVEAU, Marie-Anne (2017), *L'Analyse du discours numérique. Dictionnaire des formes et des pratiques*, Paris, Hermann.
- PELLINO, Santa, Pietro Sichera, Angelo Mario Del Grosso et Daria Spampinato (2022), « Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence », dans Fabio Ciraci, Giulia, Miglietta et Carola Gatto (dir.), *AIUCD 2022 – Proceedings*, Quaderni di Umanistica Digitale, p. 163-168.
- Praxiling - UMR 5267 (2019). *Corpus 14* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) – [En ligne : www.ortolang.fr, <https://hdl.handle.net/11403/corpus14>]
- STADLER Peter, Marcel Illtschko et Sabine Seifert (2016), "Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>", *Journal of the Text Encoding Initiative* 9. DOI : 10.4000/jtei.1433.
- STEUCKARDT, Agnès (dir.) (2015), *Entre village et tranchées. L'écriture de poilus ordinaires*, Uzès, Inclinaison.
- THIERS, Ghjacumu (1978), « Aspects de la francisation au XIX^{ème} siècle, en Corse », *Études corses*, n° 9, Ajaccio, p. 5-40.

WALTER, Richard *et al.* (2018), *L'édition numérique de correspondances. Guide méthodologique*, Consortium Cahier. [En ligne : <https://cahier.hypotheses.org/guides/guide-correspondance>]

Répertoires et dictionnaires en ligne

ACCADEMIA DELLA CRUSCA, *Lessicografia della Crusca in rete* (dizionari 1612, 1623, 1691, 1729-1738). [En ligne : <http://www.lessicografia.it/index.jsp>]

ACCADEMIA DELLA CRUSCA, *Il grande dizionario della lingua italiana* (1961-2002, 2004, 2009), UTET. [En ligne : <http://www.gdli.it/>]

ALFONSI, P. Tommaso (1932), *Il dialetto còrso nella parlata Balanina*, Livorno, Raffaello Giusti. [En ligne : [Full text of “Il Dialetto Corso Nella Parlata Balanina” \(archive.org\)](#)]

Banque de données Langue Corse. [En ligne : <https://bdlc.univ-corse.fr/bdlc/corse.php>]

GIAGNACOVO, Maria (2014), *Appunti di metrologia mercantile genovese*, Firenze, Firenze University Press. [En ligne : https://media.fupress.com/files/pdf/24/2833/2833_21514]

Glosbe - Dictionnaire corse-français. [En ligne : <https://fr.glosbe.com/co/fr>]

INFCOR - Banque de donnée en langue Corse (adecec.net). [En ligne : <https://adecec.net/infor/about.php>]

Lessichi (adecec.net) – [En ligne : <https://adecec.net/adecec/lessichi.php>]

MONTI, Anton Dumenicu (1982), *Essai sur les anciennes unités de mesure utilisées en corse avant d'adoption du système métrique*, ADECEC. [En ligne : [anciennesunitesmesure \(adecec.net\)](http://anciennesunitesmesure.adecec.net)]

TOMMASEO, Nicolò (1861), *Dizionario della lingua italiana*. [En ligne : <http://www.tommaseobellini.it/#/>]

TRECCANI, *Vocabolario*. [En ligne : <https://www.treccani.it/vocabolario/>]

TLFi : *Trésor de la langue française informatisé*, ATILF - CNRS & Université de Lorraine. [En ligne : <http://www.atilf.fr/tlfi>]

Tous les sites Internet ont été consultés le 28 mars 2024.

TITRE: LES ENJEUX D'UN CADRE DE RÉFÉRENCE EN LIGNE POUR LE CORSE

AUTEURS: NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET ALAIN DI MEGLIO,
UNIVERSITÀ DI CORSICA PASQUALE PAOLI

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES
PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT,
UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 103-120

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22397](http://hdl.handle.net/11143/22397)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22397](https://doi.org/10.17118/11143/22397)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS
ATTRIBUTION 4.0 INTERNATIONAL.

Les enjeux d'un cadre de référence en ligne pour le corse

Nicolas Sorba, Università di Corsica Pasquale Paoli

Alain di Meglio, Università di Corsica Pasquale Paoli

Résumé :

À travers l'évaluation des enjeux concernant un projet de création d'un cadre de référence linguistique, cet article traite de la situation actuelle de la langue corse et de ses défis en matière d'équipement, d'élaboration et de reconnaissance. Il souligne que l'usage du corse a évolué ces dernières années et a connu une diversification de ses pratiques. Le texte mentionne également la nécessité d'une description et d'un recensement des termes et expressions nouvelles utilisés dans de nouveaux domaines (médias, politique et école notamment) ainsi que la question, plus générale, de la revitalisation linguistique. Le projet porté par l'UMR CNRS 6240 LISA, de l'Université de Corse, et qui est l'objet principal de l'article, tente de répondre à toutes ces problématiques.

Mots-clés :

Langue corse, numérisation, normativisation, revitalisation linguistique, élaboration linguistique

Abstract :

This article evaluates a project aimed at creating a linguistic reference framework for the Corsican language. The Corsican language has evolved in recent years, with a diversification of its practices. The article highlights the challenges faced by the language in terms of resources, development, and recognition. It also emphasizes the need to describe and catalog new terms and expressions used in emerging fields, as well as the more general question of linguistic revitalization. The main focus of the article is a project led by the research laboratory at the University of Corsica, which aims to address these challenges.

Keywords :

Corsican language, digitization, normativization, language revitalization, language by development

1. Introduction

Le projet de Cadre commun polynomique de référence de la langue corse¹ (Sorba, 2023²) a pris la forme, dans un premier temps, d'un dictionnaire en ligne corse/français³ issu essentiellement d'*U Muntese*, dictionnaire papier connu et reconnu par le grand public. Les enjeux autour de ce projet sont nombreux. D'un point de vue sociolinguistique, depuis le début du mouvement de revitalisation du corse initié dans les années 1970, la langue des insulaires est un sujet socialement vif et la situation linguistique peut être définie comme conflictuelle. Ce projet a comme objectif de répondre à une attente sociale importante à propos du corse puisqu'il existe désormais sur l'île un large consensus autour de son emploi à l'avenir.

Deux grandes enquêtes sociolinguistiques dévoilent que près de 90 % des personnes interrogées sont en faveur d'un bilinguisme corse/français (Collectivité Territoriale de Corse, 2013 et Collectivité de Corse et MSC, 2021). D'un point de vue intrinsèquement linguistique, la majeure partie du patrimoine immatériel de ce que l'on pourrait appeler la langue corse traditionnelle a été abondamment décrite comparativement au corse appliqué dans des espaces langagiers qui se sont ouverts à la suite du mouvement de revitalisation débuté dans les années 1970. Autrement dit, il y a une carence analytique concernant les productions langagières réalisées dans les médias, les écoles ou à l'Assemblée de Corse, pour ne citer que des lieux où la présence du corse est régulière et emblématique. Décrire le corse du XXI^e siècle est incontestablement une lacune à combler ; le projet de Cadre commun polynomique de référence ambitionne d'y répondre. Enfin, d'un point de vue numérique, la conversion d'un dictionnaire du format papier au format numérique apparaît comme une étape vers un déploiement qui profitera des divers atouts qu'offre l'espace digital. Une esquisse des fonctionnalités envisagées sera également présentée pour justifier le choix et exposer les potentialités des outils de ce type en ligne.

2. Une question préalable : où est le corse ?

Aujourd'hui, si le corse ne souffre plus aucune ambiguïté sur sa qualification de « langue », il convient pour autant d'atteindre un consensus plus large sur ce que cette langue est et représente désormais. Par exemple, tout en respectant les atlas linguistiques qui ont fortement contribué à décrire la langue corse, nous soutenons qu'un découpage en sous-systèmes définis, territorialisés et comptés est de nos jours insuffisant, voire caduc, dans une Corse actuelle où les lieux géographiques tendent à être

1. S'inscrivant dans un projet plus large intitulé B3C (Boost cultural competence in Corsica), bénéficiant d'une aide CPER (Contrat de plan État-région), le travail d'élaboration du cadre a débuté en janvier 2021, pour une fin de la première phase en décembre 2023, suivie d'une mise en ligne.

2. L'article (Sorba, 2023) est le premier consacré au projet présenté ici. Il s'appuie notamment sur une enquête de terrain dont l'objectif est de comprendre l'utilisation des rares dictionnaires de corse en ligne et d'en saisir les attentes.

3. Au moment de l'écriture de cet article, la mise en ligne est en cours. Le dictionnaire sera accessible via la plateforme M3C : <https://m3c.universita.corsica/s/fr/page/home>.

dépassés par les lieux sociaux et professionnels et où la mobilité des locuteurs a fortement brouillé le lien dialecte/territoire.

À l'instar de nombre de langues minoritaires et/ou minorées, l'épreuve du dépassement de la diglossie qui consiste à investir des domaines d'usages nouveaux est loin d'être un parcours aisé. Dans les cas où les langues minoritaires se dotent d'une planification linguistique⁴, on peut se heurter à un problème d'acceptabilité. Faute de réel centre de décision, la promotion d'une langue adaptée à de nouveaux domaines, comme les médias ou l'enseignement, est le plus souvent contestée par la ou les partie(s) des locuteurs qui estiment ne pas être respectées dans leur variété ou tout simplement dans l'usage traditionnel qu'ils font de la langue. On se trouve alors face à un paradoxe : comment faire sens avec une langue minoritaire qui, par définition, n'a pas vocation à répondre à des fonctions utilitaires qu'une ou d'autres langues présentes et parlées/écrites occupent déjà ? En même temps, les revendications logiques de nombre de défenseurs, promoteurs ou militants de la langue incitent à ces fonctions de vitalité linguistique voire de normalisation, selon les acceptions les plus courantes de ces notions.

Jacques Thiers, s'appuyant sur l'exemple d'une radio publique locale, dès les premières années d'usage d'une langue corse « socialement circulante », pointe d'emblée que « la spectacularisation soudaine d'un corse jusqu'alors contenu dans les contextes et usages non formels, s'est faite sous un jour insolite et déconcertant, celui d'une parole vivante, employée mais non reconnue, car niée ou sanctionnée par le discours puriste et normatif. » (Thiers, 2008 : 213)

Ce cheminement difficile s'exerce entre la reconnaissance interne d'un nouveau corpus, qui doit accepter qu'une langue qui retrouve une vitalité linguistique puisse évoluer, et la reconnaissance externe de son statut⁵ (Kloss, 1969), afin que la langue puisse trouver les conditions légales et réglementaires de son usage.

Pour le corse, ce parcours qui tente d'inverser la tendance à l'érosion de son usage courant se fait sur des bases assez fragiles mais réelles. Il est clair que le corse a, depuis l'extension à la Corse de la loi Deixonne en 1974, intégré les divers niveaux de l'enseignement (presque exclusivement public en Corse) de la maternelle à l'université. Et ce, même si l'on ne peut encore qualifier le système « d'éducation bilingue » dans la mesure où le corse peine à trouver un usage de langue d'enseignement, se cantonnant le plus souvent au statut de langue enseignée. De la même façon, le corse est présent

4. Nous faisons référence ici au document de planification « Lingua 2020 » voté par l'Assemblée de Corse en décembre 2014 et toujours en cours aujourd'hui, malgré les changements de majorité territoriale intervenus à partir de 2015 et confirmés en 2017 et 2021.

5. Il est fait référence ici à la distinction encore en usage aujourd'hui et opérée par Heinz Kloss (1969) entre *language corpus planning* (planification du corpus), qui porte sur la forme de la langue (écriture, néologie, standardisation) et *language status planning* (planification du statut), qui correspond notamment aux fonctions de la langue et ses rapports avec d'autres langues.

dans les médias, de la presse écrite à l'audiovisuel, de façon régulière, depuis la création d'antennes régionales en 1982.

Ces mutations sont en cours depuis environ cinquante ans et les facteurs agissant sur l'évolution des pratiques sont assez nombreux : statut des langues, qualité de l'offre d'enseignement, volume d'exposition et d'immersion, mouvements de populations, motivation individuelle, lieu de résidence, fréquentations, etc. Ces évolutions impliquent un déplacement de la présence du corse, lié à la mobilité et à la qualité de ses locuteurs : vers les villes, vers les bassins de populations de la plaine et du littoral (au détriment de la montagne et du rural), dans l'enseignement et les médias et dans les pratiques culturelles en général.

Il est de l'ordre de l'évidence que le corse, au cours d'une relative mais réelle revitalisation, opère désormais une déterritorialisation dialectale pour une reterritorialisation telle que nous l'avons évoquée plus haut. Au-delà des classiques mobilités (exode rural, concentrations urbaines et littorales), du déséquilibre démologique d'un régiolecte par rapport à l'autre et des évolutions démographiques⁶, ce mouvement tient compte aussi de territoires nouveaux tels que les médias ou les réseaux sociaux.

3. La tentation linguistique et les tensions sociolinguistiques

Par ses atlas, ses enquêtes et ses publications, tout le travail amorcé dès les années 1970 par la linguistique corse a pu consigner aujourd'hui la majeure partie du patrimoine immatériel de ce que l'on pourrait appeler la langue corse traditionnelle, même si l'on doit admettre que le terme ne peut répondre à une définition précise. Cela permet de connaître par le menu l'ensemble exhaustif de la variation dialectale géographique du corse et de l'expliquer par des études diachroniques et comparées. Il s'agit là d'un corpus fondamental qui constitue le socle de la langue corse.

Pour autant, à partir des premières revendications, on assiste à un ensemble de concessions de la part de l'État qui vont permettre au corse d'accéder à des domaines d'usage dont il était exclu auparavant. Nous pensons notamment à l'école et aux médias publics⁷ à partir de 1982.

Après plus de cinq décennies d'usage du corse dans ces nouveaux domaines, un certain nombre de constats s'impose :

- Il apparaît aujourd'hui que ce corse utilisé notamment par les médias, l'enseignement et la littérature contemporaine n'a pas été à ce jour l'objet d'une description ou d'un réel recensement des termes et expressions nouvelles qu'il utilise ;

6. « Au 1^{er} janvier 2021, 347 597 personnes résident en Corse. L'île enregistre la plus forte croissance démographique des régions métropolitaines avec 1,0 % de population en plus en moyenne annuelle depuis 2015. Ce dynamisme est exclusivement dû aux arrivées de nouveaux habitants plus importantes que les départs de l'île. » (Cadart et Pétilion, 2023)

7. La télévision régionale France 3 Corse et la radio locale de service public RCFM (Radio Corsica Frequenza Mora) pour les principaux médias, même si l'on doit aussi tenir compte de radios privées ou associatives qui ont fait le choix du bilinguisme français/corse (Voce Nustrale ou Alta Frequenza, par exemple)

- Ce nouvel espace de parole recourt à une pléthore de dictionnaires sans que ces derniers eussent ouvert un corpus de termes néologiques utiles et harmonisés ;
- Il s’ensuit une élaboration de termes nouveaux qui peut varier selon le média, l’auteur, le journaliste, etc. et, subséquemment, une variation qui peut s’avérer bien des fois incohérente ;
- Nous assistons donc à une redialectalisation du corse qui se fait sur des territoires non plus géographiques mais professionnels ou sociaux ;
- Le besoin d’un corpus de références lexicales plus complet et harmonisé se fait désormais sentir.

Le diagnostic le plus immédiat est porté par Stella Retali-Medori, linguiste à l’Université de Corse :

Actuellement, en raison du processus de distanciation entre corse et italien et dans la perspective d’élaboration linguistique, la variété de référence italienne qui pourrait constituer une source d’enrichissement naturelle du lexique corse devient périphérique, voire exogène. C’est désormais le français, aux structures linguistiques divergentes (prosodie, phonologie, morphologie : *genèsi* vs *gènesi*, *tupunimu* vs *topònimu*, *crisa* vs *crisi*, etc.), qui exerce une force centripète sur le corse. Ce changement de langue-toit affecte d’une part la capacité de création lexicale et, d’autre part, les processus d’assimilation de l’emprunt, en particulier du lexique d’origine grecque et latine. (Retali-Medori, 2023)

Dans son article, la linguiste donne un ensemble d’exemples sur lesquels nous ne saurions porter de jugement qualitatif. Il s’avère toutefois que ce diagnostic, au demeurant historiquement fondé, pose intuitivement le principe de « langue-toit » (ici l’italien⁸) comme une condition *sine qua non* de la possibilité actuelle d’une néologie pour le corse, du moins afin que l’équipement nécessaire d’un corse moderne demeure cohérent.

Or le fait de rendre le corse dépendant d’une « langue-toit » fait débat. Prenons le contrepoint de la sociolinguistique par la plume de J. Thiers qui vient largement nuancer le propos :

Tout porte donc à penser que pendant une longue période qui prend naissance au Moyen âge, la relation du corse à l’italien a été perçue sous l’angle d’un rapport complexe qui ne pourra être élucidé qu’avec les progrès des études linguistiques et historiques, mais qui apparaît irrésistiblement marqué d’un rapport contrasté d’inclusion et d’exclusion. Cette contradiction s’explique tant par les vicissitudes géopolitiques de cette partie de la Méditerranée tyrrhénienne que par la volonté récurrente des Corses de se constituer en entité culturelle autonome. (Thiers, 2008 : 45)

Tout l’appareil critique et conceptuel forgé par la sociolinguistique corse à partir des années 1970 ne vient pas contredire cette assertion. Le concept phare de cette période, le corse langue polynomique (Marcellesi, 2003 ; Di Meglio, 2020), se construit en grande partie sur la base de l’appareil théorique de

8. Le toscan s’est imposé comme norme standard de l’italien moderne grâce à l’éclat littéraire du parler de la ville de Florence, incarné notamment par Dante, Pétrarque et Boccace.

Jean-Baptiste Marcellesi et notamment sur le concept d'individuation⁹ qui marque bien cette notion d'autonomie linguistique déclarée par les Corses¹⁰. J.-B. Marcellesi ajoute que « son opposé est la satellisation, qui est l'acceptation de la subordination de la langue régionale » (2003 : 169).

Doit-on pour autant en déduire que le « processus de distanciation entre corse et italien » évoqué par S. Retali-Medori est à prendre dans une totalité stricte ? Cela implique-t-il l'exclusion de tout recours à l'italien dans la recherche d'un corpus de néologismes pour le corse ? Nous notons pour notre part que les entreprises ou les démarches en ce sens sont beaucoup plus nuancées.

Certes le recours systématique à la « langue-toit » n'est pas la règle absolue. Mais, pour autant, certains néologismes à rapprocher de l'italien sont aujourd'hui entrés dans la langue corse élaborée ou professionnelle (deux exemples parmi bien d'autres, le *cahier/u quaternu* en corse, *il quaderno* en italien ; l'échantillon/*u campione* en corse, *il campione* en italien).

Lorsque l'on observe les propositions néologiques pour le corse, il est facile de constater que l'italien est loin d'être exclu et continue de constituer une référence et un recours importants. Pour cela, il suffit de s'intéresser à un travail en cours, concernant la création de néologismes, de la part du réseau public Canopé de Corse¹¹. Par exemple, dans son outil proposé en ligne, si l'on recherche les équivalents de certains termes, voici le résultat : pour « amidon », *amidu* apparaît, plus proche de l'italien *amido* que *amidonu*, qui était une possibilité et qui aurait été plus proche du français ; pour « Arabie Saoudite », l'accent tonique est placé comme en italien *Aràbia Saudita*, évitant ainsi l'influence du français qui le déplacerait (*Arabia Saudita*), de même pour « arbitre » (*àrbitru* et non *arbitru*) ; pour « crocodile », *cuccudrillu* est choisi, plus proche de l'italien *cocodrillo*, alors que *cruccudillu* était aussi une possibilité ; pour « loupe », *lente* apparaît, emprunté à l'italien. Entre proximité linguistique originelle et démarche de propositions modernes, une enquête pourrait démontrer que le rapport à l'italien est bien présent même s'il n'est pas systématique et encore moins unique. Ce constat serait encore plus évident si l'enquête se concentrait dans les lieux évoqués précédemment, l'école et les

9. L'individuation est l'identification, par une communauté, d'une langue comme distincte des autres langues : « l'individuation sociolinguistique est le processus par lequel une communauté ou un groupe tend à systématiser ces différences, à les sacrifier, à les considérer comme déterminantes, à en faire un élément de reconnaissance » (Marcellesi, 2003 : 169).

10. On consultera à ce sujet l'enquête ATP (action thématique programmée) CNRS 91 1164, de 1984, qui définit les indicateurs de corsité par une enquête auprès des locuteurs corses et affirme l'individuation sociolinguistique du corse.

11. Son recensement de termes corses en ligne, utiles à l'enseignement bilingue, est présenté ainsi : « Ce référentiel lexical met à la disposition des enseignants, des élèves, et plus largement de tout public, un corpus de mots spécialisés dans plusieurs domaines disciplinaires (histoire, géographie, informatique, mathématiques, chimie, art...). Ces milliers des mots résultent de choix, et parfois de créations, effectués par les chargés d'études du CRDP et du Canopé de Corse et par les Inspecteurs pédagogiques régionaux en langue corse, lors de la traduction des différents manuels pour l'enseignement bilingue en langue corse. Ce lexique sera régulièrement complété, au fur et à mesure que d'autres domaines seront abordés dans le cadre des missions du Canopé de Corse. » (Extrait de : <https://www.reseau-canope.fr/nous-trouver/academie/corse.html>) (consulté le 30/03/2024).

médias. Dans les discussions informelles, le recours au lexique français semble, effectivement, plus marqué.

Évitant la notion de « langue-toit » qui maintient tout de même la langue corse à l'état de dialecte d'une autre langue (ou de satellisation), la sociolinguistique préfère avoir recours ici à des concepts tels que langues collatérales (Eloy, 2004), langues voisines ou langues proches. Une pratique féconde et respectueuse peut aussi s'inscrire dans le cadre des « langues partenaires¹² » (Matthey et Conti, 2015).

Pour notre part, nous estimons que le concept de langue polynomique, tout en affirmant une individuation et une autonomie du corse, permet d'assumer une juste proximité et un rapport à l'italien totalement dédramatisé et normalement ouvert :

Nous sommes arrivés à un moment de l'histoire de la langue corse où les locuteurs, les usagers, tout utilisateur du corse, par les attitudes tirées de la polynomie sociolinguistique, peuvent établir des passerelles vers les langues proches dans la mesure où les interactions et/ou les fréquentations culturelles l'exigent. (Di Meglio, 2019)

4. Le besoin d'un cadre de référence pour la langue corse polynomique

Le pari d'une langue corse équipée et apte à servir dans tous les domaines de la communication, quel que soit le registre recherché, est donc posé. Le projet de Cadre commun polynomique de référence de la langue corse s'inscrit résolument dans une approche sociolinguistique qui consiste à considérer la langue corse telle qu'elle est pratiquée de nos jours et notamment depuis son développement dans les domaines d'usage d'où elle était exclue dans son rapport diglossique avec le français, langue dominante et hégémonique.

C'est une approche *élaborationniste*¹³ qui est annoncée dans le projet. Dans la dynamique de son officialisation et en perspective d'une coofficialité, l'objectif visé par ce cadre est d'offrir un espace accessible aux usagers et aux professionnels qui trouveront une base de référence, notamment en matière de néologie et d'adaptation de la langue corse aux besoins professionnels actuels. Ce projet de portail gratuit en ligne est mené au sein de l'UMR 6240 CNRS LISA par des chercheurs de l'Université de Corse.

12. « La notion de partenariat permet d'envisager une relation égalitaire entre les langues, alors que l'histoire enseigne plutôt que, en situation de contact, la langue la plus prestigieuse absorbe ou fait reculer les autres, ou que des conflits linguistiques apparaissent dans le cas où les langues en contact bénéficient d'un même statut. » https://liseo.france-education-international.fr/index.php?lvl=notice_display&id=39172 (consulté le 30/03/2024)

13. La vision « élaborationniste » promue par J. Thiers dans les années 1980 prend appui sur la théorie de l'élaboration linguistique forgée par le linguiste H. Kloss en 1967 qui pose un modèle de « développement de la langue » en fonction des domaines d'usages investis.

La numérisation de ce cadre s'appuie dans un premier temps sur un dictionnaire historique du corse, *U Muntese*, qui a connu deux éditions papier (1985 et 2014). Ce dictionnaire étant le fruit d'un travail d'actualisation au sein de l'Université de Corse, il s'est agi de partir d'une base assez complète de la langue corse, augmentée de ce travail universitaire. Cette première base, d'environ 40 000 lemmes, a été incrémentée dans une phase liminaire qui a consisté à mettre au point un formulaire de saisie ayant vocation à accueillir l'ensemble des données lexicales et d'apurer les lemmes d'éventuelles scories (redondances, incohérences, répétitions, etc.).

D'autres phases sont prévues afin d'atteindre un recueil de données significatif et proposer un outil fonctionnel. Le recensement des données existantes et des recueils concernant la néologie s'avère alors important. Ce travail a déjà débuté et s'est conclu par un partenariat avec le Canopé de Corse qui a fourni un travail conséquent sur la création de néologismes. L'intégration de ce corpus dans le projet est désormais effective. Enfin, grâce à la mise en place d'une équipe mixte composée d'experts de l'université et de professionnels (médias, enseignement, littérature), il s'agira de faire des choix pertinents et argumentés pour atteindre l'objectif d'harmonisation visé. Le projet comprend aussi une numérisation de textes pertinents, issus d'un usage professionnel et littéraire de la langue corse, entre 1970 et nos jours, à des fins de requête pour évaluer les usages, les occurrences, les fréquences afin d'en tirer les choix les moins arbitraires possibles.

Au-delà des discussions ou débats concernant l'adossement de la langue corse à une autre langue, le projet de cadre commun de référence a le mérite de poser la question de la normativisation du corse, langue polynomique. Autrement dit de poser la question de son corpus augmenté des termes utiles à la société moderne. Il est vrai que, d'un strict point de vue démocratique, la question de la norme imposée ou partagée se pose. On l'aura compris, il s'agit ici de lancer un processus qui consiste à bâtir du commun dans un cadre respectueux des usages et des initiatives en cours.

Dans l'accompagnement de la mutation d'un corse traditionnel ancré dans ses territoires géographiques vers une langue corse commune, présente dans tous les domaines d'usage, il s'agit de trouver le juste compromis à une alternative : faut-il se doter d'outils institutionnels centralisés de normalisation linguistique (imposition d'une norme commune et standard à partir d'un centre politique faisant autorité avec un risque d'académisme, de surnorme ou d'acceptabilité difficile) ou bien laisser la langue se construire uniquement par l'usage (avec le risque de perte de commun et de normes anarchiquement territorialisées) ?

Une première réponse peut être trouvée ici :

En sociolinguistique, la « variété légitime » ou la « langue standard » sont considérées comme des surnormes. Ce code dominant, nécessairement arbitraire, se définit par un certain nombre de prescriptions en matière de phonologie, de lexique, de syntaxe et de style. En Corse, les débats, pour savoir quelle était la forme la plus « légitime », étaient nombreux et n'ont jamais débouché sur quelque chose de concret. (Sorba, 2018)

Le projet de Cadre commun polynomique de référence de la langue corse part du principe que la construction du commun peut revendiquer une méthode qui va exactement à l'inverse de celles qui prescrivent depuis un centre pour imposer au commun. Ici, ce sont les domaines du besoin de langue qui transforment, font évoluer, font interroger les locuteurs sur leurs propres choix, inventent et/ou proposent et c'est ensuite un centre qui devient demandeur de ces expériences normatives pour en tirer un examen, une synthèse et tenter de stabiliser du commun. C'est une démarche qui va de l'usage vers l'institution et non l'inverse.

5. Les enjeux numériques

Comme indiqué précédemment, *U Muntese*, le document servant de base au projet de cadre commun, est un dictionnaire papier. Si la première phase s'est attachée en grande partie à la numérisation de ce document de référence, les enjeux à venir reposent aussi sur la conception d'un « véritable » outil en ligne qui profite de toutes les potentialités du numérique.

5.1. Un choix comme une évidence

Le dictionnaire du corse le plus consulté actuellement est numérique (Sorba, 2023), il s'agit d'*Infor*, utilisable en ligne depuis 1999, créé par l'association ADECEC, fondée en 1970. Pour le projet du cadre commun de référence, dans un premier temps, le choix s'est porté sur un portail en ligne pour des raisons d'accessibilité assez évidentes. L'objectif d'un tel outil est, effectivement, d'être le plus répandu possible et que les utilisateurs puissent le consulter le plus souvent possible. Un dictionnaire en ligne, comparativement à une version papier, bénéficie, effectivement, d'une disponibilité constante, à la seule condition, bien sûr, d'avoir une connexion internet. Toujours concernant l'accessibilité, le coût de l'accès au dictionnaire peut aussi être un argument. En effet, rares sont les dictionnaires papier gratuits alors que les supports numériques proposent généralement des versions gratuites. Même les versions payantes des dictionnaires en ligne sont fréquemment moins chères que leurs équivalents papier. Pour le cadre de référence du corse, l'accès sera totalement gratuit. Enfin, l'accessibilité peut être aussi rapprochée de la facilité à procéder à la recherche. Les carences des compétences dans l'utilisation des dictionnaires semblent toujours plus prégnantes (Tremblay et al., 2018 ; Bishop, 2000). Dans un dictionnaire en ligne, pour les utilisateurs, la barre de recherche facilite, effectivement, la démarche. Très rapidement, bien d'autres facteurs que l'accessibilité se sont dévoilés et ont renforcé la décision du dictionnaire en ligne. Sans être exhaustif, voici les six principaux : des fonctionnalités supplémentaires ; des mises à jour rapides ; un espace rédactionnel illimité ; la personnalisation ; l'interactivité ; l'intégration d'autres outils numériques.

Concernant les fonctionnalités supplémentaires, les dictionnaires en ligne peuvent offrir des fonctionnalités que le format papier ne peut proposer. Un des exemples concrets est la possibilité de disposer d'une prononciation audio avec notamment les différentes prononciations du corse poly-

nomique. Celle-ci accompagnerait la transcription phonétique souvent peu maîtrisée par le grand public. Le corse ne bénéficiant pas pour le moment de lecteur automatisé, un enregistrement sera donc nécessaire. D'autres exemples peuvent être avancés ici comme les illustrations de phrases, les hyperliens (renvoyant d'un article à l'autre mais aussi à d'autres textes en ligne, dictionnaires et bases de données), les synonymes et les antonymes, les traductions en différentes langues en quelques clics.

Les mises à jour ne sont pas un élément négligeable. Elles peuvent, en effet, refléter l'évolution constante de la pratique langagière ou simplement corriger une coquille. Un dictionnaire en ligne peut être mis à jour facilement et rapidement pour refléter ces changements, ce qui n'est pas le cas d'un dictionnaire papier.

L'espace rédactionnel illimité offre de nombreuses perspectives tant au niveau des explications étymologiques que pour toutes les autres explications que l'utilisateur viendrait rechercher ou que les chercheurs veulent mettre à disposition du public.

Certains dictionnaires en ligne proposent de personnaliser l'expérience utilisateur, en enregistrant des mots favoris ou en créant des listes de vocabulaire. Cette personnalisation permet aussi à l'utilisateur de mieux cibler les résultats attendus. Elle laisse également des fonctions potentielles très variées qui pourront évoluer.

Les dictionnaires en ligne peuvent inclure également des éléments interactifs (comme des quiz ou des jeux) pour aider à l'apprentissage et à la mémorisation des mots. Dans le cadre d'un projet comme celui qui nous intéresse ici, qui plus est pour une langue dont l'objectif principal demeure sa revitalisation, ce type d'outils peut s'avérer utile et performant.

Pour conclure avec les six principaux facteurs qui accompagnent l'accessibilité pour justifier le choix d'un format numérique, nous pouvons ajouter l'intégration d'autres outils. Les dictionnaires en ligne peuvent être intégrés à d'autres outils numériques, comme les traitements de texte ou les navigateurs web, ce qui permet de vérifier rapidement la définition d'un mot sans avoir à quitter l'autre tâche en cours d'exécution.

Enfin, même si, en raison de l'évolution sociolinguistique du corse, les représentations diglossiques sont décadentes, certaines subsistent encore. Le corse peut, en effet, être encore caractérisé, notamment dans les discours épilinguistiques, comme un élément du passé. Nous pouvons donc ajouter à tous ces arguments pour le format numérique, l'apport de représentations positives pour une langue minoritaire, et encore souvent minorisée, comme le corse (Sorba, 2023).

5.2. Un travail diligent

Si le projet est porté par des linguistes et sociolinguistes, concernant la transformation du format papier au format numérique, l'apport d'un spécialiste de ces supports, notamment un développeur informatique, a été indispensable. Le dictionnaire servant de base ayant été conçu uniquement pour le format papier, sans imaginer une transformation numérique, le travail effectué a été fastidieux. Et ce, notamment parce que les nombreuses coquilles numériques, invisibles sur le format papier, le sont ostensiblement lorsque la version devient digitale.

Le passage d'un dictionnaire papier à un dictionnaire en ligne, via un fichier au format Adobe InDesign remis par l'éditeur, implique plusieurs étapes, dont certaines sont spécifiques au développement web, d'autres à la recherche d'éléments dans un document balisé. Voici un aperçu des tâches principales réalisées par le développeur dans le cadre de ce projet : conversion des fichiers InDesign au format HTML ; identification de la structure permettant de décrire un lemme (classe grammaticale, définitions, usages, etc.) ; mesure des variations (changements de formalismes, balises différentes, etc.) afin d'évaluer l'homogénéité.

Si les données du dictionnaire *U Muntese* étaient déjà numérisées, une structuration des données telles que voulues dans la version numérique a été nécessaire afin que celles-ci puissent être utilisées dans un dictionnaire en ligne. Cela a impliqué notamment de convertir les données en un format adapté aux fonctionnalités préalablement identifiées pour la version numérique du dictionnaire. Il a fallu aussi, rapidement, créer une interface utilisateur pratique afin d'individualiser les différentes données relatives aux lemmes et les rubriques associées aux lemmes (celles choisies en fonction du support papier existant sont : *Definizioni* – Définition ; *Detti* – Adages/Proverbes ; *Citazioni* – Citations). Dans le même temps, les nombreux tests et débogages ont permis un déploiement sur un serveur interne. L'élaboration d'un formulaire de saisie pour l'incrémentation de nouveaux lemmes, ou pour la modification de lemmes existants, a également fait partie des tâches importantes. L'intégration des néologismes de la base de données du Canopé de Corse a dû connaître une mise à niveau des informations relatives à ces nouveaux mots. C'est à partir de ce formulaire de saisie que ce travail s'est réalisé. Parmi les tâches importantes également, la recherche par autocomplétion ou la compatibilité avec les supports mobiles, notamment smartphones. La création d'une application mobile à partir du portail en ligne existant devrait, en effet, voir le jour rapidement. Lors de la conception du dictionnaire numérique, cette transformation a été anticipée. Dans l'optique d'une diffusion la plus large possible, cette conversion est évidemment fondamentale.

Au moment de l'écriture de cet article, l'intégration du dictionnaire au sein de la M3C (Médiathèque culturelle de la Corse et des Corses) est en cours. Le dictionnaire sera alors visible par le plus grand nombre. La M3C est une plateforme de science ouverte du laboratoire Lieux, Identités, eSpaces, Activités (UMR CNRS 6240 LISA) :

Ses données, entièrement consacrées à la Corse, sont en accès libre. La M3C se déploie au sein d'une plateforme numérique en ligne, d'une salle de consultation documentaire, d'une salle de numérisation ainsi que d'une salle d'immersion virtuelle. Son objectif est de diffuser des savoirs scientifiques et culturels à destination du plus grand nombre, tant bien au sein des acteurs du monde universitaire que de la société civile. Elle constitue un axe de recherche transversal auquel s'adosse la production interdisciplinaire de l'UMR 6240 LISA.¹⁴

6. Travaux en cours et perspectives d'avenir

La mise en ligne des premiers travaux effectués (numérisation du dictionnaire *U Muntese* et ajouts de néologismes émanant de la base de données du Canopé de Corse) ne signifie nullement la fin du projet. Il reste de nombreuses tâches à accomplir tant concernant le remaniement de la base initiale du dictionnaire *U Muntese* qu'au niveau du développement des fonctionnalités du dictionnaire en ligne.

6.1. Remaniement du support initial

Comme évoqué précédemment, le dictionnaire *U Muntese* n'a pas été conçu ni pensé comme un outil numérique. Par conséquent, cela engendre un certain nombre de choix faits qui ne correspondent pas aux besoins d'un dictionnaire en ligne. C'est essentiellement pour cette raison qu'un remaniement et des ajouts de données descriptives des mots sont nécessaires avant de se projeter vers d'autres fonctionnalités.

Un élément original de la normativisation du corse demeure la réponse polynomique donnée à la problématique de la variation linguistique (Sorba, 2019). Or les entrées du dictionnaire *U Muntese* appartiennent dans leur grande majorité au régiolecte septentrional, laissant généralement de côté les régiolectes central et méridional. Il convient donc, afin d'être en conformité, à la fois, avec le choix normatif des Corses et avec les usages informels comme formels du corse, de donner une version polynomique du dictionnaire en ligne. Par exemple, cela revient à dire qu'il ne faut pas que *fiume* (régiolecte septentrional) soit la seule entrée pour désigner « fleuve ». Les entrées *fiumi* (régiolecte central) et *fiumu* (régiolecte méridional) doivent donc être créées afin d'être en conformité avec les principes du concept de polynomie. Cette différenciation morphologique n'est pas la seule qui distingue les régiolectes entre eux. Le plan lexical est aussi concerné et ne doit pas être négligé. En effet, pour le grand public, il s'agit du domaine qui « retient le plus l'attention » (Comiti, 2011 : 38), et ce, malgré le fait que celui-ci soit « moins performant que la phonologie ou la morphologie dans le processus de différenciation linguistique » (Comiti, 2011 : 38). Ainsi, pour désigner « narine », le terme *nara* (régiolectes central et septentrional) ne doit pas être la seule entrée, comme c'est le cas actuellement. La proposition *nilfoghja* (régiolecte méridional) doit aussi apparaître, et ce, bien que la version papier ait déjà pris souvent en considération la variation d'un point de vue lexical. Par exemple, on

14. <https://m3c.universita.corsica/s/fr/page/home>

constate que pour « chien », *ghjacaru* et *cane* sont proposés, comme pour « grotte », *sapara* et *grotta* le sont aussi. Or *ghjacaru* et *sapara* (tous deux d'origine prélatine) appartiennent majoritairement aux régiolectes central et méridional alors que *cane* et *grotta* sont, majoritairement, du régiolecte septentrional. Nous indiquions précédemment que les entrées du dictionnaire *U Muntese* appartiennent dans leur grande majorité au régiolecte septentrional, sur le plan lexical, cette hégémonie est donc plus relative.

Cette tâche, de donner une version polynomique à la version numérique, renvoie directement à un autre travail qui doit être mené. En effet, n'ayant pas été au bout des propositions concernant la variation linguistique, il est souvent mal aisé, dans la version papier de *U Muntese*, de distinguer les variantes et les synonymes. Il est vrai que la différence peut parfois être subtile, notamment pour une langue polynomique. Lorsque la variation est phonologique ou morphologique, il n'y a pas de difficultés majeures, il s'agit alors de variantes. En revanche, lorsqu'elle est lexicale, il ne serait pas incongru de la considérer comme un synonyme. En effet, dans le contexte polynomique et la situation sociolinguistique qui s'instaure, notamment les tendances au mixalecte qui se renforcent (Sorba, 2021), la frontière devient de plus en plus ténue entre un synonyme et une variante. Il faudra donc opérer des choix qui répondent à la fois à des considérations linguistiques et à des facteurs extérieurs, notamment sociaux. À partir de cette nouvelle répartition, la création de renvois vers les variantes et les synonymes nous projette déjà dans un exemple de la phase à venir de développement du dictionnaire et de ses fonctionnalités.

Le pluriel du lemme est une autre donnée importante. La variation linguistique sera ici aussi à intégrer. Nous pouvons illustrer notre propos à travers deux exemples. Le premier, le mot *donna* qui connaît deux formes possibles au pluriel, la version du régiolecte septentrional *donne* et celle des régiolectes central et méridional *donni*. Second exemple, le mot *locu* qui peut connaître trois formes : *lochi* pour le régiolecte central et septentrional, *loca* pour le régiolecte méridional et *loche*, qui est une forme irrégulière que l'on peut rencontrer dans le régiolecte septentrional, qui compte donc deux tournures possibles.

Un autre ajout envisagé est l'étymologie du lemme. Cette action pourrait d'ailleurs s'appuyer sur l'élaboration en cours du *Dictionnaire dialectal et étymologique des parlers corses*, travail mené au sein du même laboratoire de recherche :

En 1915, treize ans après le décès de F. D. Falcucci, P. E. Guarnerio a réalisé l'édition du *Vocabolario dei dialetti, geografia e costumi della Corsica*. C'est ainsi que F. D. Falcucci dotait le corse de son premier dictionnaire. Cent ans après l'édition du *Vocabolario*, mais aussi après plus d'un siècle et demi d'études réalisées sur les parlers corses dont Falcucci a été l'initiateur (1875), il nous a semblé que la richesse des matériaux collectés, ainsi que les analyses qui ont été effectuées depuis, pourraient justifier l'émergence d'un nouveau projet lexicographique. Nous allons exposer les orientations ainsi que les choix méthodologiques que nous entendons donner au *Dictionnaire Dialectal et Étymologique des Parlers Corses* (DDEPC). (Retali-Medori et Tognotti, 2018 : 241)

Comme indiqué précédemment, un autre ajout important envisagé est l'écriture phonétique des mots adjoint à sa prononciation sonore. Là aussi, la portée polynomique de la langue corse devra trouver toute sa place avec plusieurs propositions de prononciations qui respectent les différents régiolectes du corse.

Enfin, le dernier remaniement envisagé est la traduction inversée, un élément essentiel pour le public visé par le projet. En effet, le dictionnaire *U Muntese* ne propose que des entrées à partir du corse. L'objectif du dictionnaire numérique est de pouvoir rechercher des mots autant à partir du corse que du français.

6.2. Développement du dictionnaire et de ses fonctionnalités

Dans un premier temps, c'est à partir de la même base de données que des perspectives s'ouvrent. C'est pourquoi, on peut considérer que le développement du dictionnaire et de ses fonctionnalités est un prolongement des remaniements et ajouts présentés dans la partie précédente.

Afin d'offrir un aspect ludique et convivial au dictionnaire, dans la partie « Définition », deux éléments pourraient être ajoutés. Les deux encadrés proposeraient les mots les plus recherchés du mois (le nombre de mots restant à être défini) et la définition d'un « mot du jour », qui serait ici un des mots les moins recherchés par les utilisateurs. Une partie dédiée à la synonymie et une autre à l'antonymie devraient également trouver leur place au sein de la plateforme pour accompagner la définition du mot. Ces choix d'extension sont, bien évidemment, guidés par les attendus des utilisateurs potentiels du dictionnaire (Sorba, 2023).

Dans un autre temps, et surtout à partir d'autres données que celles déjà acquises par l'Université de Corse, des perspectives nouvelles peuvent être envisagées. La partie « *Detti* », dédiée aux adages et proverbes, devrait connaître un renforcement parémiologique. Ces ajouts passeraient par l'acquisition de nouvelles bases de données. Pour le corse, les travaux réalisés à ce sujet ne manquent pas. Dans la même optique de renforcement, l'incrémentation de nouveaux lemmes, notamment des néologismes (là aussi par l'achat de données nouvelles ou par la création de nouveaux partenariats) fait partie des objectifs à long terme du projet. Ce dictionnaire en ligne s'inscrivant dans le projet de cadre commun de référence pour le corse, il doit répondre à diverses carences patentes au sujet du corse. C'est pourquoi, au-delà de l'aspect lexicographique, une portée plus large pourrait être donnée. Inspirée par la Banque de dépannage linguistique, outil proposé par l'Office québécois de la langue française, une banque d'aide linguistique pourrait être envisagée pour traiter des demandes récurrentes concernant la morphosyntaxe, l'orthographe, la grammaire, l'utilisation des majuscules ou la ponctuation. Cet outil serait essentiellement nourri par les différents travaux menés au sein de l'UMR CNRS 6240 LISA (projets divers concernant le corse ou travaux de thèses). Pour compléter cette banque d'aide linguistique, un conjugueur pourrait trouver sa place spécifique avec la possibilité d'un renvoi à partir du dictionnaire d'usage.

Parmi les éléments défailants actuellement à l'équipement de la langue corse, un outil de traduction de phrases ou de textes courts serait, à plus long terme, un élément important. Plusieurs pistes sont explorées pour répondre à ce manque mais, au moment de l'écriture de cet article, rien d'abouti ne peut être présenté ici. Nous pouvons faire le même constat, tant au niveau du manque que des réponses potentielles apportées, concernant un outil de traduction opérée à partir d'exemples de contextes.

Enfin, pour conclure cette partie sur les perspectives de développement, un enrichissement des exemples et citations illustreurs des différents emplois des lemmes fait également partie des objectifs affichés.

7. Conclusion

Le projet de Cadre commun polynomique de référence de la langue corse est un projet ambitieux qui vise à répondre à une attente sociale importante concernant la langue corse. Le projet se concentre sur une partie du corse lacunaire dans sa description et ses outils de référence. Le travail se focalise essentiellement sur les productions langagières réalisées dans les nouveaux espaces langagiers qui se sont développés depuis une cinquantaine d'années. Cette diversification des fonctionnalités du corse a engendré, dans le même temps, son élaboration linguistique et le recul du rapport diglossique avec le français. Ce projet prend donc en considération aussi bien les enjeux sociolinguistiques, linguistiques que numériques. En somme, le Cadre commun polynomique de référence de la langue corse est un pas important vers la reconnaissance et la valorisation de la langue corse dans la société contemporaine.

Bibliographie

- ATP CNRS 91 1164 (1986), *L'individuation sociolinguistique corse*, IEC-GRECSO-IREC, Corte, Université de Corse.
- BISHOP, Graham (2000), « Developing learner strategies in the use of dictionaries as a productive language learning tool », dans *The Language Learning Journal*, volume 22, n°1, p. 58-62. <http://dx.doi.org/10.1080/09571730085200261>.
- CADART, Nathalie et Xavier Pétilion (2023), « En Corse, 347 597 habitants au 1^{er} janvier 2021 », dans *INSEE flash Corse*, n°83, décembre, INSEE.
- Collectivité de Corse et MSC (2021), *Réalisation d'une enquête sociolinguistique sur la langue corse. Rapport final*. Collectivité territoriale de Corse.
- Collectivité Territoriale de Corse (2013), *Inchiesta sociolinguistica nant'à a lingua corsa : cumpetenze, usi è ripresentazione / Enquête sociolinguistique sur la langue corse : compétences, usages et représentations*.
- COMITI, Jean-Marie (2011), *A pratica è a grammatica*, Ajaccio, Albiana. [1996].
- DI MEGLIO, Alain (2020), « La langue polynomique, le concept phare de l'appareil critique corse », dans *International Journal of the Sociology of Language*, n° 261, février, p. 45-65.
- DI MEGLIO, Alain (2019), « Le corse, langue synaptique », dans Amane Gogorza et Wanda Mastor (dir.), *Les langues régionales et la construction de l'État en Europe*, Paris, LGDJ éditions, p. 187-198.
- ELOY, Jean-Michel (dir.), (2004), *Des langues collatérales*, Paris, L'Harmattan.
- KLOSS, Heinz (1969), *Research Possibilities on Group Bilingualism. A Report*, Quebec, CIRB.
- KLOSS, Heinz (1967), « Abstand languages and Ausbau languages », dans *Anthropological linguistics*, n° 9, p. 29-71.
- MARCELLESI, Jean-Baptiste, Thierry Bulot et Philippe Blanchet (2003) *Sociolinguistique : épistémologie, Langues régionales, Polynomie*, Paris, L'Harmattan.
- MATTHEY, Marinette et Virginie Conti (dir.), (2015), *Cohabitation des langues et politique linguistique. La notion de « langue partenaire »*. Actes du colloque de Champéry (2014), Neuchâtel, Délégation à la langue française.
- RETALI-MEDORI, Stella (2023), « Mots savants et néologie en corse, un lexique en devenir ? », dans *Neologica*, n° 17, *Néologie et langues régionales*, p. 97-114.
- RETALI-MEDORI, Stella et Aurelia Ghjacumina Tognotti (2018), « Un nouveau projet lexicographique pour la Corse : le dictionnaire dialectal et étymologique des parlers corses. » dans Stella Retali-Medori et Aurelia Ghjacumina Tognotti (dir.) *Actes du colloque de lexicographie dialectale et étymologique en l'honneur de Francesco Domenico Falcucci (Corte – Rogliano, 28-30 Octobre 2015)*, Alessandria, Edizioni dell'Orso.

SORBA, Nicolas (2023), « Un cadre commun de référence linguistique en ligne pour le corse », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86, (2023-1).

SORBA, Nicolas (2021), « Conceptualisation de productions linguistiques en mouvement : le mixalecte » dans Nicolas Sorba (dir.), *Transmettre les langues pourquoi et comment ? Questions politiques, familiales et migratoires*, Louvain-la-Neuve, EME Éditions, p. 345-266.

SORBA, Nicolas (2019), *Parlons polynomie*, Paris, L'Harmattan.

SORBA, Nicolas (2018), « Interventions sociolinguistiques en situation conflictuelle : la polynomie comme réponse au standard », dans Carmen Alen Garabato, Henri Boyer, Ksenija Djordjevic et Bénédicte Pivot (dir.), *Identités, conflits et interventions sociolinguistiques*, Limoges, Lambert-Lucas, p. 433-447.

THIERS, Jacques (2008), *Papiers d'identité(s)*, Ajaccio, Albiana. [1989].

TREMBLAY, Ophélie, Catherine Fréchette-Simard et Isabelle Plante (2018), « Les enseignants et le dictionnaire : sentiments, attitudes motivationnelles, connaissances déclarées et pratiques personnelles d'utilisation », dans *Formation et profession*, n° 26(3), janvier. DOI:[10.18162/fp.2018.452](https://doi.org/10.18162/fp.2018.452)

TITRE: LES DICTIONNAIRES DE LANGUE DES SIGNES FRANÇAISE À L'ÈRE DU NUMÉRIQUE :
UN PREMIER ÉTAT DES LIEUX

AUTEURS: AMÉLIE JOSSELIN-LERAY, CLLE, CNRS ET UNIVERSITÉ TOULOUSE – JEAN JAURÈS ET
JÉRÉMIE SEGOUAT, CLLE, CNRS ET UNIVERSITÉ TOULOUSE – JEAN JAURÈS

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES
PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT,
UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 121-152

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22399](http://hdl.handle.net/11143/22399)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22399](https://doi.org/10.17118/11143/22399)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS
ATTRIBUTION 4.0 INTERNATIONAL.

Les dictionnaires de langue des signes française à l'ère du numérique : un premier état des lieux

Amélie Josselin-Leray, CLLE, CNRS et Université Toulouse – Jean Jaurès

Jérémie Segouat, CLLE, CNRS et Université Toulouse – Jean Jaurès

Résumé :

Alors qu'il existe une littérature sur les dictionnaires électroniques de diverses langues des signes, aucune étude métalexigraphique ne s'est penchée sur la langue des signes française (LSF). Cet article vise donc à recenser et analyser méthodiquement les ressources lexicographiques électroniques existantes de la LSF, en lien avec les nouvelles possibilités offertes par le numérique dans la conception, la création, la diffusion et l'utilisation de ces ressources. Il s'organise en trois temps : la première partie énonce quelques particularités de la LSF et leur impact sur la lexicographie ; la seconde livre un premier état des lieux des dictionnaires numériques de LSF existants, et la troisième s'attache à présenter, au moyen de premières enquêtes exploratoires, les points de vue des concepteurs de ces dictionnaires et de certains de leurs utilisateurs. La conclusion fournit quelques pistes offertes par le numérique pour une potentielle amélioration de ces dictionnaires.

Mots-clés :

Langue des signes française, dictionnaire électronique, langue minoritaire, études empiriques, constitution de ressources lexicographiques

Abstract :

While the design and use of electronic dictionaries have been documented for many sign languages, no metalexigraphic study has focused so far on French Sign Language (LSF). The aim of this paper is therefore to analyze existing electronic lexicographical resources for LSF, in relation to the new possibilities offered by digital technology in the design, compilation, dissemination and use of these resources. It is divided into three parts: the first one outlines some of the specificities of LSF and their impact on dictionary-making; the second one provides a first overview of existing LSF e-dictionaries; and the third one presents, through exploratory surveys, the points of view of the compilers of these dictionaries and of some of their users. The concluding section outlines some of the ways in which digital technology can be used to improve these dictionaries.

Keywords :

French Sign Language, e-dictionaries, minority language, empirical studies, dictionary-compiling

1. Introduction : langue des signes française et métalexicographie

Lors de ces dix dernières années sont apparues pour diverses langues des signes (LS) de nouvelles ressources lexicographiques disponibles en ligne (McKee et McKee, 2013; Brosens *et al.*, 2021 : 51), qui ont fait suite au travail lexicographique séminal de Trevor Johnson sur la langue des signes australienne (Auslan¹). On note toutefois qu'alors qu'il existe une littérature assez récente et conséquente sur les dictionnaires électroniques de plusieurs LS (LS flamande : Brosens *et al.*, 2021; LS néo-zélandaise : McKee et McKee, 2013; LS danoise : Kristoffersen *et al.*, 2012...), aucune étude métalexicographique, à notre connaissance, ne s'est penchée de manière systématique sur la lexicographie électronique de la langue des signes française (LSF). Une seule étude, la thèse de doctorat de Cédric Moreau (Moreau, 2012), a fait un recensement en 2012 des « 105 sites internet à travers le monde qui présentent un lexique en langue des signes » (Moreau, 2012 : 20), parmi lesquels figurent des « dictionnaires » de LSF, mais la perspective adoptée dans son travail n'est pas à strictement parler métalexicographique². Moreau fait toutefois le constat suivant dès le début de son analyse :

Parmi les 105 sites étudiés, un grand nombre emploie l'appellation « dictionnaire », or seulement 17 d'entre eux proposent des définitions écrites. Les outils à disposition sont donc pour la grande majorité des lexiques et non des dictionnaires. Parmi ces 17, seulement 7 donnent des définitions en langue des signes. La quantité de dictionnaires en langue des signes est donc extrêmement faible. De plus le nombre de définitions ne dépasse pas 5000, nous sommes bien loin des 135 000 proposées par le dictionnaire Larousse en ligne (Larousse, s.d.). Enfin parmi ces 7 sites, uniquement 2 proposent un accès aux contenus via les paramètres linguistiques³ de la LSF.

Dans cette thèse se trouvent résumées un certain nombre de problématiques liées à la lexicographie de la LSF : nombre limité des ressources, aspect protéiforme de l'objet « dictionnaire » pour cette langue avec une très grande variabilité en termes de macrostructure (taille et portée de la nomenclature) ou de microstructure (types d'information consignée, présence minimale d'une définition...), mode d'accès à l'information (via le français ou la LSF – avec les difficultés que cette dernière implique d'un point de vue technique).

1. La ressource en ligne Auslan Signbank (<https://auslan.org.au/>, consultée le 12 juillet 2024) est l'édition multimédia mise à jour et élargie du dictionnaire d'Auslan produit originellement en version papier en 1984 par Trevor Johnson.

2. La thèse de Cédric Moreau visait essentiellement à répondre à la question suivante : comment permettre aux locuteurs de diverses langues des signes, dont la LSF, de pouvoir accéder aux « structures de transfert » (qui font intervenir des éléments linguistiques infra-lexicaux, n'appartenant pas au lexique), qui n'apparaissent pas dans les dictionnaires de langue de signes, à travers un outil adapté qui offrirait cette opportunité (Moreau, 2012 : 16-17). L'objectif final était d'offrir des perspectives pour la constitution d'un lexique/dictionnaire à entrée directe en langue des signes.

3. Selon Sallandre (2020) qui cite Cuxac, les unités lexicales peuvent être décomposées en quatre éléments manuels : (1) configuration manuelle, (2) emplacement, (3) mouvement, (4) orientation de la paume de la main. À ces paramètres s'ajoutent des paramètres non manuels que constituent les expressions du visage, la direction du regard, l'orientation du buste, et le mouvement des lèvres (Sallandre, 2020), qui permettent aux unités d'être instanciées en discours.

Au vu des nouvelles possibilités offertes par le numérique dans la conception, la création, la diffusion, l'utilisation des ressources répertoriant le lexique en LSF, la présente étude vise principalement à actualiser, en le complétant, le recensement méthodique des dictionnaires de LSF effectué par Bonnal-Vergès (Bonnal-Vergès, 2006 ; 2012). De manière générale, les travaux de cette autrice se penchent en effet uniquement sur les dictionnaires papier, et adoptent une perspective diachronique, retraçant l'histoire de la lexicographie de la LSF, en se focalisant sur les dictionnaires des XVIII^e et XIX^e siècles : les premiers dictionnaires dits « des Sourds et Muets » (dictionnaire de l'abbé Jean Ferrand, publié en 1784⁴ ; dictionnaire de l'abbé de l'Épée, publié en 1787, par exemple). Parmi les treize dictionnaires de LSF du XX^e siècle analysés par Bonnal-Vergès (2006), les derniers sont les dictionnaires papier d'IVT (1983-1990 et 1997-1998) – qui constituent une référence, comme le sous-entendent Sennikova et Garcia (2018 : 142⁵). Ces études de 2006 et 2012 amorcent toutefois la transition vers l'ère numérique puisque Bonnal-Vergès mentionne également l'unique dictionnaire de LSF sur CD-ROM, le *Fournier* (2000), que nous avons décidé d'inclure dans notre corpus, et qu'elle avait choisi d'analyser « pour ses qualités particulières et pour les perspectives qu'[il] offre, en vue d'une nouvelle dictionnaire » (Bonnal-Vergès, 2006 : 166).

Notre étude, qui s'inscrit dans le cadre de l'approche sémiologique telle qu'elle a été initialement développée dans les travaux de Christian Cuxac (Cuxac, 2000), s'organise en trois temps : la première partie énonce quelques particularités de la LSF et leur impact sur la lexicographie de cette langue ; la seconde livre un premier état des lieux des dictionnaires numériques de LSF existants, et la troisième s'attache à présenter, au moyen de premières enquêtes exploratoires, les points de vue des concepteurs de ces dictionnaires et de certains de leurs utilisateurs. Nous livrons en conclusion quelques pistes offertes par le numérique pour une potentielle amélioration de ces dictionnaires.

4. Le dictionnaire de Ferrand est probablement l'un des tout premiers dictionnaires de langue des signes, d'après Kristoffersen *et al.* (2012 : 294). Fischer (2014 : 199) affirme également qu'on « considère le dictionnaire de l'abbé Charles Michel de l'Épée (1712-1789) comme le plus ancien dictionnaire de signes pour sourds ». Bonnal-Vergès (2014 : 242) explique toutefois qu'on « peut discuter le fait que ces dictionnaires soient les premiers dictionnaires de langue des signes en France, puisqu'existaient depuis le XI^e siècle des listes de signes monastiques utilisées par les Cisterciens » (la plus longue ne comprenant que 365 signes).

5. Ces autrices mentionnent que le *Dictionnaire bilingue de la LSF* (Girod *et al.*, 1997 ; Galant *et al.*, 2013), c'est-à-dire le dictionnaire IVT, constitue le corpus lexicographique le plus substantiel existant pour la LSF dans un format papier, recensant environ 4500 signes lexicaux. À propos d'IVT (International Visual Theatre), Bonnal-Vergès (2012 : 280) indique : « Troupe de théâtre fondée en 1977 par des sourds et des entendants américains et français, IVT a été la première association à organiser des cours de LSF pensés selon une didactique de type FLE, à une vaste échelle, avec des professeurs sourds. La création de leur dictionnaire de LSF, conçu également dans une collaboration entre sourds et entendants sur une longue période, est aussi pionnière : leur dictionnaire, qui a déjà connu trois éditions et des dictionnaires spécialisés annexes, fait autorité ».

2. Particularités de la LSF et impact sur la lexicographie

Nous exposons ici brièvement les enjeux sous-jacents à l'élaboration de ressources lexicographiques pour la LSF, puis les défis, de taille, auxquels sont confrontés les concepteurs de telles ressources, avant d'établir quel impact cela a sur le paysage lexicographique actuel.

2.1. Enjeux liés à la constitution de ressources lexicographiques en LSF

La constitution de ressources lexicographiques en LSF doit être envisagée au regard de différents enjeux, qui sont au moins au nombre de trois.

Premièrement, du fait que la LSF est une langue minoritaire à l'histoire douloureuse (Encrevé, 2008; Cantin et Encrevé, 2022), la compilation d'une ressource lexicographique la décrivant répond avant tout à une volonté de reconnaissance de cette langue par la communauté non signante. Ce n'est en effet qu'en 1999 que la LSF a été pour la première fois considérée, quoique de manière implicite, comme une « langue de France⁶ », et ce n'est que depuis 2005 qu'elle a été reconnue comme une langue à part entière permettant de disposer des enseignements (inscrite dans la loi d'accessibilité de 2005⁷). À la différence des autres langues de France comme l'occitan, le corse, ou encore le breton, la LSF n'est pas une langue dite régionale. En effet, c'est une « *langue minoritaire sans territoire géographique* : les Sourds⁸ pratiquant la LSF se regroupent de manière volontaire au moyen d'actions *communautaires, associatives et éducatives*⁹ présentes sur tout le territoire » (Boutora et Fusellier-Souza, 2009). Comme toute langue minoritaire, elle se situe en revanche toujours par rapport à une langue dominante, en l'occurrence la forme écrite et parlée du français. Les autres langues des signes qui sont typiquement utilisées par une communauté minoritaire au sein d'une plus large communauté possédant une langue écrite et orale sont dans le même cas : c'est le cas du British Sign Language (BSL) par rapport à l'anglais britannique par exemple comme l'écrivent Kristoffersen *et al.* (2012 : 294).

6. Dès 1999, sans l'inclure explicitement, le rapport de Bernard Cerquiglini mentionne ainsi en note 7 que la LSF pourrait être considérée comme « langue de France », en tant qu'elle serait, comme par exemple le berbère, une « langue sans territoire » (<https://www.cairn.info/revue-etudes-et-documents-berberes-2000-1-page-215.htm>; consulté le 12 juillet 2024). Ce n'est toutefois qu'en 2002 qu'il l'ajoute officiellement à sa liste (Kremnitz, 2020).

7. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT00000809647/> [Consulté le 10 décembre 2023]

8. Nous utilisons la majuscule à *Sourd* dans la même optique que Sennikova et Garcia (2018 : 132) : « La majuscule [...] réfère à l'appréhension de la surdité et de la pratique d'une LS comme une réalité anthropologique et culturelle, par opposition à la minuscule (*sourd*) qui renvoie à la surdité dans sa seule dimension physiologique : la déficience auditive. »

9. Nous avons ajouté les italiques.

Deuxièmement, un dictionnaire de LSF veut également répondre à un besoin de transmission dans un cadre pédagogique. Suite au « réveil sourd » dans les années 1970¹⁰, et au développement de l'enseignement de la LSF, des besoins en termes de matériel pédagogique ont émergé : c'est dans ce cadre qu'est né, notamment le dictionnaire d'IVT évoqué plus haut (Moody *et al.*, 1998 : 3637) ; les besoins pédagogiques sont toujours présents, et vont croissant¹¹.

Troisièmement, en corollaire à ce dernier besoin, la constitution d'une ressource lexicographique répond, de manière générale, à un besoin de standardisation dans le cadre de l'apprentissage, qu'il convient toutefois de nuancer dans le cas de la LSF. En effet, la constitution de ressources lexicographiques en LSF vise actuellement plus à rassembler des données existantes qu'à tenter de les uniformiser ou de les standardiser : la variation est très présente pour la LSF (ce qui transparait dans les ressources : le dictionnaire *Elix* propose par exemple plusieurs signes pour [MAMAN]¹²). Ceci n'est toutefois pas considéré comme un problème par la communauté sourde française, mais plutôt par d'autres catégories de locuteurs de la LSF, les entendants (parmi lesquels les interprètes), et ce constat vaut pour d'autres LS, comme l'écrivent Napier *et al.* : « Deaf people find their signed languages and variation unproblematic, while hearing teachers of deaf children, second language learners, and interpreters are the ones who have difficulties.¹³ » (2015 : 280)

2.2. Les défis d'une consignation lexicographique

Un certain nombre de défis (dont certains s'appliquent en réalité à toutes les langues des signes) viennent par ailleurs complexifier la possibilité de constitution de ressources lexicographiques pour la LSF. Zwitserlood fait en effet le constat suivant : « The life of a lexicographer is not an easy one. I dare say that the life of sign language lexicographers is even more difficult: they have to cope with a number of additional challenges to the ones that all lexicographers face.¹⁴ » (2010 : 444)

10. On appelle communément réveil sourd le mouvement social et culturel « pour la réhabilitation de la LSF [qui] émerge dans les années 1970-1980 à travers une double dialectique modifiant les conditions d'existence des sourds [qui] ont pris publiquement la parole en réinvestissant leur langue séculièrement interdite. Ce mouvement a enclenché le renversement du stigmate de la déficience auditive à l'expression d'une 'culture sourde'. La revendication pour le bilinguisme synthétise la portée de ce mouvement. » (Kerbourc'h, 2006).

11. En atteste le projet actuel de création d'un master MEEF (Métiers de l'enseignement, de l'éducation et de la formation) mention premier degré avec perfectionnement en LSF et pédagogie pour les jeunes Sourds, pour lesquels l'INSPE (Institut National Supérieur de Professorat et de l'Éducation) de Toulouse serait pilote.

12. Convention de notation : à l'intérieur des crochets et en majuscules est noté le mot de la langue vocale écrite (le français) auquel le signe en LSF est censé correspondre.

13. Notre traduction : Les personnes sourdes ne trouvent pas problématique la variation dans leur langue des signes; ce sont les entendants qui enseignent aux enfants sourds, ceux qui apprennent la langue de signes comme une langue seconde et les interprètes qui la perçoivent comme une difficulté.

14. Notre traduction : Un lexicographe n'a pas la vie facile. J'oserais avancer que la vie des lexicographes qui travaillent sur les langues des signes est encore plus difficile : ils doivent faire face à un certain nombre de défis qui s'ajoutent à ceux auxquels tous les lexicographes sont confrontés.

Nous avons pu dénombrer cinq de ces défis :

(1) Mise par écrit d'un signe qui est en quatre dimensions

La quadridimensionnalité de la LSF s'inscrit fondamentalement dans l'espace puisqu'elle est produite par « de nombreux articulateurs corporels, les mains et les bras bien sûr, mais aussi le visage, les épaules, la tête, le regard et plusieurs composantes faciales, qui peuvent être activés plus ou moins simultanément et perçues [sic] par les yeux » (Braffort, 2022). L'écriture de la LSF, et de toutes les LS, constitue un défi majeur en lexicographie, comme le constatent parmi bien d'autres Kristoffersen *et al.* (2012 : 295). Étant une langue visuo-gestuelle, la LSF, comme toutes les autres LS, ne dispose en effet pas de véritable forme écrite, même s'il existe divers systèmes de notation (Dalle et Dalle, 2013 : 192). Aucun système, toutefois, n'est véritablement satisfaisant, comme l'indiquent parmi d'autres Cuxac et Antinoro Pizzuto, « surtout pour ce qui concerne la transcription de séquences entières de signes, et en particulier lorsqu'il s'agit de structures linguistiques complexes, fortement iconiques et multilinéaires, qui n'ont pas d'équivalents dans les L[angues] V[ocales].¹⁵ » (2010)

L'avènement du numérique a permis l'utilisation de vidéos, qui est le moyen auquel les dictionnaires électroniques en LS ont désormais massivement recours, comme notre analyse le montrera plus bas. La vidéo permet en effet de montrer la production de signes d'une manière naturelle et précise, et elle peut être appréhendée directement par l'utilisateur, qui n'a pas besoin d'avoir des connaissances poussées, par exemple, sur les différents paramètres des signes (Kristoffersen *et al.*, 2012 : 296). Les nouvelles technologies présentent en effet des avantages majeurs pour les LS (voir, par exemple, Vanbrugghe *et al.*, 2010 ; McKee et McKee, 2013). Elles permettent tout d'abord de passer du signe dessiné sur papier, parfois accompagné de descriptions rédigées sur la manière de l'exécuter, tel qu'on peut le trouver dans les dictionnaires de LSF (la Figure 1 montre ainsi comment le signe pour [ENSEIGNER] est représenté dans divers dictionnaires du XIX^e siècle), à de la vidéo (la Figure 2 est une capture d'écran de la vidéo qui présente le signe pour [ENSEIGNER] dans le dictionnaire en ligne *Elix*).

15. Pour une présentation détaillée des différents systèmes de notation utilisés dans les dictionnaires de LSF du XIX^e, voir l'analyse exhaustive de Bonnal-Vergès (2012 : 272-279). D'autres systèmes de notation existent, et ont été expérimentés par le passé comme ceux de l'Américain Stokoe, le système allemand HamNoSys et le système SignWriting. Voir à ce sujet la synthèse présentée par Kristoffersen *et al.* (2012 : 296-298).

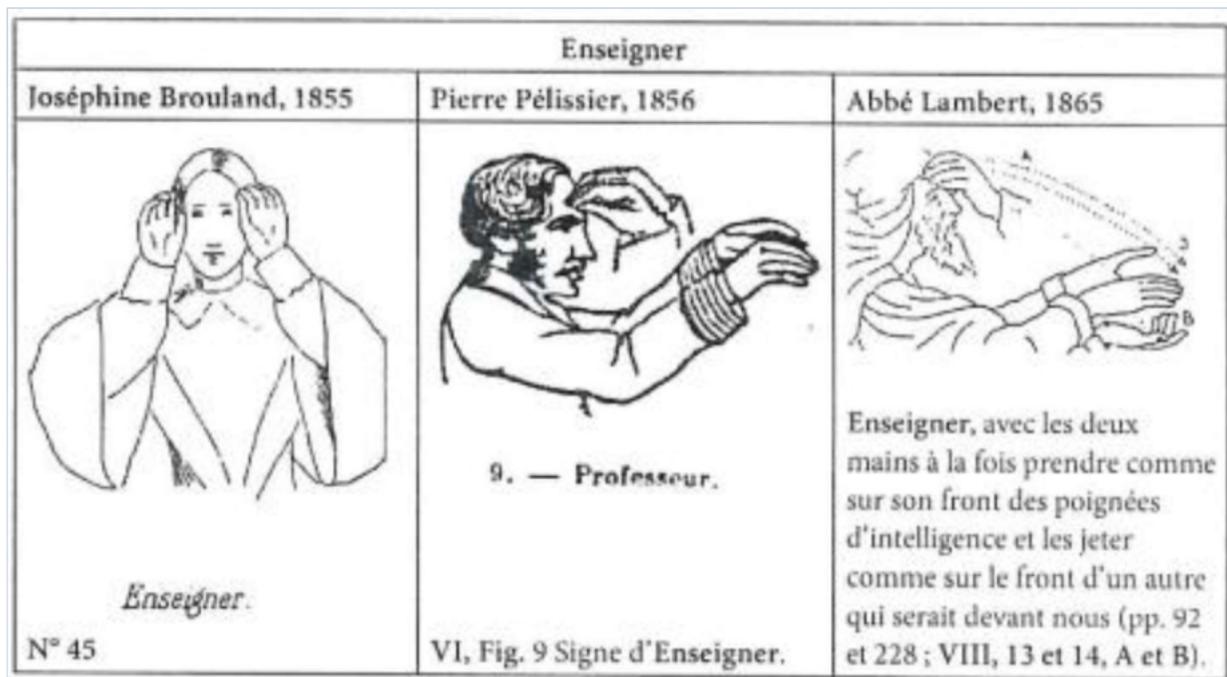


Figure 1 : Divers moyens de représenter graphiquement le signe pour [ENSEIGNER]

Source : Bonnal-Vergès, 2014 : 255

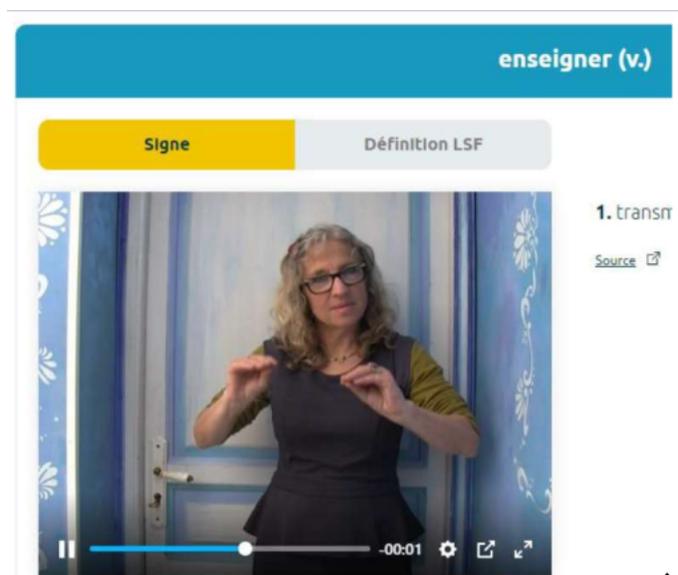


Figure 2 : Copie d'écran d'un extrait de vidéo correspondant au signe pour [ENSEIGNER] sur Elix¹⁶

Un des avantages principaux de la vidéo est de permettre le passage du statique au dynamique (McKee et McKee, 2013), la dynamique étant auparavant représentée par divers moyens comme l'utilisation de flèches ou de dessins ou photographies superposés. Les nouvelles technologies per-

16. Source : <https://dico.elix-lsf.fr/dictionnaire/enseigner> [Consulté le 10 décembre 2023]

mettent aussi une meilleure prise en compte de la quadridimensionnalité, et permettent tout à la fois de capter les LS et de stocker des vidéos parfois extrêmement volumineuses. Elles rendent aussi possibles, dans certains cas, des méthodes de recherche du signe au mot (cf. partie 5) et des méthodes de recherche comme celle des hypersignes (Dalle et Dalle, 2013)¹⁷.

(2) Faible standardisation

La LSF est une langue caractérisée par une grande variation, en particulier topolectale¹⁸ (avec, notamment, une forte différenciation Paris / « province »), qui s'explique principalement par des raisons historiques, abondamment documentées notamment par Yves Delaporte (Delaporte, 2005), mais également par le fait que « les conditions exceptionnelles d'apprentissage et de transmission [de la LSF] sont la norme, au moins pour la partie la plus importante des signeurs » (Cuxac et Antinoro Pizzuto, 2010). Millet *et al.* affirment ainsi que les variantes régionales en LSF sont « assez nombreuses ». Selon eux, cela s'explique « par le fait que la LSF ayant été interdite pendant plus de cent ans, elle s'est dialectalisée [...]. [D]ans les instituts accueillant les jeunes sourds, des signes se sont créés indépendamment » (2023 : 108). Il existe ainsi, par exemple, plusieurs signes pour [MAMAN] ou [MÉDECIN], ou bien encore plusieurs signes pour renvoyer aux différents jours de la semaine. Cette absence de standardisation vaut aussi pour d'autres LS, par exemple la langue des signes danoise (Kristoffersen *et al.*, 2012 : 294) ou encore la langue des signes française de Suisse¹⁹.

(3) Problèmes du type de vedettes

En lexicographie, le lemme est la forme sous laquelle les unités lexicales sont classées et décrites dans un dictionnaire. Dans les langues vocales (LV), il s'agit typiquement de la forme canonique d'une unité lexicale (considérée par convention comme la forme non fléchie d'un nom, d'un adjectif, d'un verbe, etc.), laquelle est nécessairement présentée en version écrite. Pour les dictionnaires de LS, la question de la « voie d'entrée » dans le dictionnaire est particulièrement épineuse, du fait de l'absence de forme écrite de ces langues, ce qui a également un impact sur le classement, qui est habituellement alphabétique dans les dictionnaires de LV, voire thématique (classement qui repose également sur de l'écrit).

17. La technologie des hypersignes consiste à permettre la navigation directement dans un contenu visuel, qu'il soit une vidéo en une langue des signes, une image, un schéma. À l'instar des liens hypertextes qui permettent la navigation depuis un texte vers un autre texte, les hypersignes permettent d'interagir avec un élément du contenu (une partie de la vidéo) pour afficher un nouvel élément visuel, respectant ainsi une logique de navigation davantage respectueuse d'une structuration visuelle de la pensée.

18. La variation diachronique est également très présente. Nous renvoyons le lecteur aux travaux de Delaporte sur ce point. On note aussi une très forte variation diachronique dans d'autres LS : par exemple en ASL, en langue des signes australienne (Auslan) ou encore en NZSL comme cela a été documenté par McKee et McKee (2011).

19. Le dictionnaire trilingue de la langue des signes suisse élaboré par la Fédération suisse des Sourds intègre ainsi les différents dialectes régionaux, qui sont clairement identifiés. Les signes pour [TANTE] par exemple diffèrent d'un canton à l'autre (<https://www.sgb-fss.ch/signesuisse/fr/lexikon/124785/tante>, consulté le 21 juillet 2024).

Traditionnellement, l'approche dans les dictionnaires de LS se fait par le biais de la LV majoritaire correspondante qui, elle, dispose d'une forme écrite. La sélection des vedettes qui constituent la nomenclature des dictionnaires de LS s'est donc habituellement basée sur des listes de mots tirés des LV ou sur des listes thématiques (nourriture, couleur, sport, etc.), elles-mêmes reposant sur une forme écrite de la LV correspondante (Zwitserslood *et al.*, 2022 : 231). Cela a un impact sur la nature même des dictionnaires, qui sont très majoritairement bilingues, mais également monodirectionnels (LV > LS), et non bidirectionnels, à l'inverse de la majorité des dictionnaires bilingues des LV, comme l'explique par exemple Fischer : « [d]'abord, c'est une lexicographie strictement monolingue sur une base bilingue. Il est commun, pour les langues minoritaires, de se voir représentées lexicographiquement dans une seule direction, allant du lemme en langue majoritaire à sa traduction en langue minoritaire ». (2014 : 214)

Toutefois, poursuit-elle pour les LS, le problème se trouve encore « aggravé » (Fischer, 2014 : 214), car l'équivalent de traduction dans la langue minoritaire (la LS) ne peut être exprimé dans le dictionnaire dans ladite langue minoritaire (en LS). En effet, dans tous les dictionnaires papier, du XVIII^e jusqu'au XX^e siècle, l'équivalent en LS ne peut en effet véritablement être fourni en LS (du fait de sa quadridimensionnalité), et se trouve réduit à une forme qui emprunte soit le canal visuel en 2D (dessin, photographie), soit le canal écrit de la LV correspondante (par le biais d'une description de l'exécution du signe, principalement). C'est ce que résumait Sennikova et Garcia en ces termes : « Une difficulté particulière pour qui veut s'atteler à décrire pour lui-même le lexique d'une LS tient au fait que, n'ayant pas de forme écrite instituée, ces langues ont été appréhendées pendant plus de deux siècles par la médiation de la forme écrite de la LV environnante. » (2018 : 133)

Pour Fischer, cela va jusqu'à changer le statut même du lemme lexicographique.

Le lemme est d'habitude fourni graphiquement en langue majoritaire, mais son « équivalent » aussi. Du coup, le lemme change de statut : il n'est plus simplement un lemme qui cherche son équivalent – il est aussi contenu sémantique, ou traduction, du signe qui va suivre. Le signe qui suit le lemme, de son côté, est représenté en cette même langue majoritaire, et il apparaît dans la seule description de son exécution – d'où l'extrême pauvreté microstructurelle de ces dictionnaires de signes. (Fischer, 2014 : 214)

Si l'on tente d'entrer dans le dictionnaire en utilisant le signe comme point d'entrée (et non le mot français), les solutions envisagées jusqu'à présent reposent très majoritairement, à notre connaissance, sur l'utilisation des paramètres, lesquels sont souvent réduits à la configuration manuelle (dont la liste des possibilités va au-delà de la soixantaine (Vanbrugghe *et al.*, 2010)). Un défaut majeur de cette approche réside dans son imprécision, puisqu'on considère qu'il faut parfois jusqu'à huit paramètres pour décrire un signe si l'on inclut, par exemple, le paramètre de l'expression faciale qui a pu, depuis Baker et Padden (1978), être considéré comme intervenant, dans certains cas, pour distinguer entre elles deux unités lexicales. C'est ce que notent Vanbrugghe *et al.* : « Cette approche est d'autant plus inefficace qu'un signe ne se limite pas à une seule configuration mais à un ensemble de paramètres (configuration, orientation, emplacement dans l'espace, mouvement, expression

faciale, ...) » (2010). Cette approche présuppose par ailleurs que l'utilisateur possède une certaine maîtrise des paramètres, qu'il doit pouvoir reconnaître, ce qui n'est pas nécessairement le cas de tous les locuteurs de LSF. Enfin, ces paramètres ne sont par ailleurs que la formalisation linguistique d'un phénomène qui, lui, n'est pas si facilement catégorisable : en discours, les paramètres sont rarement des unités discrètes : par exemple, à quel moment passe-t-on de la lettre U (index et majeur tendus et serrés, les autres doigts étant repliés) à la lettre V (index et majeur tendus et écartés, les autres doigts étant repliés) en dactylographie ? À quel moment passe-t-on de la configuration manuelle « bec de canard ouvert » (tous les doigts tendus, mais pouce en opposition avec les quatre autres doigts qui sont collés les uns aux autres) à celle de « bec de canard fermé » (pouce tendu et en contact avec index et majeur, les quatre autres doigts sont tendus et collés les uns aux autres) ? Le recours au numérique ouvre désormais d'autres possibilités comme la captation vidéo d'un signe exécuté face à une caméra, signe qui est ensuite reconnu par la machine, laquelle a travaillé à partir de corpus d'unités lexicales isolées (comme le dictionnaire de langue des signes belge mentionné dans la partie 5).

(4) Non correspondance systématique signe - mot de la LV et problèmes d'équivalence

Comme le notent Vanbrugghe *et al.*, la plupart des « ressources lexicographiques » (quel qu'en soit le nom) « présuppose[nt] une équivalence conceptuelle stricte entre un mot et un signe lexical, ce qui n'est pas toujours le cas » (2010). Sennikova et Garcia précisent en ce sens que les premiers « dictionnaires » ont été davantage des listes de correspondance de termes LV-LS, ce qui a entraîné une représentation faussée des LS envisagées par le prisme des unités des LV.

Ce principe de mise en correspondance unilatérale mot-signe a eu pour conséquence de fausser durablement la représentation commune de la nature et de la structure du lexique des LS, appréhendé au filtre de celui de la LV environnante. (Sennikova et Garcia, 2018 : 132)

Par ailleurs, même si, comme le reconnaissent ces mêmes autrices, « la question des contours du lexique et de ce qui doit constituer les entrées de dictionnaire dépasse largement les seules LS » (2018 : 134), la question se pose en termes particuliers pour les LS, du fait, notamment, de l'existence, en parallèle des unités lexicales, d'unités de spécificités de discours de même format, les structures de transfert²⁰, qui se caractérisent par le caractère non discret de leurs composants et leur non « lis-

20. Cuxac (2007 : 120) résume ainsi la distinction entre différents types de transferts décrite en détail dans Cuxac (2000) : « (1) les transferts de taille et/ou de forme [...] qui permettent de représenter la taille et/ou la forme, partielle ou globale, de lieux, d'objets, ou encore de personnages ; (2) les transferts situationnels [...] qui donnent à voir une action de déplacement d'un actant par rapport à un repère (locatif) généralement fixe ; (3) les transferts personnels qui reproduisent, en mettant en jeu tout le corps du locuteur, une ou plusieurs actions effectuées ou subies par un actant du procès de l'énoncé. Le narrateur « devient », pour ainsi dire, la personne (ou l'animal) dont il parle ». Ces transferts peuvent se combiner, donnant lieu à davantage de complexité.

tabilité » (Sennikova et Garcia, 2018 : 136)²¹. Sennikova et Garcia considèrent que les constructions multilinéaires qu'engendrent ces structures de transfert (« unités de transfert », UT) sont tout aussi centrales que les unités lexicales (UL) que représentent les signes établis. En outre, selon elles, il existe, en deçà des « signes », des items de nature morphémique. Les « signes » (UL) ne seraient que l'un des types d'unités du lexique, certains composants sublexicaux (comme un paramètre ou un ensemble de paramètres) « ayant, par leur productivité et leur rôle dans l'économie du lexique, pleinement droit sinon à un plein statut d'unités lexicales, du moins à figurer comme entrées de dictionnaires ». Bonnal (2005) a fait cette proposition dans sa thèse de doctorat.

(5) Faible disponibilité des données linguistiques sur lesquelles s'appuyer

De manière générale, malgré l'avancée des recherches en linguistique portant sur la LSF (Sallandre et Garcia, 2021), la LSF reste une langue peu dotée : peu de ressources existent en général, les lexiques sont limités, les connaissances sur sa grammaire sont encore partielles, et il existe très peu d'ouvrages de référence. Dans son état des lieux de 2022, Braffort précise ainsi que les corpus de LSF dont on dispose sont « peu nombreux et de petite taille », par rapport aux corpus de LV. Elle indique ainsi qu'en 2022, on ne dénombre que 22 entrées pour les corpus de LSF répertoriés sur des plateformes d'archivage et de diffusion de corpus telles que Cocoon²² ou Ortolang²³. Par ailleurs, ces corpus sont presque tous, du moins ceux stockés sur Ortolang, des « corpus de laboratoire, c'est-à-dire enregistrés en studio avec un objectif de recherche » (Braffort, 2022), comme le corpus LS-COLIN, premier corpus de LSF enregistré en studio spécifiquement pour la recherche en 2002. Ces corpus sont de nature très variée concernant (1) le contexte de production (contexte très contrôlé comme un monologue de brèves journalistiques, ou contexte informel d'un repas de famille (projet Signes en famille)), (2) les lieux d'enregistrement (en studio ou dans la rue), (3) les conditions techniques (« de 1 à 5 caméras synchronisées, systèmes de capture de mouvement »), (4) le degré et la nature des annotations (données non annotées, « simples gloses pour identifier les signes lexicaux », sous-titrage ou doublage en français, « descriptions fines des constructions linguistiques ou du mouvement des articulateurs »)²⁴.

21. Le statut d'instanciations de *structures de transfert* de ces constructions hautement iconiques continue à susciter des débats (voir Sennikova et Garcia (2018 : 136), pour un résumé des débats, ou encore Garcia (2010, chapitre 3) pour une analyse détaillée). Dans l'approche sémiologique de Cuxac, l'existence même de ces structures fait partie des postulats de départ. C'est une « analyse minutieuse de longs corpus de discours en LSF, appréhendés dans une perspective énonciative et fonctionnelle et intégrant d'emblée l'ensemble des paramètres, manuels et non manuels, [qui] a permis à Cuxac de montrer que [ces] constructions très iconiques ressortissent de structures linguistiques, en nombre limité » (Sennikova et Garcia, 2018 : 137).

22. <https://cocoon.huma-num.fr> [Consulté le 10 décembre 2023]

23. <https://www.ortolang.fr/> [Consulté le 10 décembre 2023]

24. Voir le tableau 1 de Braffort (2022) qui décrit précisément l'ensemble des corpus issus de projets de recherche (nom du corpus, description, taille – précisée en nombre d'heures, caractéristiques des locuteurs, sous-titrage en français, annotation, plateforme de dépôt).

2.3. Un état de fait : un paysage lexicographique encore en construction

Un constat est partagé par plusieurs auteurs pour l'ensemble des LS : la lexicographie des LS a un retard considérable à rattraper sur les LV, comme l'expliquent Kristoffersen *et al.*

Scientific sign language research is in many ways still relatively young. This is also true for sign language lexicography, and in many respects, dictionaries of sign languages still lag considerably behind those of spoken/written languages. They often offer a small number of entries and/or very little, if any, information about sign usage²⁵ (Kristoffersen *et al.*, 2012 : 294).

Même si ce point de vue peut être nuancé²⁶, en ce qui concerne la LSF, il n'en reste pas moins qu'un certain nombre d'éléments caractérisent les ressources existantes pour la LSF. Nous en avons recensé cinq principaux. Tout d'abord, de manière globale, le nombre de ressources existantes est limité (Braffort, 2022). Deuxièmement, les ressources lexicographiques contemporaines ne correspondent pas totalement à un dictionnaire tel qu'on peut le concevoir pour les LV. Bonnal-Vergès considère ainsi que :

la plupart des « dictionnaires » contemporains de LSF, aussi soigneusement faits soient-ils, ne correspondent pas du tout à ce que l'on appelle un Dictionnaire bilingue. Simples vocabulaires, glossaires, recueils de signes, qui sont plus des répertoires lexicaux, favorisant le transcodage, que de réels dictionnaires, ils établissent des équivalences lexicales entre le français et la langue des signes française qui supposent une correspondance parfaite entre le mot et le signe, au plan sémantique, voire même au plan grammatical. Ainsi une définition s'avère superflue : le signe est l'illustration du mot, ou le mot l'illustration du signe. (Bonnal-Vergès, 2012 : 279)

Troisièmement, la taille de leur nomenclature est toujours modeste²⁷, ce qui est également le cas pour d'autres langues des signes : selon Zwitserlood (2010), la plupart des dictionnaires de langues des signes comportent entre 2000 et 5000 « signes » vedettes. Le dictionnaire d'IVT, par exemple, contient 4500 signes. Quatrièmement, la compilation de ces ressources est par ailleurs essentiellement faite par des personnes qui ne sont pas des lexicographes professionnels. Cette tendance est historique : le recensement fait par Bonnal-Vergès (2006) montre qu'un très grand nombre de ressources papier ont été compilées soit par des ecclésiastiques (l'abbé Sicard, l'abbé de l'Épée, l'abbé Ferrand, etc.), soit par des enseignants de LSF (par exemple, les enseignants Sourds et entendants d'IVT). Elle est

25. Notre traduction : La recherche scientifique sur les langues des signes est, à bien des égards, encore relativement jeune. Il en va de même pour la lexicographie des langues des signes. Par ailleurs, dans bien des domaines, les dictionnaires des langues des signes accusent encore un retard considérable par rapport à ceux des langues vocales/écrites. Ils ne contiennent souvent qu'un petit nombre d'entrées et/ou très peu, voire pas du tout, d'informations sur l'utilisation des signes.

26. Cette affirmation peut toutefois être nuancée par le fait que peu de dictionnaires de LV, notamment les dictionnaires d'une LV écrite de grande diffusion comme l'est le français, ont fait l'objet d'une lemmatisation scientifiquement ancrée et fondée sur des corpus *de forme parlée*, du type de celle dont font l'objet de plus en plus de langues des signes.

27. Pour une discussion détaillée sur ce qui explique la taille modeste de la nomenclature de ces dictionnaires, voir Garcia (2021).

toutefois toujours d'actualité, comme nos premières analyses, présentées plus bas, le font ressortir. Enfin, en lien avec ce qui a été expliqué plus haut, ces ressources sont quasi-systématiquement monodirectionnelles (du français vers la LSF) et on note l'absence d'une vraie bidirectionnalité. D'après Bonnal-Vergès, cet état de fait n'est pas à imputer uniquement à la question d'absence de forme écrite de la LSF, mais également à une « situation de diglossie fort inégale, dans laquelle les entendants, communauté majoritaire, sont les utilisateurs de la langue minoritaire, la langue des signes, dans une relation que l'on pourrait qualifier de dominante, de la part des entendants » (2006 : 177). La lexicographie actuelle en LSF est toutefois assez isolée sur la question de la directionnalité par rapport à la lexicographie d'autres LS, qui s'est emparée des possibilités du numérique depuis 2000 : les dictionnaires de LS suédoise, néerlandaise, flamande, finlandaise et néo-zélandaise que l'on trouve en ligne sont bidirectionnels. Ils offrent ainsi la possibilité d'entrer dans le dictionnaire via un signe en LS, en utilisant les paramètres (notamment celui de la configuration manuelle et l'emplacement, mais également via d'autres critères plus fins) (Kristoffersen *et al.*, 2012 : 299). Dans tous les cas, le fait d'entrer majoritairement dans les dictionnaires de LSF via la LV et non la LSF implique que les utilisateurs visés, ou du moins les utilisateurs les plus à même d'utiliser les dictionnaires, sont avant tout des entendants plutôt que des Sourds. Bonnal-Vergès considérait ainsi en 2006 que « la plupart [des auteurs du XVIII^e et du XIX^e siècles [avaient] surtout réalisé leurs dictionnaires pour les enseignants et les personnes en contact avec les sourds, *i.e.* pour des entendants] et que « la situation n'a[vait] pas changé au XX^e siècle » (2006 : 177). Elle concluait ainsi que la LSF qui apparaît dans les dictionnaires, qu'elle soit objet d'étude universitaire ou objet d'apprentissage pour les entendants, « n'est pas destinée aux Sourds » (*ibid.*).

3. Bref état des lieux des dictionnaires électroniques de LSF

Nous exposons ici les conclusions provisoires livrées par une première analyse surplombante des ressources lexicographiques de LSF disponibles en ligne. Ces analyses sont amenées à être complétées par des analyses plus fines de chacun des aspects traités dans des études en cours ou à venir.

3.1. Méthodologie de l'analyse

La méthodologie utilisée dans cette étude exploratoire combine deux premières analyses : une analyse des ressources et une analyse du point de vue de leurs concepteurs et de leurs utilisateurs. Pour l'analyse des ressources, nous nous basons sur des études métalexographiques existantes : soit les quelques (rares) études métalexographiques appliquées à d'autres LS, comme, par exemple, Vale (2015) pour la NZSL ou Kristoffersen *et al.* (2012) pour la langue des signes danoise ; soit sur des études métalexographiques générales appliquées aux LV, comme celles portant sur les utilisateurs de dictionnaires (Nesi, 2008, 2015 ; Tarp, 2009), ou sur les dictionnaires numériques (Pastor et Alcina, 2010). À partir de ces études, nous avons confectionné et utilisé une grille d'analyse métalexographique en onze points, qui est détaillée un peu plus bas.

En ce qui concerne le point de vue des concepteurs et des utilisateurs, nous nous sommes interrogés sur leur profil (qui sont les concepteurs et les utilisateurs des ressources lexicographiques en LSF en ligne ?) et avons cherché à avoir une première idée de la manière dont ces ressources sont utilisées concrètement, comme cela a été fait pour les ressources en NZSL par Vale (2015 : 282). Force est de constater que, pour les LS en général, très peu d'études existent. Celles-ci sont par ailleurs souvent réalisées préalablement à la constitution d'une ressource (McKee et Pivac, 2008) et ne s'intéressent qu'aux futurs potentiels utilisateurs. Les méthodologies utilisées dans la littérature existante sont variées. Cette littérature ne concerne, à notre connaissance, que des langues autres que la LSF, pour laquelle aucune étude empirique d'ampleur n'a été menée sur l'utilisation des ressources²⁸. Ces méthodologies prennent plusieurs formes. Elles peuvent revêtir la forme, assez traditionnelle, d'enquêtes en ligne, avec par exemple Oyserman (2013) qui a mené une enquête en ligne constituée de 31 questions auprès des utilisateurs du *Dictionnaire de la langue des signes flamande*²⁹. Elles peuvent aussi se trouver sous la forme d'analyse des statistiques d'utilisation du site web concerné via Google Analytics combinée à des « Think-Aloud Protocols » auprès de participants entendants ou Sourds (pour lesquels le protocole est adapté) observés en train de consulter la ressource, que viennent enrichir des interviews préalables et postérieures à la tâche (étude empirique de Vale (2015) auprès de douze participants Sourds et entendants concernant leur utilisation du *Online Dictionary of New Zealand Sign Language* (ODNZSL))³⁰.

Pour notre étude exploratoire, nous avons choisi de recourir à l'enquête en ligne, auprès des concepteurs de huit ressources et auprès d'un échantillon d'utilisateurs contactés via les réseaux sociaux. Cette méthodologie est amenée à être enrichie d'autres méthodes empiriques, et à se focaliser sur certains types d'utilisateurs. Le travail de recherche de seconde année de Master d'Emma Bancillon (Bancillon *et al.*, 2024 ; Bancillon, 2024) que nous co-encadrons se penche en particulier sur l'utilisation que font les interprètes LSF-FR des ressources en ligne existantes. Celle-ci étudie d'un point de vue global, via un questionnaire cherchant à mettre en lumière le type d'interaction qu'ils ont avec ces ressources, la motivation de ces interactions, le degré de confiance qu'ils accordent au contenu des ressources, etc. Nous travaillons collaborativement pour mettre au point la méthodologie de cette enquête et assurer sa diffusion ainsi que l'exploitation de ses résultats. Cette étude permettra de mieux cerner les besoins de ce type d'utilisateur par rapport aux ressources.

28. Il existe quelques études faites par exemple dans des mémoires de fin d'étude de futurs interprètes en LSF, mais ces études n'ont pas été publiées et restent d'une ampleur assez limitée.

29. <https://www.vgtc.be/wp-content/uploads/2020/02/Gebruiksonderzoek-online-digitaal-woordenboek-VGT-Nederlands-Nederlands-VGT.pdf> [Consulté le 10 décembre 2023]. Cette enquête n'est disponible qu'en flamand. Nous remercions Mireille Vale de nous en avoir signalé l'existence.

30. <https://www.nzsl.nz/> [Consulté le 10 décembre 2023]

3.2. Analyse des ressources

Notre étude s'est penchée sur huit ressources numériques : une sous format CD-ROM (le *Fournier Signé*), et sept en ligne, à savoir : (1) *Elix*³¹, (2) *Ocelles*³², (3) *Dicoplus*³³, (4) *Sématos*³⁴, (5) *SpreadTheSign*³⁵, (6) *Sourds.net* (DicoLSF)³⁶, (7) *INJS Metz* (LSF Dico)³⁷.

Nous présentons ici très brièvement les onze critères qui ont présidé à l'analyse et qui ont été examinés pour chacune des huit ressources, ainsi que les tendances générales qui se dégagent pour chacun de ces points. L'analyse détaillée de chacun des critères est en cours, et les résultats de cette analyse fine feront l'objet de publications ultérieures.

1. Dénomination de la ressource donnée par les concepteurs

On note une très grande variabilité dans les dénominations (*dictionnaire, lexique...*) qui ne recouvrent pas nécessairement les mêmes réalités, et qui reflètent les hésitations sur la définition du concept de *signaire* parfois utilisé.

2. Date de création

Le *Fournier* sur CD-ROM est la ressource la plus ancienne ; ce dictionnaire date de 1999. Les informations sur les dates de création des ressources électroniques sont rarement affichées, mais nos recherches ont pu établir que la ressource la plus récente, qui a maintenant une dizaine d'années, est le dictionnaire *Elix*, créé en 2013. La ressource *SpreadtheSign*, qui regroupe plusieurs LS, a été inaugurée en 2006.

3. Statut des concepteurs de la ressource

Un trait particulièrement marquant est que ne figure aucun lexicographe professionnel parmi les concepteurs des ressources analysées, et que les structures qui portent ces projets sont plutôt des structures de petite taille, dont certaines sont privées. On trouve en effet trois associations, deux institutions, deux projets de recherche (non universitaires), et une école (Institut de Jeunes Sourds de Metz). C'est là une différence majeure avec d'autres pays où existent des projets nationaux de grande envergure de dictionnaires de langues des signes, qui sont souvent portés en partie ou intégralement

31. <https://dico.elix-lsf.fr/> [Consulté le 10 décembre 2023]

32. <https://ocelles.inshea.fr/fr/accueil> [Consulté le 10 décembre 2023]

33. <https://www.dicoplus.fr/> [Consulté le 10 décembre 2023]

34. <https://www.sematos.eu/lsf.html> [Consulté le 10 décembre 2023]

35. <https://www.spreadthesign.com/fr.fr/search/> [Consulté le 10 décembre 2023]

36. <https://www.sourds.net/category/l-s-f/dico/> [Consulté le 10 décembre 2023]

37. <https://lsfdico.injs-metz.fr/> [Consulté le 10 décembre 2023]. Il s'agit d'une ressource compilée par un Institut des Jeunes Sourds.

par des projets universitaires. Ainsi le dictionnaire néo-zélandais ODNZSL est-il porté par le « Deaf Studies Research Unit » de l'Université Victoria de Wellington.

4. Utilisateurs visés

De manière générale, le public visé par l'ensemble des ressources est très large : Sourds comme entendants, avec des niveaux de LSF variés, apprenants ou professionnels. Ces ressources sont axées majoritairement sur l'apprentissage de la langue, alors que parmi les utilisateurs réels peuvent figurer des professionnels, a priori experts de la langue, comme des interprètes ou des enseignants.

5. Type de langue incluse (langue générale / langues de spécialité)

Lorsqu'on observe l'ensemble des ressources pour la LSF, une autre caractéristique saillante est qu'il y en a peu qui relèvent strictement de la langue générale, au vu de la cohabitation de nombreux signes correspondant à des termes de spécialité avec des signes relevant de la langue générale. Même s'il est vrai que les dictionnaires généraux de LV incluent des « termes » (au sens d'unités lexicales relevant de domaines spécialisés) (Josselin-Leray, 2005), les LV font clairement la distinction, pour les ressources les décrivant, entre les bases de données terminologiques ou dictionnaires spécialisés d'une part, et dictionnaires de langue générale d'autre part. La démarcation entre ressources de langue générale et ressources de langue spécialisée semble beaucoup moins claire pour la LSF que pour les LV. L'insertion de la terminologie dans des ressources en LSF considérées comme décrivant a priori la langue générale ne semble pas relever d'une politique préétablie et réfléchiée par des lexicographes ; elle semble se faire plutôt au coup par coup, au gré de collaborations fortuites. Pour l'illustrer, on peut citer plusieurs ajouts récents sur *Elix* : celui de signes relatifs aux échecs³⁸ (début 2023), celui de 1700 signes « sur l'école » (novembre 2023), et celui de 60 signes spécialisés relevant du domaine de l'ornithologie dans le cadre d'un partenariat avec la Ligue de protection des oiseaux (60 signes en octobre 2023, cf. Figure 3, dont certains vraiment spécialisés (Giacoletti, 2024)). Sont également à venir des ajouts dans les domaines des métiers et de l'agriculture, prévus en 2024. Ces ajouts semblent s'être faits au gré des partenariats et opportunités rencontrés par la structure, comme en témoignent les différents posts d'*Elix* sur les réseaux sociaux (cf. Figure 3), et non dans le cadre d'une stratégie de développement particulière de la nomenclature du dictionnaire.

38. Une vidéo produite par *Elix* concernant un joueur d'échecs sourd contient un certain nombre de signes liés aux échecs qui figurent dans leur dictionnaire : [CAVALIER], [DAME], [ATTAQUE], [ROI] (qui est bien différent du signe [ROI] utilisé dans la langue générale). Voir : <https://www.youtube.com/watch?v=YPhipwmXwNc> [Consulté le 10 décembre 2023]

60 signes sur les oiseaux avec LeDicoElix³⁹



- aigrette garzette
- avocette élégante
- bergeronnette grise
- bernache cravant
- busard des roseaux
- canard colvert
- chardonneret élégant
- chevalier guignette
- chevêche d'Athéna
- cigogne blanche
- courlis cendré
- échasse blanche
- étourneau sansonnet
- faucon crécerelle
- faucon d'Éléonore
- faucon pèlerin
- fauvette à tête noire
- fauvette mélanocéphale
- flamant rose
- foulque macroule
- gallinule poule-d'eau
- goéland argenté
- goéland leucophée
- grand cormoran
- grand-duc d'Europe
- grive musicienne
- guépier d'Europe
- héron cendré
- héron garde-bœufs
- hirondelle de fenêtre
- hirondelle rustique
- huppe fasciée
- macareux moine
- martin-pêcheur d'Europe
- martinet noir
- merle noir
- mésange bleue
- mésange charbonnière
- moineau domestique
- moineau friquet
- monticole bleu
- mouette rieuse
- petit gravelot
- pic vert
- pie bavarde
- pigeon ramier
- pinson des arbres
- pouillot véloce
- puffin de Scopoli
- puffin yelkouan
- rossignol philomèle
- rougegorge familier
- rougequeue noir
- sterne caugek
- tadorne de Belon
- tourterelle des bois
- tourterelle turque
- troglodyte mignon
- vautour fauve
- verdier d'Europe



Figure 3 : Post Facebook Elix : Liste des ajouts de signes ornithologiques sur *Elix*

6. Domaines de spécialité répertoriés

Parmi les domaines (dont la liste est très éclectique) que l'on retrouve dans l'ensemble des ressources, les domaines relevant des sciences « dures » (STIM : sciences, technique, ingénierie, mathématiques) dominent de loin les domaines relevant des Sciences humaines et sociales (Segouat et Josselin-Leray, à paraître)³⁹. Ceci pourrait peut-être s'expliquer par le fait que les ajouts « terminologiques » correspondent à des besoins concrets et ponctuels, ou par le fait que les locuteurs impliqués dans les sciences « dures » sont plus enclins à utiliser les nouvelles technologies et donc à diffuser les signes par ce biais. Par exemple, pour certaines ressources (ex. : *Sourds.net*), les choix semblent faits en lien avec l'actualité scientifique (téléscope de Newton, Space X).

7. Taille de la nomenclature

La taille des nomenclatures est modeste dans l'ensemble, mais elle reste très difficile à quantifier, sauf si elle est expressément spécifiée par les concepteurs de la ressource. De plus, si la taille est spé-

³⁹. *SpreadTheSign* inclut ainsi divers sous-domaines STIM : chimie, physique, botanique...

cifiée, les critères de dénombrement ne sont pas indiqués (ce qui est souvent le cas également pour les dictionnaires de LV). On note un grand écart entre les ressources, le dictionnaire sur CD-ROM de Fournier contenant 1921 signes, contre 23 500 signes⁴⁰ affichés pour le dictionnaire *Elix* (qui demeure la ressource la plus fournie).

8. Fonctionnalités de recherche

Un résultat frappant est la prédominance des fonctionnalités de recherche par mot du français écrit, présente dans 7 ressources sur 8. La recherche par signes est sous-représentée (2,5 sources sur 8) et quand celle-ci est proposée, c'est uniquement par le biais des paramètres manuels (par exemple dans le *Fournier* ou dans *Dicoplus*)⁴¹, ou par le biais de l'initialisation⁴² (*Sematos*). Aucun dictionnaire ne permet d'entrer dans la ressource via une captation vidéo d'un signe qui serait exécuté face à l'ordinateur (ce système est toutefois envisagé pour la suite chez *Elix*). Très peu de ressources proposent des recherches par mot composé, ou par expression polylexicale comme « il était une fois », comme cela peut être proposé par certains dictionnaires électroniques de LV. Seul *Sematos* propose ainsi, au sein d'une liste classée par ordre alphabétique, des locutions comme « à mon avis », « à peu près », et combine plusieurs types de recherches possible, comme on peut le voir sur la Figure 4, notamment la recherche thématique et/ou alphabétique (ce dernier type étant privilégié par certaines autres ressources).

40. <https://dico.elix-lsf.fr/> [Consulté le 19 juillet 2024]

41. Dans INJS Metz, on peut chercher uniquement par configuration manuelle, en cliquant sur une icône dessinée qui représente une certaine forme de main : en griffe, bec de canard etc. Pour les signes à deux mains, il est possible de choisir entre main dominante et main dominée.

42. C'est-à-dire l'utilisation du signe figurant la lettre de l'alphabet correspondant à la première lettre du mot équivalent en LV.



Figure 4 : Page d'accueil permettant d'effectuer une recherche dans *Sematos*⁴³

9. Recherche en ligne : affichage et contenu des résultats

Nous nous sommes intéressés aux diverses formes sous lesquelles les résultats d'une recherche peuvent apparaître dans les 8 ressources analysées. Aucune ressource n'a recours à la photographie, au dessin ou à un système de notation (cf. partie 2.2.). En revanche, le recours à la vidéo est systématique pour l'ensemble des ressources. Ces vidéos sont parfois stockées sur YouTube (2 sur 8) ; leur source n'est pas toujours indiquée (statut de la personne qui signe : Sourd ou entendant ?) et elles se trouvent sous des formats divers : (1) 1 sur 8 permet l'utilisation du ralenti, (2) 2 sur 8 incluent une répétition du signe, (3) 2 sur 8 fournissent une décomposition du signe, et (4) 2 sur 8 lancent une lecture automatique de la vidéo. Les trois premiers points mentionnés montrent clairement une

43. Source : <http://www.sematos.eu/lsf.html> [Consulté le 10 décembre 2023]

orientation de la ressource vers des apprenants de LSF, qui ont besoin d'aide pour savoir reconnaître et/ou reproduire le signe.

De manière générale, on note une présence importante de variantes dans les ressources. *Elix* propose ainsi 5 vidéos pour le mot [ECUREUIL], qui illustrent un aspect ou l'autre de l'écureuil : le fait qu'il ait une queue (que le signe soit fait sur le corps du signeur ou devant celui-ci), qu'il mange des noisettes, qu'il grignote avec ses pattes de devant, etc. Un certain nombre d'écueils sont à signaler toutefois : (1) aucune explication n'est proposée sur les critères de choix de l'unique vidéo qui apparaît dans la première étape de la recherche (c'est uniquement en cliquant sur une icône « voir les autres signes » que l'on accède aux 4 variantes supplémentaires), (2) une des vidéos (la première des variantes) n'est en fait qu'une version allongée de la variante « principale » qui décompose le signe en plus de le représenter, (3) aucune information métalinguistique n'est fournie qui permettrait de différencier ces variantes (variation diatopique, diachronique, etc.), et aucune information de nature prescriptive ne figure non plus dans la ressource sous forme d'une éventuelle recommandation (signe à privilégier ou à éviter, par exemple). De manière générale, et contrairement, peut-être, à ce qui se passe dans certains cas pour les LV, au vu de sa profusion, la variation en LSF semble être perçue par ses locuteurs non comme un problème à régler mais comme une diversité ou une richesse à préserver. La question d'une variété qui serait plus prestigieuse qu'une autre ou d'un « standard » à respecter ne semble pas véritablement pertinente (affirmation qui se doit d'être étayée par des études sociolinguistiques plus poussées). On note d'ailleurs que, dans une démarche que l'on peut qualifier de lexicographie « contributive » (Sajous, 2023), le dictionnaire *Elix* s'adresse régulièrement à la communauté des utilisateurs via les réseaux sociaux (page Facebook notamment) pour collecter divers signes, comme ce fut encore le cas très récemment concernant les jours de la semaine⁴⁴ (Figure 5).

44. Il y a ainsi trois variantes pour [MERCREDI].



Figure 5 : Post Facebook *Elix* : collecte de signes pour [MERCREDI]

Il est cependant difficile de savoir quel filtre est ensuite appliqué par les concepteurs de la ressource une fois les vidéos collectées.

En ce qui concerne la présence et la forme des définitions, notre analyse a livré les tendances suivantes :

(1) les définitions sont loin d'être systématiquement présentes : elles ne figurent que dans 50 % des ressources. Le nombre affiché de définitions dans *Elix* (27 660) est supérieur au nombre affiché de signes répertoriés (23 500), ce qui montre que les unités lexicales du français répertoriées, qui sont le point d'entrée, sont en partie polysémiques⁴⁵. En revanche, pour le concepteur d'*OCELLES*, Cédric Moreau, la

45. On note également dans *Elix* des problèmes d'appariement assez récurrents entre le signe et la définition annoncés et la vidéo associée, qui correspond à une autre définition ou à un autre signe. Par exemple, le signe présenté pour [HÉLICE] est en fait celui pour [ÂME], et 3 des 5 vidéos proposées pour le signe pour [ÂME] sont des vidéos traduisant la définition, et non le signe.

définition est centrale, car elle doit « aider l'apprenant à comprendre le concept », et elle primerait quasiment sur le signe⁴⁶. C'est la raison pour laquelle, dans cette ressource, pour certains mots ou termes du français, une définition est fournie en LSF, mais il n'y a pas de signe associé. C'est notamment le cas du terme [DÉCANTATION] (le bandeau supérieur indique « 0 signe, 1 définition »), cf. Figure 6.

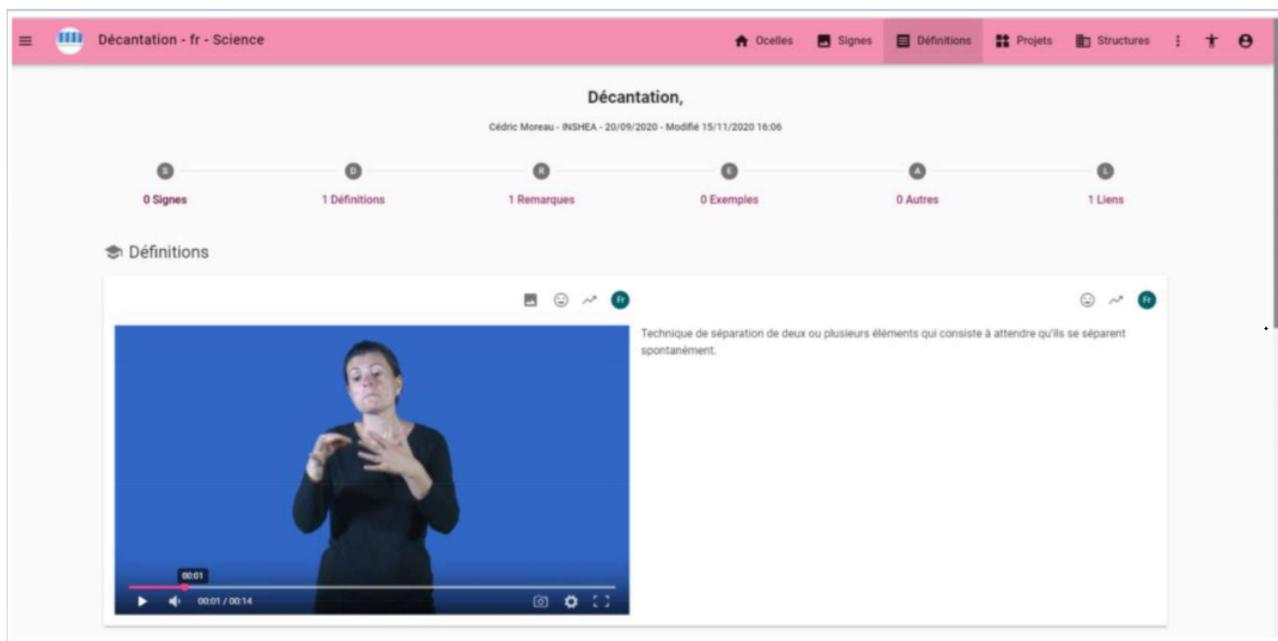


Figure 6 : Entrée [DÉCANTATION] dans *OCELLES*

(2) la langue dans laquelle est exprimée la définition, lorsqu'il y en a une, est pour moitié des ressources le français, et pour moitié la LSF (mais la définition en LSF est en réalité souvent la traduction en LSF d'une définition écrite à l'origine en français, comme dans *OCELLES*, ce qui nous a été confirmé par le concepteur de la ressource. Dans cette ressource, la définition écrite apparaît en sous-titrage de la vidéo de la définition). Une seule ressource (*OCELLES*) propose des définitions en LSF étiquetées selon un niveau donné, défini par la ressource (débutant, intermédiaire, avancé).

(3) la source des définitions est souvent problématique : soit parce que les définitions ne sont pas exprimées directement en LSF, soit parce qu'elles ne sont pas rédigées par des lexicographes professionnels (seules les définitions dans le *Fournier signé* sont extraites du *Nouveau Petit Robert*), soit parce qu'elles proviennent de dictionnaires aux contours mal définis (celles d'*Elix* proviennent du site « sensagent »⁴⁷). La présence de ces définitions tirées du français, voire « plaquées » depuis le français, fait de ces dictionnaires des ressources hybrides qui ne sont ni de vrais dictionnaires unilingues, ni de vrais dictionnaires

46. Communication personnelle de Cédric Moreau (juin 2023).

47. D'après Corbin (2020) et Corbin et Gasiglia (2020), ce dictionnaire (<https://dictionnaire.sensagent.com/> [Consulté le 10 décembre 2023]) existe en quatre versions hébergées sur des sites distincts. On ne dispose, en guise de métadonnées, que de l'identification de la société qui le met à disposition (Sensgates, spécialisée dans le Traitement automatique de la langue et la sémantique lexicale). La version disponible sur le serveur sensagent.com est, en quelque sorte, un agrégateur de contenu puisqu'elle intègre des articles du dictionnaire *Littre* et des définitions issues de l'encyclopédie *Wikipédia*.

bilingues (dans lesquels les indications sémantiques sont normalement adaptées en fonction de la langue-cible vers laquelle on traduit, et non directement des définitions tirées de dictionnaires unilingues, sauf si l'on pense à des dictionnaires du type *bilingualized learner's dictionaries* tels que ceux de Hornby (Marello, 1998), qui restent une denrée rare) ;

(4) dans une ressource (*SpreadTheSign*), la définition est en fait remplacée par un dessin, qui vient suppléer la vidéo.

Enfin, nous avons examiné quelles ressources fournissaient des exemples. Seules 3 ressources sur 8 en fournissent, mais leur finalité ou leur forme ne sont pas dénuées de problèmes. Le *Fournier* indique ainsi que « Pour un signe présenté, plusieurs exemples peuvent être proposés pour souligner certaines variantes selon la nature de l'énoncé ». L'exemple semble donc servir de moyen détourné pour présenter de la variation, plutôt que pour venir illustrer un signe en particulier. Il pourrait aussi servir à signaler que, dès que le signe est mis en discours, il ne se traduit pas nécessairement de la même manière que lorsqu'il est isolé : on retrouve également ce phénomène dans les LV, où la prise en compte du contexte peut faire varier l'équivalent de traduction. Dans *OCELLES*, on trouve certes des exemples (de manière non systématique), mais ceux-ci sont écrits en français, accompagnés d'une vidéo signée et sous-titrée (par exemple, pour [PARAMÈTRE]) qui n'est en fait qu'une traduction de l'exemple donné en français écrit.

10. Fonctionnalités d'interaction avec les utilisateurs

Cinq ressources sur les huit analysées présentent des fonctionnalités d'interaction avec les utilisateurs, qui revêtent des formes diverses. *Elix* propose ainsi aux utilisateurs de signaler, via sa page Facebook, un problème sur une vidéo, et cherche, via les réseaux sociaux, à récolter des signes, comme nous l'avons mentionné plus haut. *Sematos* et *Sourds.net* permettent de laisser des commentaires, tandis qu'*OCELLES* se veut une plateforme collaborative qui mutualise des ressources.

11. Autres potentialités du numérique

Enfin, les ressources utilisent partiellement les potentialités du numérique. La moitié des ressources utilisent des applications (*Elix*, *SpreadTheSign...*) et une (*Elix*) un plug-in, la « Bulle Elix », une extension pour navigateur qui a pour objet de « rendre le web accessible en LSF » (« Un simple clic sur un mot et une bulle apparaît pour vous donner son sens en LSF et le signe correspondant »). Pour les recherches par écrit, certaines ressources permettent la saisie prédictive, mais aucune n'utilise les possibilités du numérique pour la captation du signe par vidéo, comme déjà indiqué plus haut.

Ce premier tour d'horizon rapide de l'ensemble des ressources permet de voir qu'elles sont donc assez disparates, plutôt incomplètes, fortement orientées vers les apprenants, et qu'elles ne semblent pas obéir à des principes de constitution strictement établis en amont tels qu'ils sont préconisés en lexicographie générale : circonscription de la taille de la nomenclature, identification précise du public visé, niveau de détail de la microstructure, etc. (Atkins et Rundell, 2008 : 1744). Elles semblent plus répondre à des besoins *ad hoc*, localisés et reposent pour l'essentiel sur des bonnes volontés ou

des volontés individuelles et non sur des équipes de lexicographes professionnels. Les dictionnaires existants semblent n'entrer dans aucune catégorie des typologies lexicographiques habituelles. *Elix* par exemple est une ressource lexicographique hybride résultant de l'agrégation de définitions provenant d'une ressource dictionnaire monolingue du français, juxtaposées à des vidéos contributives enregistrées, pour certaines, par des internautes amateurs volontaires et bénévoles.

4. Du concepteur à l'utilisateur : point(s) de vue sur les ressources

Afin de pouvoir mieux caractériser les ressources lexicographiques que nous avons analysées, et au vu de leur disparité, il nous a semblé indispensable, d'une part, de mieux connaître les intentions de leurs concepteurs lors de la création des ressources, mais également vis-à-vis de leur maintenance et de leurs potentiels développements, et, d'autre part, de mieux cerner le profil des utilisateurs et leur degré de satisfaction vis-à-vis des ressources. Nous avons donc lancé de premières enquêtes d'ampleur modeste, dans la perspective d'en mener d'autres plus ciblées prochainement.

4.1. Point de vue des concepteurs

Seuls les concepteurs de trois ressources sur huit ont répondu à notre court sondage (qui comportait treize questions principales, hors questions d'information sur leur profil), lancé par courrier électronique en mai 2023, à savoir les concepteurs d'*Elix*, de *Dicoplus.fr* et d'*OCELLES*. Tous les répondants étaient entendants. Il n'a pas été possible d'obtenir d'informations précises, lors de cette première enquête, sur la méthodologie de création de la ressource (groupes, participants). En ce qui concerne le profil des personnes concevant les ressources, nous avons pu établir que pour le dictionnaire *Elix*, il s'agit d'une équipe de dix personnes qui « anime une communauté » (selon leurs propos), et que *OCELLES* rassemble des « experts » (selon leurs propos) pour créer des définitions. Une autre question portait sur la méthode de validation des signes. Du côté d'*Elix*, les signes sont collectés en « validant le niveau de LSF des contributeurs » (selon leurs propos). Pour *Dicoplus.fr*, le statut de traductrice professionnelle d'une bénévole de l'association permet cette validation.

Un premier résultat de cette enquête concerne l'adéquation entre l'utilisateur ciblé et l'utilisateur réel : pour les concepteurs des trois ressources, celles-ci s'adressent à « n'importe quel type de public » et, si l'on se base sur les données récoltées concernant le profil des utilisateurs, c'est exactement ce type d'utilisateur qui utilise la ressource. En ce qui concerne la manière de qualifier leur ressource, les concepteurs d'*Elix* considèrent qu'il s'agit d'un « dictionnaire/signaire » ; le concepteur d'*OCELLES* décrit cette ressource comme étant une « superposition de réseaux de concepts » (selon ses propos). Les objectifs inhérents à la création de la ressource ont été ainsi décrits par les concepteurs : (1) *Elix* : faciliter la communication, (2) *Dicoplus.fr* : diffuser de nouveaux signes, (3) *OCELLES* : répondre aux besoins conceptuels, linguistiques et pédagogiques des utilisateurs. En ce qui concerne de potentielles interactions avec les utilisateurs, celles-ci sont « prévues », pour « enrichir le contenu » (selon

leurs propos), mais leur forme n'est pas précisée. Les réponses à la question « globalement, êtes-vous satisfait de votre ressource ? » n'ont pas été unanimes ; toutefois, c'est plutôt la satisfaction qui semble dominer. Les concepteurs d'*Elix* se disent satisfaits car il y a déjà eu 80 000 installations de l'application, ceux de *Dicoplus.fr* se disent satisfaits sans donner de justification. Le concepteur d'*OCELLES*, quant à lui, a un avis mitigé, car cette plateforme, qui se veut une mutualisation, ne compte pas assez de dépôts. Se posent aussi des problèmes de validation pour cette ressource, qui se fait en lien avec un comité scientifique, ce qui est contraignant du fait notamment de l'ajout de procédures et de débats qui ralentissent le processus de validation. Seuls les concepteurs d'*Elix* et d'*Ocelles* expriment le souhait de faire évoluer leur ressource. Les concepteurs d'*Elix* annoncent vouloir mettre en place les évolutions suivantes : associations des signes à des lexiques métiers ; recherche inversée ; amélioration de l'expérience utilisateur ; réponse aux demandes de signes ; correction des erreurs. Le concepteur d'*OCELLES* annonce vouloir coupler sa ressource à une ressource pédagogique. Des entretiens plus complets sont envisagés avec ces concepteurs.

4.2. Point de vue des utilisateurs

Le but de cette première enquête « utilisateurs » était de collecter un premier jeu de données par rapport à l'utilisation effective des ressources existantes (motif de recherche, degré de satisfaction) et à d'éventuelles suggestions d'amélioration. Le sondage comportait treize questions rédigées en français écrit mais également disponibles en LSF par le biais de vidéos. Les réponses ont été apportées en français écrit, principalement sous la forme de cases à cocher et quelques champs de texte pour apporter d'éventuels commentaires ou remarques complémentaires. Le sondage a été diffusé en mai 2023 par courrier électronique au sein de notre réseau de formateurs en LSF, élèves/étudiants et interprètes (l'utilisation du réseau a aussi été le choix fait par Vale 2015 dans son étude empirique). Vingt-trois résultats se sont avérés utilisables. Nous en résumons ici les caractéristiques principales.

Les utilisateurs des ressources sont a priori tout autant des personnes sourdes que des personnes entendantes, et d'un âge plutôt inférieur à 40 ans, ceci correspondant peut-être à une plus grande aisance avec les ressources numériques, alors que les répondants de plus de 40 ans seraient plutôt utilisateurs de ressources papier. La grande majorité des utilisateurs sont des professionnels de la langue, principalement des enseignants ou des interprètes, et, pour un tiers, des usagers sourds.

Les ressources les plus consultées sont, par ordre décroissant : *Elix*, *Sourds.net*, *Sématos* et *SpreadTheSign*, sachant que la ressource la plus employée ne l'est que rarement (quelques fois dans le mois) : ainsi, ces ressources ne sont pas réellement utilisées autant qu'elles pourraient l'être, en tous cas pas d'après le retour des usagers.

En grande majorité, les utilisateurs y cherchent un signe, mais il est intéressant de constater qu'ils viennent aussi, pour un quart d'entre eux, vérifier un signe qui leur aurait été fourni, un signe qu'ils auraient trouvé par ailleurs, ou un signe pour lequel ils auraient un doute.

Bien que peu utilisées, les ressources donnent « plutôt » satisfaction (*Sourds.net* en tête, suivi par *Elix*, *SpreadTheSign* et *Sématos*), mais, à part pour *Elix*, il y a également de nombreuses personnes insatisfaites. Les critiques qui prédominent sont les suivantes : manque de contenus (vidéos de signes) et présence d'erreurs dans les contenus proposés.

En conclusion de cette partie centrée sur les avis de quelques concepteurs et d'un petit échantillon des utilisateurs des ressources, il semble assez frappant de noter les divergences qui peuvent exister entre les concepteurs, qui se disent plutôt satisfaits, et semblent relativement peu enclins à faire évoluer leurs ressources, et les avis des utilisateurs, qui semblent ne pas être entièrement satisfaits. On peut se demander également dans quelle mesure les retours des utilisateurs (lorsqu'ils sont autorisés par la ressource) sont pris en compte.

5. Perspectives

Cette première étude a permis de défricher un terrain encore inexploré, celui de la lexicographie numérique de la LSF. Les résultats doivent toutefois en être approfondis car chacun des onze points évalués concernant la macrostructure et la microstructure des ressources mérite une analyse plus détaillée. Les premières enquêtes portant sur les concepteurs et les utilisateurs doivent être complétées par des entretiens (auprès des concepteurs) et par des sondages diffusés auprès d'un plus large échantillon d'utilisateurs, adapté en fonction des profils (qu'il reste à définir précisément : entendants/Sourds ; apprenants/professionnels ; types de professionnels, etc.). Cerner les besoins et les usages spécifiques de certains de ces profils, notamment ceux des interprètes français-LSF, fait partie de nos objectifs à moyen terme, en lien avec les recherches entamées par Emma Bancillon (Bancillon *et al.*, 2024 ; Bancillon, 2024). En effet, comme Vale (2015 : 299-300) a pu le constater pour les utilisateurs du dictionnaire de NZSL, même si ce qui constitue l'essentiel des utilisateurs des ressources relève plutôt d'un public occasionnel, leur éventuelle amélioration devrait avant tout viser d'autres profils d'utilisateurs, qui regardent la ressource en détail et qui auraient besoin de types d'information spécifique (information métalinguistique, par exemple). Des possibilités de personnalisation des outils selon les différents types de profils sont envisagés (Vale, 2015 : 299-300).

En ce qui concerne la manière dont le numérique pourrait venir améliorer les ressources, on note quelques pistes possibles : (1) la recherche inversée (de la langue des signes vers la langue vocale), telle qu'elle a été implémentée par exemple dans le *Dictionnaire contextuel bilingue* en langue des signes belge de Namur qui repose sur l'utilisation de la captation vidéo⁴⁸, (2) les signeurs virtuels ou avatars en 3D, qui permettent l'anonymisation, (3) les possibilités de validation claire des données (qui peut augmenter la confiance que les utilisateurs accordent à la ressource), (4) la contextualisation des signes proposés.

48. <https://dico.corpus-lsfb.be/> [Consulté le 10 décembre 2023]

Des études menées sur d'autres LS que la LSF ouvrent des perspectives. La piste, à l'initiative des concepteurs du dictionnaire de NZSL, qui consiste à impliquer les utilisateurs de la ressource pour l'améliorer (Vale et McKee, 2019), par exemple à l'aide de la plateforme collaborative NZSL Share⁴⁹, semble particulièrement intéressante.

49. <https://www.nzslshare.nz/> [Consulté le 10 décembre 2023]

Bibliographie

- ATKINS, B. T. Sue et Michael Rundell (2008), *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- BAKER Charlotte et Carol Padden (1978), « Focusing on the nonmanual components of American Sign Language » dans P. Siple (dir.), *Understanding Language through Sign Language Research*, New York : Academic Press, p. 59-90.
- BANCILLON, Emma, Amélie Josselin-Leray et Jérémie Segouat (2024), « The Use of Online Terminological Resources by French Sign Language Interpreters », Poster présenté au *STEM Global summit*, du 2 au 5 mars, Université Gallaudet, États-Unis.
- BANCILLON, Emma (2024), *Les interprètes en langue des signes face à la terminologie : étude exploratoire de l'utilisation des ressources en ligne pour la LSF*, mémoire de Master 2 Traduction et Interprétation, Université Toulouse - Jean Jaurès.
- BONNAL, Françoise (2005), « Sémio-génèse de la langue des signes française : étude critique des signes de la langue des signes française attestés sur support papier depuis le XVIII^e siècle et nouvelles perspectives de dictionnaires », Thèse de doctorat, Université Toulouse 2. [en ligne : <https://www.theses.fr/2005TOU20036>]
- BONNAL-VERGÈS, Françoise (2006), « Langue des signes française (LSF) : des lexiques des XVIII^e et XIX^e siècles à la dictionnaire du XXI^e siècle », *Glottopol*, n° 7, p. 160-190.
- BONNAL-VERGÈS, Françoise (2012), « Les dictionnaires de traduction français-langue des signes française (LSF) : Une gageure pour le passé et l'avenir » dans Michaela Heinz (dir.), *Dictionnaires et traduction*, Actes des quatrième journées allemandes des dictionnaires, Frank et Timme, p. 267-304.
- BONNAL-VERGÈS, Françoise (2014), « 1784, abbé Jean Ferrand, Chartres : sémiotique du premier dictionnaire de la langue des signes française » dans Michaela Heinz (dir.), *Les sémiotiques du dictionnaire*, Frank et Timme, p. 225-260.
- BOUORA, Leila et Ivani Fusellier-Souza (2009), « La langue des signes française, langue enseignée et langue d'enseignement : un état des lieux » dans Patrick Sauzet et François Pic (dir.), *Politique linguistique et enseignement des Langues de France*, L'Harmattan, p. 209-222.
- BRAFFORT, Annelies (2022), « Langue des signes française : état des lieux des ressources linguistiques et des traitements automatiques », Journées jointes des groupements de recherche linguistique informatique, formelle et de terrain (LIFT) et traitement automatique des langues (TAL), Novembre 2022, Marseille (France), p. 131-138.
- BROSENS, Caro, Hannes De Durpel, Inez Beukeleers et Beatrijs Wille (2021), « Towards a functional label set for the online dictionary of Flemish Sign Language », *FEAST, Formal and Experimental Advances in Sign language Theory*, n° 4, p. 50-60. [En ligne : <https://doi.org/10.31009/FEAST.i4.04>]

- CANTIN, Yann, et Florence Encrevé (2022), « Perspectives : On the Historicalness of Sign Languages », *Frontiers in Communication*, n° 7. [En ligne : <https://www.frontiersin.org/articles/10.3389/fcomm.2022.801862>]
- CORBIN, Pierre (2020), « Les dictionnaires monolingues généraux du français actuel gratuits en ligne : Évolutions récentes », *Academic Journal of Modern Philology*, n° 9, p. 65-77.
- CORBIN, Pierre, et Nathalie Gasiglia (2020), « Les dictionnaires monolingues généraux du français “actuel” gratuits en ligne », *SHS Web of Conferences*, n° 78. [En ligne : <https://doi.org/10.1051/shsconf/20207805008>]
- CUXAC, Christian (2007), « Une manière de reformuler en langue des signes française », *La linguistique*, vol. 43 n° 1, p. 117-127. [En ligne : <https://doi.org/10.3917/ling.431.0117>]
- CUXAC, Christian (2000), *La LSF, les voies de l'iconicité*. Ophrys, Paris
- CUXAC, Christian, et Elena Antinoro Pizzuto (2010), « Émergence, norme et variation dans les langues des signes : Vers une redéfinition notionnelle », *Langage et société*, vol. 131, n° 1, p. 37-53. [En ligne : <https://doi.org/10.3917/ls.131.0037>]
- DALLE, Patrice et Juliette Dalle (2013), « Langues des signes, enseignement et nouvelles technologies », *La nouvelle revue de l'adaptation et de la scolarisation*, vol. 64, n° 4, p. 189-202. [En ligne : <https://doi.org/10.3917/nras.064.0189>]
- DELAPORTE, Yves (2005), « La variation régionale en langue des signes française ». *Marges Linguistiques*, n° 10 («Langues régionales», présenté par Claudine Moïse, Véronique Fillol, Thierry Bulot), p. 118-132.
- ENCREVÉ, Florence (2008), « Sourds et société française au XIX^e siècle : 1830-1905 ». Thèse de doctorat. Université Paris 8.
- FISHER, Renate (2014), « Les racines de la lexicographie des langues des signes : L'Épée, Ferrand, Jarisch et Pélissier », dans Michaela Heinz (dir.), *Les sémiotiques du dictionnaire*, Frank et Timme, p. 199-224.
- GALANT, Philippe et Ronit Laquerrière-Leven (2013), *La langue des signes*, tome 4, Éditions IVT.
- GARCIA, Brigitte (2010), *Sourds, surdité, langue(s) des signes et épistémologie des sciences du langage. Problématiques de la scripturisation et modélisation des bas niveaux en langue des signes française (LSF)*, note de synthèse rédigée en vue de l'obtention de l'Habilitation à diriger des recherches, Université Paris 8. [En ligne : <https://hal.archives-ouvertes.fr/tel-01709759>]
- GIACOLETTI, Lucas (2024), *Motivations et procédés de création d'un lexique spécifique à l'environnement, en Langue des signes française : exemple du signaire ornithologique de la Ligue pour la Protection des Oiseaux et de STIM Sourd France*, mémoire de M1 Sciences du Langage, Université de Lille.
- JOSSELIN-LERAY, Amélie (2005), *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues - Etude d'un domaine de spécialité : volcanologie*. Thèse de doctorat non publiée, Université Lumière Lyon II.
- KERBOURC'H, Sylvain (2006), « Le Réveil Sourd : d'hier à aujourd'hui (1971-2006). De l'action collective d'un mouvement culturel pour la réhabilitation de la langue des signes françaises, à l'affirmation d'une identité collective pour la participation sociale des sourds ». Thèse de doctorat, EHESS.

- KREMnitz, Georg (2020), « La problématique initiale de la liste Cerquiglini et ses effets ultérieurs », *Glottopol*, n° 34, p. 37-45.
- KRISTOFFERSEN, Jette Hedegaard, Thomas Troelsgård, Sylviane Granger et Magalie Paquot (2012), « The electronic lexicographical treatment of sign languages : The Danish Sign Language Dictionary », *Electronic lexicography*, Oxford University Press, p. 292-313.
- MARELLO, Carla (1998), « Hornby's bilingualized dictionaries », *International Journal of Lexicography*, vol. 11, n°4, p. 292-314. [En ligne : <https://doi.org/10.1093/ijl/11.4.292>]
- MCKEE, David et Alexander Pivac (2008), « NZSL Online Dictionary project 2008—2011 : User requirements survey report ». DSRU.
- MCKEE, Rachel et David McKee (2013), « Making an Online Dictionary of New Zealand Sign Language ». *Lexikos*, n° 23, p. 500-531. [En ligne : <https://doi.org/10.4314/lex.v23i1>]
- MCKEE, Rachel et David McKee (2011). « Old Signs, New Signs, Whose Signs? Sociolinguistic Variation in the NZSL Lexicon ». *Sign Language Studies*, vol. 11, n°4, p. 485-527.
- MILLET, Agnès, Marion Kobylanski et Laurent Verlaine (2023), *LSF : Grammaire pratique : comprendre, enseigner, apprendre*. UGA éditions, Université Grenoble Alpes.
- MOODY, Bill, Agnès Vourc'h, Michel Girod et Michel Duf (1997), « La langue des signes. Tome 2 : Dictionnaire bilingue LSF / français », Éditions IVT.
- MOODY, Bill, Agnès Vourc'h, Michel Girod, Dominique Hof, Rachid Benelhocine et Serge Dumartin (1998), « Historique », *La langue des signes. Introduction à l'histoire et à la grammaire de la langue des signes entre les mains des sourds*. Éditions IVT. p. 17-44.
- MOREAU, Cédric (2012), « Stratégies de reconstruction du sens en langue des signes française à partir de données incomplètes en et hors contexte : Perspectives pour la constitution d'un lexique-dictionnaire à entrée directe en langue des signes ». Thèse de doctorat, Université Paris 8 - Vincennes-Saint-Denis. [En ligne : <https://www.theses.fr/2012PA083515>]
- NAPIER, Jemina, George Major, Lindsay Ferrara et Trevor Johnston (2015), « Medical Signbank as a model for sign language planning? A review of community engagement », *Current Issues in Language Planning*, vol. 16, n°3, p. 279-295. [En ligne : <https://doi.org/10.1080/14664208.2014.972536>]
- NESI, Hilary (2008), « Dictionaries in electronic form » dans Anthony Paul Cowie, *The Oxford History of English Lexicography*, Oxford University Press. p. 458-478.
- NESI, Hilary (2015), « Researching Users and Uses of Dictionaries » dans Howard Jackson (dir.), *The Bloomsbury Companion to lexicography*. Bloomsbury Academic. p. 62-74. [En ligne : <https://www.bloomsbury.com/uk/bloomsbury-handbook-of-lexicography-9781350181700/>]
- PASTOR, Veronica et Amparo Alcina (2010), « Search Techniques in Electronic Dictionaries : A Classification for Translators », *International Journal of Lexicography*, vol. 23, n°3, p. 307-354. [En ligne : <https://doi.org/10.1093/ijl/ecq015>]

- SAJOUS, Franck (2023), « Quantité et qualité dans le Wiktionnaire : de la diversité. . . à la rigueur ? » dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [En ligne : <https://journals.openedition.org/linx/9835>]
- SALLANDRE, Marie-Anne (2020), « Comparaisons typologiques entre les langues des signes, une approche sémiologique », *Lalies*, 40, p. 9-30.
- SALLANDRE, Marie-Anne et Brigitte Garcia (2021), « Langue des signes française et linguistique : enjeux actuels ». *Le français moderne - revue de linguistique française*, vol. 2021-2, p. 281-292.
- SEGOUAT, Jérémie et Amélie Josselin-Leray (à paraître), « First Thoughts on Domain Gain in French Sign Language - the Dynamics of Specialized Sign Creation in LSF », Amsterdam University Press.
- SENNIKOVA, Yana et Brigitte Garcia (2018), « Statut et rôle des composants sublexicaux dans la structuration du lexique en langue des signes française (LSF) », *Lexique*, p. 131-159. [En ligne : <https://hal.science/hal-01980496>]
- TARP, Sven, (2009), « Reflections on Lexicographical User Research », *Lexikos*, n° 19. [En ligne : <https://doi.org/10.5788/19-0-440>]
- VALE, Mireille (2015), « A study of the users of an online sign language dictionary », *Proceedings of the eLex 2015 conference*, p. 281-303.
- VALE, Mireille et Rachel McKee (2019), « NZSL Online Dictionary 8 years on. Adapting to a changing environment through user research », *Australlex Conference*, Canberra.
- VANBRUGGHE, Anne, Cédric Moreau et Véronique Geffroy (2010), « OCELLES: Une plateforme collaborative LSF / bilingue français », *La nouvelle revue de l'adaptation et de la scolarisation*, n° 49, p. 163-169.
- ZWITSERLOOD, Inge (2010), « Sign Language Lexicography in the Early 21st Century and a Recently Published Dictionary of Sign Language of the Netherlands », *International Journal of Lexicography*, vol.23, n° 4, p. 443-476. [En ligne : <https://doi.org/10.1093/ijl/ecq031>]
- ZWITSERLOOD, Inge, Jette Hedegaard Kristoffersen et Thomas Troelsgård (2022), « Issues in sign language lexicography » dans Howard Jackson (dir.), *The Bloomsbury Handbook of Lexicography*. Bloomsbury. p. 227-249.

TITRE: POUR UNE ANALYSE QUALITATIVE *ET* QUANTITATIVE, MANUELLE *ET* COMPUTATIONNELLE, SYNCHRONIQUE *ET* DIACHRONIQUE, DES DICTIONNAIRES NUMÉRIQUES

AUTEUR: FRANCK SAJOUS, CLLE, CNRS ET UNIVERSITÉ DE TOULOUSE – JEAN JAURÈS

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 153-186

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/2241](http://hdl.handle.net/11143/2241)

DOI: [HTTPS://DOI.ORG/10.17118/11143/2241](https://doi.org/10.17118/11143/2241)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Pour une analyse qualitative et quantitative, manuelle et computationnelle, synchronique et diachronique, des dictionnaires numériques

Franck Sajous, CLLE, CNRS et Université de Toulouse – Jean Jaurès

Résumé :

Dans un contexte de profondes mutations (achevées et en cours) de l'activité lexicographique et de l'offre dictionnaire, cet article aborde la question des pratiques métalexigraphiques, et entend donner un éclairage méthodologique à l'activité centrale qu'est l'analyse des dictionnaires. Un accent spécifique est mis sur les possibilités qu'offrent les dictionnaires numériques, ainsi que les limites et contraintes qu'ils imposent à ces analyses. Après avoir inventorié les pratiques actuelles, en distinguant analyses qualitatives et analyses quantitatives d'une part et analyses manuelles et analyses automatiques d'autre part, et en montrant les difficultés afférentes à chaque type d'analyse, une approche menant de front analyses qualitatives manuelles et analyses quantitatives automatiques est proposée.

Mots-clés :

Métalexigraphie, méthodologie d'analyse des dictionnaires, dictionnaires numériques, processus lexicographique

Abstract :

Against the backdrop of past and ongoing changes of both the dictionary-making process and the offer of dictionaries, this article deals with metalexigraphic practices – especially dictionary analysis – from a methodological perspective. Particular emphasis is placed on the possibilities offered by digital dictionaries, as well as the limits and constraints they impose on such analyses. After surveying current practices, distinguishing between qualitative and quantitative analyses on the one hand, and manual and automatic analyses on the other, and highlighting the difficulties associated with each type of analysis, an approach that combines manual qualitative analyses and automatic quantitative analyses is proposed.

Keywords :

Metalexigraphy, methodology of dictionary analysis, digital dictionaries, dictionary-making process

1. Introduction

Qu'on les qualifie de « bouleversements » ou de « révolutions », de profondes mutations ont transformé – et continuent de transformer – la lexicographie aux plans théorique, technologique et économique¹ : révolution descriptive (Trap-Jensen, 2018), numérisation des dictionnaires papiers, rétroconversion et structuration sous forme de bases de données (Nagao *et al.*, 1980; Berg *et al.*, 1988), apports de la linguistique de corpus (Rundell et Stock, 1992) et automatisation par des outils de traitement automatique des langues et d'analyse de données (Rundell et Kilgarriff, 2011), diversification des supports de publication et mise en ligne (Nesi, 2008), arrêt pour certains dictionnaires de l'impression papier (Rundell, 2014)², changement de modèle économique (Kilgarriff, 2005), apparition de dictionnaires « DIY » et d'agrégateurs (Gao, 2012), émergence de la lexicographie dite « collaborative³ » et des approches par *crowdsourcing* (Sajous et Josselin-Leray, 2022), arrivée des outils fondés sur des grands modèles de langue comme ChatGPT, dont le potentiel et les limites pour la lexicographie font actuellement l'objet d'études (Rundell, 2023; Jakubíček et Rundell, 2023; Lew, 2023) qui ont vocation à être affinées, complétées et réitérées à mesure que cette famille d'outils évolue. Certains des produits dictionnaires issus de ces (r)évolutions, ou affectés par elles, ont fait l'objet d'études métalexigraphiques dans le cadre des éditions successives des colloques *Dictionnaires et culture numérique dans l'espace francophone*, initiés en 2016. Dans le présent chapitre, nous nous demandons ce qu'il en est de la métalexigraphie : la méthodologie classique des métalexigraphes – pour peu qu'une abstraction aussi vague et englobante ait un sens – est-elle affectée ou remise en cause par les changements de pratiques lexicographiques ? Nous nous demanderons notamment si l'on peut – et si l'on doit – analyser les dictionnaires électroniques comme on analyse les dictionnaires imprimés. À cette occasion, nous nous demanderons également quelle est la fiabilité des conclusions tirées des analyses métalexigraphiques portant habituellement sur une partie seulement du dictionnaire (papier comme électronique). Avant de répondre à ces questions, nous expliquons en section 2 ce que recouvrent les différentes acceptions du terme *métalexigraphie*. L'analyse des dictionnaires (description, comparaison et évaluation de leur structure et de leur contenu) étant une activité centrale de cette discipline, la suite de l'article propose une réflexion d'ordre méthodologique et épistémologique sur ses différentes mises en œuvre, indépendamment des visées applicatives. Nous tentons d'établir, en section 3, des éléments de typologie des méthodes d'analyse avant de mettre en évidence, en sections 4 à 7, les difficultés théoriques et pratiques que pose chacun des

1. L'impact et la temporalité de ces évolutions sur la pratique lexicographique et sur l'offre dictionnaire dépendent fortement des langues décrites. Sur l'évolution de la lexicographie française et sa situation au début du 21^e siècle, on peut lire Corbin (2008). Concernant l'offre dictionnaire contemporaine, lire Corbin (2020) et Corbin et Gasiglia (2020, 2022, 2023).

2. Michael Rundell annonçait en 2014 la fin de l'impression papier du *Macmillan English Dictionary for Advanced Learners*. En juin 2023, Macmillan Education Ltd annonce sa décision de mettre un terme à sa version électronique, le *Macmillan English Dictionary* (en ligne depuis 2009) et au thésaurus associé. Le communiqué ne mentionne pas l'*open dictionary*, complément participatif du dictionnaire, qui disparaît également.

3. Sajous (2023) explique pourquoi l'adjectif *contributive* est plus approprié pour la qualifier.

types de méthode identifiés et les limites qui en découlent. En section 8, nous dressons un bilan des pratiques recensées avant d'envisager une démarche reposant sur la complémentarité des méthodes les plus pertinentes.

2. Quelle(s) métalexigraphie(s) ?

Dans la littérature anglophone, *metalexigraphy*, emprunté au français *métalexigraphie*⁴, voisine avec *lexicographic research*, *academic lexicography*, *dictionary research*, *theory of lexicography/theoretical lexicography*, *dictionary criticism/dictionary review*. Ces dénominations sont utilisées tantôt comme co-hyponymes par différents auteurs pour contraster des activités de recherche connexes mais distinctes, tantôt comme synonymes (l'alternance étant parfois le fait d'une même personne). Plus que le foisonnement lexical, c'est l'instabilité des usages que regrette Hartmann (2001: 28) : « [metalexigraphy] has caught on in the published literature as an alternative to 'theoretical lexicography' or 'theory of lexicography'. Unfortunately, it can be used by different authorities to refer to potentially quite different things ». Il juge trop vague la définition de l'équivalent français formulée par Hausmann dans l'*Encyclopédie internationale de la lexicographie* : « toute activité qui fait du dictionnaire un objet de réflexion et de recherche mais qui, elle-même, ne vise pas à la production de dictionnaires » (Hausmann *et al.*, 1989 : 216). Quemada parlait à la même époque de « l'avènement d'une *lexicographie critique et théorique* dite par certains *métalexigraphie*, discipline nouvelle dont l'objectif principal est l'étude des types et des méthodes des dictionnaires de langue » (187 : 233). Béjoint décrit un usage conforme à la définition de Hausmann *et al.* : « The word *metalexigraphy* [...] is now frequently used to refer to the activities of anyone who writes about lexicography but does not write dictionaries » (2000 : 8). Contrairement à Hartmann, nous souscrivons à l'acceptation large décrite par Hausmann *et al.* et Béjoint (ainsi qu'à celle de Quemada, dont ne parle pas Hartmann), qui font de *métalexigraphie* un terme générique désignant un continuum de sous-disciplines – ou d'activités menées au sein de ces sous-disciplines – dont les recouvrements rendent difficile une démarcation franche. Ces sous-disciplines et activités se distinguent notamment par les objets d'étude, qui, comme nous le montrons en sections 2.1 à 2.4, dépendent des buts poursuivis.

2.1. Motivations et objets d'étude

À l'exception des recherches portant sur les utilisateurs et utilisatrices, abordées en section 2.2, la métalexigraphie a pour objets d'étude les dictionnaires et leurs procédés de conception. L'analyse métalexigraphique peut viser leur simple description – critique ou non – ou l'identification de voies

4. D'après Hartmann (2001), on lirait communément que c'est Rey-Debove (1971) qui aurait forgé le mot *métalexigraphie*. Béjoint (2000) attribue quant à lui la paternité du mot à Rey et Delesalle (1979). Sans préjuger de qui l'a réellement forgé, on trouve une occurrence de ce mot dans l'article de Rey et Delesalle (p. 4), mais pas, sauf erreur de notre part, dans l'ouvrage de Rey-Debove.

d'amélioration des uns et des autres⁵. Pour Gouws (2011 : 28-29), qui reprend l'opposition de Tarp (2008 : 9-11) entre approche *contemplative* et *transformative* de la théorie lexicographique, la métalexicographie ne devrait pas se contenter de décrire ce qui est fait, mais devrait préconiser ce qui devrait être fait, c'est-à-dire être force de proposition pour améliorer le processus lexicographique. Nous ne partageons que partiellement ce point de vue. S'il est louable, pour un métalexicographe, d'accompagner ses observations et analyses de propositions de modification des procédés et traitements lexicographiques, l'absence de prescription n'enlève rien à l'utilité d'une description pertinente. D'une part, parce que les destinataires des descriptions métalexicographiques ne sont pas uniquement les lexicographes. À condition de produire, outre des écrits scientifiques, des discours vulgarisés et publiés dans des supports suffisamment accessibles, les descriptions métalexicographiques peuvent trouver leur public dans les acquéreurs (la critique de dictionnaires comme guide d'achat) et usagers (comment interpréter le discours lexicographique) des dictionnaires, les enseignants du primaire (enseignement du vocabulaire et éducation à l'utilisation du dictionnaire) et de langue seconde ou étrangère⁶. D'autre part, parce que les lexicographes (non nécessairement métalexicographes) peuvent s'emparer des analyses des métalexicographes (non nécessairement lexicographes) pour envisager des modifications des procédés de conception. En effet, « il faut voir dans chaque critique correctement fondée une proposition constructive d'amélioration des dictionnaires », comme l'écrit Corbin (1984 : 114), qui considère les métalexicographes jouant un rôle de *lexicographe-conseil* comme « des prestataires de services bénévoles mettant leurs travaux à la disposition des lexicographes professionnels » (*ibid.*). D'après Corbin (1995a), le travail fragmenté au jour le jour des lexicographes ne leur permet d'avoir une vision ni distanciée ni globale du dictionnaire. Dans ces conditions, l'image des dictionnaires que leur renvoient les métalexicographes universitaires est d'autant plus précieuse. Béjoint va dans le même sens : « [L]inguists enjoy a luxury that lexicographers do not have: time, which makes it possible to think, to try out solutions, to make mistakes and start all over again » (2000 : 177). Des « accusés de réception » des analyses critiques sont relevés par Corbin (1984) dans le discours des lexicographes. Concernant la prise en compte effective de ces critiques et leurs retombées lexicographiques réelles, qui s'observent – ou non – dans le dictionnaire, Corbin relativise cependant : « [L]es suggestions des métalexicographes n'occupent pas le premier rang dans les critères de décision des multinationales de l'édition. » (1991 : 167).

2.2. Études sur les utilisateurs

Les recherches sur les utilisateurs et sur leurs usages du dictionnaire s'intéressent à l'identification du profil de ces personnes (compétences et lacunes), leurs besoins et utilisations effectives (quelle

5. « Amélioration » peut s'entendre en termes de qualité des dictionnaires (contenu, structure, utilisabilité, adaptation aux tâches et publics visés) ou de rentabilité, par « rationalisation » des coûts de production (temps de développement, main d'œuvre, etc.).

6. Nous avons conscience du décalage entre l'énoncé de cette potentialité, l'état de la lexicographie française et celui de l'enseignement et de la formation pédagogique.

information recherchée pour effectuer quelle tâche) ainsi que leurs stratégies de consultation. Ces études, qui visent l'amélioration de l'utilisabilité des dictionnaires par l'établissement de correspondances adéquates entre types de dictionnaire, types d'utilisateurs et types d'utilisation, recourent à des méthodes particulières : enquêtes par questionnaires ou interviews, expériences en laboratoire plus ou moins sophistiquées (observation, protocole « *think-aloud* », *eye-tracking*) et, pour les dictionnaires en ligne, analyse des fichiers de log et des interactions (enregistrement des frappes au clavier et mouvements/clics de la souris). Ces méthodes sortant du cadre de ce chapitre, nous renvoyons à la littérature abondante sur ce champ d'étude. Pour un panorama synthétique, lire par exemple Lew (2015) et Nesi (2022).

2.3. Analyse des procédés de conception

Considérant la prolifération d'ouvrages consacrés à la métalexigraphie et le peu de citations qu'ils suscitent, de Schryver (2022) se demande⁷ si cette discipline traverse une crise existentielle. En le lisant, et en considérant les titres qu'il mentionne, on comprend que de Schryver s'intéresse au versant de la métalexigraphie qui théorise les procédés de conception des dictionnaires, en vue de les améliorer (les procédés et, éventuellement, les dictionnaires). Ses préoccupations rejoignent celles de Hartmann, pour qui une approche réflexive sur la pratique lexicographique peut permettre d'améliorer cette dernière : « Whether we call this 'metalexigraphy' or 'dictionary research', a better understanding of the theoretical foundations of lexicography is likely to raise standards in both dictionary making and use » (2001 : 28). Pour de Schryver comme pour d'autres métalexigraphes-lexicographes, la connaissance des procédés de conception est supposée établie, soit qu'elle découle d'un consensus entre praticiens, soit qu'elle émane des discours (paratexte, communication à destination des usagers et usagères via les sites et blogs des maisons d'édition ou par presse interposée, publications scientifiques, etc.) qui les décrivent. Or, selon Coleman et Ogilvie (2009), ces discours de première main, bien que précieux, ne doivent pas être pris pour argent comptant car ils sont nécessairement subjectifs, voire soumis à certaines motivations à déformer la réalité. Pour ces autrices, les preuves les plus fiables sont celles contenues dans le corps même du dictionnaire et son analyse est à même de mettre au jour des faisceaux d'indices permettant de reconstituer la pratique lexicographique. Cette démarche de reconstitution est également au cœur de la série d'études *Le monde étrange des dictionnaires* qui « consiste à refaire le parcours des lexicographes de façon à dégager les principes mis en œuvre dans leur pratique » (Corbin, 1995b : 125). Ainsi, pour améliorer les procédés de conception, il faut les connaître, et cette connaissance repose – notamment – sur celle des dictionnaires.

7. Plus exactement, comme on le comprend dans un article ultérieur, il feint de se le demander. Qualifiant de polémique son questionnement, il y répond par la négative (de Schryver, 2023a).

2.4. Analyse des dictionnaires

Qu'il s'agisse de décrire un dictionnaire spécifique, de comparer des dictionnaires différents ou deux éditions d'un même dictionnaire, que la description porte sur la globalité du dictionnaire, une partie spécifique du vocabulaire ou le traitement d'un fait linguistique particulier, l'analyse du ou des dictionnaire(s) constitue une activité centrale de la métalexigraphie, pour au moins deux raisons :

- d'une part, c'est la seule manière fiable de produire des connaissances sur ce que contient réellement le dictionnaire. Les discours lexicographiques sont souvent lacunaires et parfois trompeurs. Les distorsions entre discours lexicographique et faits dictionnaires peuvent être volontaires, comme le relèvent Coleman et Ogilvie (2009), mais elles peuvent également résulter de la vision parcellaire du dictionnaire qu'en ont les lexicographes « que la division des tâches et les lacunes des programmes rédactionnels mettent dans la situation de ces acteurs qui ne peuvent comprendre quel film ils ont tourné qu'une fois qu'ils le voient projeté » (Corbin, 1984 : 114).
- d'autre part, l'analyse des dictionnaires est au moins complémentaire des autres activités métalexigraphiques. Nous avons en effet mentionné en section 2.3 que l'analyse d'un dictionnaire permet de mettre au jour des indices sur le processus lexicographique qui l'a produit. L'analyse des procédés passe donc, notamment, par l'analyse du produit qui en résulte. Concernant les études sur les utilisateurs, l'analyse d'un dictionnaire permet de vérifier, dans une certaine mesure, si celui-ci est susceptible de remplir les besoins du public visé.

Donnons ici un exemple qui illustre d'une part la nécessité d'analyser les dictionnaires plutôt que de se fier aux « discours de première main » et d'autre part la méconnaissance que peuvent avoir les lexicographes du dictionnaire qu'ils produisent. Dans sa thèse portant sur le *Trésor de la langue française* (TLF), Radermacher (2004 : 400-413) étudie le nombre moyen d'exemples par page dans le premier et le dernier tome du dictionnaire. Cette étude est motivée par l'annonce faite par Paul Imbs, dans la rubrique « Au lecteur » du tome II, de la diminution du nombre d'exemples – diminution par la suite « constatée » et déplorée par plusieurs linguistes. L'étude quantitative de Radermacher montre que cette diminution est un mythe⁸.

Malgré la place centrale qu'occupe l'analyse des dictionnaires, il n'existe à notre connaissance aucun ouvrage de référence sur la pratique de cette activité – vide bibliographique qui contraste avec l'abondance de manuels de lexicographie et autres ouvrages de lexicographie théorique relevée par de Schryver (2022). Du côté des recherches sur la « revue » ou la « critique » de dictionnaires (*dictionary review/dictionary criticism*⁹), les réflexions se contentent généralement de regretter que les analyses ne se fondent sur aucune liste de critères préalables et de conclure qu'il est nécessaire d'en

8. Imbs n'aura pas connu le dernier tome du TLF, mais concernant les tomes I à VII qu'il a dirigés, une étude (en cours) menée automatiquement sur l'intégralité du *Trésor de la langue française informatisé* (TLFi) montre que le nombre médian d'exemples reste stable sur les quatre premiers, avant d'augmenter.

9. Voir notamment les travaux du « cercle de Tokyo » et le volume de *Lexicographica Series Maior* intitulé *Dictionary Criticism* (Bielińska et Schierholz, 2017).

établir une. Ces recherches débouchent au mieux sur l'énumération de tels critères mais ne donnent jamais de procédures opérationnelles pour vérifier si un dictionnaire les satisfait. Même un article intitulé *Methods in dictionary criticism* (Swanepoel, 2017), qui énumère ce qu'il convient de faire (le *quoi*, i.e. à quelles questions la critique doit répondre), reste totalement silencieux sur le *comment*. Plusieurs articles de Pierre Corbin, publiés entre la fin des années 1980 et le milieu des années 1990, fixent un cadre pour une lecture critique des dictionnaires. Il y est notamment question de « l'examen d'un faisceau de critères que l'on peut utiliser pour caractériser et évaluer les dictionnaires » (Corbin, 1995a : 90-93). Ces critères sont énumérés et discutés de manière convaincante mais relativement succinctement. Malgré sa concision, ce passage sur les critères de caractérisation et d'évaluation des dictionnaires ouvre la voie à l'élaboration des fondements d'une méthodologie d'analyse, dont il constitue un point de départ stimulant, qui reste à poursuivre. La présente étude ne prétend nullement s'atteler à un tel chantier, certainement trop vaste pour un seul auteur et un unique chapitre. Il s'agit plutôt ici de recenser des pratiques observées et de les commenter, d'un point de vue méthodologique, et avec un regard particulier : celui d'un informaticien-métalexigraphe pratiquant essentiellement l'analyse des dictionnaires électroniques, outillée à chaque fois que cela est possible et semble pertinent. C'est à ces questions de possibilité (de mise en œuvre) et de pertinence (de méthode, pour mener l'analyse d'un phénomène donné dans un dictionnaire donné) que la suite de l'article est consacrée.

3. Éléments de typologie des méthodes d'analyse

Nous nous intéressons dans cette section à la catégorisation, d'un point de vue méthodologique, des différents types d'analyse des dictionnaires qui guidera, en sections 4 à 7, notre recension des pratiques métalexigraphiques. L'ensemble des traits que nous retenons est le suivant :

1. Motivation : visée descriptive et/ou prescriptive et/ou applicative
2. Portée de l'analyse
 - a) ensemble du dictionnaire
 - b) vocabulaire spécifique
 - c) phénomène linguistique ou traitement lexicographique particulier
3. Nature de l'analyse
 - a) quantitative vs qualitative
 - b) synchronique vs diachronique
4. Possibilité de mise en œuvre : manuelle vs automatique
5. Profil des métalexigraphes
 - a) lexicographes ou non
 - b) degré de compétences en informatique (et en statistiques)

6. Caractéristiques du dictionnaire

- a) support : papier vs électronique
- b) mise à jour : à intervalles réguliers, continue ou inexistante
- c) contraintes légales : domaine public, licence libre, *copyright*, conditions d'utilisations...

Certains des traits recensés ci-dessus sont interdépendants. Par exemple, les métalexigraphes dépourvus de compétence informatique ne peuvent pas mener d'analyses automatiques et ne peuvent donc raisonnablement pas étudier l'ensemble du dictionnaire. Les possibilités de mise en œuvre et la nature de l'analyse sont en outre conditionnées par certaines caractéristiques du dictionnaire étudié. Par exemple, une analyse automatique n'est possible que pour un dictionnaire électronique (ou dictionnaire papier rétroconverti et encodé dans un format exploitable) dont le statut légal est suffisamment permissif.

3.1. Commentaires

3.1.1. Visée de l'analyse

Nous l'avons vu en section 2, l'analyse des dictionnaires peut être descriptive, accompagnée ou non d'une dimension prescriptive (*transformiste*). Cette activité consiste à décrire la structure et le contenu du dictionnaire ou, à travers l'analyse de ce dernier, à déceler les mécanismes du processus lexicographique. L'analyse peut également avoir une visée applicative. Par exemple, le travail de Docherty et Heid (1998), que ces linguistes computationnels qualifient de « métalexigraphie computationnelle pratique » consiste à analyser automatiquement l'ensemble d'un dictionnaire issu d'une maison d'édition pour le mettre à jour semi-automatiquement en recourant à un corpus.

3.1.2. Études diachroniques

Nous accorderons une moindre importance à l'opposition entre études synchroniques et diachroniques dans la mesure où, du point de vue de la mise en œuvre, une étude diachronique consiste souvent à comparer deux études synchroniques menées sur des données produites à des distances plus ou moins longues sur l'axe temporel. Selon la périodicité de mise à jour d'un dictionnaire, une étude diachronique peut consister en une comparaison des différents millésimes de dictionnaires tels que le *Petit Larousse illustré* (PLI) et le *Petit Robert* (PR) ou des éditions successives d'un dictionnaire comme le *Dictionnaire de l'Académie française* (DAF). Elle peut également porter sur la comparaison des différents tomes d'un dictionnaire multivolume, rédigés chronologiquement, afin de mettre en évidence une éventuelle évolution du processus lexicographique au cours de sa rédaction, comme le fait par exemple Radermacher (2004) avec le TLF. Pour les dictionnaires en ligne, la mise à jour peut être effectuée à intervalles réguliers, comme c'est le cas pour le *Oxford English Dictionary* (OED), généralement mis à jour trimestriellement, ou en continu, comme pour *Usito* ou *Wiktionary*. Lorsque l'objet de l'étude est défini avec anticipation, une étude diachronique consiste, à défaut de version

datée du dictionnaire, à répéter, selon la temporalité choisie, la même analyse d'un « instantané » du dictionnaire (dont il convient d'archiver tout ou partie du contenu, selon le phénomène à étudier, les contraintes légales et les moyens techniques disponibles). Une étude diachronique décidée *a posteriori*, en revanche, pose le problème de l'accès à des données évanescentes¹⁰. Notons à cet égard la spécificité de *Wiktionary*, qu'il n'est pas nécessaire de « photographier » régulièrement pour l'étudier en diachronie. En effet, une particularité du dictionnaire contributif est d'offrir un accès manuel (en consultation) et automatique (via la mise à disposition de ses « *dumps* historiques ») à l'historique de l'ensemble des révisions pour tous les articles¹¹. Cette fonctionnalité constitue un moyen privilégié à la fois pour l'étude du changement linguistique et pour celle du processus éditorial¹².

3.1.3. Profil des métalexigraphes

On trouve dans la littérature, de manière plus ou moins récurrente, un débat – parfois polémique – sur la légitimité des métalexigraphes à produire des analyses critiques des dictionnaires, selon que ces métalexigraphes sont ou non lexicographes. Pour Akasu (2022 : 31), un prérequis à toute évaluation réaliste et raisonnable d'un dictionnaire, au jugement sensé, sans exagération, est la compréhension de ses conditions réelles de conception. Or, d'après Akasu, cette compréhension passe nécessairement par l'expérience pratique de la lexicographie. Ainsi, on ne devrait pas produire d'analyse métalexigraphique sans avoir été au moins une fois lexicographe. Cette position nous semble quelque peu caricaturale, d'autant qu'Akasu n'envisage pas, réciproquement, une possible exagération dans l'indulgence avec laquelle un lexicographe peut mener une analyse métalexigraphique. La critique aveugle et l'indulgence excessive existent pourtant toutes les deux, comme l'écrit Béjoint : « lexicographers tend to excuse all mistakes because they know how difficult it is to write a dictionary, whereas linguists will typically find fault within everything » (2000 : 107). Corbin (1984) s'oppose à un possible parti pris négatif de dénigrement tout en récusant le plaidoyer *pro domo* qui consistait, pour les lexicographes des années 1970 exerçant une « métalexigraphie de praticiens », à brandir les contraintes commerciales leur étant imposées comme « bouclier tous azimuts » excusant par avance toute imperfection du dictionnaire (et, partant, de la pratique lexicographique). Il nous semble qu'une posture rigoureuse consistant à analyser les faits dictionnaires, à les interpréter à travers le prisme d'un champ disciplinaire et éventuellement à les confronter aux discours lexicographiques, sans dénigrement ni indulgence (bref, une posture scientifique) est possible. Étant donné les contingences financières et temporelles du monde de l'édition et la vision parcellaire qu'ont les

10. Voir à ce sujet la conclusion et le post-scriptum de Corbin et Gasiglia (2023), qui parlent de « déperdition d'archive à répétition induite par le flux numérique ».

11. Cette fonctionnalité a notamment été exploitée par Wolfer et Müller-Spitzer (2016) pour décrire la dynamique des contributions, par Sajous (2022) pour étudier les entrées néologiques les plus révisées et par Sajous (2023) pour analyser le taux de révision des pages et des définitions dans le dictionnaire contributif.

12. Le comparatif de versions successives d'articles ne permet pas toujours de déterminer avec certitude si un changement dans le dictionnaire reflète un changement linguistique ou relève du processus éditorial (c'est également le cas avec la comparaison d'éditions successives de dictionnaires imprimés).

lexicographes des produits qu'ils développent (cf. sections 2.1 et 2.4), les métalexicographes universitaires sont certainement les plus à même de mener ces analyses.

Un autre trait distinctif des métalexicographes qui, à notre connaissance, n'est jamais mentionné dans la littérature, est la capacité à analyser automatiquement (informatiquement) les dictionnaires. Or, sauf à disposer d'une équipe vraiment nombreuse sur un temps long (cela arrive-t-il ?), cette aptitude est un prérequis à la mise en œuvre de certaines analyses, particulièrement celles qui portent sur l'ensemble du dictionnaire.

3.1.4. Nature des dictionnaires

Parmi les traits distinctifs concernant les dictionnaires, nous n'avons conservé que ceux qui ont une incidence sur la mise en œuvre des analyses *i.e.* le support (papier ou électronique), la fréquence des mises à jour et les contraintes légales. Nous ne cherchons pas à inclure, dans cette ébauche de typologie, les oppositions traditionnelles sur la nature des dictionnaires telles qu'on les trouve encore mentionnées dans certains manuels. Les frontières entre catégories étant poreuses, certaines oppositions binaires sont désormais caduques (Rundell, 2012 : 78–79), à supposer, pour certaines, qu'elles aient déjà été totalement valides : il y a un demi-siècle, Rey-Debove critiquait déjà l'opposition entre dictionnaire « de langue » et dictionnaire « encyclopédique » (1971 : 29-34). Par ailleurs, de nouveaux traits distinctifs apparaissent, qui sont peu ou pas encore décrits. Par exemple, évoquer l'opposition entre dictionnaires conçus par des humains (éventuellement à l'aide d'outils permettant l'automatisation totale ou partielle de certaines tâches) et ceux générés entièrement automatiquement aurait pu prêter à sourire quelques mois avant l'écriture de ces lignes, ou, du moins, sembler prématuré. À lire l'état de l'art de Schryver (2023b) – publié à peine un an après le lancement de ChatGPT – sur l'analyse des entrées de dictionnaires générées entièrement automatiquement, il est probable que la pertinence de cette opposition s'impose à la date de publication du présent chapitre.

3.2. Principales oppositions retenues

Parmi les différents traits distinguant les types d'analyse décrits plus haut, nous retenons comme axes principaux de classification les oppositions entre études qualitatives et études quantitatives d'une part et entre études manuelles et études automatiques d'autre part (cf. figure 1). Notons d'emblée que ces deux axes, orientés respectivement verticalement et horizontalement par convention, ne correspondent pas à des caractéristiques réellement orthogonales (les études qualitatives sont le plus souvent manuelles et les études quantitatives sont le plus souvent automatiques), mais ils ne se confondent pas non plus : on peut mener des études quantitatives manuellement et, dans une certaine mesure, automatiser des études qualitatives.

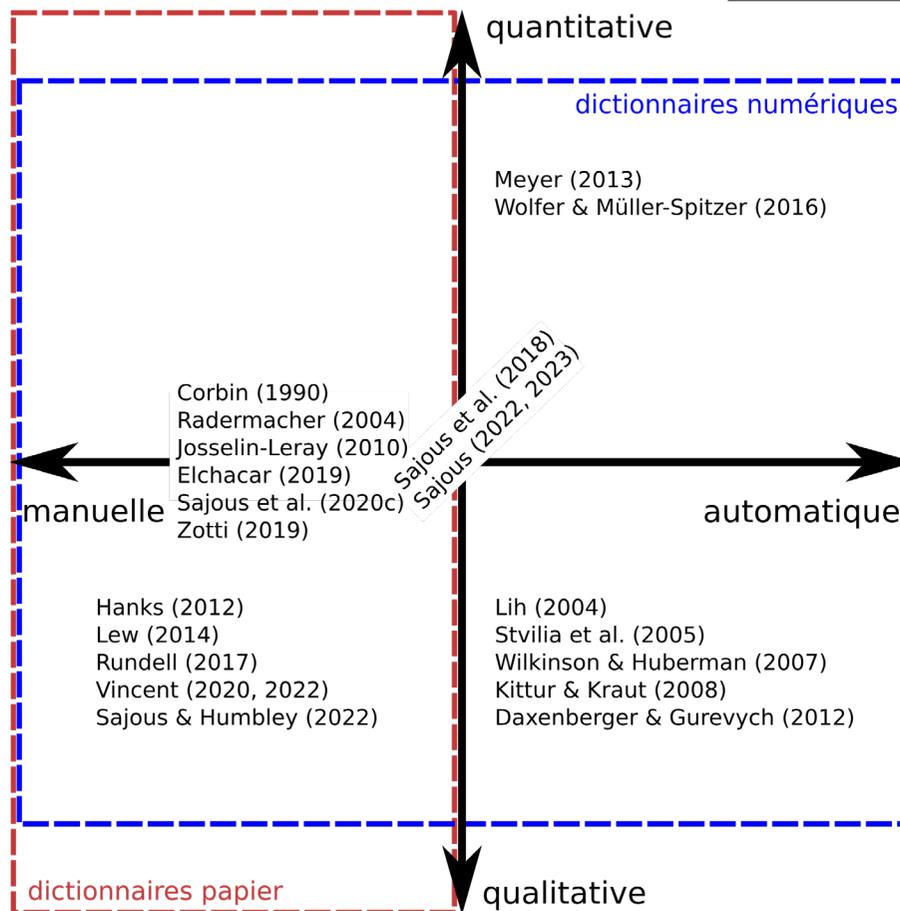


Figure 1 : Axes principaux de classification des analyses de dictionnaires

Dans le repère de la figure 1, certains travaux se situent clairement dans une des quatre zones délimitées par les deux axes classificateurs. D'autres, qu'ils correspondent à des études mixtes ou à une série d'études de natures différentes, chevauchent plusieurs catégories. Les études de Corbin (1990), Josselin-Leray (2010) et Sajous *et al.* (2020c), présentées en section 5.1, sont menées manuellement et associent analyses quantitatives et analyses qualitatives, complémentaires l'une de l'autre. Elles consistent, après constitution d'un corpus d'étude, à analyser qualitativement un phénomène observé, puis à quantifier les observations produites dans l'étude qualitative. Elles sont donc situées à cheval sur la ligne de partage *qualitative/quantitative* dans la partie gauche correspondant aux analyses manuelles. Celles de Radermacher (2004), Sajous *et al.* (2018), Elchacar (2019), Zotti (2019) et Sajous (2022; 2023) comprennent également, chacune, une analyse qualitative manuelle et une analyse quantitative, manuelle pour Radermacher (2004), Elchacar (2019) et Zotti (2019), computationnelle pour Sajous *et al.* (2018) et Sajous (2022; 2023). Si, comme dans le cas précédent, les analyses quantitatives et qualitatives de ces cinq études sont complémentaires, elles sont indépendantes du point de vue de leur mise en œuvre : l'une des analyses n'exploite pas les résultats de l'autre.

Que les analyses soient qualitatives ou quantitatives, l'étude des dictionnaires n'existant qu'au format papier s'effectue nécessairement manuellement. Les dictionnaires numériques, eux, peuvent être analysés manuellement ou automatiquement. Autrement dit, toute analyse menée sur les dictionnaires papier peut également l'être sur un dictionnaire numérique, aux détails près suivants :

- étudier les caractéristiques typo-dispositionnelles d'un article a autant de sens dans un dictionnaire numérique que dans un dictionnaire papier, mais les études consacrées aux caractéristiques des pages (*e.g.* nombre moyen d'articles par page...) n'a pas d'équivalent dans un dictionnaire numérique ;
- les techniques d'échantillonnage qui reposent sur les numéros de page sont sans objet ;
- le parcours de l'intégralité de la nomenclature n'est possible que si celle-ci est mise à disposition (ou reconstituable par un moyen quelconque) ;
- les études diachroniques sont possibles, leur mise en œuvre dépendant de la fréquence des révisions et de la communication qui les accompagne ou non.

Nous examinons dans la suite de l'article les difficultés afférentes aux principales catégories d'analyse représentées dans la figure 1, tout en étudiant l'incidence des autres traits distinctifs identifiés lorsque qu'ils se révèlent pertinents.

4. Études qualitatives manuelles

Les études qualitatives manuelles portent généralement sur un nombre restreint, voire très restreint d'articles. C'est le cas des analyses menées par Hanks (2012), Lew (2014) et Rundell (2017), qui étudient les définitions de *Wiktionary*. Lew, par exemple, fonde sa comparaison entre les définitions de *Wiktionary* et celles du *Longman Dictionary of Contemporary English* sur l'étude d'une seule entrée, le verbe *handle*. Sajous et Humbley (2022) étudient quant à eux le traitement des entrées relatives aux mesures d'isolement sanitaire dans le *Wiktionnaire* et la *Wikipédia* en analysant 8 entrées du dictionnaire et 16 entrées de l'encyclopédie.

La taille modeste des observables examinés n'empêche pas les études mentionnées ci-dessus d'être pertinentes. Les analyses qualitatives portant sur un nombre limité de cas peuvent en effet servir à illustrer des observations récurrentes ou, au contraire, servir de contre-exemples pour infirmer des hypothèses. Notons par ailleurs que les études qualitatives n'ont pas nécessairement vocation à être généralisées et que, par ailleurs, une analyse fine n'est possible que sur un nombre restreint de cas. Par exemple, afin d'illustrer la nécessité d'un traitement polyphonique d'emplois polémiques, Vincent (2022) analyse, dans plusieurs dictionnaires, le traitement d'une même unité lexicale – le mot *woke* – avant de proposer son propre traitement (polyphonique). Il n'est pas question ici de donner une description d'ensemble du traitement des emplois polémiques dans chaque dictionnaire, et encore moins de proposer pour chacun un traitement plus adapté, mais bien de montrer, à travers l'examen d'un exemple particulier, qu'un traitement polyphonique de ce type d'entrée est plus approprié que les traitements qui ignorent la polyphonie.

Ces remarques étant faites, il convient de garder à l'esprit qu'un nombre limité de cas examinés proscrit toute possibilité de généralisation. En outre, l'interprétation des phénomènes observés dans les dictionnaires étudiés peut nécessiter une connaissance approfondie des mécanismes qui sous-tendent le processus éditorial. Nous illustrons les deux assertions précédentes en commentant ci-dessous deux études menées dans les versions anglaise et française de *Wiktionary*.

Hanks, ayant remarqué l'absence de *rogue elephant* dans le dictionnaire contributif, crée l'entrée à titre d'expérience et rédige la définition de son sens figuré « someone or something that is large, dangerous, and unpredictable » (2012 : 81). Il écrit être impressionné par la rapidité (quelques minutes) avec laquelle un contributeur intervient pour mettre en forme l'article et ajouter le sens littéral de l'entrée. À propos de cette expérience, Rundell (2017) écrit que le niveau de contrôle éditorial dans *Wiktionary* est impressionnant et que les contributions inappropriées y sont rapidement éliminées. Sajous (2023) montre pourtant que les révisions ne sont ni systématiques (les articles ne sont que très rarement révisés, même en cas d'erreur manifeste) ni rapides (elles interviennent, le cas échéant, avec un délai médian de 3 à 4 ans). L'erreur d'interprétation de Rundell est imputable à deux facteurs : la généralisation hâtive à partir d'un exemple unique et une connaissance imparfaite des spécificités du processus éditorial du dictionnaire contributif. Une certaine habitude de l'« écosystème *Wiktionary* », de ses articles et des agissements de ses contributeurs, mais également la consultation de son historique, nous permettent d'avancer une explication à l'observation atypique de Hanks. Différents types d'administrateurs de *Wiktionary* sont chargés de veiller au grain et, notamment, de lutter contre les actes de vandalisme. Dans une moindre mesure, ils peuvent vérifier que des modifications récentes ne contreviennent pas à la ligne éditoriale du dictionnaire, fût-elle floue. Des alertes automatiques peuvent être envoyées à ces administrateurs lorsque certains événements se produisent, comme la création d'une page, ou une suppression/modification importante de contenu. Pour alléger leur charge de travail, les administrateurs peuvent décider de faire confiance à certains wiktionnaristes identifiés et jugés fiables, qui pourront ainsi effectuer certaines actions sensibles en dehors de toute surveillance, alors que ces mêmes actions auront toutes les chances de déclencher des vérifications si elles sont effectuées par des contributeurs inconnus. Dans le cas de Hanks, il y a fort à parier que la création d'une page par un contributeur inconnu (des wiktionnaristes) n'est pas passée inaperçue, ce qui expliquerait la réactivité du contributeur qui a pris la relève. On peut lire sur la « page utilisateur » du contributeur qui a révisé la page créée par Hanks : « As an administrator I patrol "Recent changes", look out for vandalism, and spruce up any new articles that need it ».

Vincent (2020) constate une inclusion précoce des néologismes dans le *Wiktionnaire*, au regard des critères habituels de fréquence et de distribution. Elle explique que les articles peuvent être révisés après création et illustre cette caractéristique du dictionnaire contributif en signalant une modification récente (13/11/2018) de l'article *vapoter*, qu'elle interprète comme une « bonification ». Or, dans le *Wiktionnaire*, révision n'implique pas toujours bonification et, dans ce cas précis, l'historique montre que la révision en question ne vaut ni bonification ni dégradation mais *statu quo* : le 13/11/2018, l'article a subi deux modifications, chacune ayant été révoquée. Au total, l'article aura donc été révisé

quatre fois, pour être finalement inchangé. Cela ne contredit pas le propos de Vincent (on peut effectivement réviser des articles dans le *Wiktionnaire* et ces modifications peuvent – éventuellement – bonifier les articles), mais l'exemple choisi n'en est pas une illustration. Parmi les entrées étudiées, les révisions qui ont conduit au remplacement du sens initial de *vapoteuse* « Féminin singulier de vapoteur » par « Utilisatrice d'une cigarette électronique » (20/06/2019) et à l'ajout du sens « Cigarette électronique » (26/02/2014) sont en revanche des exemples d'authentiques bonifications.

Les études de Vincent (2020) et Hanks (2012) montrent que l'interprétation des observations impose parfois la connaissance préalable de certains mécanismes du processus éditorial. Dans le cas de *Wiktionary*, la possibilité d'accéder à l'historique de l'ensemble des révisions d'un article permet d'étudier ces mécanismes *a posteriori*. Cet outil, qui permet d'étudier en diachronie la manière dont se construisent les articles est, pour l'étude du dictionnaire contributif, un moyen privilégié d'investigation et d'interprétation dont il serait dommage de se priver. L'étude de Hanks montre également qu'à défaut d'étude quantitative, toute généralisation d'une étude qualitative est hasardeuse.

5. Études quantitatives manuelles

Seules les études quantitatives permettent d'apprécier l'importance d'un phénomène et d'en évaluer la systématisme. Les études quantitatives d'un dictionnaire général, menées manuellement, portent presque toujours sur un sous-corpus d'observables¹³, l'ensemble du dictionnaire étant trop volumineux pour être étudié dans son intégralité. On peut subdiviser, parmi ces études, celles qui portent sur un sous-corpus d'observables répondant à des critères spécifiques, que nous étudions en section 5.1, et celles qui portent sur un phénomène susceptible d'apparaître n'importe où dans le dictionnaire, imposant le recours à des techniques d'échantillonnage, que nous abordons en section 5.2.

5.1. Sélection de sous-corpus d'observables répondant à des critères déterminés

La sélection d'observables peut se fonder sur des critères internes à la microstructure (e.g. marques du diasystème) et à la macrostructure (e.g. index thématiques) ou sur des critères externes (e.g. corpus, ressources lexicales). Ces différentes options sont étudiées en sections 5.1.1 à 5.1.4.

5.1.1. Requête via interface de recherche

Les dictionnaires numériques disposent d'interfaces de recherche plus ou moins sophistiquées qui peuvent, selon les cas, être mises à profit pour la constitution d'un corpus d'étude. La barre de recherche d'*Usito* ne permet d'effectuer des requêtes que sur les vedettes, avec recherche d'appariement exact et suggestions par complétion automatique. À l'opposé, le TLFi permet d'effectuer des

13. On parlera ici d'*observable* pour désigner toute délimitation cohérente d'un segment textuel relativement au cadre d'une étude donnée, quelle que soit la granularité : page, article, rubrique, division de sens, définition, etc.

requêtes multicritères complexes. À mi-chemin, les versions en ligne du PR et des différentes éditions du DAF offrent chacune un formulaire intitulé « Recherche avancée » qui offre plus de fonctionnalités que la barre de recherche d'*Usito*, tout en étant plus intuitif que le système de requêtage (plus puissant) du TLFi. Ces possibilités d'interrogation peuvent s'avérer utiles, à condition d'y recourir avec circonspection. Coleman et Ogilvie (2009) mettent en garde contre la distorsion des résultats qui peut découler de la combinaison d'une pratique lexicographique incohérente (*i.e.* d'un manque de systématisme) et de l'utilisation d'outils de recherche trop frustes, surtout lorsque les dictionnaires n'ont pas été initialement conçus pour une exploration informatique. Les autrices donnent l'exemple du comptage du nombre total d'emprunts au français dans la version en ligne de l'OED. Une telle recherche devrait commencer par la vérification d'une utilisation cohérente des noms de langue (orthographe et abréviations) dans la rubrique étymologique du dictionnaire. Au moment de leur analyse, la version en ligne de l'OED combinait les deuxième et troisième éditions du dictionnaire. Si le codage de la langue d'origine des étymons est systématique dans la troisième, ce n'est pas le cas dans les entrées non révisées de la deuxième, où alternent *French*, *Fr.* et *F.* Or, chercher *F.* dans le moteur de recherche sélectionne également *f.*, abréviation de *from*. Chercher *French* dans la rubrique étymologique génère du bruit, *e.g.* « Of uncertain attribution; perh. the name of Maximilien Ringelmann (1861-1931), French scientist ». Si l'alternance de trois formes codant une même langue relève d'une asystématisme lexicographique, l'impossibilité de formuler une requête sur un élément particulier de la microstructure (ici, la langue d'un étymon et pas seulement la rubrique étymologique) vient de la rétroconversion des dictionnaires imprimés au format numérique. Le plus souvent, seul un balisage à gros grain des rubriques a été opéré en sus du balisage typographique (ce qui constitue déjà un travail considérable).

Donnons également l'exemple de la marque LITTÉR. dans le PR, problématique en soi puisqu'utilisée à la fois comme marque de domaine identifiant les « termes des études littéraires » et comme marque d'usage identifiant les mots de « la langue écrite élégante ». Du point de vue du système de requêtage, un autre problème se pose : la cohabitation de cette marque avec LITTÉRATURE et LITT., parfois dans un même article. Par exemple, le sens 1, 5 de *harmonie* est marqué GRAMM., LITTÉRATURE et la seconde nuance de sens « La poésie » (sous *harmonie*, 1, 5) est marquée LITTÉR. Les marques alternatives à LITTÉR. – non documentées – sont très rares : LITTÉRATURE n'apparaît que dans les entrées *aléatoire*, *grotesque*, *harmonie* et *hors-d'œuvre* et LITT. uniquement dans *tragédie*. Mais rien ne dit que, pour d'autres marques (ou d'autres éléments de la microstructure), les incohérences ne soient pas plus massives et que les résultats des requêtes portant sur ces éléments soient totalement fiables. Les deux exemples tirés de l'OED et du PR montrent en effet qu'il faut s'attendre aussi bien à du bruit qu'à du silence lors de la constitution d'un corpus par requêtage, sauf dans le cas où le moteur de recherche s'appuie sur un encodage systématique et cohérent de la macro- et de la microstructure du dictionnaire. Il appartient donc aux métalexigraphes de s'en assurer ou, à défaut, de tenir compte dans leurs analyses de la fiabilité relative du corpus constitué.

5.1.2. Recensions internes

On trouve dans certains dictionnaires des pages indexant les vedettes correspondant à une catégorie donnée, chaque entrée mentionnée renvoyant à l'article qui lui est consacré. C'est le cas d'*Usito*, dont la page d'accueil est composée, sous la barre de recherche, de liens pointant vers la liste des entrées correspondant à une partie du discours donnée, des particularismes, des emplois critiqués ou officialisés, etc. C'est aussi le cas de *Wiktionary* qui contient de nombreuses pages « lexique », « catégorie » et autres. Ces recensions peuvent constituer un point de départ à partir duquel sélectionner des articles à étudier qualitativement. Lorsqu'on envisage une étude quantitative, une attention particulière devrait être portée aux mécanismes qui sous-tendent ces éléments macrostructurels, en l'occurrence à la manière – systématique ou non – dont ces listes sont constituées.

Dans un travail portant sur les appellations des identités de genre non traditionnelles dans les dictionnaires professionnels et profanes, Elchacar (2019) compare, à travers une étude qualitative, les traitements lexicographiques opérés par plusieurs dictionnaires. Elle compare également, à travers une étude chiffrée, les nomenclatures des différents répertoires. Si, avec ses 133 entrées, le *Wiktionnaire* est présenté initialement comme « le plus complet » du point de vue de la nomenclature, l'étude restreint ensuite la comparaison à 18 unités lexicales, sur la base des fréquences (en nombre de documents) calculées dans un corpus de presse. Pour le champ lexical étudié, et pour cette sélection d'unités lexicales, Elchacar conclut que « [a]dmittant [...] que c'est un dictionnaire professionnel, le GDT, qui présente la nomenclature et le traitement lexicographique les plus complets, on peut se demander quelle est la plus-value des dictionnaires profanes » (2019 : 189-190).

Nous nous intéressons ici à l'étude quantitative et laisserons le traitement lexicographique de côté pour nous concentrer sur la comparaison des nomenclatures. Le *Wiktionnaire* entre dans la catégorie des « dictionnaires profanes » étudiés par l'autrice qui, pour le comparer aux autres ressources, se fonde sur son « vocabulaire LGBTIQ¹⁴ en français¹⁵ ». Elchacar écrit qu'« [i]l s'agit donc d'entrées indépendantes de celles auxquelles on peut avoir accès directement à partir du *Wiktionnaire* » (2019 : 179). Mais la page intitulée « vocabulaire LGBTIQ en français » n'est pas indépendante des entrées du *Wiktionnaire*. Il s'agit simplement d'une page d'index qui liste les entrées dont l'article contient, dans son *wikicode*, une balise catégorisante¹⁶. Cette balise, ajoutée manuellement par les contributeurs, produit à la fois, dans l'article concerné, l'apparition d'un lien qui pointe vers la page *Vocabulaire LGBTIQ en français* et le référencement (*i.e.* l'ajout d'un lien vers l'article) légèrement différé de l'entrée dans la page du vocabulaire LGBTIQ. Ainsi, hors délai de synchronisation, la relation entre les ve-

14. Nous reproduisons le sigle en maintenant l'ordre des initiales tel qu'elles apparaissent dans le *Wiktionnaire*.

15. https://fr.wiktionary.org/wiki/Catégorie:Vocabulaire_LGBTIQ_en_français. L'étude de Elchacar, publiée en 2019, a fait l'objet d'une communication le 19/10/2018. La page du vocabulaire LGBTIQ n'ayant pas été modifiée entre le 3/08/2017 et le 6/02/2022, la version utilisée dans l'étude est très probablement celle du 3/08/2017. Pour la consulter, on peut accéder à l'historique (lien *Voir l'historique* en haut à droite de la page) et cliquer sur cette date.

16. En l'occurrence : [[Catégorie:Vocabulaire LGBTIQ en français]]

dettes des articles balisés et les unités lexicales recensées dans les pages « vocabulaire » est bijective : toute vedette d'un article contenant la balise figure dans la page « vocabulaire » et toute unité lexicale recensée dans la page « vocabulaire » a son article dans le *Wiktionnaire*. La méconnaissance de ce fonctionnement de la macrostructure ne porte cependant pas préjudice à l'étude. En revanche, le fait qu'une unité lexicale ne figure pas dans la page du « vocabulaire » ne signifie pas qu'elle est absente du dictionnaire. Des 18 unités lexicales sélectionnées par Elchacar, 6 sont indiquées comme absentes de la nomenclature du *Wiktionnaire*. Or, au moment de l'étude, si les entrées ne figuraient pas dans la page d'index consultée, 5 étaient bien présentes dans le dictionnaire contributif¹⁷ : le tableau 1 donne, pour chaque vedette, sa date d'ajout à la nomenclature ainsi que les éventuels « lexiques » ou « vocabulaires¹⁸ » dans lesquels l'entrée est référencée à la période de l'étude de Elchacar (≤ 2018) et lors de l'écriture du présent article (février 2024).

Tableau 1 : Dates d'inclusion d'entrées liées aux appellations des identités de genre non traditionnelles dans le *Wiktionnaire*

Vedette	Date d'inclusion	Lexique (l.) / Vocabulaire (v.)	
		≤ 2018	février 2024
<i>asexuel</i>	29/08/2006	l. sexualité	l. sexualité, v. LGBTIQ
<i>bispirituel</i>	19/10/2018	-	l. genres humains et identités de genre
<i>cisgenre</i>	15/11/2007	l. genres humains et identités de genre	l. transitude, l. genres humains et identités de genre
<i>intersexué</i>	30/02/2011	l. genres humains et identités de genre	v. LGBTIQ
<i>pansexuel</i>	16/04/2014	-	l. sexualité
<i>Trans</i>	22/02/2006 sens <i>transsexuel</i> : 25/09/2008	-	l. transitude

17. *Bispirituel*, seule entrée absente au moment de l'analyse, a été ajoutée « en direct » par un wiktionnariste présent dans l'auditoire du colloque où l'étude a été présentée. Voir à ce sujet l'épilogue de Sajous *et al.* (2019) ou l'introduction de Sajous (2023).

18. Contrairement aux « lexiques », l'indexation du « vocabulaire » LGBTIQ est construit non par l'apposition d'une marque de domaine dans la définition (e.g. `{{lexique|genres humains et identités de genre|fr}}`) dans la définition de *cisgenre*, mais par celle d'une balise catégorisante en fin d'article (e.g. `[[Catégorie:Vocabulaire LGBTIQ en français]]`) dans l'article *intersexué*).

On voit que lors de l'étude de Elchacar, les entrées sélectionnées existaient sans pour autant être référencées dans une catégorie particulière ou en étant indexées dans d'autres pages *lexique* ou *vocabulaire* que celle examinée (on dénombre dans les deux colonnes de droite du tableau 1 trois « lexiques » et un « vocabulaire » qui auraient pu rentrer dans le cadre de l'étude. Il en existe d'autres encore, comme le *Lexique du transféminisme*).

Pour cette partie de l'analyse, le recours à une recension interne aboutit donc à une conclusion erronée : ce n'est pas le GDT, avec ses 15 unités lexicales présentes dans sa nomenclature sur les 18 sélectionnées pour examen, qui présente la nomenclature la plus complète, mais le *Wiktionnaire*, avec 17 unités sur 18.¹⁹

Au-delà des considérations sur l'infrastructure particulière du *Wiktionnaire*, nous conseillons, si ce n'est une certaine méfiance, du moins une vigilance particulière quant aux listes d'entrées issues d'une catégorisation donnée, que ces listes soient constituées manuellement ou reposent sur l'exploitation automatique d'un quelconque marquage dans la microstructure. Plusieurs problèmes se posent en effet :

- d'abord, un problème de cohérence dans le traitement lexicographique, notamment l'attribution des marques. Étudier, par exemple, le domaine de l'informatique dans le PR n'est pas seulement étudier les articles portant la marque INFORM. (Sajous *et al.*, 2020c). De la même manière, Sajous (2022) montre qu'on ne peut se contenter d'utiliser la catégorie *Coronavirus* de *Wiktionary* pour étudier les néologismes liés au Covid-19.
- ensuite, pour les dictionnaires rétroconvertis, gardons en tête que, souvent, seule la mise en forme a été encodée et qu'aucun balisage logique (structurel) ni aucune normalisation n'ont été opérés. Une même information peut être codée de différentes manières (et un même code être porteur de sens différents, cf. section 5.1.1).
- enfin, pour les dictionnaires nativement numériques, il ne faut pas présupposer qu'ils sont toujours structurés et soumis à un traitement informatique duquel résulterait quelque logique systématique. Le contenu du *Wiktionnaire*, par exemple, n'est pas structuré dans une base de données qui serait exploitée pour générer automatiquement articles, pages d'index et autres. Construit en partie par import d'articles de dictionnaires anciens, le *Wiktionnaire* en reproduit les incohérences. Mais même le contenu créé en propre dans le dictionnaire contributif n'est pas exempt de redondances, lacunes et incohérences, loin s'en faut. Sajous et Hathout (2015) dénombrent par exemple 5 manières de coder la marque *œnologie* dans le *Wiktionnaire*. Rien n'indique *a priori* lesquelles seraient prises en compte pour générer un « lexique de l'œnologie ». Outre la pluralité des formes existantes pour encoder une même information, il existe dans le *Wiktionnaire* une arborescence pléthorique de marques : 20 marques sous-catégorisent, par exemple, le domaine de l'informatique (Sajous *et al.*, 2020c).

19. Nous n'avons vérifié les données que pour ce qui concerne le *Wiktionnaire* car il offre la possibilité d'observer l'état antérieur de chaque article, quelle que soit la date choisie. Nous nous prononçons donc uniquement sur la comparaison GDT/*Wiktionnaire* et non sur les autres répertoires. À toutes fins utiles, signalons pour les études de la nomenclature de *Wiktionary*, la ressource WIND (Sajous *et al.*, 2020a), qui contient les dates d'inclusion des entrées dans les éditions française et anglaise du dictionnaire contributif, librement disponible à l'adresse <http://redac.univ-tlse2.fr/lexiques/>.

5.1.3. Parcours exhaustif

Il est possible de sélectionner un ensemble d'articles à étudier en parcourant une nomenclature de manière exhaustive. C'est ce qu'ont fait Corbin (1990) et 45 étudiants, qui ont parcouru la nomenclature de cinq grands dictionnaires généraux monolingues « de langue » et encyclopédiques à la recherche de noms de végétaux dérivés en *-ier* et de leur base apparente. Le matériau collecté sur l'ensemble des dictionnaires totalise 249 noms de végétaux, chaque dictionnaire en comportant entre 171 et 201. Le traitement opéré par chacun des cinq dictionnaires est ensuite étudié pour l'ensemble des noms. Indépendamment de l'analyse des descriptions lexicographiques, le nombre de personnes à l'œuvre pour le parcours des nomenclatures donne une idée de l'ampleur de la tâche.

Si l'on s'intéresse aux entrées ajoutées plus ou moins récemment aux dictionnaires généraux français, la ressource *DiCo* (Sajous et Martinez, 2022) peut s'avérer utile : y figurent notamment les mots ajoutés aux PLI et PR depuis 1998, ainsi qu'aux 8^e et 9^e éditions du DAF. Sajous *et al.* (2020c) s'appuient sur cet inventaire pour parcourir les 3334 ajouts au PR sur la décennie 2008-2017 afin de sélectionner les entrées susceptibles de relever du domaine de l'informatique, indépendamment du marquage. Après vérification manuelle de l'appartenance effective au domaine dans les cas litigieux, le corpus à étudier compte 120 articles, dont un peu moins de la moitié contiennent la marque INFORM.

5.1.4. Critères externes

On peut également fonder la sélection d'observables sur des critères externes, en utilisant par exemple une liste d'unités lexicales existante (nomenclature de terminologie, recueil de variantes diatopiques, de faux-amis, glossaire d'argot, etc.). À défaut d'inventaire prêt à l'emploi, une solution envisageable est de recourir à un corpus. Par exemple, Josselin-Leray (2010) extrait 110 termes français et 110 termes anglais d'un corpus bilingue de vulgarisation en volcanologie avant d'en étudier l'inclusion et le traitement dans deux dictionnaires monolingues français, deux dictionnaires monolingues anglais et deux dictionnaires bilingues. Zotti (2019), quant à elle, extrait d'un corpus littéraire 10 unités lexicales désignant des *realia* gastronomiques québécoises avant de quantifier leur présence dans la nomenclature d'une variété de dictionnaires. Cette quantification est suivie d'une étude qualitative des descriptions lexicographiques, notamment les définitions et le marquage diatopique.

5.1.5. Sélection de sous-corpus d'observables : bilan

Les incitations à la prudence et les réserves émises en sections 5.1.1 et 5.1.2 ne nous conduisent pas à proscrire l'utilisation des outils de recherche et des listes d'unités lexicales toutes faites fournies par les dictionnaires (imprimés ou numériques) : ces deux points d'entrée constituent une manière rapide et pratique d'initier une étude qualitative, dont il serait dommage de se priver. En revanche, mener une étude quantitative à partir des outils de recherche impose de connaître leur fonctionnement précis, de la même manière que se fier aux recensions internes impose de décrypter les arcanes du processus (automatique ou manuel) qui les a produites. Dans les deux cas, il faut avoir conscience que l'encodage de la macro- et de la microstructure (diasystème compris) ne présente pas toujours la systématisme que l'on peut légitimement attendre d'un dictionnaire numérique.

Le seul problème que semblent poser les études citées en sections 5.1.3 et 5.1.4, qui combinent analyses manuelles quantitatives et qualitatives, est la somme de travail qu'elles nécessitent²⁰. Ce travail est néanmoins nécessaire dès lors que l'on veut constituer un corpus d'entrées présentant des caractéristiques spécifiques, sur la base de critères externes objectifs, et non sur un marquage largement incohérent.

5.2. Sélection d'observables par échantillonnage

Les statistiques inférentielles consistent à décrire – quantitativement – les caractéristiques d'une population restreinte (un échantillon) dans le but d'estimer celles de la population générale. Concernant l'analyse des dictionnaires, Bukowska (2010) observe que, si beaucoup d'énergie est consacrée à l'analyse des échantillons, très peu de réflexion semble être accordée par les métalexigraphes aux mécanismes de sélection des échantillons eux-mêmes. Pourtant, la méthodologie d'échantillonnage mérite réflexion, sous peine d'introduire des biais importants. C'est l'objet du chapitre intitulé « Études métalexigraphiques par analyse d'échantillons : toute représentativité est-elle illusoire ? », qui montre que :

- l'échantillonnage par zone contiguë est la méthode la plus utilisée par les métalexigraphes bien que ce soit celle qui présente le plus de biais ;
- la taille des échantillons examinés en métalexigraphie ne garantit aucunement leur représentativité.

Une autre étude (en cours) montre qu'une même analyse menée sur deux échantillons distincts peut conduire à des conclusions opposées. Cette étude, menée sur l'intégralité des tomes I et XVI

20. Sans même parler de l'analyse du traitement lexicographique des articles sélectionnés, Corbin (1990) note que la sélection elle-même des noms de végétaux dans un dictionnaire est plus complexe que le « simple » fait de parcourir la nomenclature. Les deux listes de 110 termes retenus par Josselin-Leray (2010) ont été élaborées à la main, selon des critères de rétention et de rejet prédéfinis, parmi 7095 termes candidats français et 7836 termes candidats anglais, préalablement extraits automatiquement d'un corpus.

du TLFi montre que la conclusion de Radermacher (2004), mentionnée en section 2.4, est juste (le nombre moyen d'exemples par page n'est pas inférieur dans le dernier tome du TLF que dans le premier), mais qu'un nombre non négligeable d'échantillons auraient pu, à l'inverse, amener l'autrice à confirmer ce qu'elle qualifie de mythe. Par ailleurs, cette étude montre qu'on peut aller plus loin encore que la conclusion de Radermacher : non seulement le nombre moyen d'exemples par page ne diminue pas, mais il augmente. Et cette augmentation est *statistiquement* significative, ce que ne permettent pas nécessairement de montrer les échantillons de Radermacher.

Devant l'ensemble des problèmes méthodologiques que posent les études par échantillonnage, nous concluons que les analyses quantitatives menées sur l'intégralité des observables (pages, articles, etc.) sont à privilégier. Ces études et leurs conditions de mise en œuvre font l'objet de la section 6.

6. Études quantitatives computationnelles

Les analyses automatiques permettent d'étudier un phénomène sur la totalité d'un dictionnaire en s'affranchissant des problèmes de représentativité évoqués en section 5.2. Leur mise en œuvre est conditionnée par :

- la nature du dictionnaire : on ne peut traiter informatiquement un dictionnaire que s'il existe au format numérique, et à condition qu'il soit libre de droits, ou, du moins, diffusé sous des conditions légales qui permettent l'analyse envisagée ;
- le profil (connaissances théoriques et compétences techniques) des métalexigraphes.

Ces deux aspects sont développés dans les sections 6.1 et 6.2.

6.1. Limites légales

Certains dictionnaires tels que *Wiktionary*, téléchargeables en intégralité et diffusés sous licence libre, constituent un objet d'étude de choix pour les analyses automatiques. Concernant les dictionnaires commerciaux ou institutionnels, la situation est souvent moins idéale. Les dictionnaires tombés dans le domaine public, qui ne posent donc aucun problème légal, ne sont analysables automatiquement que s'ils ont été numérisés et convertis dans un format exploitable, comme le *XMLLittré*, une version du *Littré* au format XML²¹. Pour les dictionnaires institutionnels sous droits, il arrive que des accords de recherche puissent être conclus. Après avoir longuement gardé le TLFi dans son giron, l'ATILF semble disposé à signer des conventions de recherche avec d'autres laboratoires²². Si l'on s'inscrit dans une démarche *transformiste* (cf. section 2), on peut s'interroger sur la pertinence d'analyser des dictionnaires d'un autre âge (un siècle et demi pour le *Littré*, un demi-siècle à trois décennies pour

21. <https://www.littre.org/>. Voir la FAQ pour obtenir les liens de téléchargement.

22. Pour CLLE (ERSS à l'époque), c'est le cas depuis 2005.

le TLF, selon les volumes), qui n'ont pas ou plus vocation à être actualisés. Quant aux dictionnaires institutionnels du français, généraux et contemporains, le pluriel ne s'impose pas : on ne voit guère qu'*Usito*²³. Une question préalable à l'analyse automatique des dictionnaires en ligne se pose : vaut-il mieux passer par la grande ou la petite porte ? En dehors de toute considération légale, pratique ou scientifique, la simple correction indique la première option. Mais, faute de réponse institutionnelle en temps raisonnable²⁴, nous suggérons de lire les mentions légales (licence, conditions d'utilisation, etc.) : on a parfois la bonne surprise de découvrir qu'elles sont moins restrictives que ce à quoi on aurait pu s'attendre (cette dernière remarque vaut également pour les dictionnaires commerciaux). Sur l'importance de consulter les mentions légales dans le détail, terminons cette section par une mise au point :

- le fait qu'un dictionnaire (comme toute autre ressource) soit disponible en ligne ne signifie pas qu'il est libre de droits ;
- un dictionnaire sous droit n'est pas nécessairement inexploitable. Il arrive qu'il soit interdit de le redistribuer, mais pas de le télécharger, de le stocker, de l'interroger automatiquement, etc. Il convient donc de ne pas prendre la dimension légale à la légère, mais il ne s'agit pas non plus de se censurer *a priori* plus que nécessaire.

6.2. Profil des analystes, choix des observables et interprétation

Nous avons vu que l'accès aux données peut être conditionné par le support du dictionnaire (papier ou numérique) et par des contraintes légales. Il l'est également par la capacité des analystes souhaitant pratiquer une métalexigraphie outillée à traiter et analyser automatiquement la ressource étudiée. En effet, le traitement automatique d'un dictionnaire numérique dans son intégralité impose un minimum de compétences informatiques et son analyse, des compétences statistiques. Réciproquement, les chercheurs suffisamment à l'aise avec la manipulation automatique de données et les statistiques ont rarement une connaissance approfondie des dictionnaires et du processus lexicographique. Cela ne pose pas de problème tant qu'il s'agit de manipuler des données et de les quantifier – encore faut-il identifier des données pertinentes. En revanche, l'analyse et l'interprétation des observations requiert une connaissance disciplinaire qui fait souvent défaut aux métalexigraphes outillés. Meyer (2013) donne dans sa thèse une description des éditions anglaise et allemande de *Wiktionary* en adoptant une perspective TAL (traitement automatique des langues) et métalexigraphique. Comme le relèvent Sajous *et al.* (2020b), son approche est entièrement quantitative et menée totalement automatiquement. Elle ne porte donc que sur des observables

23. Nous n'ironerons pas sur la contemporanéité de la 9^e édition du DAF, encore inachevée à ce jour. À supposer que ces mots satisfassent les critères d'inclusion, il est inutile d'y chercher *agroécologie*, *AMAP*, *antispécisme* ou *antivax* : l'attestation la plus ancienne, celle d'*agroécologie*, date de 1986 (d'après le PR, qui l'inclut en 2016), date de parution du premier tome du dictionnaire. L'addenda ne comble pas ces manques.

24. Une requête envoyée au Service des communications de l'Université de Sherbrooke en juillet 2021 aurait été portée, après relance en novembre de la même année, à l'attention du comité de gouvernance pour « recommandation ». Nous ne doutons pas, lors de l'écriture de ces lignes (en février 2024), de l'obtention prochaine d'une réponse.

quantifiables automatiquement, e.g. la couverture lexicale, le degré de polysémie des entrées ou le nombre de marques lexicographiques. Pour mener ses analyses automatiques, Meyer choisit de comparer les caractéristiques de *Wiktionary* à celles de ressources disponibles au format électronique telles que *WordNet*, destinées à un usage psycholinguistique et largement utilisé en TAL, mais pas à celles de dictionnaires généraux. Il ressort de cette étude quelques tendances très générales : le degré de polysémie des entrées des ressources comparées est plus ou moins corrélé ; *WordNet* affiche une plus grande proportion d'entrées marquées comme relevant des sciences humaines et sociales, etc. En bref, est calculé ce qui est calculable. Mais que conclure de ces observations qui ne répondent à aucune problématisation métalexographique préalablement formulée ? Rien n'est dit sur la pertinence réelle de l'inventaire des sens des unités lexicales ou sur la cohérence de l'attribution des marques lexicographiques. On l'aura compris, la description que Meyer donne de *Wiktionary* adopte une perspective « surtout TAL ». Cette orientation est possiblement imputable au profil de l'auteur, mais pas uniquement. Si rien n'est dit dans l'étude de Meyer, par exemple, sur la qualité des définitions, c'est aussi parce que les approches computationnelles peinent à appréhender des propriétés qualitatives complexes. Cette difficulté fait l'objet de la section 7.

7. Études qualitatives computationnelles

Mener automatiquement une analyse de phénomènes qualitatifs consiste à approximer des caractéristiques qualitatives par des indices calculables. Or cette caractérisation passe par l'établissement de critères déterminés *a priori* qui sont souvent discutables. Sajous (2023) décrit par exemple une série d'initiatives visant à étudier automatiquement la qualité des articles de *Wikipédia* et, dans une moindre mesure, de *Wiktionary* (Lih, 2004; Stvilia *et al.*, 2005; Wilkinson et Huberman, 2007; Kittur et Kraut, 2008; Daxenberger et Gurevych, 2012). Ces études tentent de mettre en lien la qualité des articles, estimée en utilisant des labels attribués par la communauté des contributeurs, et des métriques fondées sur des caractéristiques calculables de ces articles, censées prédire la qualité de ces derniers. Ces études sont problématiques à plusieurs titres. En premier lieu, parce qu'elles recourent, pour estimer la qualité des articles, à une évaluation qui émane des contributeurs et des contributrices et qui s'appuie sur des critères de qualité définis par la communauté (critères discutables et souvent inconnus des internautes ou délibérément négligés). Recourir à cette évaluation interne est d'autant plus contestable que ni son bien-fondé ni sa mise en œuvre ne sont discutés dans les recherches. Concernant les prédicteurs de qualité élaborés dans ces travaux, on peut s'interroger non seulement sur la manière de les calculer, mais aussi sur leur pertinence. La « rigueur » de Lih (2004), qui correspond au nombre de révisions d'une page, est souvent utilisée. Pourtant, Sajous (2023) dément la pertinence de cette mesure comme prédicteur de qualité lorsqu'elle est calculée sur l'ensemble d'une page. Outre les mesures proposées par Lih, Stvilia *et al.* (2005) tentent de prédire la qualité des articles de *Wikipédia* en s'appuyant sur le calcul de la longueur des articles, leur fraîcheur (*currency*), i.e. le délai écoulé entre leur dernière mise à jour et la date de consultation, et le degré de formalité de la langue. Mais peut-on penser qu'il y aurait-il une « bonne » longueur commune à tous

les sujets traités ou que l'absence d'une mise à jour récente d'un article traitant d'une information historique et factuelle soit problématique ? Concernant le degré de formalité de la langue, outre que la pertinence de la procédure automatique choisie pour l'estimer est douteuse, s'agit-il réellement d'un critère de qualité pour une encyclopédie ? Si nous avons vu, en section 6.2, que tout ce qui est calculable n'est pas nécessairement digne d'intérêt, nous voyons ici que les propriétés calculables ne sont pas toujours des propriétés pertinentes pour caractériser le phénomène qu'il s'agit d'étudier.

Revenons à l'analyse des dictionnaires « de langue » et à celle de ses définitions, dont nous regrettons en section 6.2 l'absence de l'étude de Meyer (2013). On peut difficilement imaginer qu'un algorithme puisse statuer sur le fait qu'une définition soit précise, complète, claire et compréhensible pour un public donné. Une procédure automatique pourra déceler un ensemble de problèmes, mais pas identifier une « bonne » formulation. On peut vérifier automatiquement, par exemple, la fermeture du dictionnaire (*i.e.* le fait que les mots du *definiens* soient tous présents dans la nomenclature), identifier d'éventuelles circularités ou encore calculer des scores de complexité syntaxique. Aucune procédure automatique ne sera en revanche capable d'évaluer la qualité d'une définition tant que les propriétés d'une « bonne » définition ne seront pas formellement explicitées. Or, si certaines recensions des critères à prendre en compte pour la critique des dictionnaires mentionnent que les définitions doivent être « appropriées » et « adaptées stylistiquement au public visé » (Swanepoel, 2008), ou (plus laconiquement) « correctes » (Tarp, 2017: 123), aucune procédure opérationnelle (manuelle ou automatique) n'est décrite. Pour ce type d'étude, il peut être sage de renoncer à l'automatisation et se tourner vers des analyses manuelles.

8. Bilan et proposition

La présente étude, motivée par l'évolution de la nature des dictionnaires et celle de leurs conditions de production, porte spécifiquement sur l'analyse des dictionnaires (particulièrement sur celle des dictionnaires numériques) et entend donner un éclairage méthodologique et épistémologique à cette activité. Nous avons affirmé la place centrale de l'analyse de dictionnaires dans la pratique métalexigraphique, qu'il s'agisse de fournir une connaissance sur la structure et le contenu réels du dictionnaire, de reconstituer le processus lexicographique lorsque celui-ci est peu ou pas décrit ou de déterminer s'il y a correspondance ou divergence entre discours lexicographique et faits dictionnaires.

En affirmant la centralité de l'analyse des dictionnaires en métalexigraphie, nous avons évoqué leur utilité pour différents publics (acquéreurs, utilisateurs, enseignants, lexicographes) qui pourraient en bénéficier, directement ou indirectement. Nous ne serions pas choqué que cette assertion soit qualifiée d'utopique ou de naïve. L'utilité des analyses métalexigraphiques serait effective dans le cas (idéal ?) où l'activité lexicographique, l'offre dictionnaire et les efforts consentis dans le domaine de l'enseignement ne seraient pas simultanément réduits à la portion congrue. Envisager la lexicographie *de langue française*, et non la lexicographie française, permet de reconsidérer une

situation peu enthousiasmante. En effet, l'activité lexicographique et terminographique au Québec et, plus largement, les besoins en matière de dictionnaire dans la francophonie pourraient ouvrir de nouveaux horizons métalexigraphiques.

Une réflexion sur les méthodes d'analyse des dictionnaires nous a paru d'autant plus nécessaire qu'il n'existe pas d'ouvrage de référence traitant la dimension méthodologique de cette activité métalexigraphique. Notre étude ne prétend pas combler ce manque, mais, en inventoriant les pratiques courantes et en les commentant, elle pose des jalons pour l'identification des correspondances les plus appropriées entre type de démarche, type de phénomène à décrire et type de dictionnaire étudié. Loin de pouvoir établir une procédure à suivre systématiquement, cette recension a permis de mettre en avant les écueils à éviter et les limites des différents types d'analyse :

- les études qualitatives manuelles s'appuient le plus souvent sur l'examen d'un nombre très restreint d'exemples (au regard de la nomenclature globale) qui permet des analyses fines mais dont il est hasardeux de tirer des généralisations ;
- les études qualitatives automatiques reposent sur l'approximation de caractéristiques qualitatives par des propriétés calculables, souvent très discutables ;
- les études quantitatives manuelles dépendent de la sélection d'un ensemble d'observables. Lorsque qu'il s'agit de décrire ces observables, sélectionnés sur des critères internes (système de requêtage, pages d'index) ou externes (liste préétablies, corpus), une vigilance particulière est nécessaire pour contrôler la validité des modalités de sélection. Lorsque qu'il s'agit de généraliser au dictionnaire les caractéristiques calculées sur un échantillon, la méthode d'échantillonnage et la taille requise de l'échantillon sont rarement maîtrisées ;
- les analyses quantitatives automatiques portant sur l'ensemble d'un dictionnaire numérique sont conditionnées par les droits d'accès à ce dictionnaire.

À l'aune des travaux étudiés et des limites afférentes aux différentes catégories d'analyse, nous pouvons affirmer la nécessité de mener de front analyses quantitatives et analyses qualitatives, analyses computationnelles et analyses manuelles, en privilégiant le couple analyses qualitatives manuelles + analyses quantitatives automatiques. Cette préférence est justifiée par deux raisons. La première réside dans les difficultés, énoncées plus haut, que posent les autres types d'analyse. D'une part, il est très difficile de mettre en œuvre automatiquement des analyses qualitatives pertinentes. Pour pouvoir en tirer des conclusions, il faut pouvoir valider ces analyses manuellement, ce qui limite le nombre d'observables à étudier et réduit l'intérêt de ce type d'étude. D'autre part, dans le cas où l'accès automatique aux données est possible, il y a tout intérêt à remplacer les études quantitatives manuelles par des études computationnelles (sauf, bien sûr, si les phénomènes à quantifier sont qualitatifs et trop difficilement calculables automatiquement). La seconde raison de favoriser le couplage entre analyses quantitatives automatiques et analyses qualitatives manuelles est la complémentarité de ces deux approches, notamment au stade de l'interprétation des résultats. Les résultats numériques, interprétés séparément, peuvent se révéler faussement convaincants. Ce ne sont pas nécessairement les manipulations informatiques ou les descriptions statistiques qui posent

problème, mais la suggestion d'une relation d'implication – un simple *donc* – entre les observations et une interprétation proposée. Les statistiques peuvent établir des corrélations, mais pas de liens de causalité. Elles peuvent mettre en évidence des incohérences, mais pas les expliquer. C'est à travers les études qualitatives manuelles que les métalexicographes peuvent interpréter et expliquer les résultats quantitatifs, à la lumière de leur connaissance disciplinaire. Nous proposons donc la démarche suivante, dont les étapes sont récapitulées dans la figure 2, qui reprend les axes de classification de la figure 1 :

1. en phase exploratoire, pratiquer manuellement des analyses qualitatives des dictionnaires afin de se forger des intuitions et de vérifier ou infirmer empiriquement des hypothèses ;
2. quantifier ces observations en menant une analyse automatique sur l'ensemble du dictionnaire, qui permettra en outre de dégager tendances générales et cas particuliers ;
3. observer et analyser, manuellement, cas typiques et cas particuliers ; interpréter ces analyses ;
4. vérifier, automatiquement (si la vérification peut reposer sur des critères calculables), si les interprétations sont valides ;
5. affiner éventuellement hypothèses et analyses.

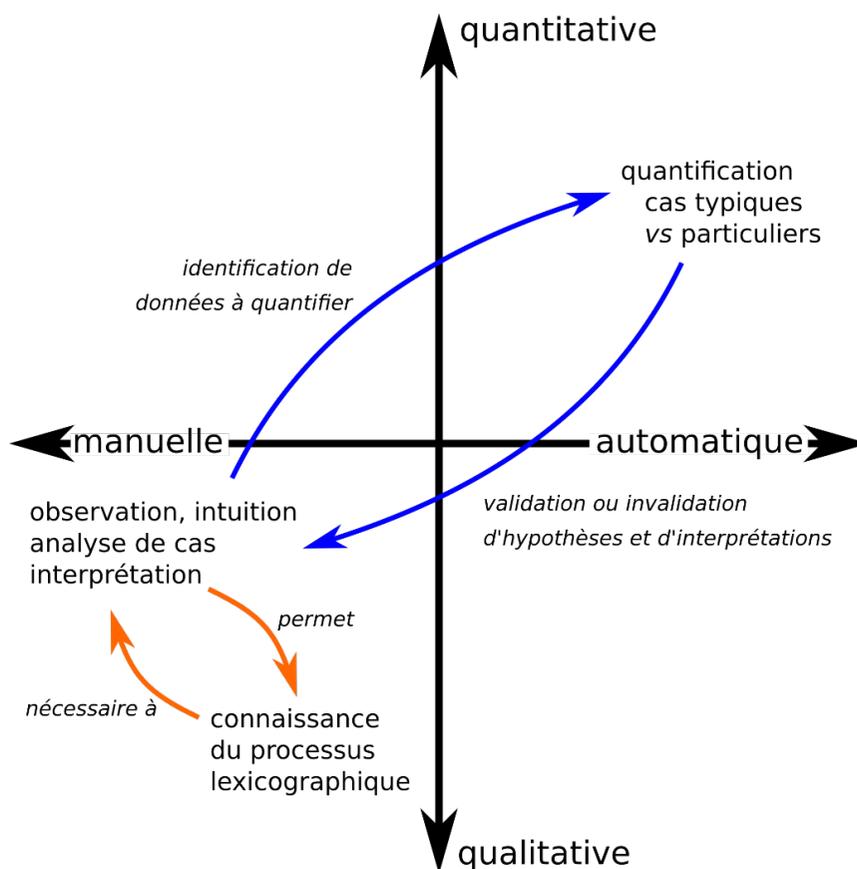


Figure 2 : Complémentarité des méthodes d'analyse des dictionnaires

Chemin faisant, les faits dictionnaires analysés peuvent apporter des indices permettant de reconstruire le processus lexicographique, la connaissance de ce dernier pouvant à son tour se révéler utile à une interprétation pertinente des faits dictionnaires²⁵. On pourrait critiquer la circularité apparente de la démarche. Il faut cependant la considérer comme incrémentale plutôt que circulaire (l'image d'une spirale serait donc plus appropriée) : plus on étudie les dictionnaires et plus on acquiert des connaissances sur le processus lexicographique, et mieux on connaît ce processus, plus on est apte à fournir des interprétations pertinentes des analyses du contenu et de la structure du dictionnaire.

Reste à répondre, par avance, à une objection légitime : les métalexigraphes sont rarement informaticiens et ne peuvent donc mener des analyses automatiques. C'est largement vrai, au moins à l'heure actuelle, pour la métalexigraphie française. Et nous ajoutons que les quelques spécialistes de l'informatique qui s'aventurent en terrain métalexigraphique ne s'intéressent que moyennement à ce champ disciplinaire²⁶, connaissent peu leur objet d'étude et la « tradition lexicographique ». Il leur est difficile, dans ces conditions, de produire des interprétations pertinentes des analyses quantitatives menées. Cette situation n'est nullement une fatalité : en attendant de former une relève polyvalente, il est possible et souhaitable de faire collaborer métalexigraphes non spécialistes en informatique (souvent linguistes) et informaticiens de bonne volonté. À bon entendeur...

25. Nous avons effectivement montré que, si l'analyse du contenu et de la structure du dictionnaire – et de l'évolution éventuelle de ce contenu et de cette structure, révélée par des études diachroniques – permet de mettre au jour certains mécanismes du processus lexicographique ou de vérifier/infirmier le discours qui l'accompagne (cf. l'évolution supposée du nombre d'exemples par page dans le TLF, section 2.4), la connaissance du processus lexicographique est parfois réciproquement un prérequis pour l'interprétation des observations du contenu et de la structure du dictionnaire (cf. l'explication de la construction de l'article *rogue elephant* dans *Wiktionary*, section 4).

26. Intérêt que n'encourage pas l'accès difficile aux publications d'époques plus ou moins lointaines, souvent introuvables, ou à certaines, plus récentes, (in)accessibles sur des supports non libres, et dont aucune version auteur n'est mise à disposition.

Bibliographie

- AKASU, Kaoru (2022), « Methods in Dictionary Criticism », dans Howard Jackson (dir.), *The Bloomsbury Handbook of Lexicography*, London, Bloomsbury, p. 31-42. [En ligne : <http://dx.doi.org/10.5040/9781350181731.ch-004>]
- BÉJOINT, Henri (2000), *Modern Lexicography : An Introduction*. Oxford, Oxford University Press.
- BERG, Donna L., Gaston H. Gönnet et Frank W. Tompa (1988), The New Oxford English Dictionary Project at the University of Waterloo. *Technical Report OED-88-01*, Centre for the New Oxford English Dictionary, University of Waterloo.
- BIELIŃSKA, Monika et Stefan J. Schierholz (2017), *Wörterbuchkritik - Dictionary Criticism, Lexicographica Series Major*, volume 152. Berlin, Boston, De Gruyter.
- BUKOWSKA, Agnieszka A. (2010), Sampling techniques in metalexigraphic research. Dans Anne Dykstra et Tanneke Schoonheim (dir.), *Proceedings of the 14th EURALEX International Congress*, Leeuwarden/Ljouwert, The Netherlands, p. 1258-1269. [En ligne : <https://euralex.org/publications/sampling-techniques-in-metalexigraphic-research/>]
- COLEMAN, Julie et Sarah Ogilvie (2009), « Forensic Dictionary Analysis: Principles and Practice ». *International Journal of Lexicography*, n° 22(1), p. 1-22. [En ligne : <https://doi.org/10.1093/ijl/ecn048>]
- CORBIN, Pierre (1984), « Lexicographe-conseil ». *Lez Valenciennes. Cahiers de l'UER Froissart*, n° 9, p. 113-121.
- CORBIN, Pierre (1990), « Le monde étrange des dictionnaires (7). Logique linguistique et logique botanique : problèmes posés par la définition d'une classe de mots dérivés français ». *Cahiers de lexicologie*, n° 56-57, p. 75-108.
- CORBIN, Pierre (1991), « La lexicographie moderne : contrepoint. Réflexions sur la contribution de Josette Rey-Debove ». *Travaux de linguistique. Revue internationale de linguistique française*, n° 23, p. 161-169.
- CORBIN, Pierre (1995a), « L'articulation entre la pratique lexicographique, la métalexigraphie et la linguistique dans la formation et dans l'activité des lexicographes », dans Jean Pruvost (dir.), *Les dictionnaires de langue. Méthodes et contenus. La journée des dictionnaires*, Université de Cergy-Pontoise: Centre de Recherche Texte / Histoire, p. 81-112.
- CORBIN, Pierre (1995b), « Le monde étrange des dictionnaires (8). Du *Petit Robert* (1967) au *Micro Robert* (1971) : le recyclage de citations ». *Lexique*, n° 12/13, p. 125-145.
- CORBIN, Pierre (2008), Quel avenir pour la lexicographie française ? Dans *Actes du 4e Congrès mondial de linguistique française (CMLF 2014)*, Berlin, p. 1127-1250.
- CORBIN, Pierre (2020), « Les dictionnaires monolingues généraux du français actuel gratuits en ligne : évolutions récentes (2020) ». *Academic Journal of Modern Philology*, n° 9, p. 65-77. [En ligne : <https://doi.org/10.34616/ajmp.2020.9>]

- CORBIN, Pierre et Nathalie Gasiglia (2020), « Les dictionnaires monolingues généraux du français “actuel” gratuits en ligne (début 2019) », dans *Actes du 7e Congrès mondial de linguistique française (CMLF 2020)*, Montpellier. [En ligne : <https://doi.org/10.1051/shsconf/20207805008>]
- CORBIN, Pierre et GASIGLIA, Nathalie (2022), L'offre de dictionnaires monolingues généraux du français imprimés et électroniques au début de 2022 : comparaison de caractéristiques en fonction des supports. Dans *Actes du 8e Congrès Mondial de Linguistique Française (CMLF 2022)*, Orléans. [En ligne : <https://doi.org/10.1051/shsconf/202213804008>]
- CORBIN, Pierre et Nathalie Gasiglia (2023), « Les dictionnaires d'éditeurs dans l'offre globale de dictionnaires généraux électroniques du français (été 2022) » dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [En ligne : <https://doi.org/10.4000/linx.9774>]
- DAXENBERGER, Johannes et Iryna Gurevych (2012), A Corpus-Based Study of Edit Categories in Featured and Non-Featured Wikipedia Articles. Dans *Proceedings of COLING 2012*, Mumbai, India, p. 711-726. [En ligne : <https://aclanthology.org/C12-1044>]
- de SCHRYVER, Gilles-Maurice (2022), « Metalexigraphy: an existential crisis », *Proceedings of the 20th EURALEX International Congress*, Mannheim, Germany, p. 196-206. [En ligne : <https://euralex.org/publications/metalexigraphy-an-existential-crisis/>]
- de SCHRYVER, Gilles-Maurice (2023a), « The future of metalexigraphy: Reaching for the mesosphere ». *Lexikos*, n° 33(1), p. 136-165. [En ligne : <https://doi.org/10.5788/33-1-1803>]
- de SCHRYVER, Gilles-Maurice (2023b), « Generative AI and Lexicography: The Current State of the Art Using ChatGPT », *International Journal of Lexicography*, n° 36(4), p. 355-387. [En ligne : <https://doi.org/10.1093/ijl/ecad021>]
- DOCHERTY, Vincent J. et Ulrich Heid (1998), « Computational Metalexigraphy in Practice - Corpus-based support for the revision of a commercial dictionary » dans Thierry Fontenelle, Philippe Hilgsmann, Archibald Michiels, André Moulin et Siegfried Theissen (dir.), *Proceedings of the 8th EURALEX International Congress*, Liège, Belgium, p. 333-345. [En ligne : <https://euralex.org/publications/computational-metalexigraphy-in-practice-corpus-based-support-for-the-revision-of-a-commercial-dictionary/>]
- ELCHACAR, Mireille (2019), « Comparaison du traitement lexicographique des appellations des identités de genre non traditionnelles dans les dictionnaires professionnels et profanes ». *Études de linguistique appliquée*, n° 194(2), p. 177-191. [En ligne : <https://doi.org/10.3917/ela.194.0177>]
- GAO, Yongwei (2012), « Online English Dictionaries: Friend or Foe » dans *Proceedings of the 15th EURALEX International Congress*, p. 422-433. Oslo, Norway. [En ligne : <https://euralex.org/publications/online-english-dictionaries-friend-or-foe/>]
- GOUWS, Rufus H. (2011), « Learning, unlearning and innovation in the planning of electronic dictionaries » dans Pedro A. Fuertes-Oliviera et Henning Bergenholtz (dir.), *e-lexicography*, London, Continuum, p. 17-29.

- HANKS, Peter (2012), « Corpus evidence and electronic lexicography » dans Sylviane Granger et Magali Paquot (dir.), *Electronic Lexicography*, Oxford, Oxford University Press, p. 57-82. [En ligne : <https://doi.org/10.1093/acprof:oso/9780199654864.003.0004>]
- HARTMANN, Reinhard R.K. (2001), *Teaching and Researching Lexicography*. London, Routledge.
- HAUSMANN, Franz J., Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (1989), *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie*. New-York, Walter de Gruyter.
- JAKUBÍČEK, Miloš et Michael Rundell (2023), « The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? » dans *Proceedings of the eLex2023 conference*, Brno, Czech Republic, p. 518-533. [En ligne : <https://elex.link/elex2023/proceedings-download/>]
- JOSSELIN-LERAY, Amélie (2010), « Affiner la description des termes dans les dictionnaires généraux : l'apport d'un corpus de vulgarisation ». *Lexis*, n° 4, p. 65-104. [En ligne : <https://doi.org/10.4000/lexis.590>]
- KILGARRIFF, Adam (2005), « If dictionaries are free, who will buy them? ». *Kernerman Dictionary News*, no 13, p. 17-19. [En ligne : https://lexicala.com/wp-content/uploads/kdn13_2005_If_dictionaries_are_free_AK.pdf]
- KITTUR, Aniket et Robert E. Kraut (2008), « Harnessing the Wisdom of Crowds in Wikipedia : Quality through Coordination » dans *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW'08*, San Diego, California, p. 37-46. [En ligne : <https://doi.org/10.1145/1460563.1460572>]
- LEW, Robert (2014), « User-generated content (UGC) in online English dictionaries ». *OPAL*, n° 4, p. 8-26. [En ligne : <http://hdl.handle.net/10593/5011>]
- LEW, Robert (2015), « Research into the use of online dictionaries ». *International Journal of Lexicography*, n° 28(2), p. 232-253. [En ligne : <https://doi.org/10.1093/ijl/ecv010>]
- LEW, Robert (2023), « ChatGPT as a COBUILD lexicographer ». *Humanities and Social Sciences Communications*, n° 10. [En ligne : <https://doi.org/10.1057/s41599-023-02119-6>]
- LIH, Andrew (2004), « Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource » dans *Proceedings of the 5th International Symposium on Online Journalism*, Austin, Texas. [En ligne : <https://isoj.org/wp-content/uploads/2018/01/wikipedia.pdf>]
- MEYER, Christian M. (2013), *Wiktionary : The Metalexicographic and the Natural Language Processing Perspective*. Thèse de doctorat, Technische Universität Darmstadt. [En ligne : <https://tuprints.ulb.tu-darmstadt.de/id/eprint/3654>]
- NAGAO, Makoto, Jun'ichi Tsujii, Y. Ueda et M. Takiyama (1980), « An attempt to computerized dictionary data bases » dans *Proceedings of COLING 1980*, Tokyo, Japan, p. 534-542.
- NESI, Hilary (2008), « Dictionaries in electronic form » dans Anthony P. Cowie (dir.), *The Oxford History of English Lexicography*, Oxford, Oxford University Press, p. 458-478.

- NESI, Hilary (2022), « Researching users and uses of dictionaries » dans Howard Jackson (dir.), *The Bloomsbury Handbook of Lexicography*, London, Bloomsbury, p. 43-56.
- QUEMADA, Bernard (1987), « Notes sur *lexicographie* et *dictionnaire* », *Cahiers de lexicologie*, n° 51, p. 229-242.
- RADERMACHER, Ruth (2004), *Le Trésor de la langue française. Une étude historique et lexicographique*. Thèse de doctorat, Université Marc Bloch, Strasbourg. [En ligne : https://www.atilf.fr/wp-content/uploads/theses/atilf_These_Radermacher_Ruth_2004.pdf]
- REY, Alain et Simone Delesalle (1979), « Problèmes et conflits lexicographiques ». *Langue française*, n° 43, p. 4-26.
- REY-DEBOVE, Josette (1971), *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris – La Haye, Mouton.
- RUNDELL, Michael (2012), « ‘It works in practice but will it work in theory?’ The uneasy relationship between lexicography and matters theoretical » dans *Proceedings of the 15th EURALEX International Congress*, Oslo, Norway, p. 47-92. [En ligne : <https://euralex.org/publications/it-works-in-practice-but-will-it-work-in-theory-the-uneasy-relationship-between-lexicography-and-matters-theoretical/>]
- RUNDELL, Michael (2014), « Macmillan English Dictionary: The End of Print? ». *Slovenščina 2.0*, n° 2(2), p. 1-14. [En ligne : <https://journals.uni-lj.si/slovenscina2/article/viewFile/6938/6631>]
- RUNDELL, Michael (2017), « Dictionaries and crowdsourcing, wikis, and user-generated content » dans Patrick Hanks et Gilles-Maurice de Schryver (dir.), *International Handbook of Modern Lexis and Lexicography*. Berlin, Heidelberg, Springer. [En ligne : https://doi.org/10.1007/978-3-642-45369-4_26-1]
- RUNDELL, Michael (2023), « Automating the creation of dictionaries: are we nearly there? », *Proceedings of ASIALEX2023*, Seoul, Korea, p. 9-17. [En ligne : <https://korealex.org/34>]
- RUNDELL, Michael et Adam Kilgariff (2011), « Automating the creation of dictionaries: Where will it all end? » dans Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin, et Magali Paquot (dir.), *A Taste for Corpora. In honour of Sylviane Granger*, John Benjamins, p. 257-282.
- RUNDELL, Michael et Penny Stock (1992), « The corpus revolution ». *English Today*, n° 30, p. 9-14. [En ligne : <https://doi.org/10.1017/S0266078400006751>]
- SAJOUS, Franck (2022), « Using Wiktionary revision history to uncover lexical innovations related to topical events: Application to Covid-19 neologisms » dans Annette Klosa-Kückelhaus et Ilan Kernerman (dir.), *Lexicography of Coronavirus-related Neologisms*, volume 163 of *Lexicographica Series Maior*, Berlin, Boston, De Gruyter, p. 275-306. [En ligne : <https://doi.org/10.1515/9783110798081-014>]
- SAJOUS, Franck (2023), « Quantité et qualité dans le Wiktionnaire : de la diversité... à la rigueur ? » dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86. [En ligne : <https://doi.org/10.4000/linx.9835>]

- SAJOUS, Franck, Basilio Calderone et Nabil Hathout (2020a), « ENGLAWI: From Human- to Machine-Readable Wiktionary », *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, p. 3016-3026. [En ligne : <https://www.aclweb.org/anthology/2020.lrec-1.369>]
- SAJOUS, Franck, Basilio Calderone et Nabil Hathout (2020b), « Extraire et encoder l'information lexicale de Wiktionary : quel boulot pour étrangler le goulot ! ». *Lexique*, n° 27, p. 121-144. [En ligne : <https://dx.doi.org/10.54563/lexique.569>]
- SAJOUS, Franck et Nabil Hathout (2015), « GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary », *Proceedings of the eLex 2015 conference*, p. 405-426. Herstmonceux, England. [En ligne : <https://elex.link/elex2015/conference-proceedings/paper-27/>]
- SAJOUS, Franck, Nabil Hathout et Amélie Josselin-Leray (2019), « Du vin et devin dans le Wiktionnaire : neutralité de point de vue ou neutralité et point de vue ? ». *Études de linguistique appliquée*, n° 194(2), p. 147-164. [En ligne : <https://doi.org/10.3917/ela.194.0147>]
- SAJOUS, Franck et John Humbley (2022), « Mesures d'isolement sanitaire dans Wiktionnaire et Wikipédia : néologie et lexicographie ou néonymie et terminographie ? », *Estudios Románicos*, n° 31, p. 175-201. [En ligne : <https://doi.org/10.6018/ER.510631>]
- SAJOUS, Franck et Amélie Josselin-Leray (2022), « Issues in Collaborative and Crowdsourced Lexicography » dans Howard Jackson (dir.), *The Bloomsbury Handbook of Lexicography*, London, Bloomsbury Academic, p. 343-358. [En ligne : <http://dx.doi.org/10.5040/9781350181731.ch-020>]
- SAJOUS, Franck, Amélie Josselin-Leray et Nabil Hathout (2018), « The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology », *Lexis*, n° 12. [En ligne : <https://doi.org/10.4000/lexis.2322>]
- SAJOUS, Franck, Amélie Josselin-Leray et Nabil Hathout (2020c), « Les domaines de spécialité dans les dictionnaires généraux : le lexique de l'informatique analysé par les foules et par les professionnels... de la lexicographie », *Neologica*, n° 14, p. 83-107. [En ligne : <http://dx.doi.org/10.15122/isbn.978-2-406-10571-8.p.0083>]
- SAJOUS, Franck et Camille Martinez (2022), « Metalexicographical Investigations with the DiCo Database », *International Journal of Lexicography*, n° 35(1), p. 75-106. [En ligne : <https://doi.org/10.1093/ijl/ecab017>]
- STVILIA, Besiki, Michael B. Twidale, Linda C. Smith et Les Gasser (2005), « Assessing information quality of a community-based encyclopedia », *Proceedings of the 2005 International Conference on Information Quality (ICIQ 2005)*, Cambridge, Massachusetts, p. 442-454. [En ligne : <https://myweb.fsu.edu/bstvilia/papers/quantWiki.pdf>]
- SWANEPOEL, Piet (2008), « Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria ». *Lexicos*, n° 18, p. 207-231. [En ligne : <https://doi.org/10.5788/18-0-485>]
- SWANEPOEL, Piet (2017), « Methods in dictionary criticism » dans Monika Bielińska et Stefan J. Schierholz (dir.), *Wörterbuchkritik - Dictionary Criticism, vol. 152 of Lexicographica Series Major*, Berlin, Boston, De Gruyter, p. 81-112. [En ligne : <https://doi.org/10.1515/9783110547108>]

- TARP, Sven (2008), « Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography », Vol. 134 of *Lexicographica Series Maior*. Tübingen, Max Niemeyer Verlag. [En ligne : <https://doi.org/10.1515/9783484970434>]
- TARP, Sven (2017), « Dictionary criticism and lexicographical function theory » dans Monika Bielińska et Stefan J. Schierholz (dir.), *Wörterbuchkritik - Dictionary Criticism*, Lexicographica Series Maior, vol. 152, Berlin, Boston, De Gruyter, p. 113-132. [En ligne : <https://doi.org/10.1515/9783110547108-006>]
- TRAP-JENSEN, Lars (2018), « Lexicography between NLP and Linguistics: Aspects of Theory and Practice », *Proceedings of the 18th EURALEX International Congress*, Ljubljana, Slovenia, p. 25-37. [En ligne : <https://euralex.org/publications/lexicography-between-nlp-and-linguistics-aspects-of-theory-and-practice/>]
- VINCENT, Nadine (2020). « Quand l'actualité impose des néologismes, les dictionnaires professionnels peuvent-ils rivaliser avec les dictionnaires collaboratifs? », dans Giovanni Tallarico, John Humbley et Christine Jacquet-Pfau (dir.), *Nouveaux horizons pour la néologie en français. Hommage à Jean-François Sablayrolles*, Paris, Lambert-Lucas, p. 69-85.
- VINCENT, Nadine (2022), « Faut-il adapter les dictionnaires à l'air du temps ? Proposition d'un traitement polyphonique du mot *woke* », dans Geneviève Bernard Barbeau et Nadine Vincent (dir.) *Regards linguistiques sur des mots polémiques*, *Circula*, n° 15, p. 122-145. [En ligne : <https://doi.org/10.17118/11143/19982>]
- WILKINSON, Dennis M. et Bernardo A. Huberman (2007), « Assessing the value of cooperation in Wikipedia », *First Monday*, n° 12(4). [En ligne : <https://doi.org/10.5210/fm.v12i4.1763>]
- WOLFER, Sascha et Carolin Müller-Spitzer (2016), « How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions », *Lexicos*, n° 26, p. 347-371. [En ligne : <https://doi.org/10.5788/26-1-1346>]
- ZOTTI, Valeria (2019), « Ressources numériques pour la traduction des mots désignant des *realia* ». *Études de linguistique appliquée*, n° 194(2), p. 227-246. [En ligne : <https://doi.org/10.3917/ela.194.0227>]

TITRE: LA LEXICOGRAPHIE NUMÉRIQUE ET LA NORME GRAMMATICALE : LE *WIKTIONNAIRE* EST-IL AUSSI DESCRIPTIF QUE VOULU ?

AUTEUR: SOPHIE PIRON, UNIVERSITÉ DU QUÉBEC À MONTRÉAL, CRIFUQ

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 187-205

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22412](http://hdl.handle.net/11143/22412)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22412](https://doi.org/10.17118/11143/22412)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

La lexicographie numérique et la norme grammaticale : le *Wiktionnaire* est-il aussi descriptif que voulu ?

Sophie Piron, Université du Québec à Montréal, CRIFUQ

Résumé :

L'étude analyse comment 14 phénomènes grammaticaux (choix de conjonctions de subordination, choix de conjonctions de coordination, choix de mode, etc.) soumis à des tensions normatives en français contemporain et critiqués par l'Académie française sont décrits dans le *Wiktionnaire* et quel jugement normatif y est rendu. La question se posait de savoir si cette ressource lexicographique se présentant comme descriptive, et non prescriptive, atteint son objectif. Certains ouvrages de référence faisant actuellement preuve d'ouverture face à des emplois parfois en usage depuis bien longtemps, l'analyse menée dans le présent article montre que le *Wiktionnaire* opte la plupart du temps pour l'avis le plus normatif, mais que certains articles atteignent l'objectif de description linguistique.

Mots-clés :

Wiktionnaire, lexicographie collaborative, norme grammaticale

Abstract :

The study analyzes how 14 grammatical phenomena (choice of subordinating conjunctions, choice of coordinating conjunctions, choice of mode, etc.) subject to normative tensions in contemporary French and criticized by the French Academy are described in the “Wiktionnaire” and what kind of normative judgment is present. Does this lexicographic opened source, supposed to be descriptive, achieves its objective? Some other grammatical and lexicographic resources are currently showing openness towards various linguistic constructions and expressions already in use for quite a long time. The analysis carried out in the present study shows that the “Wiktionnaire” often chooses the most normative point of view, but that some lexicographic entries achieve the objective of linguistic description (instead of prescription).

Keywords :

Wiktionnaire, collaborative lexicography, grammatical norm

1. Introduction

La présente étude s'intéresse à la manière dont le *Wiktionnaire* – un dictionnaire collaboratif du français, en ligne et gratuit – traite une série de phénomènes grammaticaux soumis à des tensions normatives en français contemporain. Par exemple, que dit le *Wiktionnaire* à propos du mode à utiliser à la suite de *après que* ? Faut-il l'indicatif ou le subjonctif ? Les deux modes sont-ils équivalents ? Dans l'exposé qui suit, la description fournie par le *Wiktionnaire* sera comparée à un bassin d'ouvrages de référence contemporains, de manière à déceler quel point de vue adopte le *Wiktionnaire* : opte-t-il pour une sanction très traditionnelle ou, à l'inverse, pour une ouverture à l'innovation linguistique, ou encore, adopte-t-il une position intermédiaire entre le rejet immédiat de la nouveauté et l'acceptation enthousiaste des entorses à la tradition ? L'étude proposée envisage le dictionnaire comme un corpus au sein duquel se déploient des points de vue ancrés dans une culture (Rey, 2008). L'objet d'étude idéologique concerne ici la norme, et plus précisément la vision normative déployée au sein d'une ressource lexicographique contemporaine et numérique.

Le *Wiktionnaire* constitue probablement un des outils phares de la lexicographie francophone numérique en accès libre. Il s'agit d'un « dictionnaire libre et gratuit que tout le monde peut améliorer » (*Wiktionnaire*¹), autrement dit d'un dictionnaire collaboratif, ressortissant à la lexicographie profane, non experte (Vincent, 2019a, 2019b). Mis en ligne en 2003, ce wiki – site web dynamique – se range sous l'emblème de l'encyclopédie *Wikipédia* (et de la Wikimedia Foundation). Il s'intègre dans le projet *Wiktionary*, dont il constitue le volet en français. Le *Wiktionnaire* se veut ainsi une ressource lexicographique et, à ce titre, contient des définitions (y compris des synonymes, des antonymes et des hyperonymes), l'étymologie, la prononciation, des exemples et des citations, des expressions et des locutions. Le contenu du *Wiktionnaire* mentionne parfois des ouvrages de référence et des sites discutant de points de langue.

Pour guider les contributeurs dans la rédaction d'articles lexicographiques, une page web du site est consacrée aux conventions que les wiktionnaristes doivent suivre. Au rang des grands principes se trouvent la « vérifiabilité de l'information », la « pertinence des sources utilisées » et la « neutralité de point de vue »². À l'égard du principe de neutralité, il est précisé que les jugements sur des éléments de langue que sont, par exemple, des variantes de prononciation d'un mot ou des variantes d'usage doivent être présentés comme des points de vue, et non comme des vérités absolues.

La rédaction d'une définition doit présenter le sens d'un mot sans juger de ses usages. Une note d'usage peut préciser qu'un sens est jugé incorrect ou offensant par certaines personnes, mais la rédaction doit pouvoir montrer qu'il s'agit d'un point de vue. Tous les sens d'un mot ont leur place [...].

1. https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d'accueil [page consultée le 4 juillet 2024]

2. <https://fr.wiktionary.org/wiki/Wiktionnaire:Références> [page consultée le 4 juillet 2024]

Le Wiktionnaire ne cherche pas à défendre une prononciation correcte mais à rapporter toutes les prononciations pour tous les mots. (*Wiktionnaire*³)

La posture épistémologique de la neutralité est donc au cœur du projet lexicographique et trouve un écho dans la position descriptive – et non prescriptive – que le *Wiktionnaire* affiche très clairement face à la langue.

Un article de dictionnaire ne doit jamais spécifier qu'un usage est abusif, car c'est l'expression d'un point de vue (même si on partage ce point de vue). [...] En fait, la meilleure façon d'écrire un article de dictionnaire neutre est de décrire l'usage réel, pas de définir le *bon usage*, donc pas de prescrire ce que devrait être l'usage. (*Wiktionnaire*⁴)

Son objectif est seulement descriptif : il ne s'agit ni de défendre le français ou une autre langue, ni d'être normatif. Il ne juge donc pas la valeur des mots et n'essaie pas de leur donner ou de leur refuser son aval. (*Wiktionnaire*⁵)

L'adéquation descriptive est également prônée. Les contributeurs sont ainsi engagés à présenter des informations contemporaines, ce qui peut paraître ironique quand on sait que le *Wiktionnaire* a d'abord été constitué sur la base d'importations automatisées de deux dictionnaires anciens en libre accès, le *Littre* – datant de la seconde moitié du 19^e siècle – et l'édition de 1932 du *Dictionnaire de l'Académie française* (Sajous 2023), ainsi que du *Trésor de la langue française* (1971-1994), lui aussi accessible en ligne.

La linguistique et la lexicographie sont des disciplines qui évoluent et les méthodes et données d'hier ne sont peut-être plus valides aujourd'hui. [...] On veillera à utiliser des sources récentes en priorité, et de [sic] ne limiter l'usage de sources anciennes qu'à des fins de mises en contexte et d'historiographie lexicographique. (*Wiktionnaire*⁶)

Les contenus du *Wiktionnaire* peuvent proposer les registres de langue⁷ dont relèvent les mots décrits. Cette information est cruciale pour l'étude menée ici, dans la mesure où le classement d'un emploi dans un registre peut varier d'un ouvrage de référence à l'autre, et donc être un indicateur d'une orientation plus ou moins traditionnelle ou progressiste en matière linguistique. Par ailleurs, le système de marques étant habituellement associé à la pratique lexicographique professionnelle, on s'interrogera sur les marques utilisées par la lexicographie profane. Reproduit-elle au moins partiellement ce système ou le transpose-t-elle dans une forme non experte ? Et les non-experts adoptent-ils les mêmes sanctions que les professionnels ? Enfin, donner un registre de langue, cela revient à hié-

3. https://fr.wiktionary.org/wiki/Wiktionnaire:Neutralité_de_point_de_vue [page consultée le 4 juillet 2024]

4. https://fr.wiktionary.org/wiki/Wiktionnaire:Neutralité_de_point_de_vue [page consultée le 4 juillet 2024]

5. https://fr.wiktionary.org/wiki/Wiktionnaire:À_propos [page consultée le 4 juillet 2024]

6. https://fr.wiktionary.org/wiki/Wiktionnaire:Références#Vérifiabilité_de_l'information [page consultée le 4 juillet 2024]

7. https://fr.wiktionary.org/wiki/Wiktionnaire:À_propos [page consultée le 4 juillet 2024]

rarchiser les usages. Or la volonté affichée de faire du *Wiktionnaire* un ouvrage non normatif, neutre et descriptif pose véritablement la question de la transposition de ce principe dans ses contenus. On sait déjà, grâce à Sajous *et al.* (2019), que le *Wiktionnaire* n'applique pas systématiquement son propre principe fondateur de la neutralité de point de vue. Sajous (2023) avance plusieurs explications à ce sujet : en particulier, une rédaction par agrégation plutôt que par collaboration et avec révision, et une communauté de wiktionnaristes finalement peu nombreuse et vraisemblablement pas aussi diversifiée qu'espéré.

La présente étude vise à répondre à deux questions : à partir d'une sélection d'articles du *Wiktionnaire* consacrés à des phénomènes grammaticaux, peut-on affirmer que les contenus exposés sont conformes à l'approche descriptive recommandée ? Et quelle vision de la norme – plus spécifiquement de la norme grammaticale – le *Wiktionnaire* propage-t-il ?

Le texte qui suit expose, dans la deuxième section, les phénomènes grammaticaux soumis à l'étude, puis les références qui serviront de panel comparatif pour évaluer l'avis énoncé par le *Wiktionnaire* dans chaque situation linguistique étudiée. Il sera aussi précisé comment évolue la norme et par quelles phases elle passe. La troisième section propose une analyse systématique des phénomènes linguistiques retenus et envisage les points de vue émis par le *Wiktionnaire* et les autres ouvrages de référence du panel comparatif.

2. Méthodologie et ancrage épistémologique

Cette section, consacrée à l'exposé méthodologique, passe en revue les 14 phénomènes linguistiques retenus à l'étude, puis dresse la liste des onze ouvrages de référence choisis pour servir de comparatif au *Wiktionnaire* et évaluer le point de vue proposé par celui-ci. Cette section expose également quel est l'ancrage épistémologique de la présente étude en explicitant la vision de la norme qui sous-tend l'analyse proposée.

2.1. Phénomènes soumis à l'étude

Nous avons sélectionné 14 usages critiqués sur le site de l'Académie française *Dire, Ne pas dire*. Ils ont en commun de ne pas s'inscrire foncièrement sur le plan lexical, mais plutôt sur le plan grammatical. Nous avons ainsi retenu cinq subordonnants condamnés, soit par leur forme (*de façon à ce que*, *de manière à ce que*, *pour ne pas que*), soit dans leur existence même (*malgré que*), soit encore dans leur choix en fonction du contexte (*demande* requiert le subordonnant *que*, et non le subordonnant composé *à ce que*). Le tableau 1 expose les formes acceptées et condamnées. Les citations exemplifiant les secondes ont pour vocation de montrer, d'ores et déjà, que le rejet de ces formes peut être soumis à questionnement.

Forme acceptée	Forme condamnée par l'Académie française et citation l'exemplifiant
<i>de façon que</i>	<i>de façon à ce que</i> « Elle plaçait son éventail <u>de façon à ce qu'il pût le prendre.</u> » (Stendhal, cité dans le Grand Robert)
<i>de manière que</i>	<i>de manière à ce que</i> « Un portique disposé <u>de manière à ce qu'on trouvât de l'ombre à toute heure.</u> » (France, cité dans le Petit Robert)
<i>demandeur que</i>	<i>demandeur à ce que</i> <i>Tu demandes toujours que tout soit parfait!</i> (ou : <i>Tu <u>demandes</u> toujours <u>à ce que</u> tout soit parfait!</i>) (OQLF)
<i>bien que</i>	<i>malgré que</i> « Mais comme on déjeunait tard, <u>malgré que</u> les gens commençassent à passer dans la rue, on apportait seulement la grande tarte aux pommes [...]. » (Proust, cité dans le Grand Robert)
<i>pour que... ne... pas...</i>	<i>pour ne pas que</i> Pop. <i>pour ne pas qu'il vienne</i> (Trésor de la langue française)

Tableau 1. Subordonnants

Nous avons également retenu quatre phénomènes de construction critiqués et impliquant la morphosyntaxe. Deux d'entre eux prennent place en situation de subordination (*après que* suivi du subjonctif, *sans que* suivi d'une négation), un autre phénomène critiqué relève de la transitivité verbale (*se rappeler de quelque chose*) et le quatrième phénomène fait intervenir le choix de l'auxiliaire de conjugaison (*avoir convenu*). Ces phénomènes sont décrits dans le tableau 2.

Forme acceptée	Forme condamnée par l'Académie française et citation l'exemplifiant
<i>après que</i> + indicatif	<i>après que</i> + subjonctif « <u>Après que</u> le prévenu <u>ait été déclaré coupable</u> , le juge décide, en fonction de ce rapport, du type de probation à appliquer. » (Le Figaro, cité dans Usito)
<i>sans que</i> + phrase positive	<i>sans que</i> + <i>ne</i> <i>Je ne peux pas lui parler <u>sans qu'elle ne se sente agressée.</u></i> (Antidote, Guide grammatical)
<i>être convenu</i>	<i>avoir convenu</i> <i>Nous <u>avons convenu</u> d'un rendez-vous pour le lendemain.</i> (Usito)
<i>se rappeler quelque chose</i>	<i>se rappeler de quelque chose</i> « C'est vrai, dit le père Chapdelaine, <u>je me rappelle de ce temps-là.</u> Il n'y avait pas une seule maison en haut du lac: rien que des sauvages et quelques chasseurs » (Hémon, cité dans le TLF)

Tableau 2. Phénomènes de construction et de morphosyntaxe

Nous avons enfin retenu cinq phénomènes de grammaticalisation critiqués, au rang desquels on compte le déterminant composé *un espèce de*, les coordonnants *voire même*, *du coup* et les compositions *au final* et *des fois*. Ils sont décrits dans le tableau 3.

Forme acceptée	Forme condamnée par l'Académie française et citation l'exemplifiant
<i>par conséquent, donc,</i>	<i>du coup</i> « l'impossibilité de la rencontrer devient une évidence qui le met, <u>du coup</u> , à l'abri de l'attente. » (Laberge, citée dans <i>Usito</i>)
<i>finalelement</i>	<i>au final</i> « D'ailleurs, combien cette place vous coûtera-t-elle <u>au final</u> ? » (L'Express, cité dans <i>Antidote</i>)
<i>parfois</i>	<i>des fois</i> « Et puis, les cartes, cela faisait passer le temps... j'avais gagné <u>des fois</u> ... maintenant, jamais, jamais... » (Aragon, cité dans le TLF)
<i>voire</i>	<i>voire même</i> <i>Le chasseur (...), il lui faut étouffer le bruit de ses pas, se courber, voire même ramper si les eaux sont hautes</i> (Vidron, cité dans le TLF)
<i>une espèce de</i>	<i>un espèce de</i> <i>On ne dira donc pas, par exemple, [...] un espèce de graphique, mais bien [...] une espèce de graphique.</i> (OQLF) ⁸

Tableau 3. Formes grammaticalisées

2.2. Corpus de ressources de référence

Les phénomènes retenus à l'étude ont été recherchés d'abord dans le *Wiktionnaire*, ensuite dans un panel comparatif constitué de onze ressources de référence qui forment le paysage d'outils linguistiques en francophonie du Nord. En font évidemment partie les incontournables au sein de la francophonie : le dictionnaire grand public *Petit Larousse illustré* (2024) et le *Petit Robert 2024 électronique*. Des outils plus spécialisés ont également été consultés : *Grand Robert électronique* (2023), *Trésor de la langue française* (1971-1994), *Dictionnaire de l'Académie française* (9^e édition) et *Le bon usage* (2016). Le panel comparatif comprend également un versant québécois de cinq références, avec le site de l'Office québécois de la langue française (OQLF), le site des Ressources du portail linguistique du Canada, le *Multidictionnaire*, le dictionnaire *Usito* et l'outil *Antidote 11*. Certaines de ces références sont spécifiques au Québec, d'autres sont parfois consultées en Europe. C'est le cas d'*Antidote*, d'*Usito*, du site de l'OQLF. L'origine québécoise d'une partie des références s'explique par l'ancrage géographique de nos recherches. Nous souhaitons également évaluer comment le *Wiktionnaire* – dont les rédacteurs sont majoritairement Français – s'inscrit dans l'ensemble des outils linguistiques, y compris québécois.

8. La forme *un espèce de* est ici citée en mention, et non en usage. Comme nous le verrons plus loin, aucun ouvrage n'accepte ni n'exemplifie la forme en émergence.

Le panel de ressources de référence permettra de dresser le panorama des sanctions au sujet des 14 phénomènes linguistiques retenus (voir section 2.1). Il sera ainsi possible d'évaluer la position adoptée par le *Wiktionnaire* par rapport aux ressources professionnelles, qui font – en principe – l'objet d'une coordination éditoriale serrée.

2.3. Évolution de la norme

L'analyse proposée dans la présente étude repose sur une conception de la norme comme un ensemble de règles en évolution. Nous posons l'existence de trois phases lorsque se produit un changement linguistique à l'époque moderne. La première est celle de l'émergence. À ce stade, un point de langue est soumis à des pressions et tensions normatives, une variante apparaît dans l'usage, mais elle est condamnée et constitue donc une forme non standard. Elle n'apparaît alors que dans des listes de « fautes » et, si elle est signalée dans des ouvrages de référence, c'est uniquement à titre de séquence fortement condamnée.

La deuxième phase d'évolution est celle de la transition. À cette étape, un basculement normatif est en cours, mais n'est pas encore pleinement entériné. Les ouvrages et discours de référence émettent des sanctions différentes, les uns acceptant la nouvelle variante, les autres la refusant encore. Cette phase se caractérise également par des avis moins tranchés, naviguant entre caution du bout des lèvres et acceptation d'un usage en pleine expansion au sein de la communauté linguistique.

Enfin, la troisième phase d'évolution est celle de l'acceptation. Cette fois, la variante moderne est pleinement intégrée dans l'usage, et entérinée dans les ouvrages de référence. Désormais, la variante traditionnelle et la variante moderne coexistent au sein du standard, même s'il peut encore y avoir une hiérarchie entre les deux. La variante nouvellement acceptée peut être qualifiée de *moderne* ou de *courante*, tandis que la traditionnelle peut s'être hissée au rang de variante soutenue. Lorsque le remplacement est plus abouti, la variante la plus récente peut avoir poussé la plus ancienne vers un emploi soit littéraire soit populaire.

Les avis émis par les différents organes de référence (tantôt toujours négatifs, tantôt variables, tantôt encore systématiquement positifs) permettent d'objectiver les trois phases d'évolution qui viennent d'être exposées. Les phénomènes grammaticaux analysés dans les sections qui suivent (*après que* + subjonctif, *de façon à ce que*, *voire même*, etc.) se situent tantôt en tension normative (phase 1 d'évolution), tantôt en basculement normatif (phase 2 d'évolution). Il sera intéressant de voir comment un dictionnaire collaboratif envisage l'évolution linguistique dans les deux phases critiques où elle se met en place.

3. Analyse du Wiktionnaire

Cette troisième section analyse les 14 phénomènes linguistiques retenus à l'étude. Ils sont répartis en trois parties, qui envisagent chacune un des trois profils de couverture rencontrés dans le *Wiktionnaire* : variante linguistique non consignée, variante signalée en note et variante intégrée dans l'article⁹.

3.1. Variante non consignée

Parmi les 14 phénomènes choisis pour étude ici, trois ne sont pas présentés dans le *Wiktionnaire*, bien que ce soit, dans chacun des cas, de manière différente. Le premier est la construction *demander à ce que*. Le *Wiktionnaire* ne présente aucune des deux variantes, ni *demander à ce que*, ni sa version standard *demander que*. Dans le panel de comparaison, seul le *Petit Larousse illustré 2024* ne consigne pas non plus la variante standard *demander que*. Le *Wiktionnaire* est donc aligné sur un ouvrage très grand public, où une construction complexe (de subordination) est laissée hors champ. Les autres références du corpus tantôt ignorent la variante questionnable (*Multidictionnaire*, *Usito*), tantôt signalent qu'elle est sujette à caution, mais cela se fait alors à des degrés divers. C'est *une faute* selon le *Grand Robert*, c'est seulement *critiqué* selon le *Petit Robert 2024* et *familier* selon le TLF. Le portail linguistique du Canada adopte un point de vue ambivalent orienté positivement, puisqu'il signale que la construction est jugée fautive selon les uns et moins élégante selon les autres. Enfin, trois références (*Le bon usage*, *Antidote* et l'OQLF) présentent *demander à ce que* comme une variante non marquée à côté de *demander que*. On peut donc affirmer que la construction est au stade 2 dans le schéma d'évolution normatif que nous avons proposé dans la section 2.3.

Le 2^e phénomène absent du *Wiktionnaire* est le subordonnant *pour ne pas que*. Cette fois, la variante normée *pour que* y est signalée. Cependant, aucun des exemples proposés ne propose de négation dans la subordonnée (*pour que... ne pas*). Dans les faits, la variante de *pour ne pas que* est donc absente. D'autres ouvrages du panel comparatif ne mentionnent pas non plus la forme non standard. C'est ce que font le *Petit Larousse illustré 2024*, le *Petit Robert 2024*, le *Dictionnaire de l'Académie française* et le *Multidictionnaire*. Ainsi, le *Wiktionnaire* n'est pas isolé sur ce point, mais il ne répond pas à son objectif descriptif. *Le bon usage* (2016), tout en adoptant un point de vue ambivalent plutôt négatif, souligne que la construction « tend à passer de la langue populaire dans la langue écrite. [...] Cette construction, isolée parmi les loc[utions] conjonctives, reste généralement mal accueillie, même par les observateurs non puristes » (BU, 2016 : 1020 a3). Les autres ressources du corpus proposent une analyse qui prend place sur le continuum de la non-acceptation : incorrect (Portail du Canada), populaire (TLF), très familier (*Grand Robert*), familier (*Antidote*, *Usito*, OQLF).

9. Aucun phénomène étudié ici ne faisait l'objet d'un renvoi à un autre article.

Le 3^e phénomène linguistique, *de façon à ce que*, n'est pas traité dans la nomenclature du *Wiktionnaire*, qui propose des entrées pour *façon*, *de façon que* et *de façon à*. En revanche, *de façon à ce que* apparaît dans le *Wiktionnaire* dans une citation de David George Haskell, traduit de l'anglais, chez Flammarion en 2018 (« *de façon à ce que les abeilles puissent s'y cogner* »). Or c'est pour exemplifier *de façon à*, à propos duquel il est bien signalé qu'il se construit avec l'infinitif. C'est donc par inadvertance que cette variante apparaît dans le *Wiktionnaire*, au détour d'un exemple erronément associé à l'usage de *de façon à*. On considérera donc que ce subordonnant est ignoré, comme dans le *Petit Larousse illustré 2024*. Dans le corpus que nous exploitons, les ressources optent tantôt pour un avis ambivalent orienté négativement (*Dictionnaire de l'Académie française*, *Grand Robert*, *Multidictionnaire*), considérant que l'expression est lourde et déconseillée, même si on la trouve chez de bons auteurs ; tantôt pour un avis ambivalent orienté positivement (*Le bon usage*, *Usito*, Portail du Canada), qui signalent que l'emploi est parfois critiqué ou qu'il ne relève pas de la langue soignée ; tantôt enfin pour une acceptation totale qui présente *de façon à ce que* et *de façon que* comme deux variantes, sans marque d'usage ni remarque (*Petit Robert*, TLF, *Antidote*, OQLF).

Ainsi – hasard ou pas – l'absence de couverture descriptive constatée dans le *Wiktionnaire* à propos des trois phénomènes évoqués est alignée sur le traitement que propose le *Petit Larousse illustré 2024*.

3.2. Variante signalée en note

Cinq phénomènes parmi les 14 étudiés sont relégués en notes, et sont donc bien sûr considérés comme non normés dans le *Wiktionnaire*. En signalant ainsi une variante en note, le *Wiktionnaire* propose une hiérarchie très nette qui contrevient à l'objectif de description attendu.

Deux variantes actuellement en usage reçoivent un avis très négatif dans l'ensemble du panel comparatif. Il s'agit de *se rappeler de* + nom ou pronom et de *un espèce de* + nom masculin. Elles se situent probablement au début de la phase 2 sur le continuum d'évolution. *Un espèce de* + nom masculin présente le cas le plus clair. Aucune ressource n'accepte cette variante. Certains ouvrages (*Dictionnaire de l'Académie française*, *Petit Larousse illustré*, *Multidictionnaire*, *Usito*, TLF) ne la signalent même pas, tout en présentant expressément des exemples avec des noms au masculin : *une espèce de détective privé, d'avion* (*Petit Larousse illustré 2024*) ou en insistant sur le fait que « on dira *une espèce de*, même si le complément qui suit est masculin » (*Multidictionnaire*). Les autres ouvrages rejettent explicitement la forme *un espèce de*, en soulignant parfois qu'elle est répandue depuis longtemps (*Petit Robert 2024*, *Le bon usage*). Hormis le *Petit Robert*, qui insiste sur le fait que *un espèce de* est usité depuis longtemps et temporise en utilisant la marque *critiqué*¹⁰, les jugements normatifs des autres ressources sont forts : *fautif* (*Grand Robert*, qui dit du même souffle « familier et courant », désolidarisant ainsi fortement norme et usage), symbole signalant un mot « n'appartenant pas au

10. « REM. Bien que très répandu (depuis au moins le début du XVIII^e s.), l'emploi de *un espèce de* devant un nom masculin est critiqué : *un espèce d'idiot*. » (*Petit Robert*, *espèce*)

français général » (*Le bon usage*, 2016), « on ne dira donc pas » (OQLF), « et non » (*Antidote*, Portail du Canada). La note insérée par un collaborateur au *Wiktionnaire* semble minimiser la fréquence de l'emploi (*parfois accordé au complément qui le suit*), mais rejette indubitablement la forme en citant l'avis de Grevisse (dans son ouvrage *Le français correct*) et en signalant qu'il s'appuie sur l'Académie française. La note apparaît toutefois au sein même de l'article, juste après la forme *une espèce de*.

Au contraire, la note dans le *Wiktionnaire* sur la construction *se rappeler de quelque chose* est reléguée en fin d'article, ce qui revient à conférer un statut subalterne à la variante, à ne pas la considérer véritablement dans la description linguistique.

Par analogie avec la construction « se souvenir de quelque chose », certains disent « se rappeler *de* quelque chose » (emploi transitif indirect). Malgré un emploi bien attesté depuis le XVIII^e, l'Académie française considère que cet usage est fautif : en effet, *se rappeler* se construit d'après elle avec un complément d'objet direct. (*Wiktionnaire*¹¹)

Il faut dire que *se rappeler de quelque chose* fait l'objet de critiques nourries. À part le *Dictionnaire de l'Académie française*, qui ne la signale pas, mais rappelle les obligations linguistiques selon lesquelles « On dit *Se souvenir de quelque chose* mais on doit dire *Se rappeler quelque chose* », la grande majorité des ressources du panel comparatif rejettent explicitement la variante *se rappeler de*. C'est un *solécisme* (*Grand Robert*), elle est *familière et fautive* (*Petit Larousse illustré 2024*) ou encore *impropre* (*Antidote 11*). Certains soulignent qu'elle est toutefois fréquente dans la langue parlée (*Le bon usage*) et même très répandue (*Petit et Grand Robert*). Seul le TLF fait exception. Il s'agit en effet de la seule source acceptant les deux constructions sans marque d'usage.

Les trois autres variantes reléguées en note dans le *Wiktionnaire* reçoivent, dans le reste du corpus, des avis favorables. En plus de ne pas offrir une description de l'usage, le *Wiktionnaire* propose donc une vision très étriquée. La variante *avoir convenu que/de* est la plus emblématique. Elle n'est rejetée que par le *Dictionnaire de l'Académie française*, qui l'aborde dans une remarque au sein de l'article *convenir*. L'emploi y est dit *fautif*, devant être refusé tant à l'oral qu'à l'écrit : « on ne doit pas dire et moins encore écrire *nous avons convenu de*¹² ». La note dans le *Wiktionnaire* tente toutefois de respecter le principe de neutralité en montrant les différentes facettes de la réalité : si la règle traditionnelle est « peu respectée », y compris en littérature (pour preuve, une citation de Simenon est reproduite dans la note), l'emploi reste « incorrect selon les puristes ». En d'autres termes, la sanction demeure par le fait que l'emploi est relégué en note et par la présentation qui accorde la primauté à la règle traditionnelle¹³, malgré une tentative de produire une description qui laisse une place à un emploi différent. On relèvera surtout que presque toutes les autres ressources du corpus inscrivent l'emploi

11. https://fr.wiktionary.org/wiki/se_rappeler [page consultée le 26 août 2024]

12. <https://www.dictionnaire-academie.fr/article/A9C4082> [page consultée le 26 août 2024]

13. « La règle traditionnelle utilise l'auxiliaire *être* [...]. On dira donc [...] « Nous sommes convenus de ». « Nous avons convenu de » est incorrect selon les puristes. Cette règle est néanmoins peu respectée. » <https://fr.wiktionary.org/wiki/convenir> [page consultée le 4 juillet 2024]

en phase 3 de l'évolution linguistique. En effet, les deux constructions sont décrites comme normées et ne se différencient que par leur registre : *être convenu* est vieilli ou littéraire (*Petit et Grand Robert*), littéraire (*Usito*), soutenu (*Petit Larousse illustré*, *Antidote*, Portail du Canada), spécifique à la langue recherchée (*Multidictionnaire*). Enfin, *Le bon usage* oppose *avoir convenu* à l'emploi classique *être convenu*, sans même lui avoir accolé de marque. L'OQLF, pour sa part, ne propose aucune fiche sur le verbe *convenir*. Seul le TLF – ouvrage plus ancien dans le corpus étudié ici – est un peu plus frileux et signale que la règle traditionnelle « n'est pas toujours suivie, et l'usage tend à employer *avoir* comme seul auxiliaire ».

Autre emploi relégué en note dans le *Wiktionnaire*, *après que* + subjonctif y était encore présenté comme un *abus* en juillet 2024, et ce, dès le bloc entrée :

Après que \a.pʁɛ kə\ — **Note** : Suivi de l'indicatif. Se trouve souvent en pratique suivi du subjonctif par abus (voir ci-dessous).¹⁴

Cette entorse au principe de neutralité est désormais corrigée.

Après que \a.pʁɛ kə\ — *invariable* — **Note** : Traditionnellement de l'indicatif. Se trouve souvent en pratique suivi du subjonctif (voir ci-dessous).¹⁵

L'emploi du subjonctif n'est explicité qu'en note de fin d'article, sans exemplification toutefois. De plus, l'ouverture à la variation que contient la référence au *Bon usage* de 1993 et le constat d'un emploi en progression semble annulée par l'avis négatif de l'Académie française, qui clôt la note.

Malgré tout, l'usage du subjonctif s'est fortement répandu, et cette tendance est même devenue irrésistible selon la 13^e édition du *Bon Usage* (*Grevisse*) de 1993, qui précise cependant que l'usage de l'indicatif n'est nullement périmé. À cet égard, le 6 octobre 2011, l'Académie française rappelle l'usage de l'indicatif après cette locution.¹⁶

Environ la moitié des ressources consultées suivent cette voie du rejet. Le *Dictionnaire de l'Académie française* est clair : « *Après que*, suivi de l'indicatif ou du conditionnel, mais non du subjonctif¹⁷ ». L'OQLF et le Portail du Canada jouent de détours rhétoriques, mais critiquent le subjonctif. Quant aux deux ouvrages très attachés à la norme traditionnelle que sont le *Petit Larousse illustré* et le *Multidictionnaire*, ils rejettent la variante au point de ne même pas la signaler. Ils insistent, en creux, sur la nécessité de l'indicatif à la suite de *après que*. Les autres références du corpus sont parfois encore ambivalentes. Ainsi *Antidote* n'exemplifie-t-il que l'indicatif alors qu'il précise que « Le mode subjonctif après la locution *après que* est également accepté, mais il est parfois critiqué. » Le *Petit Robert* accorde également la marque *moderne* mais *critiqué*. La critique est affadée dans le *Grand*

14. https://fr.wiktionary.org/wiki/apr%C3%A8s_que [page consultée le 4 juillet 2024]

15. https://fr.wiktionary.org/wiki/apr%C3%A8s_que [page consultée le 15 septembre 2024]

16. https://fr.wiktionary.org/wiki/apr%C3%A8s_que [page consultée le 26 août 2024]

17. <https://www.dictionnaire-academie.fr/article/A9A2289> [page consultée le 26 août 2024]

Robert, où elle est reléguée en remarque. L'emploi du subjonctif *y* est bien consigné aux côtés de l'indicatif. Cette option transparaît aussi dans *Le bon usage* et *Usito*. Le TLF place, quant à lui, les deux variantes sur un pied d'égalité, ordonnant simplement la description en commençant par *après que* + indicatif.

La construction *sans que* + *ne* reçoit aussi une sanction négative dans le *Wiktionnaire* (« Cette locution n'appelle pas l'emploi de la négation. [...] Il conviendra donc de dire *Il vient sans qu'on l'en ait prié* et non *Il vient sans qu'on ne l'en ait prié*¹⁸), reproduisant, encore une fois, le choix du *Dictionnaire de l'Académie française*. Dans le corpus comparatif, seules deux références – d'ailleurs québécoises – adoptent aussi ce point de vue, à savoir l'OQLF et le *Multidictionnaire*. La variante est ignorée par le Portail du Canada, le *Petit Larousse illustré* et le *Petit Robert*, bien que ce dernier s'engage sur des questions semblables dans des situations syntaxiques plus spécifiques avec *ne*. Tous les autres ouvrages (TLF, *Grand Robert*, *Le bon usage*, *Antidote*, *Usito*) autorisent la variante, en ne voilant toutefois pas les critiques dont elle fait l'objet.

3.3. Variante présentée dans l'article

Dans cette section sont analysées les six variantes que le *Wiktionnaire* consigne au sein même de l'article plutôt que de les reléguer en note. Contrairement au choix que fait habituellement la lexicographie professionnelle, les séquences figées ne sont pas insérées dans l'article du mot noyau (*du coup* dans *coup*, *de manière à ce que* dans *manière*, etc.), mais font l'objet d'une entrée à part entière.

La séquence figée *des fois* fait l'objet d'une notice à part entière dans le *Wiktionnaire*, et n'est pas consignée au sein de l'article *fois*, montrant ainsi la grammaticalisation bien perçue par le grand public. Le groupe figé reçoit, dans le *Wiktionnaire*, la marque très dépréciative *populaire*. Celle-ci semble faire écho au rejet de l'Académie dans la 9^e édition de son dictionnaire et sur son blogue *Dire, Ne pas dire*, où il est affirmé sans détour : « Aux adverbes de temps *Parfois* et *Quelquefois*, on ne doit pas substituer la locution adverbiale *Des fois*.²⁰ » et à trois ouvrages (le *Petit Robert*, le TLF, *Le bon usage*) qui lui accolent également la mention *populaire*. Les autres références du corpus de comparaison ne sanctionnent pas aussi durement la locution. En effet, cinq ouvrages (*Grand Robert*, *Petit Larousse*, *Multidictionnaire*, *Antidote*, *Usito*) l'ont hissée au statut de *familier*. L'OQLF et le Portail linguistique du Canada ne mentionnent rien au sujet de *des fois*.

18. https://fr.wiktionary.org/wiki/sans_que [page consultée le 26 août 2024]

19. L'emploi du *ne* est autorisé en contexte négatif dans la principale (*Il ne vient jamais sans qu'on l'en ait prié / sans qu'on ne l'en ait prié*), ce qui n'est pas la situation envisagée ici.

20. <https://www.academie-francaise.fr/des-fois> [page consultée le 4 juillet 2024]

Du coup, refusé par le blogue *Dire, Ne pas dire* de l'Académie, est absent de son *Dictionnaire*, l'emploi étant peut-être un peu plus récent et oral. Cette fois, le *Wiktionnaire* s'affranchit de l'avis négatif de l'Académie, signale celui-ci en note, hisse la variante sémantique condamnée (sens « de ce fait ») au sein de l'article consacré à la locution et lui appose la marque *familier*, ce qui est un classement normatif moins dur. On trouve ce classement ailleurs aussi (*Petit Robert, Antidote, Usito*) tandis que d'autres ouvrages ignorent la variante en question (OQLF, Portail du Canada, TLF, *Le bon usage, Grand Robert*). Seul le *Petit Larousse illustré* n'appose aucune marque à *du coup*, « en conséquence (de quoi) » et l'intègre par conséquent pleinement à l'usage.

Voire même est un phénomène à part dans la liste de ceux qui ont initialement été retenus. En effet, il ne figure pas dans les condamnations du site *Dire, Ne pas dire* de l'Académie. La 9^e édition du *Dictionnaire de l'Académie française*, comme la 8^e édition déjà, accepte la séquence.

VOIRE. Adv. 1. Vraiment. Ne s'emploie plus guère qu'ironiquement, pour marquer le doute au sujet d'une assertion. « *C'est le plus grand écrivain de notre époque. – Voire.* »

2. Avec la valeur d'une conjonction de coordination, pour marquer un renchérissement. Et même. *Des opinions différentes, voire contraires. Il faudra des semaines, voire des mois, pour tout réparer. Cela n'arrive que rarement, voire jamais. Attendons que le soleil se couche, voire qu'il fasse nuit.* Renforcé par *Même. Ce remède est inutile, voire même néfaste.*²¹

Or ce tour a très mauvaise presse. À cet égard, le site *Langue française* du *Figaro* accepte *voire même*, bien qu'il lui faille user d'un tour de passe-passe sémantique : « lorsque *voire* a le sens de *vraiment* et non pas celui, plus moderne, de *même* »²². Toutefois, vu le « risque de provoquer l'ire de vos interlocuteurs », le site conseille d'utiliser *voire* seul ou de le remplacer par *et même*. Dans le corpus de ressources à l'étude, seul le *Petit Larousse illustré* ignore la variante *voire même* et consigne seulement *voire*. Si le *Petit* et le *Grand Robert* précisent « tour critiqué comme pléonasmie », le reste des ouvrages se rangent du côté de l'acceptation, franche dans le TLF et sur le Portail du Canada (« on peut donc employer *voire même* en toute quiétude »), plus posée dans *Antidote* et *Le bon usage*. Ce dernier résume bien les faits : « *voire même*, parfois condamné comme pléonastique, mais qui peut alléguer en sa faveur son ancienneté et l'approbation de l'Académie (depuis 1835) ainsi que celle de Littré » (*Le bon usage* 2016, par. 1094). Au contraire, trois ressources québécoises rejettent la séquence, c'est « un pléonasmie qu'il convient de remplacer » (OQLF), « l'expression *voire même* est redondante » (*Usito*), « critiquée » (*Multidictionnaire*). Face à ce portrait relativement ouvert à la variante *voire même* en Europe, il est étonnant de constater que le *Wiktionnaire* appose la marque *familier*, qu'on ne trouve nulle part ailleurs... si ce n'est dans les éditions du *Dictionnaire de l'Académie française* de 1835 et de 1878, mais plus dans celle de 1932-1935. Autre fait à relever, le *Wiktionnaire* propose deux articles, l'un consacré à *voire*, l'autre à *voire même*. Tous deux assignent à *voire même* la marque *familier*. Dans la

21. <https://www.dictionnaire-academie.fr/article/A9V1095> [page consultée le 30 novembre 2024]

22. <https://www.lefigaro.fr/langue-francaise/expressions-francaises/2018/08/09/37003-20180809ARTFIG00011-voire-meme-un-pleonasmie-a-bannir.php> [page consultée le 4 juillet 2024]

notice de *voire*, *voire même* ne se trouve pas dans le corps de la description, mais est relégué en note, lui conférant un statut moins acceptable, même s'il y reçoit le sceau du Littré et du *Dictionnaire de l'Académie*. Le *Wiktionnaire* inscrit donc une condamnation – quoique relativement légère – qui va à l'encontre des avis récents de l'Académie.

Le *Wiktionnaire* propose, pour les trois dernières variantes à l'étude ici (*malgré que*, *de manière à ce que*, *au final*), un traitement plus conforme au principe de neutralité, et surtout un affranchissement de l'Académie française. La notice consacrée à *malgré que* est intéressante de ce point de vue. En 2023, le *Wiktionnaire* accolait à ce mot la marque *familier* et remarquait en note que la locution conjonctive était « considérée comme à éviter par l'Académie française ». En 2024, la marque *familier* a disparu de la description. La note en fin d'article a également été révisée au point de prendre le contre-pied de la version précédente.

L'utilisation au sens de *bien que* [...] est considérée comme à éviter par l'Académie française.²³ (*Wiktionnaire*, *malgré que*, 24 mai 2023)

L'utilisation au sens de *bien que* est considérée comme à éviter dans un langage soutenu par des puristes tels que l'Académie française, l'Office québécois de la langue française, Littré et d'autres.²⁴ (*Wiktionnaire*, *malgré que*, 3 juillet 2024)

On peut déduire de la description de 2024 que *malgré que* serait d'un usage courant, même pour des puristes. De fait, l'OQLF – cité en 2024 – intègre explicitement *malgré que* dans la langue courante, tout en soulignant que ce lexème est absent du style soutenu.

La conjonction composée *malgré que* est suivie d'un verbe au subjonctif. Elle est souvent employée, dans la langue courante, pour exprimer la concession. Toutefois, dans un style soutenu, on utilise préférentiellement les expressions *bien que*, *encore que* et *quoique*.²⁵

On relève surtout que la notice *malgré que* du *Wiktionnaire* établit une dissociation explicite avec l'Académie (« des puristes tels que »), ce que l'on n'a vu nulle part ailleurs dans les situations analysées. En cela, le *Wiktionnaire* se rapproche de l'avis positif que partagent un certain nombre de ressources du panel comparatif, au rang desquelles on compte *Le bon usage*, le TLF, le *Grand Robert* et le Portail linguistique du Canada. Quant au *Multidictionnaire* et à *Antidote*, ils font les mêmes propositions stylistiques que l'OQLF, bien que ce ne soit pas pour les mêmes raisons. Le *Multidictionnaire* considère *malgré que* comme littéraire et vieilli, tandis qu'*Antidote* s'appuie sur les critiques dont la conjonction fait l'objet. Ces critiques sont également signalées par le *Petit Robert* et le *Petit Larousse illustré*, et par *Usito* dans une moindre mesure puisqu'il y est dit que « *malgré que* est parfois critiqué comme synonyme non standard de *bien que* », relativisant ainsi le poids de ces opinions négatives.

23. https://fr.wiktionary.org/wiki/malgr%C3%A9_que [page consultée le 24 mai 2023]

24. https://fr.wiktionary.org/wiki/malgr%C3%A9_que [page consultée le 3 juillet 2024]

25. <https://vitrlinguistique.oqlf.gouv.qc.ca/22935/la-syntaxe/les-conjonctions/emploi-de-la-conjonction-malgre-que> [page consultée le 26 août 2024]

On conclura que la marque *familier* proposée par le *Wiktionnaire* en 2023 n'est présente dans aucun des ouvrages que nous avons consultés. La description de l'usage que propose le *Wiktionnaire* est donc passée d'une marque de sanction modérée (*familier*), renseignant un usage sujet à caution, à une acceptation d'emploi.

De manière à ce que ne reçoit aucune marque d'usage dans l'article que lui consacre le *Wiktionnaire*. Après une définition sommaire, quatre citations au poids symbolique très positif (*Le Monde*, George Sand, Honoré de Balzac et Victor Hugo) exemplifient l'emploi du subordonnant. Une brève note en fin d'article signale que « Cette locution est critiquée et considérée comme incorrecte par certains grammairiens²⁶ ». Ici encore, les wiktionnaristes ont respecté le principe de neutralité, ont même temporisé les critiques et n'ont pas transposé l'avis de l'Académie, qui juge l'expression incorrecte dans son dictionnaire et la qualifie même de « étrange monstre [...] lourd et inutile [...] à éviter²⁷ ». Dans le panel d'ouvrages du corpus, la variante est encore parfois ignorée et seule est alors présentée la forme *de manière que*. C'est ce choix plus puriste que font *Le Petit Larousse illustré* et le *Multidictionnaire*. Les ouvrages qui signalent les critiques sans leur accorder trop de poids sont *Usito* et *Le bon usage*, d'autres endossent un moindre attrait envers *de manière à ce que* pour sa lourdeur (*Grand Robert*) ou son inadéquation avec un style soigné et soutenu (TLF, Portail linguistique du Canada). Enfin, trois références franchissent le pas de consigner la variante sans produire de signalement à son égard (*Petit Robert*, *Antidote*, OQLF).

La séquence figée *au final* reçoit, elle aussi, un traitement dans lequel le *Wiktionnaire* s'affranchit de l'Académie. Aucune marque d'usage n'est inscrite dans l'article, et les citations – contemporaines – proviennent d'essais scientifiques. Il faut attendre une note pour que soit signalée une référence au site *Dire, Ne pas dire* : « L'Académie française considère *au final* comme une "construction grammaticalement fautive"²⁸ ». C'est une référence au site *Dire, Ne pas dire*. Pour sa part, le *Dictionnaire de l'Académie* et aussi le TLF ignorent la variante, probablement trop récente pour avoir été consignée au moment de la rédaction des volumes concernés. Pourtant, des ouvrages même plus récents comme *Le Petit Larousse illustré 2024*, le *Multidictionnaire* et le Portail linguistique du Canada ne la consignent pas non plus. Les autres ressources du corpus critiquent presque unanimement *au final* (*Petit Robert*, *Grand Robert*, *Antidote*, *Usito*). Sa prégnance dans l'usage – notée par *Antidote*, l'OQLF et *Le bon usage* – n'est pas forcément synonyme d'acceptation. Dans le corpus, seules deux références accueillent plus favorablement l'emploi de *au final* : l'OQLF, mais il le « déconseill[e] en langue standard ou soutenue, notamment à l'écrit²⁹ », et surtout *Le bon usage*, qui n'abaisse aucunement « la locution *au final*, qui concurrence, depuis la fin du xx^e s., *finalement*, *en fin de compte*, *au total*, etc. » (*Le bon usage*, 2016, par. 1034g).

26. https://fr.wiktionary.org/wiki/de_mani%C3%A8re_%C3%A0_ce_que [page consultée le 26 août 2024]

27. <https://www.dictionnaire-academie.fr/article/DNP0059> [page consultée le 4 juillet 2024]

28. https://fr.wiktionary.org/wiki/au_final [page consultée le 4 juillet 2024]

29. <https://vitrlinguistique.oqlf.gouv.qc.ca/25429/le-vocabulaire/tours-dhorizon-et-autres-explications/usages/ladverbe-compose-au-final> [page consultée le 4 juillet 2024]

4. Conclusion

Nous avons retenu 14 phénomènes linguistiques sur la base du fait qu'ils sont en situation de tensions normatives plus ou moins fortes en français contemporain. Leur instabilité en faisait des objets linguistiques intéressants pour analyser l'attitude adoptée par la communauté de wiktionnaristes, œuvrant à un projet lexicographique non coordonné, et plutôt cumulatif (dans le sens où cette ressource est construite par couches successives de contributions tantôt se superposant et s'annulant, tantôt s'adjoignant les unes aux autres) que véritablement collaboratif.

Nous avons vu qu'un phénomène du corpus ne faisait pas partie de la couverture lexicographique proposée par le *Wiktionnaire* puisqu'aucune des deux variantes en concurrence (*demandeur que*, *demandeur à ce que*) n'était exposée. Deux autres variantes (*pour ne pas que*, *de façon à ce que*) sont absentes du *Wiktionnaire*, alors que leurs versions non condamnées sont consignées (*pour que*, *de façon que*). Pour les onze autres phénomènes du corpus, le *Wiktionnaire* aborde les variantes condamnées. Il est en accord avec les ouvrages de référence à propos de *un espèce de* et *se rappeler de*, que tous ou presque tous rejettent. Les phénomènes restants peuvent recevoir des avis d'acceptation au sein des autres ouvrages de référence alors que le *Wiktionnaire* choisit la ligne dure (*avoir convenu de*, *après que* + subjonctif, *sans que* + *ne*, *des fois*, *voire même*). Ce n'est que dans quatre cas (*du coup*, *malgré que*, *de manière à ce que*, *au final*) que le *Wiktionnaire* se fait plus ouvert à l'évolution de la norme. Ainsi, sur treize phénomènes consignés, deux variantes critiquées n'ont pas droit de cité et, pour les onze autres, le *Wiktionnaire* est réticent à l'évolution de la norme dans sept cas, soit dans la majorité des cas.

Il ressort du parcours d'analyse que l'approche descriptive recommandée dans les principes de rédaction n'est pas systématiquement mise en place, certains mots sont condamnés très rapidement et durement, et ils le sont sur l'avis de l'Académie française alors que d'autres voix s'élèvent pour un accueil plus favorable. On perçoit le poids et l'influence de cette institution sur l'imaginaire des linguistes non experts, qui reproduisent une vision plutôt fermée à l'évolution de la norme. Le tableau n'est cependant pas complètement sombre puisque, comme nous avons pu le constater, certaines variantes sont présentées de manière beaucoup plus favorable. L'Académie française peut alors être reléguée au rang des puristes jugés un peu trop pointilleux. C'est toutefois plus rare. Nous avons également pu constater que des modifications apportées aux articles (par exemple, *après que*) gommant certaines entorses à une rédaction descriptive. La lexicographie collaborative présente cet avantage d'une évolution rapide, mais malheureusement aléatoire et incontrôlable.

Cette étude ouvre la réflexion sur ce que l'on peut attendre d'une description lexicographique qui corresponde pleinement au principe de neutralité et à la vision descriptive prônée en linguistique, face à la vision prescriptive. À cet égard, on se questionnera sur l'interprétation à donner aux marques d'usage (*populaire*, *courant*, *soutenu*, etc.). Celles-ci ont initialement vocation à être descriptives, du moins pour le lexicographe qui les met en place. Il faut toutefois prendre conscience que les caté-

gories les plus utilisées dans le marquage s'inscrivent dans une hiérarchie – souvent voilée et peu avouée – qui organise la variation diastratique (notamment avec la marque *populaire*) et diaphasique (avec les marques *familier*, *courant*, *soutenu*). Par exemple, un lexème qui reçoit la marque *populaire* est issu de contextes d'emploi spécifiques. Lorsque le lexicographe appose une telle marque, il décrit l'espace d'emploi d'une forme linguistique. Cette description est cependant teintée d'une perception des usages.

De fait, il arrive que des ouvrages de référence proposent des marques différentes pour un même lexème (par exemple, *populaire* dans un ouvrage, *familier* dans un autre). C'est d'abord le signe d'une évolution linguistique en cours, mais c'est aussi un indice de positionnement normatif, purisme ou ouverture – discrète ou franche – face à l'évolution linguistique. Les lexicographes non professionnels que sont les wiktionnaristes n'abordent certainement pas les marques d'usage avec la même perspective que les professionnels. En revanche, la présente étude a permis de constater qu'ils n'innovent pas sur ce plan et reproduisent les marques issues des ouvrages de référence en place. Ils se rangent également la plupart du temps dans une attitude normative plutôt fermée à l'évolution de la norme.

Au-delà de l'usage du marquage en *production* lexicographique (où le marquage est avant tout un outil descriptif exploité par le lexicographe), il faut s'intéresser à la *réception* de ce marquage. L'utilisateur du dictionnaire cherche des réponses à des questions très pratiques en matière d'orthographe, de sens et de validation d'emploi. Dès lors, les marques d'usage sont, pour lui, des autorisations ou des interdictions. À cet égard, le *Wiktionnaire* joue souvent le rôle de gardien d'une norme plutôt traditionnelle.

Bibliographie

Académie française (1986-2024). *Dictionnaire de l'Académie française*, 9^e édition.

Antidote 11 (2021). Montréal, Druide informatique.

GREVISSE, Maurice et André Goosse, (2016) [1986], *Le bon usage*, Louvain-La-Neuve, De Boeck Supérieur.

Le Grand Robert de la langue française (2023), En ligne, Paris, Éditions Le Robert.

Le Nouveau Petit Robert 2024 : dictionnaire alphabétique et analogique de la langue française (2023), Édition numérique, Paris, Éditions Le Robert.

Le Petit Larousse illustré 2024, (2023), Paris, Éditions Larousse.

[OQLF] Office québécois de la langue française, en ligne, *Vitrine linguistique*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/>

Ressources du Portail linguistique du Canada, en ligne, <https://www.noslangues-ourlangues.gc.ca/fr/index>

REY, Christophe (2008), *Dictionnaire et société*. Paris : Honoré Champion.

SAJOURS, Franck (2023), « Quantité et qualité dans le *Wiktionnaire* : de la diversité... à la rigueur ? », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86, (2023-1), URL : <http://journals.openedition.org/linx/9835> ; DOI : <https://doi.org/10.4000/linx.9835>.

SAJOURS, Franck, Nabil Hathout et Amélie Josselin-Leray (2019), « Du vin et devin dans le *Wiktionnaire* : neutralité de point de vue ou neutralité et point de vue ? », *Études de linguistique appliquée*, 194(2), 147–164.

VILLERS, Marie-Éva de, (2021), *Multidictionnaire de la langue française*, 7^e édition, Montréal, Québec/Amérique.

VINCENT, Nadine (2019a), « Qu'est-ce qu'un dictionnaire ? Perspectives de la lexicographie au 21^e siècle », dans Anne Dister et Sophie Piron (dir.) *Les discours de référence sur la langue française*, Presses de l'Université St-Louis, Bruxelles (Belgique), p. 127-148.

VINCENT, Nadine (2019b), « La lexicographie profane dans la presse écrite », dans Wim Remysen et Sabine Schwarze (dir.) *Idéologies sur la langue et médias écrits : le cas du français et de l'italien/ Ideologie linguistica e media scitti : i casi francese e italiano*, Frankfurt, Allemagne, Éditions Peter Lang, p. 37-57.

Wiktionnaire, dictionnaire libre, en ligne, https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d'accueil

TITRE: LA RUBRIQUE *DIRE, NE PAS DIRE* DU DICTIONNAIRE NUMÉRIQUE DE L'ACADÉMIE FRANÇAISE. UN CHANGEMENT DE PARADIGME ?

AUTEUR: CAROLINE SALAGNAC, UNIVERSITÉ DE MONTPELLIER PAUL VALÉRY III, PRAXILING

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 206-238

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22413](http://hdl.handle.net/11143/22413)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22413](https://doi.org/10.17118/11143/22413)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

La rubrique *Dire, Ne pas dire* du dictionnaire numérique de l'Académie française. Un changement de paradigme ?

Caroline Salagnac, Université de Montpellier Paul Valéry III, Praxiling

Résumé :

La possibilité désormais offerte aux lexicographes de bénéficier d'un territoire numérique pour partager leur travail est l'occasion de s'interroger sur la nature du discours tenu sur les usages de la langue, dès lors que l'on s'éloigne de l'étude du dictionnaire. La rubrique *Dire, Ne pas dire* du dictionnaire numérique de l'Académie française est à ce titre un terrain d'étude privilégié, tant l'Académie française porte le rôle de gardien de la langue. La question qui se pose ici est la suivante : en quoi l'Académie française s'est-elle servie du numérique pour élargir sa mission qui transcende désormais les seules productions langagières et leurs normes et en quoi cette rubrique y participe-t-elle ? Que dit cette rubrique de l'usage et de la manière dont il y est convoqué ? Certes, l'Académie se pose en locutrice experte, gardienne du bon usage, dans un rôle attendu. Mais le discours normatif laisse place et la place à un autre type de discours, plus préventif augurant d'un changement de paradigme, qui ne prend plus seulement en compte la langue et ses usages mais également l'utilisateur.

Mots-clés :

Académie française, rubrique Dire, Ne pas dire, discours normatif, discours prescriptif, lexiculture

Abstract :

The fact that lexicographers now have access to a digital territory in which to share their work provides an opportunity to examine the nature of the discourse on language usage, once we move away from the study of the dictionary. The *Dire, Ne pas dire* section of the Académie française's digital dictionary is an ideal place to look, given the Académie française's role as guardian of the language. The question that arises here is: how has the Académie française used digital technology to extend its mission, which now transcends just language production and its norms, and how does this section contribute to this? What does this section say about usage and the way it is used? Admittedly, the Académie is the expert speaker, the guardian of correct usage, in an expected role. But the normative discourse is giving way to another type of discourse, one that is more preventive, heralding a paradigm shift that takes into account not only the language and its uses, but also the user.

Keywords :

Académie française, section Dire, Ne pas dire, prescriptive discours, preventive discourse, lexiculture

1. Introduction

Il serait faux d'imaginer que le passage au numérique, au demeurant très récent, du *Dictionnaire de l'Académie française* ne constitue qu'une proposition de mise à disposition d'un outil, fût-il moderne et innovant. Certes, comme l'indique la présentation du portail numérique¹, la consultation dynamique de toutes les éditions du dictionnaire devient un objectif, aujourd'hui atteint. Certes, les ressources linguistiques se multiplient et la neuvième édition a pu bénéficier de beaucoup d'avancées techniques (Catach, 2020). Mais plus intéressante d'un point de vue culturel et idéologique est la manière dont la quête du bon usage est réaffirmée grâce à la multiplication de rubriques numériques spécialisées. C'est à l'une d'entre elles que nous nous intéressons ici : la rubrique *Dire, Ne pas dire*².

Revenant sur le projet de numérisation du dictionnaire, Laurent Catach rappelle que « L'Académie souhaitait depuis longtemps donner davantage de visibilité à son travail de réalisation du Dictionnaire » (2020 : 337). Le paradoxe entre une diffusion très limitée du dictionnaire en édition imprimée, l'accès matériel limité à l'œuvre et la mission de « service public » pour reprendre les termes de Catach ont nourri la préoccupation d'une diffusion plus large du dictionnaire. Dès 2004, les deux premiers tomes de la neuvième édition avaient été mis à disposition du public via un partenariat avec le laboratoire INALF du CNRS. En 2014 l'Académie a entrepris une refonte de son site Internet. Catach souligne que dans le domaine des dictionnaires, le support numérique est une véritable révolution et constitue une remise en question de l'ensemble des paradigmes : macrostructure (nomenclature), microstructure, présentation des informations, navigation, mises à jour des contenus, documentation, liens vers d'autres ressources. La nouvelle édition numérique du *Dictionnaire de l'Académie française* intègre en effet des liens vers d'autres ressources lexicographiques. Ces ressources sont de deux natures : terminologique (liens hypertextes vers le site *FranceTerme* de la Délégation générale à la langue française et aux langues de France (DGLFLF) qui regroupe les termes recommandés par la Commission d'enrichissement de la langue française) et francophone (liens hypertextes depuis le Dictionnaire vers la *Base de données lexicographiques panfrancophone* (BDLP), qui est une importante base lexicale inventariant les variétés du français dans de nombreuses régions et pays francophones), à quoi s'ajoute une ressource particulière sur les difficultés du français : la rubrique *Dire, Ne pas dire*. Il s'agit d'un ensemble de notices publiées par l'Académie française depuis 2010. Elles proposent aux lecteurs des éclaircissements sur les difficultés courantes de la langue française ou des « curiosités ». Organisées en quatre catégories : *Emplois fautifs*, *Extensions de sens abusives*, *Néologismes & anglicismes*, *Bonheurs & surprises*³, elles se trouvent reliées à environ 1100 mots du dictionnaire par des liens hypertextes qui fonctionnent également dans le sens inverse permettant alors de circuler du dictionnaire vers les notices.

1. Texte de présentation du portail numérique de l'Académie française en date du 2 février 2019, <https://www.academie-francaise.fr/actualites/dictionnaire-de-lacademie-francaise-en-ligne> [consulté octobre 2022]

2. Désormais les articles concernés du Dictionnaire incluent des renvois à cette rubrique.

3. <https://www.academie-francaise.fr/dire-ne-pas-dire> [consulté en octobre 2022]

Dans le cadre du présent article et de l'interrogation autour des dictionnaires numériques comme objets porteurs de culture et d'idéologies, nous nous poserons la question suivante : en quoi l'Académie française s'est-elle servie du numérique pour élargir sa mission qui transcende désormais les seules productions langagières et leurs normes et en quoi cette rubrique y participe-t-elle ? Notre problématique est donc la suivante : que dit cette rubrique de l'usage et de la manière dont il y est convoqué ? Est-ce seulement l'usage « simplement et suprêmement » comme l'annonce Maurice Druon dans les mots issus du discours de la préface de la neuvième édition du dictionnaire⁴ ? Cette rubrique n'est-elle qu'une prolongation du dictionnaire ou l'espace numérique offert à côté du dictionnaire constitue-t-il autre chose dans l'esprit des académiciens ? Pour répondre à ces questions, nous commencerons par étudier dans quel cas l'Académie se pose en locutrice experte, gardienne du bon usage, dans un rôle attendu. Puis nous verrons que, comme elle l'indique elle-même, elle est loin de n'être que cela et nous montrerons que certaines notices de la rubrique révèlent un « sentiment de la langue » pour reprendre la formule de Gilles Siouffi (2007a). Nous verrons également que cette rubrique est l'occasion donnée à l'Académie d'investir un territoire plus grand que celui du dictionnaire pour partager des éléments de lexiculture. Enfin, puisque c'est une spécificité d'une rubrique internet aujourd'hui, nous nous poserons la question de la relation au public dans le cadre du « courrier des internautes ». En réalité, il est ici question à la fois de garder en ligne de mire le dictionnaire chaque fois que cela est possible, mais également d'opter pour une étude transversale de la rubrique.

2. Quelques mots sur la macrostructure de la rubrique et le corpus choisi

2.1. Macrostructure de la rubrique

La rubrique *Dire, Ne pas dire* est adossée au site institutionnel de l'Académie. Elle s'inscrit dans un environnement numérique autour de la langue française dont la lexicographie et le dictionnaire ne constituent qu'une partie. Il va sans dire que le titre de la rubrique annonce déjà le programme : *Dire, Ne pas dire* s'affirme comme une injonction à laquelle il semble recommandé de se soumettre et est loin du choix opéré par Larousse qui sur son site parle plutôt de *difficulté*⁵.

4. <https://www.academie-francaise.fr/le-dictionnaire-les-neuf-prefaces/preface-la-neuvieme-edition> [consulté en octobre 2022] – « *Le Dictionnaire de l'Académie est celui de l'usage, simplement et suprêmement, le Dictionnaire du bon usage, qui par-là sert, ou devrait servir de référence à tous les autres. Telle est l'ambition, mesurée mais persévérante, qui guide les académiciens français* ».

5. <https://www.larousse.fr/dictionnaires/francais-monolingue> (page d'accueil du site Larousse) – [consulté en octobre 2022]



Figure 1. Page d'accueil de la rubrique *Dire, Ne pas dire* sur le site de l'Académie française.⁶

On sait à quel point les titres sont importants et représentent des marques identitaires fortes et qu'ils disent par conséquent beaucoup de l'intention et de la volonté de produire un effet.

L'arborescence générale est la suivante : la rubrique *Dire, Ne pas dire* fait partie d'un grand ensemble nommé « la langue française ». Le Dictionnaire bénéficie lui d'un espace séparé. Le grand ensemble se décompose en rubriques :

- *Le français aujourd'hui*
- *Questions de langue*
- *Terminologie*
- *Dire, Ne pas dire*
- *Contact Dire, Ne pas dire*
- *Actualité*

La rubrique *Questions de langue*⁷ apporte son lot de confusions, car elle est aussi dévolue à des questions précises sur la langue et synthétise par ordre alphabétique un ensemble de lexies problématiques. En cliquant sur les liens proposés par cette liste, l'internaute peut déployer un texte explicatif.

Le logo de la rubrique *Dire, Ne pas dire*, ci-dessous reproduit, va dans le même sens mettant l'accent graphique sur la partie « Ne pas dire » dont la locution adverbiale négative NE PAS apparaît sur un fond rouge.

6. <https://www.academie-francaise.fr/dire-ne-pas-dire> [consulté en octobre 2022]

7. <https://www.academie-francaise.fr/questions-de-langue> [consulté en janvier 2023]



Figure 2. Le logo de la rubrique *Dire, Ne pas dire*⁸

L'empan de la rubrique *Dire, Ne pas dire* se décline ensuite en huit notices qui sont autant de programmes et en droite ligne avec un pacte prescriptif, au moins en apparence, comme nous le verrons.

1. *Accueil*
2. *Emplois fautifs*
3. *Extensions de sens abusives*
4. *Néologismes & anglicismes*
5. *Bonheurs & surprises*
6. *Bloc-notes*
7. *Courrier des internautes*
8. *Dire, Ne pas dire*

Nous nous concentrerons dans le cadre du présent article sur les rubriques 2 à 5 et 7, la notice *Accueil* reprenant synthétiquement l'ensemble des autres et les notices 6 et 8, offrant soit la possibilité d'envoyer un message et de poser une question au service du dictionnaire de l'Académie française, soit celle de lire des remarques normatives qui n'auraient pas fait l'objet d'un classement dans les rubriques 2 à 5.

En réalité, la profusion d'informations accessibles à des endroits différents, par des chemins variés sans explication préalable peut parfois apparaître comme étant de nature à entretenir une forme de confusion.

8. <https://www.academie-francaise.fr/dire-ne-pas-dire> [consulté en octobre 2022]

Il est à noter que *Dire, Ne pas dire* fait également l'objet d'une publication papier pour « confier à la mémoire du papier » (Rey, 2023)⁹ la langue française complétant ainsi celle des « nuages » depuis 2014¹⁰. Enfin, il est possible d'accéder à la rubrique de manière autonome à partir de la page d'accueil du site internet de l'Académie française¹¹.

2.2. Élaboration du corpus d'étude

Avant d'entrer dans le vif du sujet, quelques mots sur la méthodologie et le corpus qui servira de base à la présente étude : nous avons sélectionné un corpus très restreint correspondant aux mois de janvier à avril 2023 – au plus près de la date du colloque à l'origine de cette étude (soit 4 mois seulement faute de temps pour une analyse d'un corpus plus étendu). Et nous extrayons ici les faits qui nous semblent les plus saillants pour chacune des notices, bien consciente que les conclusions auxquelles nous aboutissons mériteraient d'être confrontées à un corpus plus large et ne peuvent constituer une position définitive sur la question.

La rubrique *Dire, Ne pas dire* permet d'effectuer une recherche élargie par titre, rubrique (chacune des notices) ou date de parution (par mois) qui permet de sélectionner un certain nombre d'entrées. C'est en utilisant la recherche élargie que le corpus a été sélectionné en choisissant pour chacune des notices les mois de janvier à avril 2023¹².

3. L'expert de la langue : un rôle attendu de gardien de la norme

Si, comme nous l'évoquions au début de cet article, « Le *Dictionnaire* de l'Académie est celui de l'usage, simplement et suprêmement¹³ », on peut s'attendre à ce que la rubrique *Dire, Ne pas dire* soit l'ambassadrice de cette ambition. Et c'est en effet ce qui se produit. En étudiant la manière dont l'Académie parle de la norme et des entorses faites à la langue, nous allons voir que, contrairement à ce qui se produit dans le dictionnaire, où l'exemple lexicographique sert de point d'ancrage à la mise en usage d'un mot, le choix est fait ici d'un discours explicatif (notamment grammairien) autour du bon usage.

9. *Dire, Ne pas dire - du bon usage de la langue française*, Paris, Philippe Rey, dernière édition de l'intégrale augmentée mars 2023

10. En juin 2023, l'institution présentait la version papier de la dernière édition augmentée.

11. <https://www.academie-francaise.fr/> [consulté en octobre 2023]

12. <https://www.academie-francaise.fr/actualites/dire-ne-pas-dire-janvier-2023#:~:text=La%20s%C3%A9rie%20du%20mois%20de,le%20fran%C3%A7ais%20contemporain%2C%20est%20accessible>. Exemple du moteur de recherche élargie.

13. <https://www.academie-francaise.fr/le-dictionnaire-les-neuf-prefaces/preface-la-neuvieme-edition> [consulté en mai 2023]

3.1. Un discours d'expert fait d'impératifs : norme et registre spécialisé

C'est assez naturellement la notice *Emplois fautifs* sur laquelle s'adosse principalement le discours d'expert. La liste des lexies ou locutions ayant fait l'objet de commentaires dans la notice *emplois fautifs* est la suivante :

- janvier 2023¹⁴
 - Absence de liaison après *quand*
 - *Croire à, Croire en*
 - *Sure* ou *Sûre*
- février 2023¹⁵
 - *Des pâtisseries faites maison* ou *faites maisons*
 - *Elle s'est forgé un corps d'athlète*
 - *Ils le harcèlent* ou *Ils l'harcèlent*
- mars 2023¹⁶
 - *Nous soussignés* ou *Nous soussigné*
 - *Détoxer*
 - *Les conditions sont glissantes*
 - Omission de l'article : exemples de *garder contact* et *sur base de*
- avril 2023¹⁷
 - *Malaisant*
 - *Sourcer*
 - *S'empiffrer* ou *S'empiffrer, piffer, ou piffrer?*

14. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-01-05&form_build_id=form--swbhLa8zop_kSoC9N_EQM6ZnmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher – [consulté en mai 2023]

15. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-02-02&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZnmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher – [consulté en mai 2023]

16. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-03-02&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZnmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher – [consulté en mai 2023]

17. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-04-06&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZnmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher [consulté en mai 2023]

Ce qui nous intéresse ici ce n'est pas tant le choix des occurrences que la manière dont le discours est tenu pour chacun des emplois fautifs retenus.

Ce que l'on peut remarquer, c'est que le discours accompagnant chacun des exemples de la notice *emplois fautifs* se veut argumenté. À titre d'exemple, lorsqu'il s'agit d'emplois fautifs liés aux phénomènes de prononciation, comme sur la question de liaison, la correction est annoncée au sein d'un texte qui le plus souvent reprend l'histoire du phénomène dans une démarche diachronique plus ou moins élaborée. C'est le cas pour l'absence de liaison après *quand*.

En un siècle et demi, nous avons perdu quelques liaisons, en particulier celles de plusieurs mots se terminant par *-nd*. Dans son *Dictionnaire*, Littré écrivait en effet que le « d » final de *fécond* et de *profond* se lie, prononcé comme un « t », au mot suivant et que l'on doit dire *un fécon-t-écrivain* et *un profon-t-archéologue*. Il nous apprend aussi que le « d » de *blond* ne se lie que dans la prononciation soutenue « Le blond Apollon, dites : le blon-t-Apollon », et qu'il en va de même pour *brigand* : « Le d ne se lie pas dans le parler ordinaire; dans le parler soutenu on dit : un brigant-t armé ». Aujourd'hui, on conserve la liaison en « t », après *fond*, dans l'expression *de fond en comble*, après *grand* et après *quand*. Mais on constate un début de relâchement avec ce dernier mot et l'on entend de plus en plus « quand / il viendra », « quand / on est partis », alors que c'est « quand-t-il viendra » et « quand-t-on est partis » que l'on doit dire.¹⁸

On observera que le texte explicatif, s'il prend acte de l'évolution de la langue et du phénomène, ne se prive pas de mettre en avant le fait que certaines nouvelles omissions de liaisons sont des fautes, sur lesquelles il est important pour l'Académie de revenir sous peine que le phénomène ne prenne de l'ampleur. L'attention est donc de mise pour éviter le *relâchement* du locuteur ! L'argument axiologique à valeur morale prend ici tout son sens attestant du fait que pour les académiciens, l'évolution de la langue et de ses usages est un long chemin qu'il faut savoir orienter et maîtriser (Carrère d'Encausse, 2017 : 13¹⁹).

En revanche lorsqu'il s'agit d'orthographe, par exemple *Sure* ou *Sûre* ou *Nous soussignés* ou *Nous soussigné*, les tours font l'objet d'abord d'un texte explicatif historique puis d'un tableau sous la forme de 2 colonnes exprimant de manière très explicite le bon usage et le mauvais. La formulation utilisée est soit « on dit » - « on ne dit pas » pour des formulations que l'on utilise à l'oral ou à l'écrit, soit « on écrit » - « on n'écrit pas » pour des formulations que l'on ne retrouve qu'à l'écrit.

18. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-01-05&form_build_id=form-Qp357y0rEg-oJ7eLpIGEhtuBoiyYy5B5_SF1u8RITpo&form_id=academie_blog_search_form&op=Rechercher [consulté en mai 2023]

19. « L'Académie française a, depuis sa fondation, vocation non seulement à être « le greffier de l'usage », mais aussi à orienter cet usage quand il est hésitant et à le corriger quand il est fautif »

Malaisant

Le 6 avril 2023

Emplois fautifs

On n'emploie plus aujourd'hui en français l'adjectif verbal *aisant*, même si le verbe *aisier* s'est rencontré au Moyen Âge et si le participe passé *aisé* est, lui, toujours en usage et a servi à créer l'adjectif antonyme *malaisé*. Pour qualifier ce qui produit un malaise, ce qui gêne, on évitera de recourir à l'adjectif *malaisant*, que l'on entend aujourd'hui assez fréquemment, puisque de nombreux adjectifs ou locutions, bien ancrés dans la langue, peuvent déjà traduire cette idée.

on dit	on ne dit pas
<i>Un silence gênant, embarrassant</i>	<i>Un silence malaisant</i>
<i>Cet homme met tout le monde mal à l'aise</i>	<i>Cet homme est très malaisant</i>

Figure 3. Formulation des académiciens pour corriger l'emploi fautif *malaisant*²⁰

on écrit	on n'écrit pas
<i>Nous soussignées, Valérie et Sophie, déclarons...</i>	<i>Nous, soussignées Valérie et Sophie, déclarons...</i>
<i>Je soussigné, Maître Cornille, meunier...</i>	<i>Je, soussigné Maître Cornille, meunier...</i>

Figure 4. Formulation des académiciens pour corriger l'emploi fautif *Nous soussignées*²¹, soulignant non seulement le nécessaire accord du participe passé en genre et en nombre avec les signataires mais également le placement de la virgule après le participe soussigné et non avant

L'idéal puriste et la mise en évidence du tour valable sont ceux exprimés en premier à l'inverse de ce que font tendanciellement les locuteurs qui, comme le rappelle Sonia Branca-Rosoff, vont de la forme déviante à la forme correcte (2007 : 32). La force de l'acte illocutoire traduit par *on écrit* versus *on n'écrit pas* montre à quel point le discours est énoncé avec un sens destiné à produire un effet. La posture de l'observateur impartial mais doté de connaissances est de rigueur. Comme le rappelle Cécile Canut : « ces discours incluent des jugements d'acceptabilité, traduits par les tours "on dit" - "on ne dit pas" sur la formule à adopter et se signalent d'un côté par leur réflexivité et de l'autre par leur dé-subjectivation » (2007 : 49).

20. Recherche | Académie française https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-04-06&form_build_id=form-8kNB6wYQV-kQrFk9ZVscFzBI4WiY8tO77nEDH593ltI&form_id=academie_blog_search_form&op=Rechercher [consulté en mai 2023]

21. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-02-02&form_build_id=form-KqB8vPQOrUqKA6ogE61-UDtLAZ0t_q-whhke7OQA08Y&form_id=academie_blog_search_form&op=Rechercher [consulté en mai 2023]

Puisqu'il s'agit d'être direct, aux fins de faire adopter la norme, l'implicite dispose d'une place réduite, contrairement à ce qui se produit dans l'espace lexicographique. Marc Picone rappelle à cet égard que :

Le contenu sémantique généré par l'inférence est le résultat de l'interprétation de données non formellement exprimées. Par conséquent, celui qui consulte un dictionnaire peut, lorsqu'il fait appel à la déduction pour parvenir à dégager des informations implicites, se méprendre sur la véracité de ces déductions. (2011 : 188)

La démarche est d'orienter le lecteur/internaute vers un usage prescrit, en l'occurrence ici une démarche morphosyntaxique acceptable et acceptée, le discours accompagnant ayant une obligation de résultat, celui d'aller vers le bon usage et donc de réduire au maximum la part autorisée à la déduction. Il s'agit ici de dire *comment (bien) dire X*. L'ambiguïté réside cependant dans le fait que, alors que l'on peut considérer être dans une démarche qui vise un résultat, l'Académie recourt à un lexique spécialisé pour décrire les règles de grammaire, de prononciation ou tours valables. En témoigne cet exemple relatif à l'emploi fautif *faites maison* ou *faites maisons* et le commentaire de l'Académie :

La locution adjectivale *fait maison* s'emploie pour qualifier ce qui est fabriqué par l'artisan qui en fait commerce, et diffère donc de ce qui est produit de façon industrielle. Cette locution est **une ellipse** de *fait à la maison*. Il s'ensuit que le nom *maison* reste invariable tandis que le participe passé *fait* s'accorde, en genre et en nombre, avec le nom du produit concerné. On dira et on écrira donc *des pâtisseries faites maison* (et non *des pâtisseries faites maisons* ou *des pâtisseries fait maison*), comme on dira, encore **par ellipse**, *des pâtisseries maison*. Il en va de même pour la locution *fait main* : *des pulls faits main*.²²

Les académiciens font appel dans cet exemple à un registre/lexique particulier (grammaire, lexicologie, rhétorique) qui peut sélectionner les récepteurs du message en fonction de leur aptitude à le décoder. Si chaque internaute lisant l'explication peut se rappeler les règles d'accord en genre et en nombre du participe passé, les notions de *locution adjectivale* ou encore d'*ellipse* relèvent davantage de la terminologie grammaticale (que les internautes, cibles du discours, peuvent avoir oubliée ou ne pas connaître une fois sortis de la sphère scolaire où elle est enseignée). En tout état de cause, la description métalinguistique vise à décrire un système de la langue et sert à justifier le modèle de référence à adopter. Cette stratégie discursive d'un locuteur spécialisé (pas seulement par sa légitimité à dire la norme mais aussi par le registre discursif de grammairien qu'il utilise) peut obérer les chances de réussite non seulement du décodage du message mais aussi de l'adhésion du discours de la part du lecteur/internaute. Quoi qu'il en soit, il est possible de se questionner sur le choix paradoxal fait ici entre l'obligation de résultat (le bon usage) et la nature des ressources linguistiques mises en œuvre pour y parvenir (le registre spécialisé) dès lors que le discours n'est pas à destination de spécialistes mais d'une audience supposée être plus large.

22. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-02-02&form_build_id=form-XP_IyOH6C86QmIoN9FmF3Fu_zHgLIU5pXMSgWR_bGiQ&form_id=academie_blog_search_form&op=Rechercher [consulté en mai 2023]

Pour autant, c'est parce que le discours est orienté vers un résultat que l'on peut le qualifier de discours d'apprentissage selon une organisation discursive déployée autour de la mise en discours de la règle à suivre, de l'usage et de l'exemple. Le discours produit, pour chacune des lexies ou locutions choisies dans la rubrique, un changement : celui qui conduit l'internaute d'un mauvais usage à un bon. Le discours tenu sur les lexies mises en avant dans cette rubrique est enfin un discours complémentaire à celui du dictionnaire, qui n'entre pas en concurrence directe avec la définition lexicographique quand bien même on revient à la définition et au partage de sens.

La stabilisation d'un état de la langue et de l'usage est ici envisagée par le biais d'un « prêt à parler » (Collinot et Mazière, 1997 : 5) fondé sur des règles de grammaire qui ne souffrent pas de fluctuations et qui précisément parce qu'elles sont immuables ou appréhendées comme telles, peuvent représenter une référence à suivre. L'usage n'est pas ici abordé comme un élément possiblement évolutif mais bien au contraire comme un élément stabilisateur dont la vertu est de « combattre le caprice de la mode » et de faire acte d'utilité dans l'intérêt des locuteurs qui ont en partage la langue française (Carrère d'Encausse, 2017 : 5). Et il est évident qu'une notice comme celle qui traite des emplois fautifs apparaît comme un espace supplémentaire offert à l'Académie française pour exprimer cette ambition.

3.2. Un discours de prévention fait d'explications

À côté du discours d'expert, on trouve notamment dans la notice *Extensions de sens abusives* qui concerne particulièrement les confusions sémantiques un discours moins normatif. Nous présentons ci-dessous les lexies qui ont été retenues pour la période de janvier à avril 2023 :

- janvier 2023²³
 - *Vigie, Vigile*
 - « *Mondial* » pour « *International* »
- février 2023²⁴
 - « *Méridional* » pour « *Méridienne* »
 - « *Vexatoire* » pour « *Vexant* »

23. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-01-05&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZNmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

24. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-02-02&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZNmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

- mars 2023²⁵
 - « *Génance* » pour « *Gêne* »
 - « *Louanger* » pour « *Louer* »
- avril 2023²⁶
 - *De belle facture*
 - « *Belle journée* » pour « *Bonne journée* »

Deux cas principaux apparaissent :

- L'Académie met en concurrence les deux occurrences problématiques (par exemple *vexatoire* pour *vexant* ou *mondial* pour *international*)
- L'Académie met l'accent sur une formulation problématique (par exemple *de belle facture*).

Comme pour la notice précédente, on retrouve dans la majorité des cas un texte explicatif. Compte tenu du domaine de cette notice (le partage de sens), cette fois en concurrence directe avec celui du dictionnaire, il n'est pas rare de trouver une référence à la définition lexicographique. C'est le cas par exemple pour *mondial* et *international* :

Ces deux adjectifs n'ont pas exactement le même sens; il convient donc de les employer avec justesse. *International* **signifie « qui a lieu de nation à nation, entre plusieurs nations, qui concerne les rapports entre les nations »**, tandis que *mondial* **a pour sens « qui intéresse, concerne le monde entier qui est répandu dans l'ensemble du monde »** et se rapporte donc à un nombre beaucoup plus important de nations.²⁷

Les termes *signifier* et *avoir pour sens* relèvent directement de la sphère lexicographique et ramènent à un objectif traditionnellement dévolu au domaine du dictionnaire.

Éléments nouveaux et en droite ligne avec la neuvième édition du dictionnaire, on trouve également des références à l'étymologie comme c'est le cas dans *méridionale* pour *méridienne* et là encore la frontière est poreuse avec le domaine lexicographique.

Les adjectifs et noms *méridional*, *méridien* et *méridienne* sont parents et remontent tous à l'adjectif latin *meridianus*, « de midi », dérivé de *meridies*, « midi », lui-même composé à l'aide de *medius*, « moyen », du milieu, central », et *dies*, « jour ». On appelait d'ailleurs au Moyen Âge « diable meridien » (*daemonium*

25. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-03-02&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZNmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

26. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-04-06&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZNmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

27. <https://www.academie-francaise.fr/mondial-pour-international> (consulté en mai 2023)

meridianum) l'ennui, l'acédie qui attaquait les moines en prière vers le milieu du jour (on nommait aussi cet ennui « démon de midi », mais ce n'est que bien plus tard que cette expression en vint à désigner la tentation sentimentale et sexuelle de l'âge mûr).

S'ils ont une même étymologie, les mots *méridional*, *méridien* et *méridienne* n'ont pas le même sens.²⁸

Parfois la position de l'Académie ne s'adosse nullement à des justifications linguistiques. C'est le cas dans l'exemple pour *belle journée* vs *bonne journée* et c'est aussi ce qui rend le discours moins prescriptif (en apparence !) et de nature plus axiologique.

La formule de politesse *Belle journée* **s'entend de plus en plus** en lieu et place de *Bonne journée*. *Belle journée* n'est bien sûr pas incorrect mais on peut s'interroger sur le développement rapide de cette formule qui tend à devenir **un tic de langage**. Elle est sans doute issue du mélange de la forme exclamative *Quelle belle journée !* – qui fut aussi le titre d'une chanson interprétée par le regretté Marcel Amont – et de l'anglo-américain *Have a nice day*!²⁹

Le prédicat, basé sur le constat que la formulation n'est pas incorrecte, se construit autour de l'énonciation d'un argument à valeur psycho-sociologique : *tic de langage*, exonérant le locuteur d'un geste linguistique qu'il ne contrôle pas, mais l'accusant à bas bruit de suivre une mode. On ne sera pas surpris de la justification de l'Académie qui met l'accent sur le recours à un calque linguistique, rappelant par là-même à demi-mot la nécessité d'être vigilant quant à l'utilisation ou à la réappropriation des anglicismes comme on le verra dans la notice suivante.

Lorsqu'il reste dans le cadre d'une argumentation linguistique, le discours moins manichéen se fait plus compréhensif à l'égard du locuteur fautif. L'Académie explique que la confusion sémantique peut ne pas être surprenante sous couvert de phénomènes bien connus linguistiquement : *homophonie*, *doublés*, *étymologie*. En expliquant l'origine de la faute, cette dernière ne devient pas plus admissible, mais la position du locuteur expert transforme l'interaction qui s'inscrit davantage dans le partage de la connaissance et l'explication.

Cette notice, parce qu'elle s'inscrit moins fermement dans un cadre de correction syntaxique et grammairienne que la précédente, envisage la notion de faute d'une manière un peu différente et conséquemment la notion de correction. Alors que, lorsqu'il s'agit d'orthographe ou de syntaxe, la correction est exclusivement normative, tendant à réduire l'écart entre la production d'un énoncé fautif et la norme de référence, dès lors qu'il s'agit de corrections liées aux impropriétés du vocabulaire, comme c'est le cas dans cette notice, il est question d'augmenter la compétence linguistique du locuteur en faisant appel à ses aptitudes cognitives dans le but de ne pas altérer la communication

28. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=365&date=2023-02-02&form_build_id=form-jJFG8pKtlv-fwfaKJlkqUfwTpXNbeGxBxtM2bc6czHQ&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

29. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=365&date=2023-04-06&form_build_id=form-S5EoDQqky5gJwKjOuE-YrQHYqIrEDcNfUWoxsyQ1FtM&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

dans l'exercice du partage du sens et du décodage du message. On peut en effet considérer qu'une faute de liaison ou d'orthographe n'obère pas le sens d'un énoncé quand bien même il ne respecte pas les règles. En revanche la confusion sémantique pose problème dans l'exercice de la communication. En témoigne l'exemple des deux adjectifs que sont *vexatoire* et *vexant* qui conduisent à deux attitudes axiologiquement très différentes et donc à des effets sur la communication et sur l'allocutaire : *vexatoire* qualifiant comme l'indique l'Académie « une décision perçue comme abusive ou, plus souvent aujourd'hui, comme humiliante, et émanant d'une autorité brutale³⁰ », quand le verbe *vexer*, d'où est tiré le participe présent et adjectif *vexant*, « s'emploie au sens de "froisser, blesser quelqu'un dans son amour-propre" » (*ibid.*). Dans le premier cas, un jugement de valeur est porté par le locuteur pour lui-même indépendamment de son allocutaire; dans le second, le locuteur cherche à produire un effet sur l'éthos de son allocutaire.

Alors que pour *l'emploi fautif*, les modalités déontiques se réfèrent à un ordre moral ou social pour exprimer ce qui doit être (obligation) ou ce qui peut être (permission), l'injonction ne porte plus sur un ordre strict *on dit* vs *on ne dit pas* mais sur un conseil formulé de manière plus pédagogique et plus appréciative mais parfois également ironique.

Il **convient donc de tenir compte du contexte** dans lequel on rencontre ces adjectifs afin de ne pas employer l'un à la place de l'autre.³¹

On se gardera bien de confondre ces homophones entre eux, mais aussi avec un autre mot, *Vigie*, qui a la même origine.³²

Ce dernier a plus d'un demi-millénaire d'existence, aussi **n'est-il peut-être pas nécessaire de lui adjoindre l'inutile doublet** qu'est *gênance*.³³

Il **faut donc éviter de se servir de ce verbe en lieu et place de louer**, qui, lui, est beaucoup plus neutre et ne l'employer que pour moquer ou dénoncer l'attitude de quelque flagorneur à l'âme courtisane.³⁴

Dans son étude consacrée au discours normatif, Berrendon indique que : « La bienséance et la prudence étant de ne pas heurter la susceptibilité du destinataire, il convient, lorsqu'on prescrit de lui éviter de perdre la face. Le grammairien s'efforce donc en général d'atténuer un ordre trop brutal. » (1982 : 35)

Le discours normatif avance donc « masqué » dans bien des cas. C'est la raison pour laquelle à côté de formules jussives, on trouve dans notre corpus des formules plus atténuées sous forme de

30. <https://www.academie-francaise.fr/vexatoire-pour-vexant> (consulté le 23 mai 2023)

31. <https://www.academie-francaise.fr/vexatoire-pour-vexant> (consulté le 23 mai 2023)

32. <https://www.dictionnaire-academie.fr/article/DNP1214> (consulté le 23 mai 2023)

33. <https://www.academie-francaise.fr/genance-pour-gene> (consulté le 23 mai 2023)

34. <https://www.dictionnaire-academie.fr/article/DNP1234> (consulté le 23 mai 2023)

conseils ou d'actes de mise en garde avec des énoncés du type : « il faut donc éviter », « on se gardera bien de... », « on veillera donc à ».

Ces formulations, comme le rappelle Danielle Bouverot, orientent le discours vers une démarche de prévention plus que celle de la sanction d'un usage incorrect.

Les lecteurs attendent une prescription, un ordre, un peu comme une ordonnance médicale, c'est-à-dire LA norme. Ce que les rédacteurs leur proposent c'est une description de l'usage plutôt qu'une répression à la manière des « puristes ». Pour concilier ces divergences, s'esquisse la prévention faite d'explication au lieu d'impératifs. (2007 : 122)

Le discours de prévention est un discours traditionnellement très présent dans le domaine de la santé. Dans le domaine linguistique, il vise à prévenir le lecteur dudit discours d'un danger potentiel et à lui donner les armes pour l'éviter. On retrouve ici ces deux axes mais l'ordre du paradigme danger/solution n'est pas toujours le même. Dans le cas de la confusion entre *vexatoire* et *vexant*, les conditions d'existence du discours de prévention reposent sur le besoin de bâtir un cadre nécessaire à la prévention en répondant à la question primordiale « que devrais-je faire pour être dans le bon usage ? ». La réponse des académiciens se trouve formulée par « il convient de tenir compte du contexte ». L'ordre a toute son importance, puisqu'ici les académiciens offre la solution avant même l'expression du danger.

Dans le domaine de la prévention, la modélisation du discours repose sur des messages destinés à favoriser le maintien d'un comportement dit « normal » ou de modifier un comportement dit « à risque » (Romeyer et Moktefi, 2013). En ce sens, il peut être considéré comme un discours normatif, visant le normal/la norme. Pourtant il s'agit ici de proposer au lecteur/internaute la possibilité de prendre une décision éclairée. Et ce qui éclaire la décision, c'est précisément la possibilité de mettre en œuvre une réflexion autonome et personnelle (l'analyse du contexte) plutôt que l'application d'une règle normative.

La mention de la solution comme axiome premier de l'équation, plutôt que la mention du danger responsabilise le lecteur/internaute. D'un point de vue discursif, le thème du discours n'est plus la faute mais le moyen d'y échapper en tenant compte du contexte. La thématization du discours autour de la solution montre que le paradigme du discours a changé par rapport à celui rencontré dans la notice *Emplois fautifs*.

Les « risques », au cœur de la démarche de prévention, en cas de défaillance linguistique s'organisent autour de deux thématiques principales dans cette notice.

- Les risques, que prend la langue elle-même pour elle-même : risque pour son esthétique (lourdeur), sa pérennité (oubli de formes validées et anciennes), sa musicalité et son rythme.

Il y a, en français, quelques noms terminés par *-ance* qui sont concrets, comme *ambulance*, *balance*, *excroissance*, *garance*, *laitance*, *lance*, *pitance*. Mais l'immense majorité d'entre eux sont des noms abstraits et ce trait peut leur donner un petit côté savant qui séduit et amène à la création de néologismes inutiles. Ainsi rencontre-t-on depuis quelque temps la forme *gênance*, dérivée de *gêner*, qui lui-même l'était du nom *gêne*. **Ce dernier a plus d'un demi-millénaire d'existence**, aussi n'est-il peut-être pas nécessaire de lui adjoindre l'inutile doublet qu'est *gênance*.³⁵

- Les risques que prend le locuteur (manquer de sobriété, de neutralité, de justesse, de cohérence, céder aux modes et à la facilité)

Au sujet du verbe *louanger*, on lisait déjà, dans la première édition du *Dictionnaire de l'Académie française* : « Il ne se dit guère qu'en raillerie. » Plus tard, Littré adoucissait un peu le propos en écrivant : « Il ne se dit guère qu'en plaisantant. » Les deux dernières éditions de notre *Dictionnaire* indiquent que ce verbe signifie « louer avec excès ». Il faut donc éviter de se servir de ce verbe en lieu et place de *louer*, **qui, lui, est beaucoup plus neutre et ne l'employer que pour moquer ou dénoncer l'attitude de quelque flagorneur à l'âme courtisane**.³⁶

Aux considérations purement linguistiques s'ajoutent non seulement la défense d'un patrimoine langue mais plus subtilement l'idée que la langue est aussi un médium éthique, moral que le travail de prévention doit conforter. L'usager/locuteur est au cœur de la préoccupation de l'Académie. Et il l'est à la fois pour lui-même mais également dans son rapport à l'autre.

Dans d'autres cas, par exemple,

[o]n se gardera bien de confondre ces homophones entre eux, mais aussi avec un autre mot, *Vigie*, qui a la même origine.³⁷

Le discours porte sur le danger en premier (danger de sortir de la norme et de faire une faute, ou de faire échouer la communication). La mise en discours du danger est suffisamment convaincante avec l'expression « On se gardera bien de » pour que l'internaute s'oblige à rester dans le cadre.

Dans tous les cas, il y a une forme de double injonction dans le discours de prévention : celle de donner les clés d'une compréhension plus générale sur la langue qui poussera le locuteur à appliquer une règle non plus dans le cas précis d'un usage mais dans tous les cas où il sera opportun d'être dans un état de vigilance par rapport à la langue. En réaxant le discours non plus sur la langue et sa norme, mais sur l'intérêt individuel du récepteur du message, ce sont les axes de vigilance, de prudence et d'attention qui se répondent et non plus ceux liés au bon/mauvais usage.

35. <https://www.academie-francaise.fr/genance-pour-gene> (consulté le 23 mai 2023)

36. <https://www.dictionnaire-academie.fr/article/DNP1234> (consulté le 23 mai 2023)

37. <https://www.dictionnaire-academie.fr/article/DNP1214> (consulté le 23 mai 2023)

Cette démarche de prévention n'empêche pas l'expression d'un discours prescriptif, tel que nous l'avons déjà croisé, preuve supplémentaire que la frontière est poreuse entre le discours prescriptif et le discours préventif. À titre d'exemple, l'explication relative à l'utilisation de *De belle facture*.

De belle facture

Le 6 avril 2023

Extensions de sens abusives

La locution *de bonne* (ou *de belle*) *facture* s'est d'abord employée pour qualifier une œuvre littéraire ou artistique, puis des instruments de musique. Elle peut aussi qualifier aujourd'hui des objets artisanaux ou, pour prendre un mot de la même famille que *facture*, des objets manufacturés : *un meuble, un bijou de belle facture*. L'intervention de la main de l'homme est contenue dans cette expression, aussi ne fera-t-on pas de celle-ci, dans tous les cas, un synonyme de *de qualité*. On ne dira donc pas *des cerises de bonne facture* (quand bien même le travail de l'arboriculteur aura contribué à ce résultat), mais *de belles cerises*. Et on se gardera davantage encore de confondre ce nom *facture*, emprunté du latin *factura*, « façon, fabrication, œuvre », avec son homonyme, dérivé, lui, de *facteur* pris au sens d'« agent commercial », et qui désigne le mémoire établi par un vendeur dans lequel sont détaillés la quantité, la nature et le prix des marchandises livrées, puisqu'un *produit de bonne facture* n'est pas un produit dont le prix est élevé ou avantageux.

on dit	on ne dit pas
<i>Un beau chêne, un chêne magnifique</i>	<i>Un chêne de bonne, de belle facture</i>
<i>Une prestation coûteuse, bon marché</i>	<i>Une prestation de belle facture</i>

Figure 5. Formulation des académiciens pour expliquer et corriger l'emploi de *de belle facture*³⁸

La démarche est alors identique à celle que l'on retrouvait dans les *emplois fautifs*. La reformulation du discours fautif ramène à l'opposition correct/interdit. La formulation comparatiste de la question induit la formulation manichéenne de la réponse *on dit - on ne dit pas*.

On voit donc bien que l'Académie oscille toujours entre ces deux versants que sont le discours prescriptif et le discours de prévention. Mais la présence d'un discours moins déontique montre l'évolution et la préoccupation du récepteur du message. Le discours de prévention annonce également la mission de l'Académie en tant que gardienne non plus de la langue mais du patrimoine langue.

Le discours de prévention, moins normatif, ouvre par ailleurs la porte à des remarques plus axiologiques en droite ligne avec ce que nous allons voir maintenant, c'est-à-dire le sentiment de la langue qui est, rappelons-le, « la relation subjective qu'un locuteur éprouve avec une "langue", qu'il s'agisse d'un parler vernaculaire, d'une langue faisant l'objet d'une reconnaissance politique, maternelle ou étrangère, ou d'une "langue de culture" » (Siouffi, 2007a : 267).

38. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=365&date=2023-04-06&form_build_id=form-0gNvJGzWDEP87nsW8O-oJtgAyPskXt1IUqTrD1u9aOg&form_id=academie_blog_search_form&op=Rechercher (consulté en octobre 2022)

4. L'appauvrissement et l'enrichissement de la langue : le « sentiment de la langue »

La rubrique *Dire, Ne pas dire* est un territoire supplémentaire offert à l'institution pour réaffirmer sa position vis-à-vis de la langue et ouvre donc des perspectives plus axiologiques.

4.1. Le français d'abord

Dans cette rubrique, la position de l'Académie reste la même à l'égard des néologismes et plus particulièrement à l'égard des anglicismes que celle affichée dans la préface de la neuvième édition :

Nous ne faisons place aux mots étrangers qu'autant qu'ils sont **vraiment installés dans l'usage, et qu'il n'existe pas déjà un honnête mot français** pour désigner la même chose ou exprimer la même idée. Nous sommes d'ailleurs plus accueillants qu'on ne le prétend, considérant que la langue est moins menacée par l'extension du vocabulaire que par la détérioration de la syntaxe. Nous sommes **assez rigoureux à l'égard des néologismes, dont beaucoup ne doivent leur apparition qu'à l'ignorance ou l'oubli de bons termes existant depuis fort longtemps**; nous sommes généralement **impitoyables** s'ils sont formés d'une manière qui insulte au génie de la langue.³⁹

Il apparaît que les dangers qui sont supposés menacer la langue française sont priorisés par l'institution : d'abord la syntaxe et le risque de sa « détérioration », puis le vocabulaire. Cette hiérarchisation des dangers explique et justifie la proportionnalité de la réaction (accueillants, rigoureux, impitoyables) aux dangers de la part des académiciens qui va de la capacité d'accueil au rejet impitoyable d'un terme.

L'échantillon relevé et les commentaires de l'Académie sur ces questions sont en droite ligne avec le programme annoncé dans la préface :

- janvier 2023⁴⁰
 - *Crazy Monday*
 - *Sticker*
- février 2023⁴¹

39. <https://www.academie-francaise.fr/le-dictionnaire-les-neuf-prefaces/preface-la-neuvieme-edition> (consulté en octobre 2022)

40. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=364&date=2023-01-05&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté en mai 2023)

41. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=364&date=2023-02-02&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté en mai 2023)

- *Wishlist*
- « *Vrai* » ou « *fake* » ?
- mars 2023⁴²
 - *Merci Monsieur Fabri*
 - « *Dark* » pour « *Sombre, obscur, inquiétant* »
- avril 2023⁴³
 - *Wave light*
 - « *Toping* » pour « *Nappage, glaçage* »

On note que sur l'échantillon, les remarques normatives de l'Académie portent essentiellement sur des anglicismes :

On aide les tout jeunes enfants à développer leurs capacités en motricité fine en leur proposant de coller sur différents supports des vignettes autocollantes ou des gommettes. Ces exercices plaisants sont d'une grande utilité, **mais il n'est sans doute pas nécessaire de remplacer**, dans les livres ou les cahiers qui sont destinés à ce jeune public, ces formes – ***gommette, vignette autocollante ou, simplement, autocollant*** – par l'anglais *sticker*, qui **désigne le même type d'objet**.⁴⁴

On lit dans ***La Chanson de Roland***, parue vers l'an 1100, les syntagmes *fals jugement*, « jugement injuste », et *false lei*, « fausse loi ». Ce sont les plus anciennes attestations de l'adjectif *faux*. [...] **Plus de neuf cents ans de bons et loyaux services devraient lui valoir reconnaissance et lui éviter de se voir préférer, comme cela commence à se faire, l'anglicisme *fake***, puisque, conformément à son étymologie, il peut renvoyer à ce qui est erroné mais aussi à ce qui est produit frauduleusement pour tromper (on parle de *faux tableau*, de *faux papiers*, etc.). **Pour éviter cet anglicisme, on pourra aussi user** de l'adjectif *fallacieux*, à l'instar de la Commission d'enrichissement de la langue française, qui a proposé de remplacer la locution *fake news* par le groupe nominal *information fallacieuse* ou par le mot-valise *infox*.⁴⁵

4.2. *L'honnête mot français, mais pas seulement*

Le discours de réprobation du recours à des anglicismes est, dans les exemples qui sont les nôtres, un discours modéré où le locuteur est invité à privilégier l'utilisation du mot français correspondant.

42. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=364&date=2023-03-02&form_build_id=form-yVDsgXnJsD5l_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté mai 2023)

43. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=364&date=2023-04-06&form_build_id=form-yVDsgXnJsD5l_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté mai 2023)

44. <https://www.academie-francaise.fr/sticker> (consulté en mai 2023)

45. <https://www.academie-francaise.fr/vrai-ou-fake> (consulté en mai 2023)

Mais la modération n'est qu'apparente car les modalités du discours s'inscrivent quand même dans la nécessité avec des tours du type :

- « aussi peut-on sans doute parler de X - X étant le terme français à privilégier »
- « mais il n'est sans doute pas nécessaire de remplacer X - X étant le terme à utiliser en français ».

Aussi peut-on sans doute parler de *lundi fou* en lieu et place de « crazy monday » ou même, pour élargir le propos et rappeler un de nos grands dramaturges, de *folle journée*.⁴⁶

À l'approche des fêtes de fin d'année, les jeunes enfants écrivent au père Noël pour lui communiquer la liste des cadeaux que leur bonne conduite devrait leur valoir. La pratique est ancienne puisque, en 1946 déjà, Raymond Vincy faisait chanter à Tino Rossi dans son fameux *Petit Papa Noël* : « Il me tarde tant que le jour se lève / Pour voir si tu m'as apporté / Tous les beaux joujoux que je vois en rêve / Et que je t'ai commandés. » Cette *liste de cadeaux* ou, à tout le moins, *cette liste de souhaits*, **est bien installée dans l'usage ; aussi n'est-il pas nécessaire de la remplacer, comme cela commence à se faire dans quelques slogans publicitaires, par l'anglais wishlist**. Sans doute faut-il y voir l'influence de formes moins récentes et déjà déconseillées dans cette rubrique, comme la *check-list* ou la *to-do-list*.⁴⁷

Dans l'exemple de *vrai ou fake*, le contexte justificateur est le texte de la *Chanson de Roland* et son ancienneté. Cet argument de légitimité littéraire est accompagné de l'attitude de respect qui est alors demandée au locuteur fautif. Le patrimoine linguistique vient au secours de la démarche :

Plus de neuf cents ans de bons et loyaux services devraient lui valoir reconnaissance et lui éviter de se voir préférer, comme cela commence à se faire, l'anglicisme *fake*.⁴⁸

La position de l'Académie est clairement énoncée ici : la langue française est un patrimoine au même titre que la littérature, qu'il convient alors de respecter et de pérenniser. Parfois l'Académie procède par comparaison (latinismes inutiles autrefois vs anglicismes inutiles aujourd'hui) comme c'est le cas dans les quelques lignes qui ont pour titre *Merci Monsieur Fabri*. Le texte a vocation à montrer que le danger est toujours présent, même si les ennemis ne sont plus les mêmes.

En 1521 paraissait à Rouen *Le grant et vray art de pleine rhétorique utile, profitable et nécessaire à toutes gens qui désirent a bien elegantement parler et escrire*, de Pierre Fabri. Cet auteur y condamnait **les latinismes inutiles**. [...] Quelque cinq siècles plus tard, le texte de Fabri est **toujours d'actualité, car ce qu'il écrit des latinismes de son temps peut s'appliquer aux anglicismes** du nôtre. [...]. Et pourtant, comme l'écrivait notre auteur, « L'on doit toujours prendre **les termes et les mots les plus communs que l'on puisse trouver et les mettre à leur signification à tous intelligible** ». ⁴⁹

46. <https://www.academie-francaise.fr/crazy-monday> (consulté en mai 2023)

47. <https://www.academie-francaise.fr/wishlist> (consulté en mai 2023)

48. <https://www.academie-francaise.fr/vrai-ou-fake> (consulté en mai 2023)

49. <https://www.academie-francaise.fr/merci-monsieur-fabri> (consulté en mai 2023)

Cette notice consacrée aux *Néologismes & anglicismes* montre à notre avis de manière plus claire encore que les précédentes que la rubrique *Dire, Ne pas dire* n'est pas que l'espace de l'usage ou du sens mais que l'Académie a su profiter de cette opportunité d'un territoire numérique mis à sa disposition pour réaffirmer des principes de défense du patrimoine linguistique. C'est également l'occasion pour l'Académie d'introduire des éléments de lexiculture.

5. Le renforcement de la lexiculture : la « culture partagée »

La nouveauté dans cette rubrique consiste en l'apparition de références culturelles assez nombreuses. La lexiculture, notion introduite par Robert Galisson puis développée par Jean Pruvost, repose sur « une charge culturelle partagée ». Comme le rappelle Pruvost :

Cette lexiculture implicite que les dictionnaires traditionnels chassent parce qu'ils la pensent éphémère, alors même qu'elle dépose des sédiments successifs dans lesquels s'enracine le lexique, et sans lesquels on se situe lexicalement dans la fausse apparence et en définitive dans la méprise sémantique. (2009 : 133-136)

5.1. Des opportunités supplémentaires de « culture partagée »

Or il se trouve comme le souligne Siouffi (2009 : 155-176) que la neuvième édition du *Dictionnaire de l'Académie française* fait la part belle et « nouvelle » à ces références culturelles. Il apparaît donc logique que la rubrique *Dire, Ne pas dire* fasse écho de cette nouveauté. Mais en reprenant à son compte cette démarche novatrice, elle en prend aussi les travers. Si la majorité des références culturelles sont communes à un grand nombre de locuteurs :

- Géographie : « Enfin, les noms *Méridional* et *Méridionale*⁵⁰ désignent une "personne qui est originaire des régions situées au sud d'un pays, et particulièrement, **en France, qui est originaire du Midi**"⁵¹ »
- Événements sportifs : « Ainsi, le **Tournoi des six nations** est une compétition internationale de rugby tandis que les Jeux olympiques sont une compétition mondiale.⁵² »
- Histoire : « Eugène Pottier en fait, en le substantivant, le titre d'une célèbre chanson révolutionnaire, **L'Internationale**⁵³ »

50. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=363&date=2023-02-02&form_build_id=form-swbhLa8zop_kSoC9N_EQM6ZNmBfW5MPLUQySUCOdWXY&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

51. <https://www.academie-francaise.fr/meridionale-pour-meridienne> (consulté en mai 2023)

52. <https://www.academie-francaise.fr/mondial-pour-international> (consulté en mai 2023)

53. <https://www.academie-francaise.fr/mondial-pour-international> (consulté en mai 2023)

D'autres sont d'ordre générationnel ou élitiste et donc inaccessibles aux générations présentes :

- Chanteur français : le titre d'une chanson interprétée par le regretté **Marcel Amont** (« belle journée » vs « bonne journée »⁵⁴)

Le « sédiment successif » dont parle Pruvost est bien présent mais opaque pour certains internautes, notamment aux plus jeunes générations. Rappelons que Marcel Amont est un artiste de variétés français qui a connu un succès dans les années 1950 à 1970. À l'inverse, cet exemple qui peut ne parler qu'à une certaine partie des internautes raconte quelque chose d'un patrimoine culturel qui, justement parce qu'il appartient à l'histoire, n'a pas vocation à être nécessairement actuel ou contemporain.

On notera que certaines de ces références culturelles sont absentes du dictionnaire : c'est le cas de *Marcel Amont* et d'*Eugène Pottier*. Le *Tournoi des six nations* est lui donné en exemple sous l'entrée *rugby*.

Il est nécessaire de souligner l'effort de mettre en valeur le patrimoine culturel français et international et on peut imaginer et/ou espérer que ces territoires numériques additionnels offriront à la lexiculture un espace supplémentaire.

5.2. La notice *Bonheurs & Surprises* : l'audace d'une culture qui prend son temps

La notice *Bonheurs & surprises*⁵⁵ surprend par sa dénomination axiologique et annonciatrice d'émotions heureuses et inattendues. Il ne s'agit plus de circuler au sein d'un espace normatif du correct ou de l'interdit. Il ne s'agit pas non plus de naviguer dans un espace du légitime et du vrai. Et même si l'alibi et la justification restent toujours la langue, cette notice apparaît comme une forme de « bol d'air frais » au milieu d'une architecture numérique de la rubrique *Dire, Ne pas dire*, un peu sèche et austère.

Cette notice est l'occasion pour l'institution d'un partage sur la langue comme un élément fondamental du patrimoine culturel. Les longs textes narratifs sont tous dotés de titres qui suscitent nécessairement la curiosité de l'internaute. Ainsi un titre comme « Changement de genre⁵⁶ » est l'occasion pour l'Académie de revenir sur le genre de certaines lettres de notre alphabet, aujourd'hui au masculin mais qui ont au cours de l'histoire changé de genre (la lettre F par exemple). D'autres permettent de faire le lien entre les contes de notre enfance (Cendrillon en l'occurrence) et la matière des pantoufles de l'héroïne éponyme avec un titre comme « De quelle matière étaient faites les pantoufles de Cendrillon ?⁵⁷ ».

54. <https://www.academie-francaise.fr/belle-journee-pour-bonne-journee> (consulté en mai 2023)

55. <https://www.academie-francaise.fr/dire-ne-pas-dire/bonheurs-surprises> - (consulté en octobre 2022)

56. <https://www.academie-francaise.fr/changement-de-genre> (consulté en mai 2023)

57. <https://www.academie-francaise.fr/de-quelle-matiere-etait-faites-les-pantoufles-de-cendrillon> (consulté en mai 2023)

Ces textes très longs (1 page en moyenne) sont construits comme des textes narratifs structurés en général autour d'une introduction, d'un développement et d'une conclusion qui forment un tout cohérent. Presque poétiques, ils nous racontent une histoire et apparaissent en rupture avec la dureté et la brièveté des discours tenus dans les autres notices. Ils jouent parfois sur la drôlerie d'un phénomène linguistique comme l'homophonie par exemple « Les coronaires du coroner⁵⁸ » qui explique que le substantif *coroner* (officier de justice) est apparu dans deux éditions, celles de 1835 et de 1878 puis a disparu quand l'adjectif *coronaire* est entré dans la quatrième édition (1762) et signifie « disposé en couronne ». Il s'agit ici de prendre son temps et l'Académie se montre très audacieuse en proposant de longs textes dans l'univers de la toile qui commande plutôt une esthétique de la brièveté et de la réactivité.

La posture de l'Académie n'est ici pas celle du gendarme de la langue mais plutôt celle du sachant qui cherche à partager des connaissances en retraçant l'histoire d'un mot ou en convoquant des éléments de notre patrimoine.

Si l'Académie reste détentrice d'un savoir, elle cherche à le partager dans une langue qui n'est plus seulement terminologique et qui n'a plus seulement l'objectif de l'usage, mais dans un style certes précis, clair et neutre qui, d'un seul coup, semble accessible au plus grand nombre. S'appuyant sur des ressources extra lexicographiques (littéraires, historiques), cette notice est probablement celle qui, dès lors que l'on veut bien avoir la curiosité d'y entrer, propose à l'internaute un voyage des plus rafraichissants. Elle contribue à élargir la lexiculture qui ne devient plus seulement un sédiment supplémentaire autour d'un seul mot, mais elle rattache chaque internaute à un tout plus grand que le mot lui-même et lui offre un espace de distraction et de voyages. Ainsi, porte-t-elle parfaitement son nom et son étiquette dénomminative s'avère être une promesse de surprises et de bonheurs tenue.

58. <https://www.academie-francaise.fr/les-coronaires-du-coroner> (consulté en mai 2023)

Cette notice est également ce qui fonde l'intérêt d'une culture numérique qui offre un espace à un public qui, aussi hétérogène⁵⁹ soit-il, notamment eu égard à son pays d'origine, pourra à un moment ou un autre trouver un intérêt à une culture partagée. Cette notice est enfin, pour reprendre les dires de Madame Carrère d'Encausse d'illustrer « la richesse de la langue » et la volonté de ne pas passer que « pour des pères fouettards » (2017 : 13), montrant s'il en était nécessaire que le territoire numérique sert également l'image de l'Académie.

6. La question du locuteur/internaute : tout public ?

6.1. Modernité et modernisation ?

À l'issue de cette étude, une question se pose : à quel public s'adresse cette rubrique *Dire, Ne pas dire* ? Clémentine Darmestädter souligne l'actualité de la rubrique :

Dans la rubrique *Dire, Ne pas dire*, sont en revanche répertoriées les fautes du moment : par exemple, l'emploi de la préposition sur en dépit de son sens de supériorité et de domination (« j'habite sur Paris»), et celui de la locution AU NIVEAU DE, en dehors du sens spatial («au niveau du salaire») (2011 : 285-306)

Dans cet exemple, Darmestädter semble mettre en avant la modernité des échanges et des fautes signalées. L'étroitesse de notre corpus n'a pas permis de confirmer cette modernité des fautes relevées, c'est même l'inverse qui semble se produire sur notre petit échantillon.

En 2007, Siouffi se posait la question du « dictionnaire français face à la compétence des locuteurs » faisant référence au dictionnaire de Richelet. Il indiquait alors que

59. Un an après le lancement de la rubrique, Yves Pouliquen partageait quelques statistiques de consultation de la rubrique *Dire, Ne pas dire*, notamment sur la fréquence de consultation et l'origine géographique des internautes. « Une étude statistique de la fréquentation de *Dire, Ne pas dire* nous confirme qu'à la curiosité que suscita l'annonce par les médias de son lancement a succédé une consultation régulière, attentive de ses pages. Une analyse précise de son audience au cours des dix premiers mois de son existence, du 1^{er} novembre 2011 au 31 août 2012, met en évidence les éléments suivants : 45 395 visiteurs uniques nous ont fréquenté, qui furent responsables de 63 483 visites en consultant 199 387 pages (3,14 pages par visiteur) avec, ce qui est notable, une durée moyenne de visite de 2 minutes 49 secondes et un taux de rebond de 55 %. 72 % d'entre eux nous rendaient visite pour la première fois tandis que 28 % consultèrent plusieurs fois notre site. Il est intéressant de noter que, parmi les rubriques proposées, c'est celle des *Emplois fautifs* qui fut la plus consultée, suivie à égalité par *Extensions de sens, Bonheurs et surprises* et *Néologismes et anglicismes*. Les bloc-notes sont régulièrement lus. C'est donc, en moyenne, 4 500 internautes qui s'informent chaque mois des propositions de notre site *Dire, Ne pas dire*. Ils sont une fraction des 33 429 internautes qui, entre le 22 octobre et le 21 novembre derniers, ont rendu visite au site de l'Académie française. Une majorité d'entre eux est naturellement d'origine française (23 044) à laquelle s'ajoutent deux à trois mille francophones originaires en parts égales du Canada, de Suisse, de Belgique et d'Algérie. Il en vient aussi des États-Unis, d'Allemagne, d'Italie et d'Espagne, environ huit cents pour chacun de ces pays », Yves Pouliquen (janvier 2013), <https://www.academie-francaise.fr/dire-ne-pas-dire-un-apres> (consulté en mai 2023)

Fondamentalement, la question de l'appel à la compétence des locuteurs [auteurs] peut être appréhendée de deux manières. On peut tout d'abord la relier à la question générale de « l'usage », laquelle commande une vision de la langue, et de ce qui commence à être décrit comme sa « vie » ; on peut aussi y voir un reflet des incertitudes qui régnaient au moment de chercher à asseoir l'autorité d'un dictionnaire ou d'un ouvrage de langue. (2007b : 11-19)

Or d'une certaine manière cette compétence des locuteurs/auteurs commande celle des lecteurs/internautes et, pour ce qui nous intéresse ici, celle des correspondants qui dans cette notice vont échanger avec le service du dictionnaire de l'Académie française.

En tout état de cause, pour que le succès soit au rendez-vous, il faut une forme d'homogénéité entre les aspirations des internautes et celles de l'Académie française. D'un point de vue linguistique, un locuteur doit, pour se faire comprendre de son interlocuteur et inversement, respecter a minima des principes qui offriront les conditions d'existence d'un échange.

6.2. La notice *Courrier des internautes* : la question de la compétence

Cette notice *Courrier des internautes* est sans doute la plus singulière à étudier pour au moins deux raisons : d'abord l'opportunité d'un échange avec l'Académie n'existe pas avec la version papier du dictionnaire, en tout cas pas sous cette forme même s'il était possible d'envoyer un courrier postal à l'Académie (possibilité qui de fait existe toujours), ensuite compte tenu du développement d'internet, la question des interactions avec l'institution se pose de manière très contemporaine.

C'est d'autant plus vrai dans un exercice de correspondance qui n'est rien d'autre qu'un exercice de questions/réponses. Il faut qu'à une question donnée, la réponse soit suffisante sans excès, qu'elle soit pertinente, c'est-à-dire en relation avec la question et suffisamment claire pour que l'effort d'interprétation soit accepté par l'internaute.

Ce qui frappe d'emblée, c'est que si les domaines sur lesquels portent les questions sont, d'une manière ou d'une autre, attendus : orthographe, prononciation, sens, grammaire..., les thèmes, objets du discours appartiennent à un niveau culturel élevé. On peut imaginer que ce que dit Sonia Branca-Rosoff de la conscience linguistique vaut également pour la conscience culturelle : « [l]a conscience linguistique qu'il s'agit de développer va bien au-delà de l'exclusion de quelques formes. Ce n'est pas seulement d'un tri parmi les formes disponibles du système qu'il s'agit, mais aussi du sentiment aigu qu'il y a toujours moyen d'être plus distingué, plus relevé, plus pertinent. » (2007 : 29)

En témoigne l'échantillon recueilli sur la période de janvier à avril 2023 :

- janvier 2023⁶⁰
 - Je m'interroge sur la fonction du mot *mamie* dans la phrase *Allons manger, mamie*. Quelle est la différence avec *Allons manger mamie* ?
 - Pour parler de personnes d'une même famille, dont le nom comporte une particule, doit-on dire *Les X* ou *Les de X* ?
 - Pourriez-vous me dire comment prononcer le mot *baïle*, que l'on trouve dans **Le Secret de maître Cornille**, un des contes des *Lettres de mon moulin* ?
 - Lisant, dans les **Histoires extraordinaires** d'Edgar Allan Poe, *Descente dans le Maelström*, je trouve cette phrase : « L'explosion générale de cette magnificence terrifiante était tout ce que je pouvais voir. » « Terrifique » n'est pas dans votre *Dictionnaire*. Pouvez-vous m'en donner la signification ?
- Février 2023⁶¹
 - Pourquoi le mot *Monsieur* se prononce-t-il *Meussieur* ?
 - J'aimerais savoir si le personnage du *Grinch* tire son nom du français *grincheux* et j'aimerais connaître l'étymologie de cet adjectif.
 - Je suis lycéen et j'aimerais savoir si la tique pique ou mord.
 - Pourquoi n'y a-t-il pas de tréma sur le *i* d'*ébloui* alors qu'il en a un dans *inouï* ?
- Mars 2023⁶²
 - Pourquoi le nom *Pyrénées* est-il un féminin ?
 - L'expression « de la belle ouvrage » est-elle correcte ?
 - Pourquoi écrire *sylvestre*, *sylviculture* avec un *y* et non un *i*, comme dans le latin *silva*, « forêt » ?
 - J'ai lu dans un recueil de contes de Noël, de G. Lenôtre, cette phrase : « Les invitations affluèrent : bals, chasses, dîners, il était de toutes les fêtes, les dames les plus hautaines prenaient son jour... » Pouvez-vous m'expliquer ce que signifie « Prendre le jour de quelqu'un » ?

60. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-01-05&form_build_id=form-cLHW9B4WI2RdAm17aReqtgWMXjsBHcDwqV5abwHFm48&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

61. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-02-02&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté en mai 2023)

62. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-03-02&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté en mai 2023)

- Avril 2023⁶³
 - Je souhaiterais connaître les différences entre les expressions « c'est mon credo » et « c'est mon créneau » et j'aimerais savoir si cette dernière est correcte.
 - Est-ce que le mot *fantaisie* peut s'employer comme synonyme d'*envie* dans une phrase telle que *Ce gâteau me fait fantaisie* ?
 - Les éditions de 1762 à 1878 de votre *Dictionnaire* définissent le *caïman* comme une « espèce de crocodile ». Dans l'édition actuelle, vous précisez que ce mot apparaît au XVI^e siècle, et qu'il serait « probablement d'origine caraïbe ». Or on trouve plusieurs fois *caïmans* dans le *Journal d'un bourgeois de Paris*, paru au début de l'année 1449. On y lit par exemple : « ... en ce temps furent pris caïmans, larrons et meurtriers ». Les Caraïbes ayant été découvertes plus de quarante ans plus tard, d'où ce bourgeois tirait-il ce mot ? Connaît-on d'autres emplois de *caïman*, au XV^e siècle ou auparavant ? Quelle pourrait en être l'étymologie, qui n'aurait alors aucun lien avec une langue des Caraïbes ?
 - Je souhaiterais savoir pourquoi on ne fait pas la liaison dans l'expression *bon à rien*.

Cet espace offert à l'internaute de converser avec l'institution (on notera que dans la formulation de la réponse, c'est « L'Académie » qui répond) n'impose pas qu'une compétence linguistique élevée (surprenante alors que l'on se connecte à cette rubrique pour acquérir des compétences et corriger des manquements), les internautes font montre d'hypercorrection, non plus seulement au niveau de la langue (question sur du vocabulaire appartenant à un registre élevé, non courant ou des expressions anciennes) mais également au niveau des thèmes (questions à partir d'œuvres littéraires et non pas à partir de la vie courante, références aux éditions anciennes du dictionnaire), comme si dans leur esprit, le simple fait de communiquer avec l'institution supposait que l'objet de leur question en vaille la peine et qu'il faille mériter l'expression de la réponse institutionnelle tant sur le fond que sur la forme.

En conversant avec l'Académie, ils postulent que si la distance linguistique qui les sépare de l'institution est admissible par eux-mêmes et par cette dernière, précisément parce c'est ce qui fonde l'existence d'une rubrique comme *Dire, Ne pas dire*, la distance culturelle doit être la plus petite possible, parce que le positionnement d'expert exprimé par l'Académie va en réalité dans leur esprit probablement bien au-delà de la seule compétence linguistique. L'effort de « mise à niveau » de son interlocuteur est de mise dans l'interaction tant sur le plan linguistique que sur le plan culturel.

Dans cette posture, l'Académie peut prendre les traits de l'expert impartial et compétent. Aucune subjectivité des propos n'apparaît. Les phrases assertives organisées autour d'un temps présent s'enchaînent comme des faits non négociables.

63. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-04-06&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

La forme *c'est mon créneau* n'est pas incorrecte, si l'on donne à *créneau* le sens qu'il a dans la langue du commerce, quand il désigne un secteur encore disponible dans un marché⁶⁴

La phrase *Ce gâteau me fait fantaisie* est **incorrecte**. Ces deux mots **n'ont en effet** rien à voir l'un avec l'autre. (*ibid.*)

Il s'agit d'un tour considéré jadis comme très familier, mais plutôt senti aujourd'hui comme un peu maniéré⁶⁵

Si la liberté de la question est laissée en théorie à l'internaute, l'idéologie du bien parler associée au bien penser, bien connaître, bien cultivé, pèse ici de tout son poids sous couvert d'une démarche conversationnelle, minimale puisque construite à partir d'un tour fait d'une question et d'une réponse, librement consentie sur le fond et sur la forme.

Dans la correspondance avec les internautes, une observation s'impose : alors que le registre est rarement homogène dans l'utilisation personnelle de la langue, on observe ici un *continuum* comme si le « public » concerné adoptait la même langue et le même registre. Ce n'est évidemment pas sans faire penser à la fameuse formule de l'Académie qui dans la préface de l'édition princeps de son dictionnaire, définissait *l'honnête homme* (qui aujourd'hui ne peut évidemment plus n'être qu'un homme) comme un homme de bonne conversation (et) de bonne compagnie (Quemada, 1997 : 44), somme toute assez loin de l'objectif fixé par Yves Pouliquen de mettre en place une conversation « ouverte » et « plus spontanée »⁶⁶.

64. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-04-06&form_build_id=form-mRZcGvLtp4RTvQCZvxvzsMRnpDtcJVoLZPrkbUyQ4GI&form_id=academie_blog_search_form&op=Rechercher (consulté en mai 2023)

65. https://www.academie-francaise.fr/dire-ne-pas-dire/recherche?titre=&rubrique=368&date=2023-03-02&form_build_id=form-yVDsgXnJsD5L_ab_4NeGx5glOGFCHnyYyp7sK50_R10&form_id=academie_blog_search_form&op=Rechercher - (consulté en mai 2023)

66. <https://www.academie-francaise.fr/dire-ne-pas-dire-un-apres> (consulté en mai 2023)

7. Conclusion

Alors qu'au sein de la rubrique *Dire, Ne pas dire*, et de certaines de ses notices, le bon usage est proposé de manière injonctive, comme un idéal à atteindre, plusieurs remarques s'imposent : le discours d'expert n'est pas la seule forme de discours qui existe au sein de la rubrique. On trouve également d'autres types de discours que ce soit le discours d'apprentissage ou encore une forme certes embryonnaire mais existante de discours de prévention. Par ailleurs la rubrique n'entretient que peu de lien avec l'espace du *Dictionnaire* et affirme de ce fait une forme d'autonomie par rapport à l'espace lexicographique. Plus proche sans doute d'un territoire dictionnaire que lexicographique⁶⁷, la rubrique *Dire, Ne pas dire* reflète une orientation socio-culturelle et politique assez marquée qui en dépit de l'ouverture dont elle veut faire preuve ne s'adresse pas à tous les publics.

De toute évidence, la rubrique n'apparaît pas comme un prolongement du dictionnaire, mais plutôt comme un complément qui n'épuise pas les seules productions langagières et leurs normes. Des réalités extra linguistiques se font jour et sont autant d'occasions pour l'Académie de réaffirmer des positions plus générales sur la langue et ses missions en la matière en incluant des éléments de lexiculture plus nombreux.

L'éthos auctorial particulièrement présent permet à l'institution de se poser en experte, certes bienveillante et pédagogue. Mais, comme le dit très justement Canut,

Le sentiment d'insécurité linguistique qui devrait naturellement découler des injonctions prescriptives est paradoxal. Comme il porte souvent sur un segment de la langue, il autorise l'internaute alors qu'il s'annonce en posture d'insécurité, puisqu'en attente de corrections à ses manquements, de s'exprimer dans un registre élevé sur l'ensemble de sa correspondance exception de faite de la question qu'il pose. (2007 : 60)

Cette attitude se retrouve dans bon nombre d'interactions entre l'internaute et l'Académie comme nous l'avons signalé, et confirme que l'Académie, y compris dans cette rubrique, jouit d'une autorité certaine et d'une aura qui ne lui permet pas de s'affranchir d'un élitisme dont elle est héritière et qu'à certains égards elle continue de générer.

Cela dit, cette rubrique est l'occasion pour l'institution de faire acte de résistance et comme le rappelle Maurice Druon dans sa communication au sein de l'ouvrage publié par Bernard Quemada sur *le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*, à défaut « de

67. Comme le rappelle Pruvost, *La dictionnaire*, notion initiée par Quemada, inclut « les aspects concrets de la fabrication, de la présentation, pour un public donné, avec tous les impératifs commerciaux qui s'imposent pour plaire au public » quand la *lexicographie* « se situe en fait dans le domaine de la recherche, sans préoccupation d'une mise en valeur pour un public non initié, sans avoir le souci d'adapter le contenu à des lecteurs acheteurs d'un produit. On est en quelque sorte bien en amont du dictionnaire mis en forme pour être vendu, on se situe dans la pure recherche. » (Pruvost, 2009)

mettre un gendarme derrière chaque locuteur », d'être consulté si l'on veut parler convenablement (Quemada et Pruvost, 1998 : 460).

Elle est aussi l'occasion de surprises et de bonheurs comme le nom de la notice l'indique, qui au-delà d'une préoccupation de la langue-patrimoine d'une certaine façon modernise l'institution en lui permettant, cette fois, un partage avec des internautes curieux et passionnés tel que l'imaginait Monsieur Pouliquen, lorsqu'au sein de l'Académie française a été imaginée la rubrique *Dire, Ne pas dire*.

Cette rubrique s'inscrit donc dans une longue tradition dont il est nécessaire de ne pas nier la filiation avec les travaux issus des « remarques sur la langue française » et notamment le premier d'entre eux, celui de Vaugelas. Comme le souligne Siouffi dans un ouvrage consacré aux remarqueurs de la langue française, « [Vaugelas] a initié une pratique nouvelle du commentaire grammatical, plus souple et plus appuyé sur l'usage que ce qui était habituellement proposé dans les grammaires en bonne et due forme » (2004 : 8). Ainsi considérer que certains remarqueurs n'ont pas seulement œuvré sur la norme ou l'usage mais ont, avant l'heure, fait montre d'innovation en mêlant un discours prescriptif et préventif, notamment en s'intéressant à l'usager, n'est pas forcément un contre-sens linguistique ni historique. Cinq siècles plus tard, la rubrique *Dire, ne pas Dire* s'apparente comme l'héritière moderne et innovante de cette pratique, devenue désormais numérique.

Bibliographie

- BERRENDONNER, Alain (1982), *L'éternel grammairien – Étude du discours normatif*, Paris, Peter Lang.
- BOUVEROT, Danielle (2007), « Données syntaxiques normatives dans le TLFi » dans Gilles Siouffi et Agnès Steuckardt (dir.), *Les linguistes et la norme : aspects normatifs du discours linguistique*, Berne, Peter Lang, p. 111-122.
- BRANCA-ROSOFF, Sonia (2007), « Les normes du français en situation publique, les ambiguïtés d'une notion » dans Gilles Siouffi et Agnès Steuckardt (dir.), *Les linguistes et la norme : aspects normatifs du discours linguistique*, Berne, Peter Lang, p. 21-48.
- CANUT, Cécile (2007), « L'épilinguistique en question » dans Gilles Siouffi et Agnès Steuckardt (dir.), *Les linguistes et la norme : aspects normatifs du discours linguistique*, Berne, Peter Lang, p. 49-72.
- CARRERE D'ENCAUSSE, Hélène (2017), « Dire l'usage au XXI^e siècle », dans Hélène Carrère d'Encausse, Gabriel de Broglie, Giovanni Dotoli, Mario Selvaggio (dir.), *Le Dictionnaire de l'Académie française, Langue, Littérature, société*, Paris, Hermann, p. 5-13.
- CATACH, Laurent (2020), « Le nouveau portail numérique du *Dictionnaire de l'Académie française* », *Les Cahiers du dictionnaire*, n° 11. Dictionnaire et figement. Hommage à Salah Mejri Dictionnaires et encyclopédies. Hommage à Alain Rey, Paris, Garnier, p. 337-364.
- COLLINOT, André et Francine Mazière (1997), *Un prêt à parler : le dictionnaire*, Paris, PUF linguistique nouvelle.
- DARMESTÄDTER, Cécile (2011), « Modernité et modernisation du *Dictionnaire de l'Académie française* : quelles transformations de la huitième à la neuvième édition ? ». *Éla. Études de linguistique appliquée*, n° 163, p. 285-306.
- PICONE, Marc (2011), « Implicite du discours lexicographique : le cas de la relation synonymie/antonymie », *Synergies Pologne* n° 8, Paris, p. 183-192.
- PRUVOST, Jean (2009), « Quelques perspectives lexicographiques à mesurer à l'aune lexiculturelle », *Éla. Études de linguistique appliquée*, n° 154, p. 137-153.
- QUEMADA, Bernard, dir. (1997), *Les Préfaces du Dictionnaire de l'Académie française 1694-1992*, Paris, Honoré Champion.
- QUEMADA, Bernard et Jean Pruvost (1998), *Le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*. Paris : Honoré Champion.
- REY, Philippe (2023), *Dire, Ne pas dire. L'intégrale - Nouvelle édition augmentée, Du bon usage de la langue française*, Paris, Philippe Rey.
- ROMEYER, Hélène et Amirouche Moktefi (dir.) (2013), « Pour une approche interdisciplinaire de la prévention », *Penser la prévention : discours, acteurs et dispositifs, Communication & Langages*. n°176, p. 33-47.

SOUFFI, Gilles (2004), « Préface », *Les remarqueurs sur la langue française du XVI^e siècle à nos jours, études réunies et présentées par Philippe Carron*, La Licorne, Presse Universitaire de Rennes.

SIOUFFI, Gilles (2007a), « Du sentiment de la langue aux arts du langage », *Éla. Études de linguistique appliquée*, 147, p. 265-276.

SIOUFFI, Gilles (2007b), « Le dictionnaire françois face à la compétence des locuteurs », *L'Information Grammaticale*, n° 114, 2007. p. 11-19.

SOUFFI, Gilles (2009), « Le Dictionnaire de l'Académie française : entre bon usage et culture », *Éla. Études de linguistique appliquée*, n° 154, p. 155-176.

TITRE: ÉTUDES MÉTALEXICOGRAPHIQUES PAR ANALYSE D'ÉCHANTILLONS : TOUTE REPRÉSENTATIVITÉ EST-ELLE ILLUSOIRE ?

AUTEUR: FRANCK SAJOUS, CLLE, CNRS ET UNIVERSITÉ TOULOUSE – JEAN JAURÈS

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 239-272

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22414](http://hdl.handle.net/11143/22414)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22414](https://doi.org/10.17118/11143/22414)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Études métalexigraphiques par analyse d'échantillons : toute représentativité est-elle illusoire ?

Franck Sajous, CLLE, CNRS et Université Toulouse – Jean Jaurès

Résumé :

Le dictionnaire général étant trop volumineux pour être étudié dans son intégralité, son analyse passe généralement par l'examen d'un échantillon et par la généralisation à l'ensemble du dictionnaire des résultats observés sur cet échantillon. Si beaucoup d'énergie est consacrée à l'analyse des échantillons, les métalexigraphes semblent, en revanche, accorder peu d'importance aux mécanismes de sélection des échantillons eux-mêmes. Après avoir rappelé les principes de l'échantillonnage et les différentes techniques de sélection des échantillons, cet article montre, à travers différentes expériences, que l'examen de zones contiguës, qui a quasi-systématiquement la préférence des métalexigraphes, est à proscrire, d'autres méthodes d'échantillonnage pratiquées sur l'ensemble du dictionnaire étant toujours préférables. L'article montre également que même la méthode de sélection la moins problématique ne peut garantir totalement la représentativité d'un échantillon. Les résultats obtenus par analyse d'échantillons doivent donc être considérés avec circonspection.

Mots-clés :

Métalexigraphie, méthodes d'échantillonnage, représentativité

Abstract :

As general dictionaries are too voluminous to be studied in their entirety, their analysis generally require sample analysis. Contrary to the analysis of the selected samples, which concentrate much of metalexigraphers efforts, the sampling methods are often neglected. After recalling sampling principles and sample selection techniques, this article shows, through various experiments, that the examination of one-stretch samples, which is almost systematically preferred by metalexigraphers, is to be avoided, as sampling of the entire dictionary by using other methods always produce more reliable estimations. The article also shows that even the least problematic technique cannot fully guarantee the representativeness of a sample. Results obtained from sample analysis should therefore be treated with caution.

Keywords :

Metalexigraphy, sampling methods, representativeness

1. Introduction

L'étude des caractéristiques de la structure et du contenu du dictionnaire général, lorsqu'elle est réalisée manuellement, passe généralement par l'analyse d'un échantillon et par la généralisation à l'ensemble du dictionnaire des résultats observés sur cet échantillon, l'ensemble étant trop volumineux pour être étudié dans son intégralité¹. Bukowska (2010) observe que, si beaucoup d'énergie est consacrée à l'analyse des échantillons, très peu de réflexion semble être accordée par les métalexigraphes aux mécanismes de sélection des échantillons eux-mêmes. La méthodologie d'échantillonnage mérite pourtant réflexion, sous peine d'introduire des biais importants de sélection et, partant, d'analyse. Quelques métalexigraphes sont bien conscients de certains écueils à éviter, comme l'absence de représentativité d'un échantillon de taille trop modeste. Dans la critique des dictionnaires qu'il pratique, Corbin s'impose un ensemble de contraintes méthodologiques, notamment celle de mener des « études extensives sur des échantillons suffisamment importants pour être estimés significatifs » (1984 : 113). Nous souscrivons sans réserve à cette autodiscipline. Rappelons néanmoins qu'en statistiques, la *significativité* est un concept qui a un sens spécifique², tout comme la *représentativité* : pour qu'un échantillon puisse être estimé *représentatif* (avec une probabilité donnée), sa taille minimale (« suffisamment importante ») se calcule. Mais, plus que la taille de l'échantillon, c'est l'importance capitale de la méthode d'échantillonnage qui semble être largement méconnue des métalexigraphes.

En menant des expériences d'échantillonnage sur le dictionnaire *Usito*³, nous montrons dans cet article que l'examen de zones contiguës, qui a quasi-systématiquement la préférence des métalexigraphes, est à proscrire et que d'autres méthodes d'échantillonnage pratiquées sur l'ensemble du dictionnaire lui sont préférables. Nous montrons également qu'on ne peut, en aucun cas, garantir totalement la représentativité d'un échantillon. De telles conclusions peuvent revêtir une certaine valeur d'évidence pour les spécialistes des techniques d'échantillonnage. Nous pensons toutefois que l'étude empirique présentée dans ce chapitre se justifie par sa vocation pédagogique et par le public spécifique visé (les métalexigraphes). En effet, un parcours de la littérature sur les analyses métalexigraphiques par analyse d'échantillons suggère qu'il est nécessaire d'illustrer, par des expériences concrètes portant sur l'objet d'étude des métalexigraphes (les dictionnaires), les enjeux

1. Cette prémisse souffre quelques exceptions notables. Martinez (2013), par exemple, parcourt intégralement les éditions ou millésimes successifs des *Petit Larousse illustré* (PLI), *Petit Robert* (PR) et *Dictionnaire de l'Académie française* (DAF) afin de recenser les changements dans la macrostructure de ces dictionnaires. Corbin (1990) parcourt, à l'aide de 45 étudiants, l'intégralité des nomenclatures de cinq grands dictionnaires généraux à la recherche de noms de végétaux dérivés en *-ier* et de leur base apparente.

2. Par exemple, des tests statistiques permettent de calculer la significativité d'éventuelles différences mesurées entre deux échantillons.

3. Dictionnaire du français québécois développé par l'Université de Sherbrooke, accessible en ligne à l'adresse : <https://usito.usherbrooke.ca/>.

du choix des méthodes d'échantillonnage dans le domaine particulier de la métalexigraphie et les précautions à prendre lors de l'interprétation des résultats observés⁴.

2. Échantillon et échantillonnage : principes

Les statistiques inférentielles consistent à quantifier les caractéristiques d'une population restreinte (un échantillon) dans le but d'estimer celles de la population générale. Concernant l'analyse des dictionnaires, il s'agit de sélectionner un ensemble d'observables (pages, articles, définitions, exemples ou toute délimitation cohérente d'un segment textuel relativement au cadre de l'étude menée) et de quantifier certaines caractéristiques pour la sélection choisie. Les valeurs *réelles* des caractéristiques *mesurées* sur l'échantillon analysé servent à *estimer* les valeurs *probables* de ces caractéristiques pour l'ensemble du dictionnaire. On peut recourir à l'analyse d'échantillons en synchronie, pour décrire les caractéristiques d'un dictionnaire, comparer plusieurs dictionnaires distincts ou différentes parties d'un même dictionnaire, ou en diachronie, pour comparer différentes éditions ou différents tomes (rédigés à des périodes plus ou moins éloignées) d'un même dictionnaire. Selon les cas, il s'agit donc d'analyser un seul échantillon ou de comparer les similitudes ou les différences entre plusieurs échantillons. Mais quelle que soit la configuration, la manière de constituer les échantillons (taille et méthode de sélection) a une incidence déterminante sur la fiabilité des résultats observés et sur celle de l'estimation qui en découle. L'analyse d'échantillons porte le plus souvent sur des propriétés mesurables numériquement. C'est de ce type d'analyse dont il est question à partir de la section 2.4. Mais on trouve également des études qualitatives menées sur des corpus d'observables relativement vastes, qualifiés d'*échantillons* par les auteurs de ces études. Nous donnons trois exemples de ces cas particuliers en section 2.2. La section 2.3, consacrée aux méthodes d'échantillonnage, concerne les analyses aussi bien quantitatives que qualitatives, tandis que la section 2.4, qui aborde la question du contrôle de la représentativité, porte spécifiquement sur les analyses quantitatives (chiffrées).

2.1. Intérêt de l'échantillonnage selon le support des dictionnaires

Nous avons justifié, en introduction, l'intérêt de l'échantillonnage par la taille des dictionnaires généraux qui, le plus souvent, dissuade d'envisager leur étude exhaustive. C'est assurément le cas pour les dictionnaires imprimés. En revanche, l'intérêt de recourir à l'échantillonnage pour l'étude des dictionnaires numériques ne s'impose pas nécessairement à l'esprit, ceux-ci se prêtant aux études computationnelles qui permettent de les analyser dans leur globalité. Deux situations justifient cependant d'échantillonner sur ces dictionnaires :

4. Nous nous inscrivons en cela dans la lignée de Bukowska, dont l'article ne constitue pas un travail fondateur ou « de référence » en statistiques inférentielles, mais est l'un des rares, à notre connaissance, à problématiser la mise en œuvre et la pertinence de différentes méthodes d'échantillonnage dans le domaine spécifique de la métalexigraphie. C'est la raison pour laquelle nous référerons à ce travail au long de ce chapitre.

- lorsqu’aucun moyen (API, archive, etc.) n’est proposé pour accéder automatiquement au contenu des dictionnaires et/ou lorsque des contraintes légales empêchent leur analyse automatique ;
- lorsque l’analyse à mener est trop difficilement automatisable, comme c’est le cas pour l’étude de certains phénomènes qualitatifs. À supposer que les études mentionnées en section 2.2 aient porté sur des dictionnaires accessibles automatiquement et légalement permissifs, les analyses qualitatives menées manuellement ne seraient pas automatisables pour autant.⁵

Dans ces situations, l’étude des dictionnaires numériques est donc manuelle et requiert la constitution d’un échantillon à analyser.

2.2. Le cas particulier des études qualitatives : quantifier n’est pas – nécessairement – chiffrer (ni généraliser)

Lehmann (1995) et Corbin (1995) étudient, dans le même numéro de la revue *Lexique*, les transformations que subissent les citations littéraires d’un dictionnaire à l’autre : du *Grand Robert* (GR) au PR pour Lehmann et du PR au *Micro Robert* (MR) pour Corbin. Lehmann choisit d’analyser les articles de la lettre *N*, Corbin ceux de la lettre *F*. Afin de mettre en évidence « [l]e poids des contraintes dictionnaires sur l’évolution des marqueurs », Martinez (2011 : 44-47) se penche sur « le sort de la marque *littéraire* » à travers les dix révisions des millésimes 1997 à 2007 du PLI. Pour mener son étude, il analyse les articles de la lettre *E*. On trouve, dans les trois articles, des commentaires sur le choix des échantillons. Corbin, qui parle d’une « étude menée sur un corpus extensif dont la présente contribution va donner un aperçu nécessairement simplifié et fragmentaire, l’analyse en vraie grandeur requérant une publication plus développée » (p. 125-126), consacre un paragraphe à l’« échantillon lexicographique étudié » :

L’étude porte sur l’ensemble des entrées commençant par *F*, soit près d’un vingtième du texte lexicographique des deux dictionnaires. En valeur relative, cet échantillon paraît suffisant pour un ensemble de données très copieux à étudier : 1847 entrées dans le *PR*, 1304 dans le *MR*. Le choix de la lettre *F* n’a pas d’autre justification qu’une certaine familiarité avec cet échantillon, sur lequel a déjà été menée antérieurement une étude comparative du même ordre concernant l’utilisation des marques d’usage. (Corbin, 1995 : 126)

Lehmann consacre également une section, intitulée « Le corpus », au choix et à la description de son échantillon :

5. Pour un aperçu des problèmes que pose ce type d’analyse, lorsqu’il est appliqué aux dictionnaires (ou encyclopédies), voir dans ce collectif la section 7 de l’article intitulé « Pour une analyse qualitative et quantitative, manuelle et computationnelle, synchronique et diachronique, des dictionnaires numériques ».

Pour avoir un échantillon suffisamment vaste, on a procédé au dépouillement exhaustif d'une lettre, la lettre *N*, dans les deux dictionnaires [...]. Cette lettre a été retenue parce qu'à ce stade de la fabrication du *GR*, l'équipe rédactionnelle est constituée de manière durable et que cette même équipe est, pour l'essentiel, responsable de l'élaboration du *PR*. (Lehmann, 1995 : 108)

Elle écrit plus bas que la lettre *N* du *PR* contient 804 articles, dont 243 comportent des citations. Ces citations comportent environ un tiers de syntagmes signés (contre deux tiers de phrases), proportion qui lui semble en accord avec l'intention décrite dans la préface du dictionnaire. À propos d'une potentielle généralisation de son observation, elle écrit :

Mais l'exhaustivité du dépouillement d'une lettre reste fallacieuse et n'autorise pas une exploitation d'ordre statistique dont l'intérêt est, d'ailleurs, bien secondaire. On peut, en revanche, considérer qu'elle permet un repérage qualitatif certainement complet des différentes procédures des manipulations de la citation, l'abrègement en syntagme illustrant une de ces figures. (Lehmann, 1995 : 109)

Martinez commente lui aussi le choix de son échantillon :

[l]a lettre *E* du dictionnaire, pas plus que n'importe quel échantillon limité à une tranche alphabétique, n'est représentative de l'ensemble du texte. Nous l'avons choisie parce que les marqueurs *litt.* y foisonnent. La seule démonstration tentée étant celle du caractère désordonné et aléatoire des changements de marqueurs, les données observées et les résultats déduits ne fluctuent pas en fonction du choix de l'échantillon. (Martinez, 2011 : 45)

Martinez parcourt l'ensemble des pages de la lettre *E* de 11 millésimes (des 64 pages du *PR*1997, comprenant 170 articles marqués, aux 63 pages du *PR*2007, comprenant 197 articles marqués).

Les trois métalexigraphes ont en commun de reconnaître l'arbitraire du choix de la tranche sélectionnée, qui ne repose sur aucun calcul statistique, tout en fournissant les justifications de ce choix : une certaine familiarité pour Corbin, la stabilisation de l'équipe rédactionnelle pour Lehmann et le foisonnement des marques étudiées pour Martinez. Concernant la taille de l'échantillon analysé, Lehmann parle d'échantillon suffisamment vaste et Corbin d'échantillon suffisant et d'un ensemble de données très copieux à étudier. On peut accorder aux deux métalexigraphes que les échantillons sont vastes, et le matériau à étudier copieux, mais comment doit-on comprendre *suffisant* et *suffisamment vaste* (i.e. *suffisant* pour quoi faire ?). En statistiques, la *représentativité* d'un échantillon est une notion spécifique (cf. section 2.4) qui concerne les analyses quantitatives chiffrées. Lehmann et Corbin en ont bien conscience et ne prétendent pas démontrer mathématiquement la représentativité de leurs échantillons. Lehmann s'interdit une exploitation d'ordre statistique. Quant à Corbin, il écrit que sa contribution « va donner un aperçu nécessairement simplifié et fragmentaire » et estime que son échantillon « paraît suffisant pour prétendre à une certaine représentativité » (1995 : 126). Il n'affirme donc pas que son échantillon est représentatif. Pour autant, dire que l'assertion précédente le sous-entend ne paraît pas exagéré.

Les études de Lehmann, Corbin et Martinez, dont le but est avant tout de mettre au jour les mécanismes et les motivations qui sous-tendent certains traitements, et d'étudier les effets de ces traitements, ne tombent pas dans la catégorie des analyses quantitatives chiffrées. Lehmann écrit par ailleurs que l'intérêt d'une exploitation statistique de son échantillon serait « bien secondaire ». Après avoir précisé la taille de leur corpus (en nombre d'articles examinés), et le nombre d'observables identifiés (citations, marques), Lehmann, Corbin et Martinez mènent une étude qualitative des traitements lexicographiques, qu'ils estiment plus ou moins récurrents. Corbin, par exemple, exprime la fréquence des différentes catégories de transformation qui affectent les citations du PR1967 lors de leur insertion dans le MR1971 par des quantités relatives ou des proportions non chiffrées : « la plupart des modifications », la/les modification(s) « la plus apparente », « nombreuses et variées », « assez nombreuses », « très nombreuses », les substitutions « les plus fréquentes ». Cela confirme bien que l'établissement des proportions exactes des catégories respectives des différents mécanismes observés n'est pas l'objet prioritaire de ces études. Comme nous l'avons écrit plus haut, il s'agit avant tout de démontrer l'existence d'un phénomène qui, si ses occurrences sont « très nombreuses » dans un corpus « copieux », a de bonnes chances de ne pas être marginal. On peut néanmoins regretter, dans ces études, que le chiffrage des phénomènes observés, immédiatement disponible pour les métalexigraphes qui ont collecté et annoté leur matériau dictionnaire, ne soit pas communiqué au lectorat. En effet, publier des données chiffrées n'empêcherait aucunement les métalexigraphes de commenter les résultats avec les mêmes formulations que celles rapportées plus haut, tout en permettant au lectorat d'être en mesure de se forger sa propre opinion sur la base des quantifications fournies. Une autre clarification souhaitable de la démarche adoptée dans ces études qui n'ont pas de prétention statistique serait de renoncer à emprunter la terminologie de ce domaine (*significativité*, *représentativité*) et à abandonner le terme d'*échantillon* au profit de *sous-corpus*.

2.3. Méthodes d'échantillonnage

Avant de passer en revue les différentes familles de méthodes d'échantillonnage existantes, tentons d'en donner une idée intuitive avec une illustration par l'image. Imaginons que l'on présente à des sujets différents échantillons composés de 25 % des pixels d'une image et qu'on leur demande de se faire une idée de ce que représente l'image d'où chaque échantillon est issu. Les sujets à qui l'on soumet les échantillons représentés dans les figures 1a et 1b se feront probablement des idées assez similaires de l'image globale et ils n'auront qu'à moitié raison. Celui à qui l'on présente l'image 1c aura du mal à imaginer quoi que ce soit. Face à l'échantillon 1d, le sujet aura probablement en tête une image radicalement différente de celle que se figurent les sujets à qui l'on a présenté les échantillons 1a et 1b et, comme eux, n'aura qu'à moitié raison. L'échantillon 1e, qui correspond à un tirage aléatoire d'un nombre de pixels identique à celui des autres échantillons, donne une idée plus globale, mais néanmoins imprécise, de l'image réelle⁶.

6. L'image utilisée provient du site <https://publicdomainvectors.org>. Merci à la personne qui l'a placée dans le domaine public.

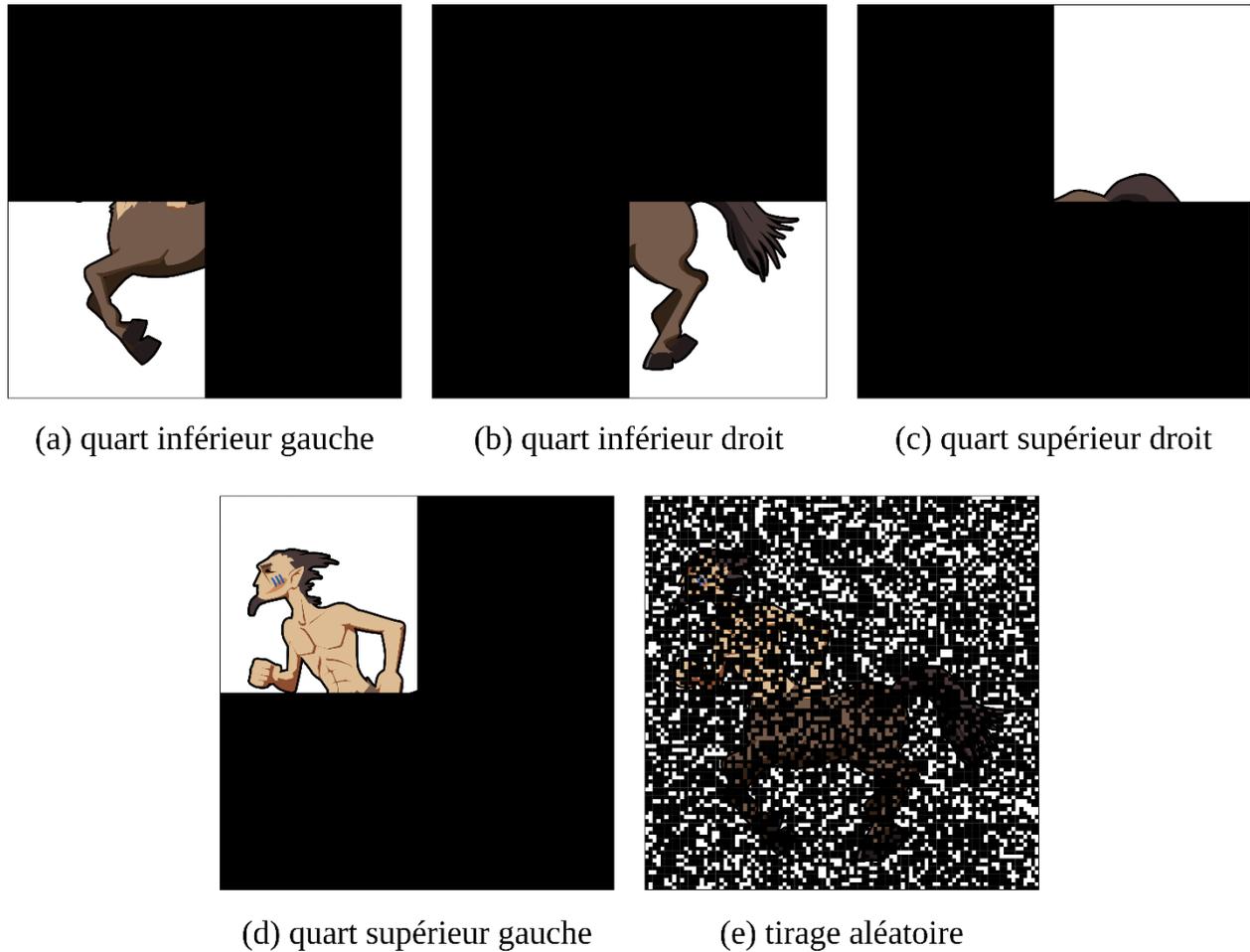


Figure 1 : Échantillons issus d'une même image (25 % des pixels)

La première leçon à tirer de cette illustration est que la méthode d'échantillonnage choisie a une incidence certaine sur l'estimation que l'on peut se faire de la réalité après examen d'un échantillon. La seconde est que même un échantillon de taille inhabituellement importante (les échantillons sont constitués ici de 25 % de la totalité des observables) n'offre pas la garantie de fournir une estimation fiable du tableau général. Notons qu'en métalexigraphie, la taille des échantillons analysés se situe généralement entre 1 et 2 % de celle du dictionnaire⁷.

Deux échantillons, constitués chacun de 1 % des pixels de notre image d'illustration sont représentés en figure 2.

7. On trouve néanmoins des études d'une plus grande ampleur. Dans son analyse du « recyclage des définitions » mentionnée plus haut, Corbin (1995) analyse par exemple 4,6 % du PR (89 pages sur 1938) et 4,8 % du MR (56 pages sur 1155). On comprend dans ce cas que, même sans en faire la démonstration mathématique, Corbin puisse être tenté de « prétendre à une certaine représentativité ».

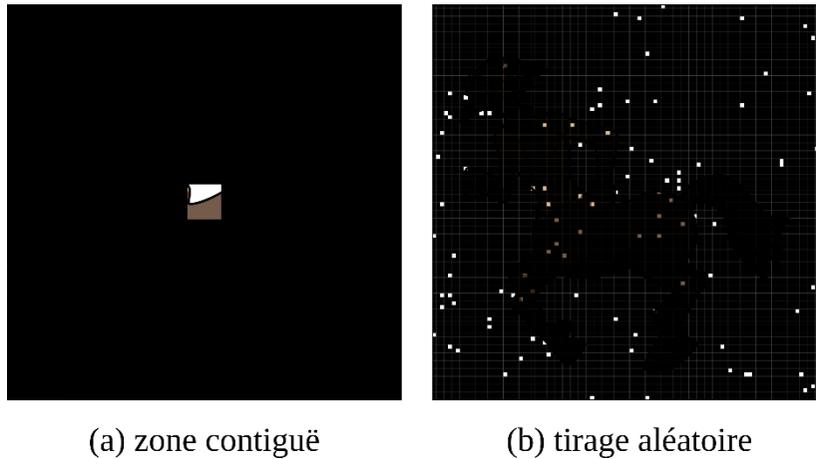


Figure 2 : Échantillons issus d'une même image (1 % des pixels)

Qu'ils soient constitués par sélection contiguë (fig. 2a) ou par tirage aléatoire (fig. 2b), il est difficile de prétendre qu'un quelconque échantillon de cette taille puisse être *représentatif* et que l'on puisse généraliser les observations effectuées.

2.3.1. Sélection par zone contiguë

La sélection d'articles contigus au sein d'une tranche donnée du dictionnaire, comme celles mises en œuvre dans les études qualitatives mentionnées en section 2.2, est également couramment pratiquée dans les analyses quantitatives chiffrées ayant vocation à généraliser les résultats mesurés sur échantillon. Pourtant, cette méthode d'échantillonnage est celle qui implique les biais les plus importants, qu'ils émanent des propriétés inhérentes au lexique de la langue étudiée ou d'artéfacts liés au processus rédactionnel. Dans les projets au long cours, différents événements peuvent advenir, comme des changements de direction éditoriale, de l'équipe de rédaction, de corpus ou de pratiques lexicographiques. Ainsi, généraliser à l'ensemble d'un dictionnaire les observations menées sur son premier (ou dernier) tome est hasardeux. C'est la raison pour laquelle certains métalexicographes considèrent que, pour étudier un dictionnaire conçu sur une période étendue et selon une chronologie suivant l'ordre alphabétique, les lettres du milieu de l'alphabet seraient les mieux adaptées pour servir d'échantillon. Les lexicographes adopteraient en effet un *modus operandi* régulier au moment où cette zone est atteinte, entre rodage initial et accélération finale due à la pression financière et temporelle. En particulier, les traitements opérés dans les articles situés en fin d'alphabet seraient moins fouillés qu'au début du dictionnaire et la sélection des mots pour inclusion dans la nomenclature deviendrait plus sévère à mesure de l'avancement du projet. C'est ce qu'Osselton (2007) nomme « alphabet fatigue ». Il observe ce phénomène sur tous les dictionnaires britanniques du XVII^e siècle, mais l'explique finalement par une autre raison qu'une évolution du processus éditorial : le fait que ces dictionnaires aient tous bâti leur nomenclature à partir d'une même liste de mots. À l'inverse de l'*alphabet fatigue*, pour les dictionnaires historiques des XIX^e et XX^e siècles (e.g. l'*Oxford English Dictionary*, mais également un dictionnaire allemand et un autre danois), l'assurance d'un finance-

ment institutionnel pérenne a permis aux lexicographes d'affiner les traitements avec l'avancée de la rédaction. L'évolution supposée du processus lexicographique n'est donc pas systématique et ne va pas toujours dans le même sens. Par ailleurs, Bukowska (2010) montre à travers l'étude du *Webster's Revised Unabridged Dictionary* (1913) que les lettres du milieu de l'alphabet ne sont pas nécessairement le reflet de pratiques rédactionnelles stabilisées (elle observe, pour la lettre *M*, une nette baisse du nombre de citations). Cette croyance en une stabilisation du processus éditorial en milieu d'alphabet – ou à mi-projet – est pourtant tenace, comme en témoigne la justification du choix de la lettre *N* dans une étude récente : « *N* was chosen as a range well on in the alphabet, at which point editorial practices, in the still pervasive alphabetical editing order, can be considered as stabilized » (Ferrett et Dollinger, 2021). Clarifions notre propos. Il ne s'agit pas de nier le fait que des changements interviennent au cours de la conception d'un dictionnaire et que, notamment, la densité de certaines informations varie parfois à mesure de l'avancement du projet⁸. Ce qu'il convient selon nous de réfuter à la suite de Bukowska, c'est qu'il existerait une zone de l'alphabet à partir de laquelle il serait particulièrement propice de tirer un échantillon.

2.3.2. Tirage aléatoire

La constitution d'un échantillon par tirage aléatoire consiste à sélectionner au hasard le nombre d'observables requis pour atteindre la taille d'échantillon souhaitée. Cette sélection peut se faire dans une zone particulière (e.g. une lettre donnée ou un tome particulier) ou sur l'ensemble du dictionnaire. La mise en œuvre peut s'appuyer, dans les dictionnaires papier, sur la sélection aléatoire d'un numéro de page, puis, au sein de la page, sur celle de l'article, etc. Par exemple, pour comparer automatiquement⁹ les styles définitoires de six dictionnaires anglais édités entre la fin du XVIII^e siècle et le début du XXI^e siècle, Kamiński (2015) constitue, pour chaque dictionnaire, des échantillons en se fondant sur un tirage aléatoire des numéros de pages. Les dictionnaires numériques présentent différents cas de figure. *Usito* et le *Trésor de la langue française informatisé* (TLFi) donnent accès à l'ensemble de leur nomenclature¹⁰, à partir de laquelle un tirage peut être effectué. D'autres, comme le *Wiktionnaire*, disposent d'un lien *Page au hasard*¹¹. Pour les dictionnaires dépourvus de ces fonctionnalités, d'autres stratégies doivent être envisagées, en fonction des contraintes légales qu'ils imposent.

8. Cela vaut non seulement pour la densité, mais également pour la nature des informations, qui peut varier tout au long du processus lexicographique, comme le montre Rademacher (2004) dans son étude du *Trésor de la langue française*.

9. La comparaison, fondée sur le calcul des fréquences de n-grammes de mots dans les définitions, et sur un clustering de ces n-grammes, est quantitative.

10. Cf. section 3 pour *Usito* et la page <http://atilf.atilf.fr/tlf.htm> pour le TLFi.

11. Le *Wiktionnaire* met en libre accès une archive de l'intégralité de ses articles (*dump*) et peut donc être analysé automatiquement dans son ensemble, sans recourir à l'échantillonnage.

2.3.3. Échantillonnage probabiliste stratifié

L'échantillonnage probabiliste stratifié est un raffinement du tirage aléatoire qui consiste d'abord à diviser le dictionnaire en zones non chevauchantes (les strates) correspondant par exemple aux lettres initiales des vedettes ou à des parties rédigées par des éditeurs différents, puis à produire aléatoirement des échantillons qui respectent certaines proportions, relativement à celles du dictionnaire (e.g. parties du discours, marquage) ou à celles calculées en corpus (e.g. parties du discours, rangs de fréquence). Sa mise en œuvre n'est pas toujours possible, dans la mesure où l'on ne connaît pas nécessairement les proportions de telle ou telle caractéristique pour l'ensemble du dictionnaire (e.g. proportions respectives des parties du discours, de la lettre initiale des vedettes...). Les expériences menées par Bukowska (2010) montrent que cette méthode n'améliore pas systématiquement les résultats obtenus par tirage aléatoire.

2.3.4. Échantillonnage systématique

L'échantillonnage systématique consiste à sélectionner un observable tous les N à partir d'un point de départ choisi arbitrairement ou tiré aléatoirement. Selon Freeman (1963), cité par Bukowska (2010), il pose un problème d'ordre théorique, que nous n'étayerons pas ici : le fait que la théorie des probabilités et les statistiques inférentielles ont peu à dire sur la confiance que l'on peut accorder à un échantillonnage non aléatoire. Il est néanmoins utilisé (quoique rarement à notre connaissance) en métalexigraphie. Par exemple, Cormier et Fernandez (2005) le mettent en œuvre afin d'évaluer la potentielle influence de la nomenclature anglaise du *Great French Dictionary* (1688) de Guy Miège sur celui du *Royal Dictionary* (1699) d'Abel Boyer. L'échantillon étudié, qui représente 5 % de la partie anglais-français du dictionnaire royal, est constitué en sélectionnant une page sur vingt, à partir de la page 5, tirée au hasard.

2.3.5. Pratique dominante

Malgré les mises en garde de Bukowska contre la constitution d'échantillons par zones contiguës, les métalexigraphes semblent continuer de privilégier cette méthode d'échantillonnage. Par exemple, pour étudier les mots composés dans deux dictionnaires du norvégien *nynorsk* et *bokmål*, Paulsen (2023) sélectionne cinq échantillons composés de zones contiguës du début de l'alphabet. Francœur (2021), qui mène une étude quantitative et qualitative sur deux dictionnaires bilingues du XVII^e siècle afin d'étudier un possible plagiat portant sur la partie anglais-français, travaille sur les 1061 articles de la lettre *F* du dictionnaire le plus récent. Le choix de cette tranche, utilisée pour son étude quantitative de la nomenclature, n'est pas justifiée par l'auteur. L'étude qualitative (chiffrée), qui consiste à comparer les éléments microstructurels des articles des deux dictionnaires, porte sur les 100 premiers articles de la lettre *F*.

Il est probable que ces pratiques ne soient pas guidées pas une remise en cause du travail de Bukowska, ou plus généralement d'un consensus en statistiques inférentielles, mais reflètent plutôt la méconnaissance de l'un et de l'autre. La sélection par zone contiguë relève plus d'un réflexe hérité d'une tradition métalexigraphique que d'un choix conscient. Le seul cas d'échantillonnage que nous avons trouvé où la méthode de sélection par zone contiguë est justifiée par les auteurs est celui de Ferrett et Dollinger (2021), cité en section 2.3.1, qui réfèrent à l'article de Bukowska et justifient leur renoncement à un échantillonnage sur l'ensemble des zones du dictionnaire par la complexité de mise en œuvre et une possible inadéquation avec l'étude menée.

Dans ce contexte, la sélection d'échantillon par tirage aléatoire fait figure d'exception. Podhajecka (2015), qui tente de mettre au jour le processus éditorial (en particulier les sources utilisées) d'un dictionnaire bilingue anglais-polonais du XIX^e siècle, réfère également à Bukowska et sélectionne un échantillon constitué aléatoirement à partir de chaque lettre de l'alphabet. Il est cependant étonnant de ne trouver aucune information sur la taille de l'échantillon, la proportion d'observables tirés de chaque lettre, etc. L'échantillon n'apportant pas de preuve probante de l'utilisation d'autres sources que celles mentionnées par l'auteur du dictionnaire, elle parcourt manuellement l'ensemble du dictionnaire à la recherche d'indices qui pourraient attirer son attention. La conclusion de l'autrice à ce stade de l'article et la démarche adoptée par la suite sont quelque peu déroutantes : « the sample alone did not provide me with strong enough evidence. As combining two different methods has invalidated the quantitative results, only the qualitative ones will be referred to in the sections below » (2015). L'abandon de l'échantillon analysé ne plaide pas contre l'échantillonnage par tirage aléatoire mais trahit un rapport problématique aux données et aux statistiques : soit, une fois une méthodologie définie, on se fie à l'échantillon constitué et analysé, soit la méthodologie doit être remise en cause. On ne peut en effet justifier l'abandon d'un échantillon par le fait qu'il ne fournit pas les preuves qui valideraient une hypothèse de départ.

2.4. Représentativité et estimation par intervalles

Bukowska critique l'absence de contrôle des métalexigraphes sur la représentativité des échantillons :

Most of the samples in current metalexigraphic research are judgmental one-stretch samples based on what metalexigraphers intuitively consider reliable and representative, usually without having tested this representativeness in any way. (2010)

Outre la méthode de sélection, le critère proposé par Bukowska pour « contrôler » la fiabilité d'un échantillon est le calcul de l'intervalle de confiance. Ce concept statistique popularisé par les sondages électoraux permet d'estimer, avec un certain niveau de confiance, que la valeur réelle d'une caractéristique de la population est comprise dans un intervalle situé autour de la valeur de cette même caractéristique mesurée sur l'échantillon. L'amplitude de l'intervalle dépend de la marge d'er-

reur. Lorsque la caractéristique à observer est une proportion (e.g. proportion d'articles marqués, intention de votes pour un candidat...), on peut formaliser l'énoncé précédent de la façon suivante¹² :

$$p_r \in [p_o - m_e; p_o + m_e]$$

$$m_e = z \times \text{erreur - type} = z \times \sqrt{\left(\frac{p_o (1 - p_o)}{n} \right)}$$

Avec :

- p_o : proportion observée dans l'échantillon ;
- p_r : proportion réelle dans la population ;
- n : taille de l'échantillon ;
- m_e : marge d'erreur ;
- z : coefficient correspondant au niveau de confiance souhaité dans la table de la loi normale centrée réduite (e.g. 1,96 pour un niveau de confiance de 95 % ou 2,58 pour un niveau de confiance de 99 %).

Nous illustrons l'estimation de la représentativité par intervalles de confiance en section 3.4.

3. Expériences : marquage FIG. et FAM. dans *Usito*

Afin de comparer les pertinences respectives des différentes méthodes d'échantillonnage décrites en section 2, nous nous intéressons ci-après à la proportion d'articles contenant les marques FAM. (familier) et FIG. (figuré) dans le dictionnaire *Usito*. Ces marques ne sont pas spécialement plus intéressantes que d'autres phénomènes qui se prêteraient à des expériences de quantification, mais elles présentent l'avantage d'être suffisamment répandues dans le dictionnaire et de constituer un observable factuel, donc aisément identifiable automatiquement.

12. Ce calcul s'applique lorsque la distribution de la variable étudiée est normale. Lorsqu'elle ne l'est pas, le calcul s'applique également sous les conditions suivantes : $n \geq 30$, $np_o \geq 5$ et $n(1 - p_o) \geq 5$. Lorsque la caractéristique observée est une σ moyenne, l'erreur-type s'obtient en divisant l'écart-type des moyennes de l'échantillon par la racine carrée de sa taille : $\frac{\sigma}{\sqrt{n}}$

En septembre 2022, la section « Tous les mots du dictionnaire¹³ » comptait 46 364 vedettes. Parmi elles, nous avons restreint notre corpus aux 31 310 noms, non pour limiter les calculs à effectuer, mais pour éliminer un potentiel facteur explicatif des divergences observées entre les échantillons, notamment entre ceux issus de différentes tranches alphabétiques (cf. section 3.1). En effet, à supposer que la probabilité qu'un article soit marqué dépende de la partie du discours de sa vedette (que cette hypothèse soit fondée ou non), on pourrait expliquer les différences de proportions d'articles marqués observées entre plusieurs échantillons issus de tranches différentes par l'inégale répartition des parties du discours dans l'alphabet. Or, dans la présente étude, seule l'incidence de la méthode d'échantillonnage sur les résultats observés nous intéresse « toutes choses étant égales par ailleurs ».

Pour chacune des deux marques, il s'agit dans les expériences qui suivent, après avoir calculé automatiquement la proportion d'articles marqués pour l'ensemble du corpus, de générer automatiquement un certain nombre d'échantillons, selon différentes méthodes (en faisant varier un seul facteur à la fois), puis de mesurer pour chaque échantillon la proportion d'articles marqués. En simulant ainsi automatiquement, de façon répétée, les résultats que l'on pourrait obtenir de manière manuelle, nous pourrions nous faire une idée des méthodes d'échantillonnage qui permettent d'atteindre les résultats les plus proches de la réalité (*i.e.* de la proportion réelle des articles marqués parmi les noms de l'ensemble du dictionnaire).

Nous comparons en section 3.1 les résultats des méthodes d'échantillonnage par zone contiguë et par tirage aléatoire, lorsque la sélection est opérée dans une tranche alphabétique donnée. Nous répétons cette expérience en section 3.2 en appliquant ces deux mêmes méthodes à une sélection pratiquée sur l'ensemble du dictionnaire afin de comparer les résultats ainsi obtenus à ceux des échantillons issus des tranches alphabétiques. Nous testons en section 3.3 les échantillonnages systématique et probabiliste stratifié, que nous comparons à la meilleure des méthodes identifiées précédemment. Enfin, en section 3.4, nous discutons de manière empirique de la possibilité de « contrôler » la représentativité des échantillons.

3.1. Sélection par zone contiguë vs tirage aléatoire

Partant du constat que les métalexicographes tirent généralement leurs échantillons au sein d'une tranche donnée, nous avons, afin de reproduire leur démarche, divisé *Usito* en tranches, préalablement à la génération des échantillons (contigus et aléatoires), chaque tranche étant composée des

13. Cette rubrique ne recense pas les sous-entrées : on y trouve *affirmative*, mais pas *dans l'affirmative*, sous-entrée présente sous *affirmative*. La liste des vedettes est disponible à l'URL suivante : <https://usito.usherbrooke.ca/index/mots>. Étonnamment, c'est la rubrique intitulée « Tous les articles du dictionnaire », disponible à l'URL <https://usito.usherbrooke.ca/index/articles> qui recense les 75 109 vedettes, sous-entrées et renvois. Corbin (2020) relève, en juillet 2020, 73 000 items indexés dans la rubrique « Tous les mots du dictionnaire », alors accessible à l'URL <https://usito.usherbrooke.ca/index/mots>. Il semble donc que les listes correspondant aux deux rubriques aient été interverties – à tort – entre les deux périodes.

noms commençant par la même lettre initiale. Le tableau 1 donne le nombre d'articles que contient chaque tranche et le pourcentage que ce nombre représente par rapport à la totalité du corpus.

Tableau 1 : Répartition par tranches des 31310 noms (vedettes) d'*Usito*

Tranche	Nb articles	% articles	Tranche	Nb articles	% articles
a	3350	10,70	n	647	2,07
b	1909	6,10	o	700	2,24
c	3855	12,31	p	3248	10,37
d	1826	5,83	q	172	0,55
e	1004	3,21	r	1669	5,33
f	1234	3,94	s	2305	7,36
g	1156	3,69	t	1667	5,32
h	842	2,69	u	154	0,49
i	1061	3,39	v	799	2,55
j	290	0,93	w	61	0,19
k	165	0,53	x	24	0,08
l	945	3,02	y	46	0,15
m	2084	6,66	z	97	0,31

Sans surprise, les tranches sont de tailles très variables. Une première question, pour un métaléxicographe menant une étude manuelle, serait de savoir de quelle tranche tirer un échantillon, et si ce choix a une incidence sur la fiabilité (la représentativité) des résultats observés (*i.e.* un échantillon issu d'une tranche donnée est-il plus susceptible d'être représentatif de l'ensemble du dictionnaire qu'un échantillon issu d'une autre tranche ?).

Le diagramme en barres de la figure 3 représente, pour chaque tranche, la proportion d'articles comportant la marque FIG.

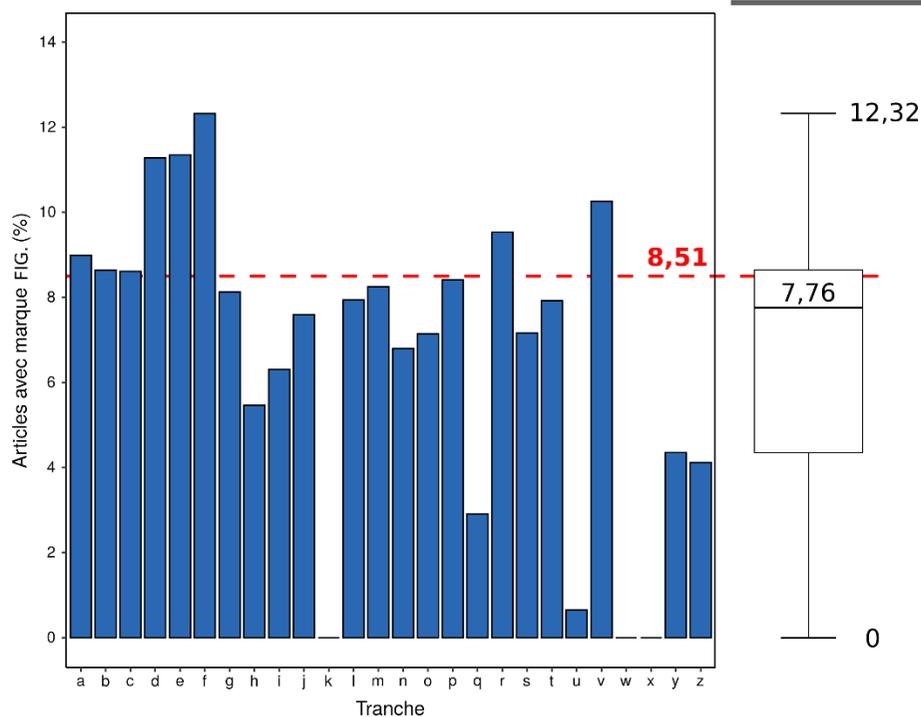


Figure 3 : Proportion d'articles du corpus contenant la marque FIG.

La boîte à moustaches représente la variabilité de ces proportions pour l'ensemble des tranches. Parce que nous traitons automatiquement le dictionnaire dans sa globalité, nous pouvons calculer la proportion réelle d'articles marqués dans notre corpus. Sur les 31 310 articles, 2663 contiennent la marque, soit 8,51 %. Cette proportion est représentée par la ligne horizontale pointillée dans la figure 3. Les proportions d'articles marqués sont très variables d'une tranche à l'autre puisqu'elles fluctuent de 0 à plus de 12 %. Les tranches affichant une proportion de 0 % sont celles qui comportent le moins d'articles. Une grande variabilité s'observe néanmoins entre les « grosses » tranches. On peut donc supposer que le choix de la tranche à partir de laquelle est tiré un échantillon aura un impact sur le résultat. Une lecture intuitive de ce diagramme indique que, pour ce dictionnaire et pour le phénomène observé, les tranches les plus représentatives (et donc, celles dont il conviendrait de tirer les échantillons) sont les tranches *b*, *c* et *p*. Or, d'une part, les métalexigraphes qui mènent une étude manuelle n'ont pas accès à cette information. D'autre part, cette lecture est également quelque peu naïve ou, du moins, vaine car non recyclable. Peut-on en effet imaginer que ces tranches soient également les plus représentatives quels que soient le dictionnaire et le phénomène étudiés ? Comparons, dans *Usito*, avec la marque FAM. : 3106 noms sont marqués, sur les 31 310 du corpus, soit 9,92 %. Le diagramme en barres de la figure 4 montre, pour chaque tranche, la proportion d'articles marqués FIG. et FAM., ainsi que les proportions réelles calculées sur la globalité des noms du dictionnaire.

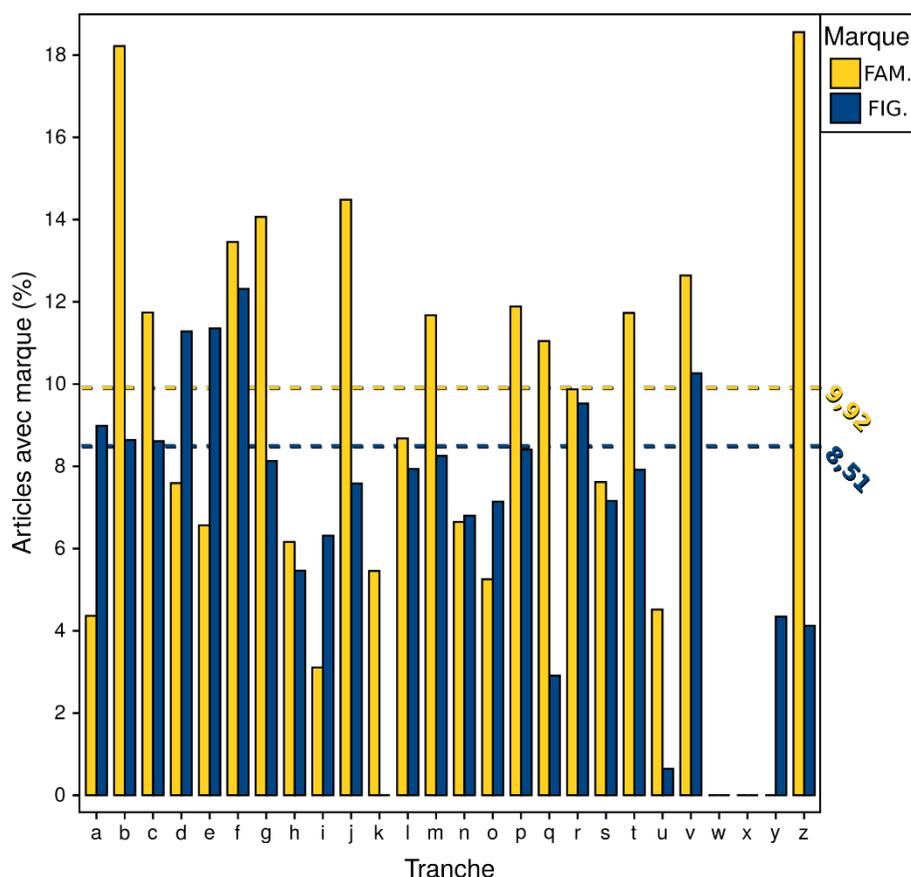


Figure 4 : Proportion, globale et par tranche, d'articles marqués FIG. et FAM.

Comme pour la marque FIG., on constate que les proportions d'articles marqués FAM. fluctuent énormément d'une tranche à l'autre : entre 3 % et 18,5 % hors « petites tranches », *i.e.* entre moins du tiers et un peu moins du double de la valeur réelle calculée sur l'ensemble du dictionnaire. On constate surtout que les « bonnes » et les « mauvaises » tranches, *i.e.* celles dont les proportions sont respectivement les plus proches et les plus éloignées de la valeur globale réelle, ne sont pas les mêmes que pour la marque FIG. Par exemple, la tranche *b*, qui est pour la marque FIG. l'une de celles dont la proportion d'articles marqués est la plus proche de celle du dictionnaire dans sa globalité, est la tranche dont la proportion est la plus éloignée de la valeur réelle pour le marquage FAM. Cette observation confirme qu'il n'existe pas de « bonne tranche » qui serait plus représentative que les autres pour tous les phénomènes observables.

Afin de comparer les méthodes d'échantillonnage, l'étape suivante consiste à produire les échantillons. Pour chaque tranche, sont générés automatiquement :

- 100 échantillons de 500 articles (maximum¹⁴) contigus, l'article de départ étant tiré aléatoirement ;
- 100 échantillons de 500 articles (maximum), tous tirés aléatoirement.

Le nombre et la taille des échantillons choisis sont arbitraires mais nous semblent raisonnables. Un échantillon d'une taille de 500 articles à analyser (pour 31 310 noms, soit 1,6 % du corpus) semble en effet cohérent avec l'ordre de grandeur des échantillons constitués dans les analyses quantitatives manuelles que l'on trouve dans la littérature. Un nombre de 100 échantillons par tranche, qui correspond à une expérience où l'on demanderait à un métalexigraphe de répéter 100 fois son analyse, avec un nouvel échantillon à chaque fois, assure en outre une variabilité suffisante pour mener les comparaisons entre méthodes d'échantillonnage.¹⁵ Pour chaque échantillon, la proportion d'articles marqués est ensuite mesurée. À titre d'exemple, le diagramme en barres de la figure 5 représente les proportions d'articles portant la marque FIG. dans les 100 échantillons composés d'articles contigus de la tranche *a* (la ligne pointillée rouge d'ordonnée 8,51 représente la proportion pour le dictionnaire ; la ligne jaune en pointillés mixtes d'ordonnée 8,99 représente la proportion de la tranche *a*).

14. Lorsqu'une tranche contient moins de 500 articles, l'échantillon est constitué de l'ensemble de la tranche. Dans ce cas, les 100 échantillons, tous identiques, afficheront la même proportion d'articles marqués, *i.e.* celle de la tranche entière.

15. Les expériences ont été réitérées en faisant varier le nombre et la taille des échantillons. Comme on peut s'y attendre, plus la taille des échantillons est importante, plus les proportions obtenues sont proches de la proportion réelle, et plus le nombre d'échantillons est élevé, plus la variabilité des distributions est grande. Pour le reste, les résultats obtenus vont toujours dans la même direction que celle des résultats présentés ci-après.

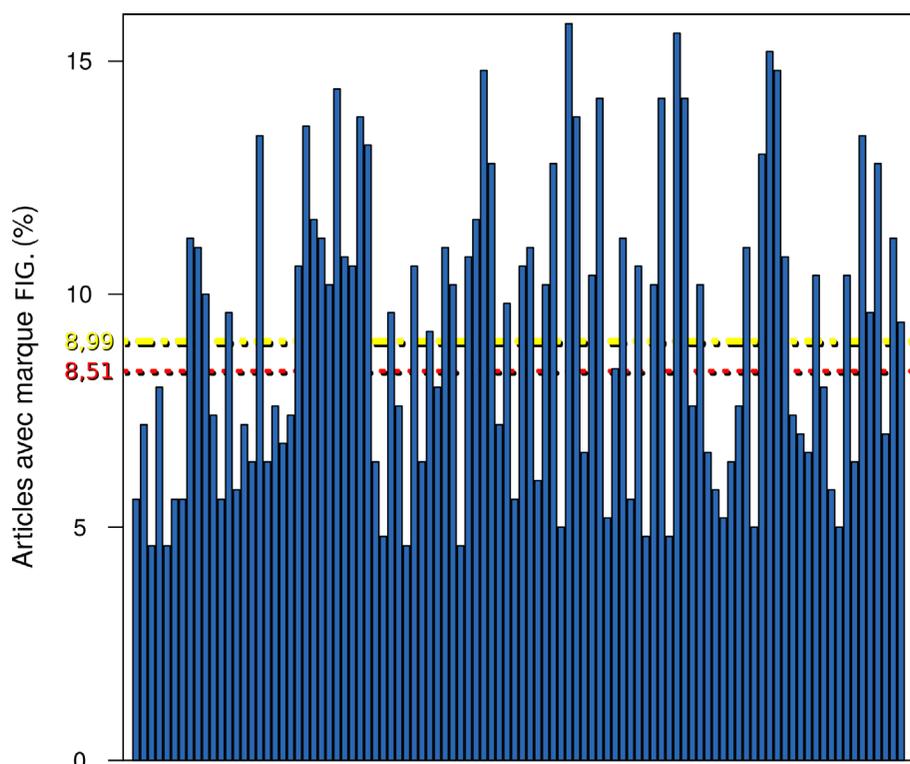


Figure 5 : Proportion d'articles marqués FIG. pour les 100 échantillons de 500 articles contigus de la tranche a

Notons pour cet exemple à quel point les valeurs obtenues fluctuent d'un échantillon à l'autre (moins de 5 % à plus de 15 %, alors que les valeurs réelles pour cette tranche et pour l'ensemble du dictionnaire sont de 8,99 % et 8,51 %), bien qu'ils soient tirés de la même tranche et par la même méthode. C'est le simple fait de sélectionner dans la tranche un point de départ différent pour constituer les échantillons d'articles contigus qui est responsable de cette importante variabilité.

Il reste maintenant à comparer les résultats obtenus par les deux méthodes d'échantillonnage. Dans la figure 6, qui reprend la figure 3, les boîtes à moustaches représentent, pour chaque tranche, la variabilité des proportions calculées pour les échantillons d'articles contigus (en marron, boîte de gauche) et pour ceux obtenus par tirages aléatoires (en jaune, boîte de droite).

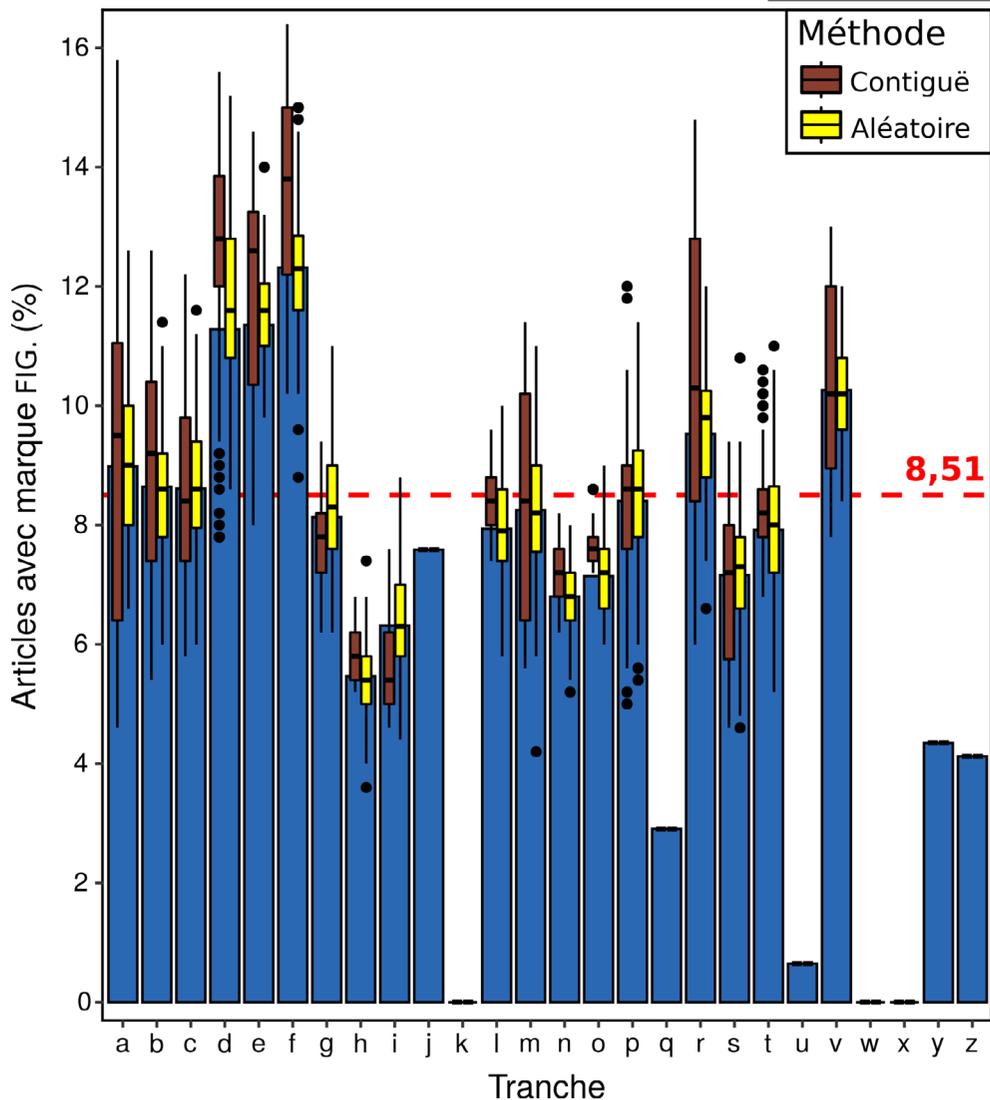


Figure 6 : Variabilité des proportions d'articles contenant la marque FIG. dans les échantillons, par tranche et par méthode d'échantillonnage

Chaque boîte à moustaches représente donc la variabilité des proportions pour les 100 échantillons générés par une méthode donnée au sein d'une tranche donnée. Par exemple, c'est la variabilité des 100 échantillons dont les proportions sont représentées en figure 5 (pour rappel : 100 échantillons de 500 articles contigus issus de la tranche *a*) qui est représentée par la boîte à moustaches marron la plus à gauche de la figure 6. Les tranches pour lesquelles les boîtes à moustaches sont réduites à un unique trait horizontal (ce qui indique une absence de variabilité) sont celles dont la taille est inférieure à 500 (et dont les échantillons extraits sont tous identiques).

Concernant les méthodes d'échantillonnage, on observe que les boîtes à moustaches marron, qui correspondent à la sélection contiguë, sont plus étalées verticalement que les boîtes à moustaches jaunes (sélection aléatoire). Les résultats sont donc moins homogènes pour la première méthode, ce qui signifie qu'en tirant deux échantillons d'articles contigus au sein d'une même tranche, la diffé-

rence entre leurs proportions respectives observées aura tendance à être plus marquée qu'entre deux échantillons tirés aléatoirement. La méthode d'échantillonnage par zone contiguë est donc moins robuste que celle par tirage aléatoire. Plus important encore, on constate que les boîtes à moustaches correspondant à la sélection contiguë sont moins centrées sur les valeurs réelles des tranches et offrent donc des estimations moins fiables (*i.e.* moins proches de la réalité). Ces deux observations nous conduisent à abandonner la méthode par zone contiguë dans la suite des expériences.

3.2. Source des échantillons : tranche vs globalité du dictionnaire

Si l'on se concentre sur la meilleure des deux méthodes étudiées en section 3.1 (le tirage aléatoire), on observe que, même pour les « grosses » tranches, la variabilité des proportions mesurée entre les différents échantillons reste importante (cf. fig. 6). La proportion obtenue pour un échantillon pourra donc conduire à une estimation éloignée de la réalité, qu'on la compare à la proportion réelle de la tranche dont est issu l'échantillon ou à la proportion réelle calculée sur l'ensemble du dictionnaire : alors que cette dernière est de 8,51 %, les valeurs obtenues pour les échantillons tirés aléatoirement, pour les « grosses » tranches, oscillent entre moins de 4 % et plus de 15 %, c'est-à-dire entre quasiment la moitié et le double de la valeur réelle.

Nous avons jusqu'ici sélectionné chaque échantillon au sein d'une même tranche car, comme nous l'avons écrit en section 3.1, c'est la manière dont les métalexigraphes procèdent la plupart du temps. Nous proposons de comparer les résultats déjà observés à ceux obtenus en générant, pour chacune des deux marques FIG. et FAM., 100 échantillons supplémentaires sélectionnés aléatoirement à partir de l'ensemble des noms du dictionnaire, indépendamment de leur initiale.¹⁶ Les caractéristiques (quartiles, valeurs minimales, médianes, moyennes et maximales) des distributions correspondantes sont données dans le tableau 2.

Tableau 2 : Tirage aléatoire sur l'ensemble du dictionnaire : caractéristiques des distributions

Marque	% réel global	Distribution des échantillons					
		Min.	Q1	Médiane	Moyenne	Q3	Max.
FIG.	8,51	5,80	7,60	8,00	8,31	9,00	12,00
FAM.	9,92	6,40	9,15	9,90	9,82	10,60	12,40

Les figures 7 et 8 représentent quant à elles la variabilité des proportions d'articles marqués dans les échantillons ainsi générés à partir de la totalité du dictionnaire, en comparaison de la variabilité de ceux tirés au sein des différentes tranches.

16. Dans le cas de l'analyse manuelle d'un dictionnaire papier, une mise en œuvre possible est de tirer aléatoirement un numéro de page, puis l'article dans la page, et d'itérer autant de fois que d'observables à sélectionner.

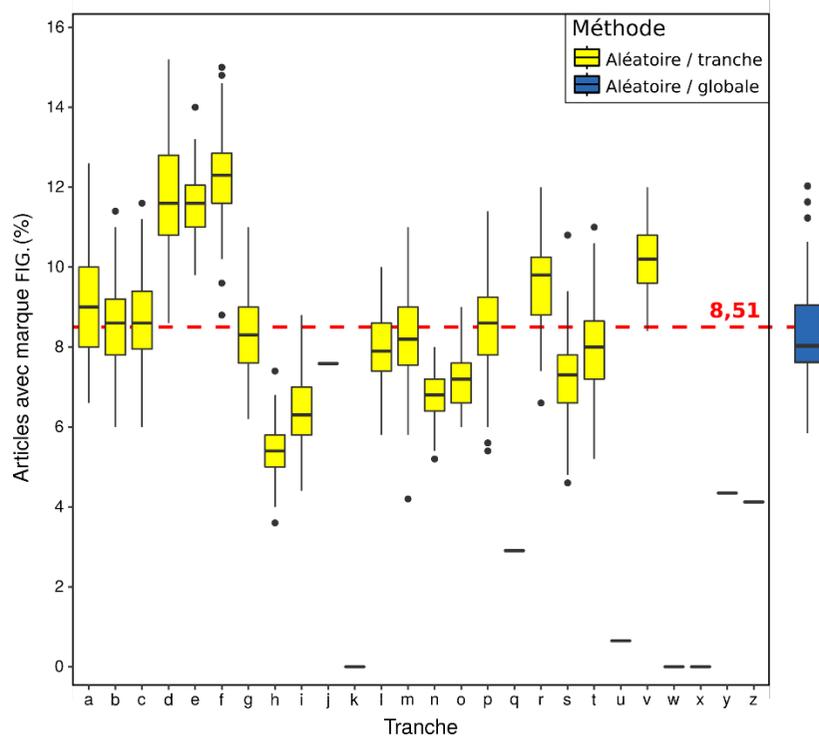


Figure 7 : Marque FIG. – échantillonnage par tranche vs sur tout le dictionnaire

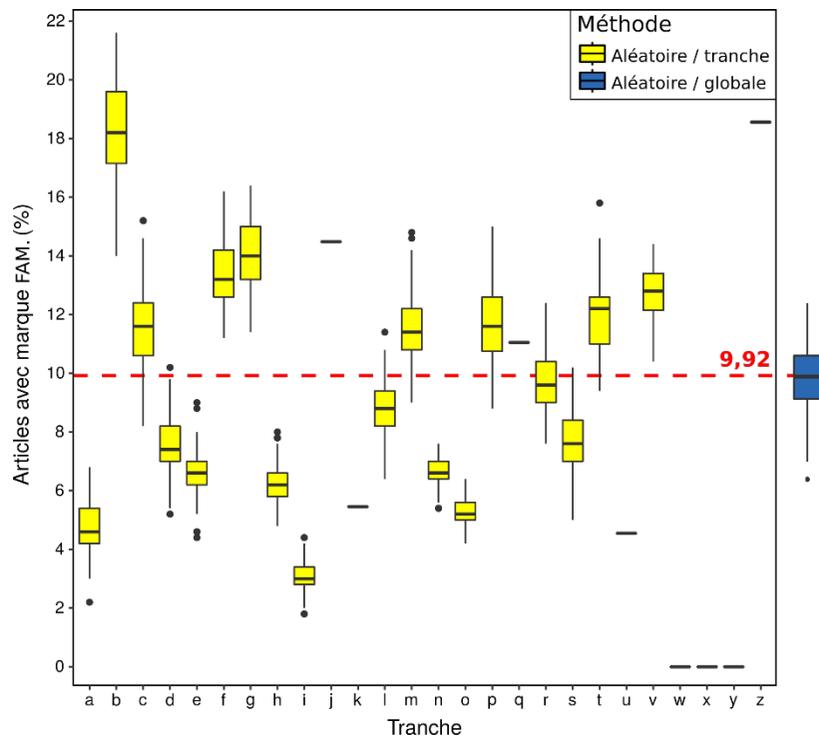


Figure 8 : Marque FAM. – échantillonnage par tranche vs sur tout le dictionnaire

Les boîtes à moustaches bleues, à droite des figures, représentent les distributions dont les caractéristiques sont données dans le tableau 2. Pour la marque FIG., 5 des 26 tranches (*b*, *c*, *g*, *m* et *p*) produisent des échantillons dont la distribution est plus centrée sur la proportion réelle d'articles marqués (cf. fig. 7). Les 21 autres tranches de la marque FIG. et l'intégralité des tranches de la marque FAM. (cf. fig. 8) produisent des échantillons dont la distribution est moins bien centrée sur la valeur réelle. Ce résultat illustre qu'il est préférable de constituer un échantillon par tirage aléatoire sur l'ensemble du dictionnaire qu'au sein d'une tranche donnée.

3.3. Tirages aléatoire, probabiliste stratifié et systématique

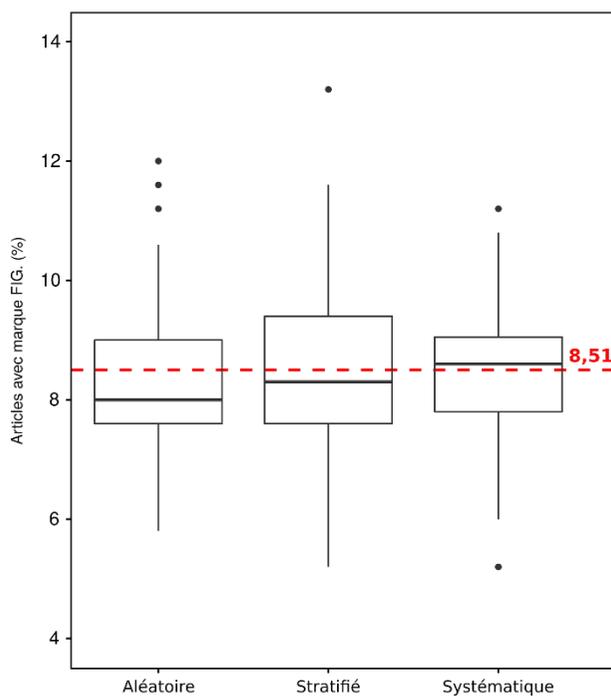
Après avoir déterminé, en section 3.1, que l'échantillonnage par tirage aléatoire est préférable à la sélection par zone contiguë et, en section 3.2, qu'un tirage aléatoire sur l'ensemble du dictionnaire est préférable à ce même tirage pratiqué au sein d'une tranche donnée, nous testons ici les échantillonnages systématique et probabiliste stratifié. Pour chacune de ces deux méthodes, 100 nouveaux échantillons de 500 articles sont générés. Pour l'échantillonnage systématique, cela consiste à sélectionner un nom tous les 62 (= 31 310/500), en décalant d'un article le point de départ pour chaque échantillon. Pour la méthode probabiliste stratifiée, l'échantillon est constitué en pratiquant un tirage aléatoire qui respecte les proportions de noms par lettre initiale données dans le tableau 1, section 3.1¹⁷. Les distributions des proportions d'articles marqués dans ces nouveaux échantillons sont représentées, à côté de celle obtenue pour les échantillons générés par tirage aléatoire, dans la figure 9 pour la marque FIG. et dans la figure 10 pour la marque FAM.

Avec une médiane de 8,3 %, l'échantillonnage probabiliste stratifié a globalement un léger avantage sur le tirage aléatoire (médiane de 8 %) pour la marque FIG. (valeur réelle : 8,51 %), mais dégrade très légèrement les résultats pour la marque FAM. (médianes respectives de 10 % pour l'échantillonnage probabiliste stratifié et de 9,9 % pour le tirage aléatoire, avec une valeur réelle de 9,92 %). Dans ses expériences, Bukowska (2010) constate de la même manière que l'échantillonnage probabiliste stratifié n'améliore pas systématiquement le tirage aléatoire. Quant à l'échantillonnage systématique, il produit les meilleurs résultats pour la marque FIG. et les moins bons pour la marque FAM. si l'on considère uniquement les valeurs médianes, les résultats obtenus avec les trois méthodes étant néanmoins, pour cette dernière marque, très proches de la proportion réelle.

Les résultats empiriques de cette comparaison ne permettent pas de conclure de manière définitive quant à la préférence à accorder à l'une ou l'autre des trois méthodes d'échantillonnage. Notons que s'il existe une réserve d'ordre théorique sur l'échantillonnage systématique, que nous avons rapportée en section 2.3.4, c'est, pour les deux marques, la méthode d'échantillonnage qui produit les distributions avec les variabilités les plus réduites. C'est donc, pour cette expérience, la méthode la plus robuste. À l'inverse, c'est l'échantillonnage probabiliste stratifié qui produit, pour les deux marques,

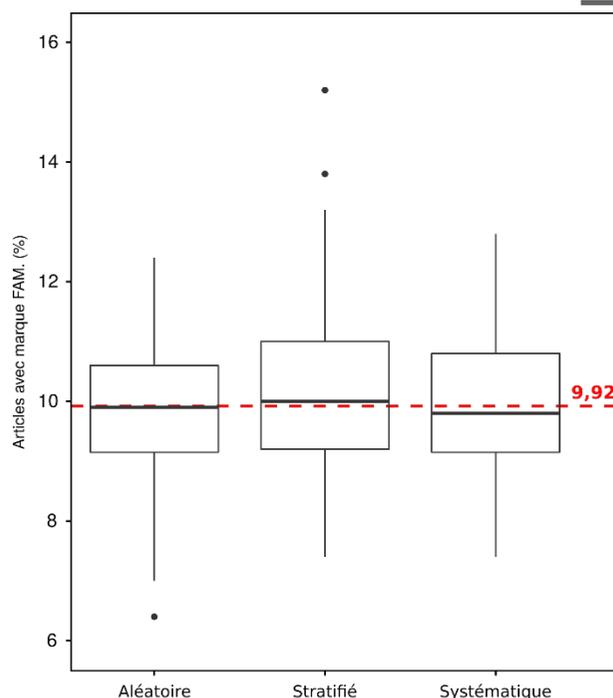
17. Il s'agit donc de tirer aléatoirement 54 (= 500 x 10,7 %) articles de la tranche *a*, 30 (= 500 x 6,1 %) articles de la tranche *b*, etc.

les distributions affichant une plus grande variabilité et qui semble donc, pour cette expérience, la moins robuste. Rappelons enfin une considération pratique : nous avons pu tester l'échantillonnage probabiliste stratifié, en choisissant comme strates les tranches alphabétiques correspondant aux lettres initiales des vedettes, parce que nous pouvions accéder (automatiquement) à la nomenclature du dictionnaire étudié. La mise en œuvre de l'échantillonnage probabiliste stratifié est conditionnée par la possibilité d'accéder à ce type d'information, ce qui n'est pas toujours possible.



Tirage	Min.	Q1	Médiane	Moyenne	Q3	Max.
Aléatoire	5,80	7,60	8,00	8,31	9,00	12,00
Stratifié	5,20	7,60	8,30	8,51	9,40	13,20
Systématique	5,20	7,80	8,60	8,52	9,05	11,20

Figure 9 : Marque FIG. – comparaison des méthodes d'échantillonnage sur tout le dictionnaire



Tirage	Min.	Q1	Médiane	Moyenne	Q3	Max.
Aléatoire	6,40	9,15	9,90	9,82	10,60	12,40
Stratifié	7,40	9,20	10,00	10,17	11,00	15,20
Systématique	7,40	9,15	9,80	9,97	10,80	12,80

Figure 10 : Marque FAM. – comparaison des méthodes d'échantillonnage sur tout le dictionnaire

3.4. Contrôle de la représentativité des échantillons

Dans l'expérience décrite en section 3.2, la méthode la plus fiable (tirage aléatoire sur tout le corpus) pour estimer le taux de marquage FAM. produit une distribution dont la médiane (9,9 %) se confond quasiment avec la valeur effective de la proportion (9,92 %) calculée sur l'ensemble des noms du dictionnaire.

Cependant, parmi les 100 échantillons de 500 articles générés aléatoirement, ceux affichant les proportions minimales et maximales (respectivement 6,4 % et 12,4 %, cf. tableau 2) donneraient une vision relativement éloignée de la réalité à une personne malchanceuse qui les aurait tirés. Mais peut-on réellement parler de malchance, *i.e.* quelle est la probabilité de tirer un échantillon conduisant à

une estimation aussi inexacte ? L'histogramme de la figure 11 représente le nombre d'échantillons répartis par tranches de pourcentages d'articles marqués¹⁸.

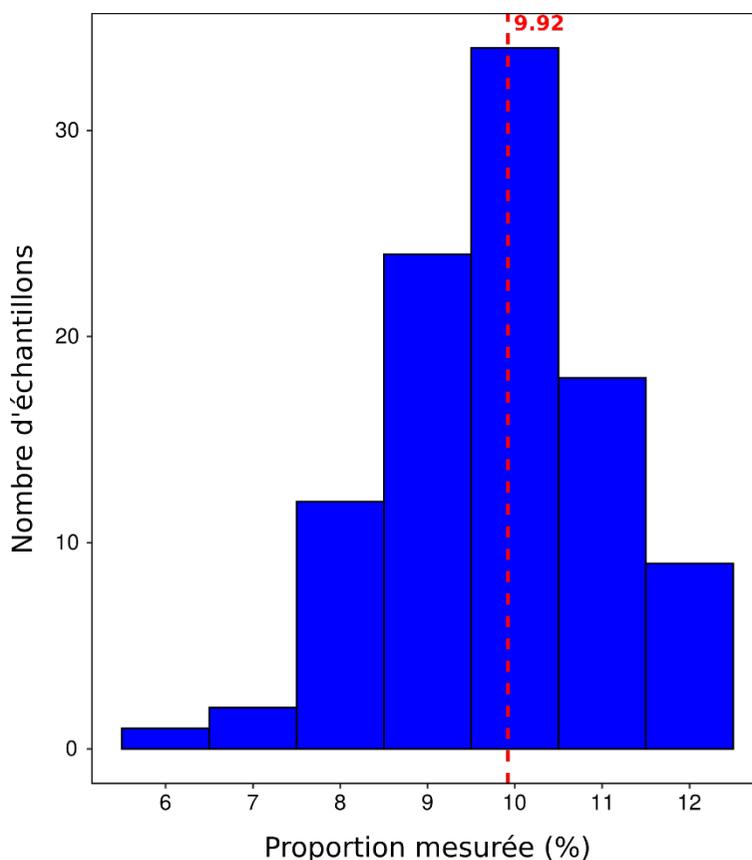


Figure 11 : Tirage aléatoire sur tout le dictionnaire – nombre d'échantillons par tranche de pourcentages d'articles marqués FAM.

Sans surprise, la distribution est normale et centrée sur la valeur réelle : la plupart des échantillons fournissent des estimations relativement proches de la réalité et quelques-uns des estimations éloignées. Mais les métalexicographes n'ont pas de moyen de déterminer avec certitude si la proportion observée (mesurée) p_o sur leur échantillon est plus ou moins fiable. La réponse de Bukowska (2010) est celle qui s'impose à toute personne pratiquant l'échantillonnage statistique : le recours aux intervalles de confiance mentionnés en section 2.4. Le tableau 3 donne, pour les deux échantillons présentant les proportions minimales et maximales observées (p_{\min} et p_{\max}), ainsi que pour un échantillon (fictif) dont la proportion observée serait strictement égale à la proportion réelle (p_r), les marges d'erreur au niveau de confiance 95 % et les intervalles de confiance correspondants.

18. Il se lit comme suit : parmi les 100 échantillons générés aléatoirement, un seul affiche une proportion d'articles marqués comprise entre 5,5 et 6,5; 2 échantillons ont une proportion comprise entre 6,5 et 7,5; 12 ont une proportion comprise entre 7,5 et 8,5; etc.

Tableau 3 : Marque FAM. – marges d’erreur au niveau de confiance 95 % et intervalles de confiance des échantillons

proportion observée (%)	marge d’erreur (%)	intervalle de confiance (%)
6,40 ($p_o = p_{\min}$)	2,15 %	[4,25; 8,55]
9,92 ($p_o = p_r$)	2,62 %	[7,3; 12,54]
12,40 ($p_o = p_{\max}$)	2,89 %	[9,51; 15,29]

Ces données se lisent de la façon suivante (un métalexicographe ne formulerait qu’un seul énoncé correspondant à l’échantillon tiré) :

- ($p_o = p_{\min}$) il y a 95 % de chances que la proportion de noms marqués FAM. dans *Usito* soit comprise entre 4,25 % et 8,55 % ;
- ($p_o = p_r$) il y a 95 % de chances que la proportion de noms marqués FAM. dans *Usito* soit comprise entre 7,3 % et 12,54 % ;
- ($p_o = p_{\max}$) il y a 95 % de chances que la proportion de noms marqués FAM. dans *Usito* soit comprise entre 9,51 % et 15,29 %.

On constate que, pour l’échantillon présentant la valeur minimale de 6,4 %, l’intervalle de confiance ne contient pas la valeur réelle (9,92 %). Une question se pose : combien d’échantillons sont dans ce cas, *i.e.* pour combien d’échantillons la valeur réelle est-elle en dehors de l’intervalle de confiance ? La figure 12 représente, pour les 100 échantillons de 500 articles générés pour notre expérience, la proportion observée pour chaque échantillon et l’intervalle de confiance au niveau de confiance 95 %. La ligne horizontale d’ordonnée 9,92 correspond à la proportion réelle calculée sur tout le corpus. On voit que la valeur réelle est en dehors des intervalles de confiance de seulement 3 échantillons (représentés en jaune). Pour cette expérience, les chances (97 %) de tirer un échantillon dont l’intervalle de confiance contient la valeur réelle sont bien supérieures ou égales au niveau de confiance choisi (95 %). Peut-on dire pour autant que l’on « contrôle » ainsi la représentativité d’un échantillon ? D’une part, les métalexicographes n’ont aucun moyen de savoir si leur échantillon fait partie des 95 « bons » pourcents théoriques ou des 5 « mauvais »¹⁹. D’autre part, même dans le cas idéal où la valeur observée correspond exactement à la valeur réelle, *i.e.* 9,92 % (information à laquelle les métalexicographes n’ont pas accès), doit-on se réjouir de pouvoir affirmer que la valeur réelle est probablement comprise entre 7,3% et 12,54 % ? Pour cette taille d’échantillon analysé, et, dans une moindre mesure, pour cette proportion observée, la marge d’erreur et l’amplitude et l’intervalle de confiance qui en résulte nous paraissent bien étendus (on s’en rend compte en comparant l’empan vertical des intervalles de confiance à celui des barres représentant les proportions mesurées).

Pour diminuer la marge d’erreur, au même niveau de confiance (95 %), la solution est d’augmenter la taille de l’échantillon examiné. Pour la proportion $p_o = 9,92$ %, obtenir une marge d’erreur qui n’excède pas 1 % nécessiterait, au même niveau de confiance, de sélectionner un échantillon constitué de 3500

19. Ou, dans la pratique, pour cette expérience, des 97 « bons » pourcents ou des 3 « mauvais ».

articles²⁰ (ce qui n'empêcherait pas le risque d'obtenir un intervalle de confiance ne contenant pas la proportion réelle, avec une probabilité de 5 %). Pour avoir 99 % de chances d'obtenir une marge d'erreur qui n'excède pas 1 %, c'est un échantillon de près de 6000 articles qu'il faudrait constituer.

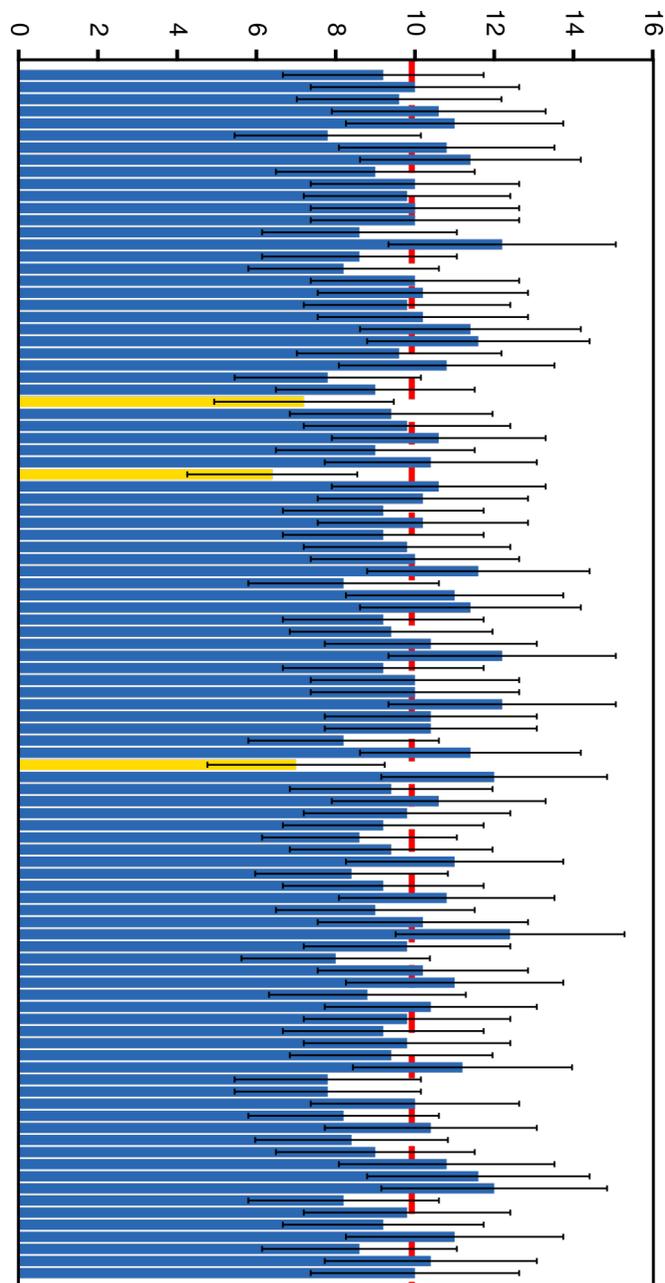


Figure 12 : Marque FAM. – intervalles de confiance au niveau de confiance 95 % des 100 échantillons de 500 articles

20. Cf. la formule donnée en section 2.4.

Revenons à nos 100 échantillons d'une taille plus réaliste de 500 articles chacun. Afin d'estimer, dans notre expérience, l'erreur probable entre la proportion mesurée sur un échantillon e et la proportion réelle, on définit le « taux d'erreur²¹ » de la manière suivante :

$$T_e = \frac{|p_0 - p_r|}{p_r}$$

Ce taux d'erreur exprime le rapport entre l'erreur d'estimation²² et la valeur réelle. La boîte à moustaches de la figure 13 représente la variabilité de ce taux d'erreur pour les 100 échantillons.

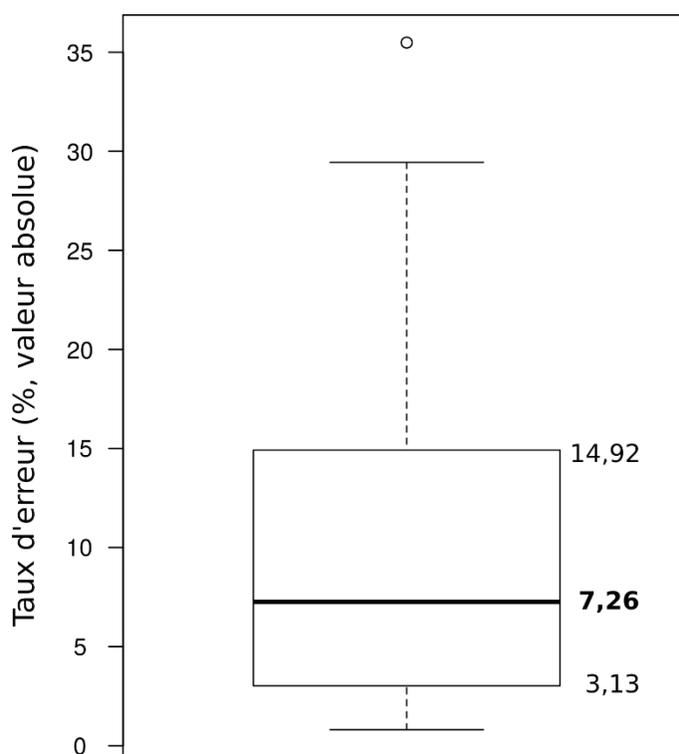


Figure 13 : Marque FAM., tirage aléatoire sur tout le dictionnaire – variabilité des taux d'erreur réels des échantillons

Elle montre que le taux d'erreur fluctue, selon les échantillons, entre 0,8 % et 35,5 %, avec une valeur médiane de 7,26 %. Une autre représentation de ces résultats est proposée en figure 14, qui illustre la probabilité qu'un des échantillons, parmi les 100 tirés aléatoirement dans notre expérience, affiche un taux d'erreur inférieur à un seuil donné.

21. Les guillemets signalent ici que cette notion, comme les représentations qui en sont données en figures 13 et 14, ne sont pas standards en statistiques.

22. Différence, en valeur absolue, entre la valeur mesurée sur l'échantillon et la valeur réelle sur l'ensemble du corpus.

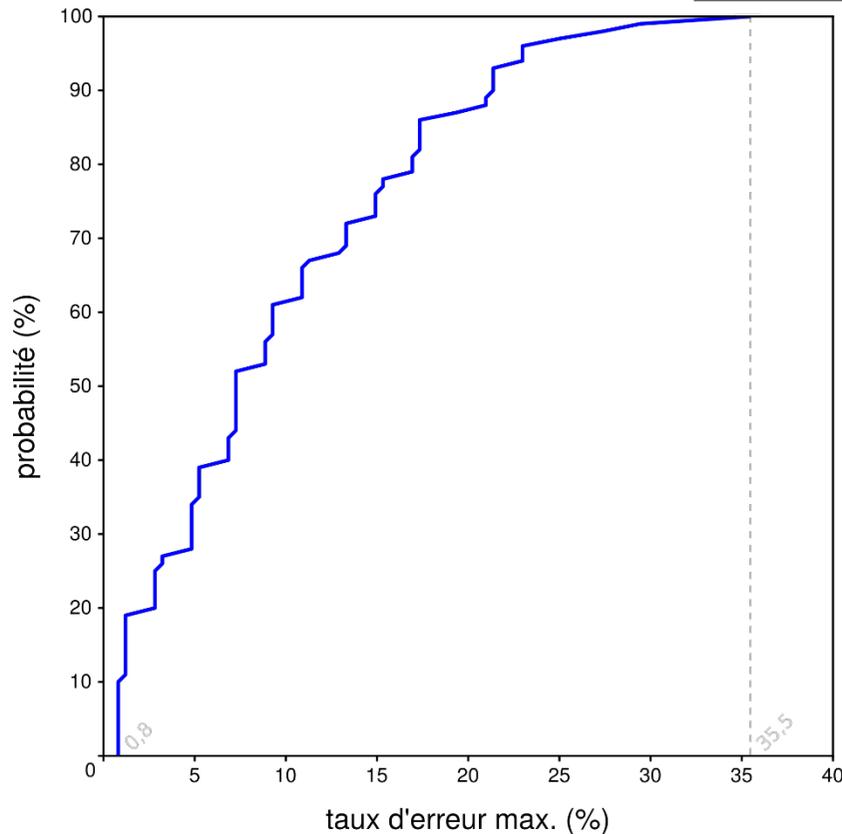


Figure 14 : Probabilité, pour un des 100 échantillons tirés, d'afficher un taux d'erreur inférieur à un seuil donné

L'interprétation que l'on peut donner de la courbe représentée est la suivante : un échantillon tiré au hasard parmi les 100 échantillons générés a :

- 10 % de chances d'afficher un taux d'erreur inférieur à 1 % ;
- 19 % de chances d'afficher un taux d'erreur inférieur à 2 % ;
- 34 % de chances d'afficher un taux d'erreur inférieur à 5 % ;
- 61 % de chances d'afficher un taux d'erreur inférieur à 10 % ;
- 87 % de chances d'afficher un taux d'erreur inférieur à 20 % ;
- 100 % de chances d'afficher un taux d'erreur inférieur à 36 %.

Cette lecture est celle qui consiste à voir le verre à moitié plein. Une autre lecture possible consiste à donner l'interprétation réciproque qu'un échantillon a, dans cette expérience :

- 90 % de chances d'afficher un taux d'erreur supérieur à 1 % ;
- 81 % de chances d'afficher un taux d'erreur supérieur à 2 % ;
- 66 % de chances d'afficher un taux d'erreur supérieur à 5 % ;

- 39 % de chances d'afficher un taux d'erreur supérieur à 10 % ;
- 13 % de chances d'afficher un taux d'erreur supérieur à 20 % ;
- aucune chance d'afficher un taux d'erreur supérieur à 36 %.

Ce « risque » pourra être jugé plus ou moins acceptable, mais gardons à l'esprit que les résultats présentés ici à titre d'illustration ne valent que pour l'expérience menée. La courbe représentée en figure 14 n'est pas une loi de probabilité générale : elle n'est pertinente que pour le phénomène étudié (marquage FAM.) observé dans le dictionnaire choisi (*Usito*) avec les paramètres de l'échantillonnage mis en œuvre (méthode de sélection et taille des échantillons), mais aussi pour le « hasard du moment » : un autre tirage aléatoire reproduisant l'ensemble de ces conditions expérimentales conduirait à des résultats différents. Or si, dans une certaine limite (à définir), un taux d'erreur peut être jugé raisonnable, c'est bien le manque de contrôle sur ce taux (et donc sur la représentativité des échantillons sélectionnés et analysés) qui ne l'est pas. Bukowska se satisfait de l'estimation par intervalles de confiance comme moyen de « contrôler » la représentativité. Dans la mesure où il n'existe pas d'autre moyen statistique qui en permette un meilleur contrôle, nous devons bien nous en contenter aussi. Mais, avant de se lancer dans une estimation quantitative par échantillonnage, il convient de bien mesurer les efforts à fournir pour analyser un échantillon dont la taille permettra de pouvoir prétendre à une certaine représentativité, cette représentativité étant – seulement – plus ou moins probable.

4. Synthèse et discussion

Cet article, destiné avant tout aux métalexigraphes, se fixait deux objectifs : d'une part, celui de montrer empiriquement, à travers des expériences portant spécifiquement sur l'analyse de dictionnaire, l'importance du choix des méthodes d'échantillonnage ; d'autre part, celui de relativiser le « contrôle » que l'on peut exercer sur les échantillons et, partant, sur les estimations que l'on peut effectuer à partir des résultats mesurés sur ces échantillons. Des expériences décrites en sections 3.1 à 3.4, nous retiendrons que :

1. la méthode d'échantillonnage par zone contiguë est la plus problématique. Un tirage aléatoire lui est dans tous les cas préférable. L'échantillonnage probabiliste stratifié, plus complexe – et pas toujours possible – à mettre en œuvre, de même que l'échantillonnage systématique, peuvent, selon les cas, améliorer ou dégrader les résultats ;
2. sauf à vouloir comparer deux parties spécifiques d'un dictionnaire, échantillonner sur l'intégralité du dictionnaire est plus fiable qu'échantillonner au sein d'une tranche donnée ;
3. même les méthodes d'échantillonnage les moins problématiques n'offrent pas de garantie absolue que l'échantillon observé soit représentatif : le calcul des intervalles de confiance n'apporte pas de solution totalement satisfaisante pour « contrôler » cette représentativité.

Faut-il alors, à l'aune de cette dernière assertion, renoncer aux analyses par échantillonnage en métalexicographie ? Sur la question de la représentativité des échantillons, on peut prêter aux métalexicographes chevronnés une certaine expérience de la discipline et une certaine familiarité avec les dictionnaires étudiés qui pourraient les alerter lorsqu'un échantillon leur paraît trop atypique de la représentation qu'ils se sont forgée du dictionnaire. Mais c'est bien le danger : on aura toujours tendance à faire confiance à un échantillon qui va dans le sens d'une intuition fondée sur une observation récurrente. Serait-il en revanche plus sage, en tirant un échantillon qui infirme son intuition, de récuser l'échantillon (et d'en tirer de nouveaux, jusqu'à ce que l'un d'entre eux se décide à rentrer dans le rang) ou de remettre en cause son intuition ? À qui pencherait en faveur de la première option, nous demanderions quel est alors l'intérêt de recourir aux statistiques.

Nous avons mentionné, en section 2.2, des études menées sur des échantillons dont la priorité n'est pas, à proprement parler, de quantifier des phénomènes, mais de montrer leur existence et leurs effets. Ces études entrent selon nous dans la catégorie des études qualitatives (quand bien même certaines portent sur une masse de données extrêmement importante). Concernant les études quantitatives, devant l'ensemble des problèmes méthodologiques que posent les techniques d'échantillonnage, nous concluons que les analyses menées sur l'intégralité des observables sont à privilégier à chaque fois que leur mise en œuvre est possible²³. C'est le cas pour les dictionnaires disponibles au format numérique et dont le statut légal est suffisamment permissif. Dans le cas contraire, il convient de mettre en œuvre les méthodes les moins problématiques (tirage aléatoire, probabiliste stratifié ou systématique sur l'intégralité du dictionnaire d'un échantillon de taille suffisante selon le niveau de confiance souhaité) et, malgré ces précautions, de considérer les résultats obtenus avec toute la circonspection qu'ils méritent.

Remerciements

L'auteur tient à remercier les évaluateurs anonymes pour leurs suggestions et commentaires constructifs.

23. Cette conclusion rejoint celle formulée en conclusion de l'article intitulé « Pour une analyse qualitative *et* quantitative, manuelle *et* computationnelle, synchronique *et* diachronique, des dictionnaires numériques », dans le même collectif.

Bibliographie

- BUKOWSKA, Agnieszka A. (2010), « Sampling techniques in metalexigraphic research » dans Anne Dykstra et Tanneke Schoonheim (dir.), *Proceedings of the 14th EURALEX International Congress*, Leeuwarden/Ljouwert, The Netherlands, p. 1258–1269.
- CORBIN, Pierre (1984), « Lexicographe-conseil », *Lez Valenciennes. Cahiers de l'UER Froissart*, n° 9, p. 113–121.
- CORBIN, Pierre (1990), « Le monde étrange des dictionnaires (7). Logique linguistique et logique botanique : problèmes posés par la définition d'une classe de mots dérivés français », *Cahiers de lexicologie*, n° 56-57, p. 75–108.
- CORBIN, Pierre (1995), « Le monde étrange des dictionnaires (8). Du *Petit Robert* (1967) au *Micro Robert* (1971) : le recyclage de citations », *Lexique*, n° 12/13, p. 125–145.
- CORBIN, Pierre (2020), « Les dictionnaires monolingues généraux du français actuel gratuits en ligne : évolutions récentes (2020) », *Academic Journal of Modern Philology*, n° 9, p. 65–77.
- CORMIER, Monique et Heberto Fernandez (2005), « From the *Great French Dictionary* (1688) of Guy Miège to the *Royal Dictionary* (1699) of Abel Boyer: Tracing Inspiration », *International Journal of Lexicography*, n° 18(4), p. 479–507.
- FERRETT, Emma et Stefan Dollinger (2021), « Is digital always better? Comparing two English print dictionaries with their digital counterparts », *International Journal of Lexicography*, n° 34(1), p. 66–91.
- FRANCŒUR, Aline (2021), « Forging a New Path in English-French Lexicography: Guy Miège in Relation to Robert Sherwood », *International Journal of Lexicography*, n° 34(4), p. 437–452.
- FREEMAN, Harold (1963), *Introduction to statistical inference*, Reading, MA, Addison-Wesley Publishing Company.
- KAMIŃSKI, Mariusz Piotr (2015), « In Search of Lexical Discriminators of Definition Style: Comparing Dictionaries through N-Grams », *International Journal of Lexicography*, n° 29(4), p. 403–423.
- LEHMANN, Alise (1995), « Du *Grand Robert* au *Petit Robert* : les manipulations de la citation littéraire », *Lexique*, n° 12/13, p. 105–124.
- MARTINEZ, Camille (2011), « Le poids des contraintes dictionnaires sur l'évolution des marqueurs dans les *Petit Larousse* (1997–2007) » dans Fabienne Baidier, Efi Lamprou et Monique Monville-Burston (dir.), *La marque en lexicographie : états présents, voies d'avenir*, Limoges, Lambert-Lucas, p. 39–50.
- MARTINEZ, Camille (2013), « La comparaison de dictionnaires comme méthode d'investigation lexicographique », *Lexique*, n° 21, p. 193–220.
- OSSELTON, Noel E. (2007), « Alphabet Fatigue and Compiling Consistency in Early English Dictionaries » dans John Considine et Giovanni Iamartino (dir.), *Words and Dictionaries from the British Isles in Historical Perspective*, Newcastle, Cambridge Scholars Publishing, p. 81–90.

PAULSEN, Mikkel Ekeland (2023), « Wheat or Chaff? A Compound Selection Model Based on Look-Up Data ». *International Journal of Lexicography*, n° 36(3), p. 306–324.

PODHAJECKA, Mirosława (2015), « Erazm Rykaczewski's *A Complete Dictionary English and Polish...* (1849) : Uncovering the Compilation Process », *International Journal of Lexicography*, n° 29(1), p. 1–30.

RADERMACHER, Ruth (2004), *Le Trésor de la langue française. Une étude historique et lexicographique*. Thèse de doctorat, Université Marc Bloch, Strasbourg.

TITRE: LES PRÉJUGÉS SEXISTES DANS LES DICTIONNAIRES DE PROFESSIONNELS ET DE PROFANES : UN PETIT ESPOIR DE CHANGEMENT?

AUTEUR: ANNICK FARINA, UNIVERSITÀ DEGLI STUDI DI FIRENZE

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 273-295

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22415](http://hdl.handle.net/11143/22415)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22415](https://doi.org/10.17118/11143/22415)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Les préjugés sexistes dans les dictionnaires de professionnels et de profanes : un petit espoir de changement?

Annick Farina, Università degli Studi di Firenze

Résumé :

Les dictionnaires profanes en ligne, et en particulier des ressources collaboratives comme le *Wiktionnaire*, sous le signe de l'« immédiateté » et de l'absence de limitations dans le choix des entrées, semblent pouvoir échapper à des contraintes qui empêchent au contraire les dictionnaires de professionnels de présenter une image actualisée de la langue. La quantité et la diversité des auteurs qui participent à leur réalisation devraient de plus permettre d'attester d'usages et de savoirs élargis sur la langue. L'observation du traitement lexicographique de certains mots et expressions se référant aux femmes, que l'on peut considérer comme participant d'une idéologie sexiste, confirme les observations du Haut conseil à l'égalité entre les femmes et les hommes qui dénonce le fait qu'en France les manifestations du sexisme, « perçues comme banales », bénéficient d'une « tolérance sociale » qui est en décalage avec une opinion négative sur l'idéologie qu'elles véhiculent. L'analyse des défauts et des qualités des différents types de ressources consultées sur ce point nous permettent de considérer qu'une influence réciproque pourrait aider à combler ce retard lexicographique par rapport à la société, pour sortir avec elle de l'écueil discriminatoire sexiste.

Mots-clés :

Lexicographie, sexisme, insultes, stéréotypes, dictionnaires collaboratifs

Abstract :

On-line profane dictionaries, and in particular collaborative resources such as *Wiktionnaire*, under the banner of “immediacy” and the absence of limitations in the choice of entries, seem to be able to escape the constraints which, on the contrary, prevent professional dictionaries from presenting an up-to-date picture of the language. On the other hand, the number and diversity of authors involved in their production should make it possible to attest to a wider range of uses and knowledge of the language. Observation of the lexicographical treatment of certain words and expressions referring to women, which may be considered to be part of a sexist ideology, confirms the observations of the French High Council for Equality between Women and Men, which denounced the fact that manifestations of sexism “perceived as banal”, benefit from a “social tolerance” that is at odds with a negative opinion of the ideology they convey. However, an analysis of the faults and qualities of the different types of resources consulted on this point suggests that a reciprocal influence could help to close the lexicographical gap with society, and help to avoid the sexist discrimination pitfall.

Keywords :

Lexicography, sexism, insults, stereotypes, crowdsourced dictionaries

1. Introduction

Les analyses comparatives de la lexicographie francophone « traditionnelle » et « profane »¹ en ligne tendent à souligner une plus grande ouverture de la lexicographie profane à tout ce qui concerne l'intégration de nouvelles lexies provenant de variétés « non hexagonales » (Vincent, 2017) ou « non conventionnelles » (Murano, 2017) et de créations néologiques (Molinari, 2021 : 37). Ce phénomène est généralement considéré comme une richesse particulière de ce type de ressources par rapport aux dictionnaires traditionnels. En effet, d'une part, des contraintes éditoriales pèsent sur les dictionnaires traditionnels, contraintes liées aux délais entre la rédaction des nouvelles entrées et leur publication, mais aussi à la restriction nécessaire de la nomenclature pour ceux qui prévoient des versions papier ou qui se plient à des politiques rédactionnelles posant la fréquence d'usage comme un critère de sélection. De plus, l'introduction de nouveaux mots dans un dictionnaire à visée commerciale l'oblige à se plier aux « lois du marché » et aux « attentes des lecteurs », comme en témoignent les réactions relayées par la presse pour des mots ou expressions « polémiques » présents dans leur macrostructure². Pour ce qui concerne les dictionnaires profanes en ligne, et en particulier des ressources collaboratives comme le *Wiktionnaire*, sous le signe de l'« immédiateté » et de l'absence de limitations dans le choix des entrées, ils peuvent échapper à ces contraintes, et la quantité et la diversité des auteurs qui participent à leur réalisation devraient permettre d'attester d'usages et de savoirs sur la langue actualisés et élargis.

La différence entre ces deux modèles de lexicographie est assimilée à un positionnement par rapport à la norme linguistique, les lexicographes professionnels travaillant sous l'égide du normatif³ et les profanes du descriptif. Ce point de vue, synthétisé par Meyer et Gurevych (2012 : 260) d'une plus

1. Pour la distinction entre ces deux types de lexicographie, nous nous basons sur l'opposition proposée par Nadine Vincent (2017) entre dictionnaires professionnels et profanes : « Les dictionnaires professionnels sont [...] produits par des lexicographes œuvrant au sein d'institutions publiques ou privées, alors que les dictionnaires profanes émanent d'entreprises privées et sont rédigés par des non-lexicographes qui peuvent venir d'horizons variés, incluant la linguistique, mais sans spécialité en lexicographie ». L'adjectif « traditionnel » tout en ne se référant qu'à la lexicographie professionnelle dans notre étude qui porte sur des dictionnaires du XX^e et du XXI^e siècle permet de rendre compte d'une continuité existant entre ces ouvrages et ceux dont ils sont en quelque sorte les héritiers, dont les auteurs – les moines de Trévoux ou Littré, par exemple –, n'étaient pas des professionnels au sens d'aujourd'hui.

2. Dans l'actualité récente, nous pouvons citer la polémique médiatisée autour de l'insertion de l'expression « travailler comme un nègre » dans le dictionnaire *Usito* en 2020 (cfr. Tommy Brochu, « "Travailler comme un n*****" : raciste ou pas ? », *La Tribune*, 4 juin 2020) ou « iel » dans le dictionnaire *Le Petit Robert* en 2021 (cfr. Alice Develey, « Le Petit Robert persiste et signe avec le "iel" », *Le Figaro*, 17 mai 2022), ou encore de « bolos » (avec un seul s) dans *Le Petit Larousse* en 2015 (cfr. « "Bolos", "boloss" ou "bolosse"... une querelle orthographique », *Le Monde* en ligne, 18 mai 2015).

3. Ce positionnement envers la norme peut être analysé comme correspondant à une volonté affichée par les auteurs de dictionnaires ou comme une « vocation » que leur attribuent leurs lecteurs : « Comme le manuel, le dictionnaire — et pas seulement le dictionnaire — est voué au didactisme, c'est-à-dire à la "reproduction" socioculturelle d'un savoir, à la diffusion d'attitudes et de jugements acquis. [...] Alors même que ses auteurs peuvent n'avoir souci que de description objective, le dictionnaire a des destinataires, pour lesquels cette description — didactiquement transmise — est la norme, la vérité. Ceci donne une signification différente, toujours dans le sens normatif au texte lexicographique » (Rey, 1983 : 566).

grande capacité des dictionnaires collaboratifs à « exprimer l'usage réel du langage, dans l'esprit du "meaning in use" de Wittgenstein » que des dictionnaires construits par des experts qui décriraient « la façon dont les gens "devraient" utiliser la langue », apparaît dans nombre d'études de linguistes qui voient dans la lexicographie profane collaborative un terrain fertile d'étude de phénomènes linguistiques émergents (par ex. Murano, 2017 et 2019; Sajous, Josselin-Leray et Hathout, 2018; Sajous et Humbley, 2022).

Cependant, de même que l'on ne peut affirmer que les dictionnaires traditionnels contemporains sont des ouvrages qui diffusent une norme subjective / prescriptive du français, alors que les politiques éditoriales qui régissent la plupart d'entre eux, du *Petit Robert* à *Usito*, en passant par le *Trésor de la langue française (TLFi)*, sont celles de la description du lexique basée sur son usage effectif (Farina, 2024), force est de constater que la rédaction des ressources profanes les plus consultées, qu'elles soient collaboratives (*Wiktionnaire*) ou participatives (*Dictionnaire de la Zone, Bob*), reste « dirigée » et basée sur des règles similaires à celles des dictionnaires traditionnels. Le modèle de rédaction *bottom-up* proposé par les ressources profanes semble en effet d'une richesse indéniable, en ce qu'elle leur donne « la potentialité de n'être jamais obsolètes » et de représenter une « réserve dynamique de la connaissance » (De Schryver, 2003 : 157). La « sagesse des foules », cependant, si tant est que la participation soit élevée, ce qui n'est pas le cas dans tous les dictionnaires collaboratifs observés ni pour tous les types de lemmes (Sajous, 2023 : 26), n'est pas toujours la garantie d'une description objective.

Nous basant sur des analyses déjà faites sur la persistance de représentations stéréotypées des femmes qui peuvent être associées à un sexisme entériné par les dictionnaires traditionnels, tant sur les monolingues français (Farina, 2005 ; Kottelat, 2009) et italiens (Farina, 2022) que bilingues français-italien (Farina, 2005) et français-anglais (Campbell, 2006), nous comparerons le traitement d'entrées potentiellement « polémiques » relatives aux femmes, en particulier les insultes les plus fréquentes proférées à leur insu dans les différents types de ressources. En effet, à la suite d'Océane Foubert (2023) qui analyse le traitement de néologismes féministes dans *Wiktionary*, nous pensons que les dictionnaires collaboratifs représentent un terrain fertile pour l'expression de la « représentation actuelle du genre et des idées féministes ». Reste à vérifier si la lutte contre le sexisme est toujours au goût du jour dans l'activisme « féministe » qui s'exprime à travers les espaces collaboratifs en ligne et si le qualitatif – en particulier, la révision d'articles à contenu sexiste – accompagne le quantitatif – la création de nouvelles entrées à visée égalitaire entre les sexes.

2. Norme sociale et dictionnaires : les bienfaits de la mixité en lexicographie

De même qu'il a fondé ses jugements de grammaticalité sur une norme linguistique, le dictionnaire, lieu de référence, doit définir ses jugements d'acceptabilité d'après une norme culturelle. Ses assertions sur l'homme doivent lui être communes avec le plus grand nombre de lecteurs : des formes extrêmement

variées d'une culture, le dictionnaire retient surtout les éléments courants et, par voie de conséquence, les stéréotypes les plus étroits, les formulations les plus banales, les images d'Epinal, que lui imposent ses lecteurs. Cette norme culturelle est conforme à l'idéologie de la classe sociale dominante et c'est à elle que vont se référer les jugements de valeur du dictionnaire. (Dubois et Dubois, 1971 : 99)

Cette affirmation de Dubois et Dubois en 1971 n'est sûrement que partiellement applicable à la lexicographie telle qu'elle s'est pratiquée depuis et qu'elle se pratique encore aujourd'hui. En effet, non seulement la recherche dans les domaines de la lexicographie et de la métalexigraphie a permis de dévoiler et de dénoncer le message idéologique porté par les dictionnaires, de développer une sensibilité à ce « message » de la part des rédacteurs de dictionnaires, mais la ligne définie par la direction de certains dictionnaires, et en particulier celle des éditions Le Robert pour ce qui concerne la lexicographie française, a aussi permis d'éviter en partie cette adhésion au diktat idéologique de la norme culturelle et sociale dominante. Le Robert affiche en effet depuis longtemps sa volonté de ne pas céder à la pression du marché et des prétendues attentes des lecteurs⁴. De plus, dès la constitution de l'équipe qui a mis au point et dirigé, pendant plus d'un demi-siècle, la rédaction du *Petit Robert*, la variété des rédacteurs et des choix des articles qui leur étaient confiés ont été pensés dans un souci de pluralité des perspectives linguistiques et culturelles.

Sans se présenter comme un défenseur de la condition féminine, en introduisant Josette Rey-Debove dans son « phalanstère de jeunes linguistes », selon la formule alors retenue par la presse, Paul Robert se révèle néanmoins sensible à un sujet jusque-là sans résonance dans le milieu lexicographique : l'importance du choix d'un lexicographe ou d'une lexicographe pour la rédaction de tel ou tel article. Pour la première fois en effet, avec le dictionnaire alphabétique et analogique de la langue française, apparaît cette dimension critique consistant à prendre conscience que la sensibilité masculine ou féminine de l'auteur de l'article ne serait pas sans importance dans la rédaction dudit article. (Pruvost, 2008 : 52-53)

L'avènement de la mixité à l'intérieur de la rédaction des dictionnaires a sans nul doute eu une grande influence sur la plus grande présence des femmes et l'évolution constante de la manière de les décrire à l'intérieur des dictionnaires et dans l'ensemble des productions lexicographiques des dernières décennies, que cela soit le résultat de décisions délibérées de la part de femmes-lexicographes comme Josette Rey-Debove pour permettre aux femmes de retrouver à l'intérieur de l'espace

4. Josette Rey-Debove résume cette politique éditoriale comme suit : « Contrairement à ce que disent certains universitaires envieux, les dictionnaires Le Robert ne sont pas "commerciaux", et ni Paul Robert, ni Alain Rey, ni moi-même, ni notre actuel directeur Pierre Varrod n'avons eu pour priorité de vendre une marchandise. Lorsqu'elle s'est bien vendue, c'était grâce à la reconnaissance de sa qualité. La philosophie de notre maison d'édition est de publier des dictionnaires utiles, innovants et "fréquentables", même dans un marché très restreint, afin que rien ne soit perdu du français contemporain. Tous nos dictionnaires sont originaux et reconnus comme tels. » (2003 : 108) Il est à noter que l'allusion aux universitaires « envieux » de Josette Rey-Debove, quoique révélatrice d'une rivalité entre lexicographie « commerciale » et lexicographie « universitaire » à la fin du XX^e siècle, par exemple entre les dictionnaires Le Robert qu'elle représente et le TLF, ne l'a pas empêchée, contrairement à nombre d'autres lexicographes responsables de la rédaction de grands dictionnaires commerciaux, de suivre de près les recherches en métalexigraphie et de dialoguer avec les universitaires, ce qui a sans nul doute nourri son travail.

symbolique du dictionnaire la place qu'elles ont dans la « vraie vie » (Rey-Debove, 2004 : 230) ou d'une simple différence de point de vue.

C'est cette même dynamique de la pluralité des voix qui s'exprime dans les ressources lexicographiques collaboratives, démultipliée par l'accessibilité de ces ressources et par la diversité des sociétés d'où proviennent ceux qui y participent, qui nous fait penser qu'elles peuvent permettre de sortir du carcan de stéréotypes et d'idéologies d'une norme sociale donnée. Et pourtant, nous verrons que la « dimension critique » apportée par les femmes dans la rédaction des dictionnaires n'a pas toujours suffi pour modifier le regard véhiculé par notre langue à leur égard, et pour empêcher qu'il ne soit ratifié génération après génération par les lexicographes. Cela vaut à la fois pour la non-révision d'informations qui perdurent d'une édition de dictionnaire à une autre dans la lexicographie traditionnelle et pour la non-modification de ces mêmes informations lors de la création de nouvelles entrées relatives à des mots anciens dans les dictionnaires de profanes.

3. Féminisme et néologie : il ne s'agit là que de la partie visible de l'iceberg

Le thème de « l'inclusion de genre », très présent dans les dernières décennies, a d'une certaine manière remplacé dans le discours institutionnel et dans celui des nouveaux adeptes du féminisme celui qui traitait du respect des droits des femmes et de la lutte contre le sexisme. Le positionnement sous-jacent à ce choix linguistique reflète certainement un positionnement plus positif et « constructif » : la recherche de l'inclusion vs la lutte contre l'exclusion, mais il déplace le débat en le sortant de sa relation avec la réalité pour en faire un programme d'action abstrait. L'utilisation du mot « genre » est emblématique dans ce contexte : comme illustré par Fraisse (1996 : 45-46), utiliser « genre » plutôt que « sexe » exprime « la volonté conceptuelle de se défaire du concret du sexe pour l'abstrait du genre », et représente un choix philosophique :

[...] la négation de la différence sexuelle (voire de la sexualité?) et le choix d'une analyse purement sociale ; la reconduction de l'opposition nature/culture (biologie/société) plutôt que sa mise en question ; la perte de la représentation de la relation sexuelle, et du conflit inhérent au profit d'une abstraction volontariste. (Fraisse, 1996 : 45-46)

Quand la réflexion sur le genre se déplace d'un contexte sociopolitique au domaine linguistique, nous pouvons noter que nous retrouvons la même « abstraction volontariste » opposée à une analyse concrète des relations entre sexes et de leur conflit, mais qui correspond aussi à une différence d'objet d'étude pour les linguistes qui s'en occupent : ceux qui s'intéressent à la « parité de genre » ou à l'« inclusion de genre » dans le langage s'arrêtent généralement au système linguistique, à la norme, et en particulier une norme « subjective » qui tend à devenir « imaginaire » (Farina, 2024).

Bonjour Sayoxime

Bonjour, j'ai vu que tu as ajouté des lexiques « non-binarité » Je voulais juste te signaler qu'il existe déjà un lexique des genres humains et identités de genre qui n'est pas très rempli et comprend déjà la non-binarité et les « identités de genre » — Victoire F., le 29 octobre 2023 à 14:01

Bonjour Sayoxime et Victoire F.,

Je venais également écrire ici à ce sujet. Il ne me paraît pas opportun de créer un nouveau lexique. Ce n'est pas un domaine sémantique, bien qu'il y ait un champ lexical. Les lexiques couvrent des ensembles de mots bien plus importants, qui sont en lien avec des activités ou disciplines. L'inventaire des lexiques est déjà très développés [sic] et il est préférable de ne pas le rendre trop complexe ou touffu. Un lexique devrait comporter au moins une vingtaine de termes communs, pas une petite poignée — Noé 30 octobre 2023 à 00:05

J'avais vu qu'il y avait le lexique du transféminisme donc par logique il devrait exister un lexique du transmasculinisme et de la nonbinarité — Sayōxime 30 octobre 2023 à 07:06

Je ne suis pas très certaine non plus de la pertinence du lexique du transféminisme d'ailleurs (en partie par son nom, qui demeure assez flou et n'est pas utilisé par toutes les féministes qui traitent des question trans).

La partie « identités de genre » du lexique « des genres humains et identités de genre » me semble justement faite pour inclure toutes ces identités de genre non-binaires.

(Quand au transmasculinisme, je ne sais pas ce que c'est. Le lexique transféminisme concerne le féminisme, pas les femmes trans.) — Victoire F., le 30 octobre 2023 à 10:45⁵

Cette discussion sur la page utilisateur Sayoxime (qui se déclare rédacteur « aroace » et « demi-boy-demienby »⁶) illustre l'activité militante qui est à la base de la création de nouvelles entrées « inclusives » (et de nouveaux « lexiques » qui les classifient) dans le *Wiktionnaire*, mais aussi la difficulté, pour les administrateurs, de maintenir systématisme et scientificité dans la rédaction : ne pas rendre l'inventaire des lexiques de spécialité « complexe et touffu » est un beau tour de langage dans la médiation de Noé⁷ pour tempérer les ardeurs néologiques de certains rédacteurs. Il n'en reste pas moins que Sayoxime et Victoire F. font partie de la minorité de rédacteurs très actifs sur le *Wiktionnaire* et ont créé et proposent de créer de nombreuses entrées néologiques qui procèdent de ce que j'ai appelé plus haut la création d'une « norme imaginaire » du français de la part des néo-féministes. On trouve la liste d'une partie des « lexiques » que ces rédacteurs participent à constituer et dont la spécialisation est discutable aux yeux de Noé (mais dont il ne peut empêcher la création, en vertu des règles démocratiques des dictionnaires collaboratifs ouverts) dans les ressources de la page de l'« utilis-

5. https://fr.wiktionary.org/wiki/Discussion_utilisateur:Sayoxime#Lexique_NB

6. Notre lecteur pourra consulter le *Wiktionnaire* pour comprendre le sens de ces termes, les deux derniers ayant le même Sayoxime comme créateur de l'entrée.

7. Nous avons conservé les pseudonymes utilisés par les utilisateurs du *Wiktionnaire*, même si l'on peut souvent remonter à leur identité réelle et que cette information peut avoir une importance pour évaluer la part du « profane » et du « professionnel » dans les ressources collaboratives et le type de professionnalité qui caractérise leurs rédacteurs. Pour ce qui concerne les administrateurs cités ici, Noé est un linguiste spécialisé en lexicographie et Darmon117 une informatrice spécialisée en intelligence artificielle.

taire » Scriptance : « écriture inclusive / genre neutre en français / double flexion figée / double flexion abrégée / LGBTIQ / Lexique en français des genres humains et identités de genre / Lexique en français de la sexualité / Lexique en français de la transitude / Lexique en français du transfémisme / Lexique en français du féminisme »⁸. Le détail de certaines entrées déjà créées ou encore à créer dans ces lexiques peut être consulté, par contre, sur la page de l'utilisatrice Victoire F.⁹

Une grande partie du travail de ces rédacteurs et qui constitue un terrain d'investigation lexicographique pour une des rares administratrices du *Wiktionnaire*¹⁰, Darmo117, est centrée sur la présentation des genres à l'intérieur du dictionnaire : si Sayoxime et Scriptance s'intéressent particulièrement à la création d'entrées relatives à des formes néologiques « neutres » d'adjectifs ou de noms (par ex. les articles *oppressæ* et *chercheur* pour Sayoxime ou *auteurice* et *nombreuse* pour Scriptance), Darmo117 semble vouloir se concentrer sur l'intégration de formes épiciques utilisant un point médian (*chercheur-e*, *musicien-ne*).

Ceci nous pousse à nuancer l'affirmation de Steffens (2017) sur le fait que, dans le *Wiktionnaire*, « la communauté d'utilisateurs est [...] suffisamment active pour assurer un contrôle par les pairs des informations encodées et leur adéquation avec les critères d'acceptabilité d'une entrée », ce qui assurerait la présence de « garde-fous efficaces » pour éviter l'intégration de « mots farfelus sans usages réels ». L'objectivité de *Wikisource* est certes protégée par l'existence de marques telles « néologisme » ou « rare » apposées par les réviseurs en deuxième instance sur nombre de ces créations néologiques qui sont censées donner plus de visibilité au « genre ». Cependant, contrairement à ce que l'on trouve dans la lexicographie professionnelle¹¹, et que Sajous a pu observer pour le traitement du lexique de l'occultisme,

8. <https://fr.wiktionary.org/wiki/Utilisateur:Scriptance>

9. https://fr.wiktionary.org/wiki/Utilisatrice:Victoire_F./Brouillon

10. Dans l'article « Gender bias » de *Wikipedia* (https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia), on trouve des données statistiques sur la mixité dans la participation à la rédaction des ressources Wikimedia : en 2018, il y avait 90 % des 3 734 participants déclarant leur sexe qui étaient des hommes, 8,8 % des femmes, et 1 % se déclaraient comme « autre ». On retrouve cette même disparité dans la rédaction du *Wiktionnaire*, même si la teneur des articles que nous avons consultés nous a permis de noter une forte condensation de la minorité de rédactrices féminines ou non binaires sur les mots reliés à des questions de féminisation.

11. Il nous semble intéressant de noter ici que la marque « néologisme » n'est pas présente dans les dictionnaires professionnels, et en particulier dans le *Petit Robert* – elle apparaît dans la liste des abréviations du dictionnaire mais pas dans les articles, comme le prouve une recherche de l'abréviation « néol. » dans l'ensemble du dictionnaire – ce qui correspond à la volonté de présenter une norme objective de la langue : seuls les mots qui sont suffisamment utilisés par les locuteurs du français entrent dans la nomenclature du dictionnaire, ils ne sont plus alors des « néologismes ». Pour ce qui concerne la marque « rare », une recherche sur cette marque nous indique que seules 30 % de ces marques portent sur des entrées et pas sur des sens d'articles polysémiques et que ces entrées sont à environ 90 % des dérivés (adverbes, adjectifs, ...) dont la basse fréquence d'emploi dépend de leur technicité. La seule exception à cette règle dont dépend la création de nouvelles entrées pour des mots peu fréquents que j'ai pu observer est le mot *iel* qui peut être probablement aussi considéré comme la seule « dérive » néologique d'influence activiste (voire « wiktionnairienne ») dans ce dictionnaire.

[...] aucune initiative de révision systématique des entrées de ces domaines dont le traitement lexicographique est visiblement délicat n'a été prise alors que «les foules» auraient pu être mobilisées pour, collaborativement, passer en revue les entrées correspondantes et réviser les définitions qui le méritaient, selon des critères discutés conjointement. (Sajous, 2023)

Tout ceci n'est en vérité que « la partie visible de l'iceberg » pour ce qui concerne la présence d'une évolution lexicographique parallèle à un militantisme centré sur le genre, si tant est que l'on puisse le considérer comme une volonté d'amélioration de la condition des femmes dans nos sociétés, en ce que cette prolifération néologique se concentre exclusivement sur la forme et non sur le fond.

4. Un passé prégnant qui a pris racine un peu partout

Qu'ils soient conçus à partir d'éditions précédentes du même dictionnaire et consistent en une simple mise à jour des informations, comme dans le cas des dictionnaires de maisons d'édition du type Larousse ou Le Robert, ou qu'il s'agisse de dictionnaires en mouvement constant, dont les nouvelles entrées naissent spontanément au jour le jour, comme le *Wiktionnaire*, tous s'inspirent d'articles déjà construits par d'autres, dans ces dictionnaires contemporains commerciaux ou dans d'autres ressources plus anciennes en libre accès sur internet comme le TLFi, les différentes éditions du *Dictionnaire de l'Académie française* voire dans des dictionnaires encore plus anciens comme le *Littre*. Dans le cas du *Wiktionnaire*, une partie conséquente de ses articles a été importée directement de ces ressources en libre accès¹² sans connaître ensuite de révision substantielle. Il est donc très fréquent que l'on trouve des éléments qui n'ont plus lieu d'être aujourd'hui mais que l'on a conservé par manque de révision ou de sens critique des réviseurs.

Pour ce qui concerne la vision de la femme, les dictionnaires ont participé pendant des siècles à la diffusion d'idées prétendument scientifiques sur sa physionomie et sa sexualité, comme celle qui concerne par exemple la relation entre l'hystérie féminine et son utérus. La description de la « fureur utérine », par exemple, présente dans les dictionnaires des XVIII^e et XIX^e siècles¹³ a disparu dans les dictionnaires du XX^e siècle. Elle reste ensuite dans les dictionnaires à nomenclature extensive comme le *Grand Robert* (1974) et le TLFi, qui l'indiquent cependant simplement comme archaïque avec un renvoi synonymique à *nymphomanie* sans se référer aux théories qui reliaient le désir sexuel féminin à une affection hystérique. Dans le TLFi apparaît une citation de Rétif de la Bretonne : « Vivement ému, je l'embrassai moi-même. Alors Nanette [sic] parut comme saisie d'une fureur utérine; elle me serra, s'empara de tout mon être, et me fit palper le sien (...). Enfin, il lui prit un tel accès d'érotisme, qu'elle

12. Une recherche de la source « Littre » dans les macrostructures de *Wiktionnaire* donne par exemple plus de 18 000 résultats.

13. La « fureur utérine » donne lieu à un article de l'*Encyclopédie* de Diderot et d'Alembert (analysé dans Farina, 2005 : 3) ; elle est présente dans l'article « utérin » du *Littre* et dans le *Dictionnaire de l'Académie française* jusqu'à la sixième édition (1835).

voulut être possédée¹⁴ ». Le *Wiktionnaire* réhabilite la *fureur utérine* dans son article relatif à l'adjectif *utérin* en lui accordant une acception qui appartiendrait au domaine de la médecine (l'absence de marque diachronique fait présupposer son actualité) et serait synonyme d'hystérie. Il recycle pour cela la citation de Rétif de la Bretonne trouvée dans le TLFi en éliminant cependant les éléments qui permettaient l'interprétation de la « *fureur* » du personnage féminin non comme un symptôme névrotique mais comme faisant partie d'un jeu érotique auquel les deux personnages participent :

2. (*Médecine*) Hystérique.

Alors, Nannette parut comme saisie d'une fureur utérine ; elle me serra, s'empara de tout mon être, et me fit palper tout le sien. — (Nicolas Rétif de la Bretonne, *Monsieur Nicolas*, 1796, bibliothèque de la Pléiade, tome 1, page 55)¹⁵

Le recyclage d'informations provenant de sources lexicographiques anciennes sans jugement critique a posteriori est évident dans une multitude d'autres articles du *Wiktionnaire*. Je n'en citerai qu'un autre, *menstruation*, tout d'abord parce qu'il me semble illustrer particulièrement bien combien la professionnalisation des lexicographes a pu correspondre à une recherche de scientificité du discours lexicographique, tant dans la rédaction des définitions que dans le choix des exemples pour le cas étudié ici. Il est de plus révélateur de la difficulté pour un lecteur non averti – et pour un rédacteur profane de dictionnaires collaboratifs – de déceler des témoignages d'une censure relative à la sexualité féminine qui va de pair avec sa méconnaissance et avec une pudibonderie encore prégnante dans notre imaginaire collectif.

menstruation \mãs.tɾy.a.sjõ\ *féminin*

(*Physiologie*) Ensemble des phénomènes qui, à des intervalles quasi mensuels, se manifestent au cours du cycle menstruel.

La menstruation marque le début de la période pendant laquelle une femme est apte à procréer. — (Danielle Haase-Dubosc, Susie Tharu, Marcelle Marini, Mary-E John, Rama Melkote, *Enjeux contemporains du féminisme indien*, 2003)

C'est la bourgeoisie qui a fait de l'argent une chose sacrée qui exige le mystère, analogue aux menstruations auxquelles on pense tout le temps et dont on ne parle jamais. — (Roger Vailland, *325.000 francs*, 1954, réédition Le Livre de Poche, page 49)¹⁶

Dans l'historique de la rédaction de cet article, l'origine de la définition est indiquée clairement : il provient de la huitième édition du *Dictionnaire de l'Académie française* (1932-1935). La définition originale¹⁷ a été légèrement remaniée par des rédacteurs successifs qui, au lieu de procurer des

14. <http://stella.atilf.fr/Dendien/scripts/tlfiv5/advanced.exe?56;s=3091724040> [page consultée le 13 août 2024]

15. <https://fr.wiktionary.org/wiki/ut%C3%A9rin> [page consultée le 13 août 2024]

16. <https://fr.wiktionary.org/wiki/menstruation> [page consultée le 13 août 2024]

17. « *T. de Médecine. Ensemble des phénomènes qui, à des intervalles mensuels, se manifestent dans les organes génitaux de la femme.* » (DAF, 1835, tome 2, p. 175)

informations nécessaires à la compréhension du processus physiologique en jeu dans la menstruation, comme par exemple la référence essentielle à l'écoulement du sang qui apparaît dans tous les dictionnaires contemporains professionnels, ont ajouté de nouvelles approximations. Ils modifient la durée des « phénomènes » qui de mensuels deviennent « quasi mensuels » et leur type : il n'est plus question de quelque chose qui se passerait « dans les organes génitaux de la femme », une information qui avait au moins l'avantage de définir la femme et ses organes génitaux comme les principaux acteurs de ce processus, mais d'un élément du « cycle menstruel » défini dans le même dictionnaire comme « cycle correspondant au temps écoulé entre le premier jour de la menstruation et le dernier jour avant la menstruation suivante »¹⁸. La citation présente dans l'article, qui compare la vision bourgeoise de l'argent à celle des menstruations comme d'« une chose sacrée qui exige le mystère » nous donne peut-être la clé de ces choix lexicographiques.

Le manque de connaissance des principes de base qui régissent la rédaction des définitions tels qu'ils ont été définis dès 1967 par Josette Rey-Debove et appliqués ensuite d'une manière systématique dans les dictionnaires Le Robert et plus généralement dans les dictionnaires généraux français contemporains, et le manque de sens critique de la plupart des rédacteurs de dictionnaires profanes par rapport à l'information qu'ils recyclent lors de la rédaction de leurs articles expliquent la prolifération d'informations surannées et souvent erronées dans les ressources collaboratives profanes – ou l'absence d'information dans le cas de définitions circulaires où chaque article renvoie à un autre –, alors que les spécialistes s'efforcent de les faire disparaître depuis plus d'un siècle des dictionnaires professionnels. Il n'en reste pas moins que la connaissance et la reconnaissance de certains éléments métalinguistiques parasites qui font partie du système que représente la microstructure du dictionnaire est souvent difficile même pour ces spécialistes. Eux aussi se font ainsi le véhicule, souvent d'une manière insidieuse, d'idéologies nouvelles et passées, en créant ou en maintenant des relations tant pragmatiques que sémantiques entre les mots et les stéréotypes qu'ils ont pu véhiculer.

Dans l'article *ménopause* du *Petit Robert*, par exemple, on trouve un lien synonymique vers *âge critique*, qui n'est heureusement plus en usage aujourd'hui. Il est normal de l'intégrer à la macrostructure d'un dictionnaire pour en permettre la compréhension à la lecture de textes anciens (décodage) mais pas de le proposer comme équivalent au mot *ménopause* pour l'encodage comme le prévoit le *Petit Robert* pour ses renvois synonymiques qui correspondent à des lemmes de différents registres, mais toujours présents dans la langue de référence (le français d'aujourd'hui). Il s'agit évidemment d'un élément maintenu par manque de révision a posteriori d'informations appartenant à des éditions passées, tout comme l'absence de marque pour indiquer la valeur offensive du mot *mal-baisée* que nous verrons plus loin, qui résulte du fait qu'il est indiqué comme collocation à l'article *baiser* qui n'a probablement pas été revu.

18. https://fr.wiktionary.org/wiki/cycle_menstruel [page consultée le 13 août 2024]

5. Le traitement de la violence sexiste dans les dictionnaires : l'exemple des injures

En France, le Haut Conseil à l'égalité entre les femmes et les hommes (2019 : 14) repère deux manifestations du sexisme particulièrement insidieuses : l'humour et les injures. « Perçues comme banales », ces manifestations bénéficieraient en effet d'une grande « tolérance sociale » alors qu'elles contribueraient à entretenir une violence symbolique, renforçant les « stéréotypes de sexe » et par là même les inégalités.

« Dans l'injure, l'homme se sert de la parole comme si elle était investie de force physique : pour offenser (étymologiquement, « heurter », « porter un coup »), pour insulter (« sauter sur »), pour commettre un affront (« frapper au front ») » (Huston, 2022 : 105). Utilisé comme une arme, le mot injurieux est ressenti comme faisant violence. Et il est de ce fait aussi sanctionné par la loi.

Dans le droit français, le régime juridique de l'injure se base sur deux principes : le respect de la liberté d'expression et la protection des droits d'autrui (droit à l'honneur, à la considération et à la dignité). Dans l'article 29 de la Loi du 29 juillet 1881 sur la liberté de la presse, l'injure est définie comme « toute expression outrageante, terme de mépris ou invective qui ne renferme l'imputation d'aucun fait¹⁹ ». La Loi du 1^{er} juillet 1972 (article 33) sur les injures spéciales ou aggravées repère comme telles tant des injures « commises [...] envers une personne ou un groupe de personnes à raison de leur origine ou de leur non-appartenance à une ethnie, une nation, une race ou une religion déterminée » qu'envers « une personne ou un groupe de personnes à raison de leur sexe, de leur orientation sexuelle ou identité de genre ou de leur handicap²⁰ ».

Le traitement lexicographique des injures peut sembler complexe si l'on en reste au niveau du sens des mots : définies par Huston comme procédant du fait d'« appeler les gens à l'aide de noms qui ne leur "conviennent" pas » (2002 : 104), tout mot qui désigne une personne ou un groupe de personnes et qui n'est pas sanctionné comme « nomination juste » dans la société à laquelle il appartient, est susceptible d'être ressenti comme une injure. Si l'on se base sur les différentes classes d'injures proposées par Huston (2002 : 110), le type auquel appartient une insulte n'influe cependant sur son traitement sémantique que dans le cas de la « nomination littérale » : les mots *youpin*, *rital* ou *boche* sont des synonymes de *juif*, *italien* ou *allemand*, et la description de leurs sens ne diffèrera donc pas ; ce n'est qu'au niveau de leur connotation que l'on peut les repérer comme injurieux. Pour ce qui concerne les autres classes, par contre, l'appellation injurieuse peut toujours être considérée comme un déplacement sémantique et valoir comme sens dérivé du mot. C'est le cas tant pour la « nomination antisphratique » (« où l'on colle sur l'ennemi l'étiquette qui désigne l'opposé diamétral de son "idéal du moi" » (Huston, 2002 : 110), appeler *putain* une femme vertueuse par exemple) que pour

19. https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000006419790

20. https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006070719/LEGISCTA000006165460/2024-01-28

la « nomination métaphorique » (« où l'on évoque des objets ou des qualités dont on prétend qu'ils ont des traits en commun avec l'adversaire » (*ibid.*), appeler par exemple des personnes *thon*, *singe* comme des animaux ou *fumier*, *ordure* comme des objets) et « métonymique » (« où l'on réduit le tout à l'une de ses parties » (*ibid.*), par exemple *trou du cul*).

Dans sa définition de ce type de nomination, le lexicographe devrait rendre compte de la non-identité entre le nom donné et la personne décrite. On peut cependant noter ce que l'on peut considérer comme un manque de distance du lexicographe dans l'apposition de certaines définitions par inclusion qui se caractérisent justement par le fait de pouvoir se substituer à l'entrée décrite. Les « Marie-couche-toi-là » et les « traînées » seraient des femmes faciles ou de mauvaise vie si l'on en croit les définitions de *Marie-couche-toi-là* dans le *Wiktionnaire* à la suite du TLFi qu'il indique comme source : « Femme facile ou débauchée²¹ » ou du *Petit Robert 2025* « Fille de mauvaise vie, prostituée ». De même, le mot *traînée* est défini dans le *Wiktionnaire* comme une « Femme de mœurs légères²² » – ou, de même acabit, « Femme de mauvaise vie (qui "traîne" avec tous les hommes) » dans le *Petit Robert 2025*. Cette identité entre un mot et sa définition peut tout à fait être évitée et ce de plusieurs manières : soit, dans une définition par inclusion, en créant une distance entre la vision de l'autre et la véracité de cette vision comme pour *pétasse* défini dans le *Petit Robert 2025* « Femme, fille que l'on trouve vulgaire ou ridicule » ou en utilisant une définition métalinguistique comme dans le cas de l'article *salope* du *Petit Robert 2025* : « Terme d'injure, pour désigner une femme qu'on méprise pour sa conduite ».

Dans tous les cas, le traitement lexicographique de l'injure ne peut cependant s'arrêter à la description de sa seule dénotation : si l'injure est considérée comme un délit, c'est en raison de l'intention du locuteur qui la profère et donc du contexte pragmatique de son élocution, et de sa réception.

Les injures ont un pouvoir performatif car elles agissent sur notre monde, que ce soit pour la personne qui les utilise ou bien pour celle qui les reçoit voire même pour l'auditoire général. La façon dont nous les percevons leur donne cette puissance. C'est notre ressenti (leur pouvoir perlocutoire) qui leur donne cette puissance d'agir. Les injures agissent et sont donc performatives. (Flory, 2016 : 113-114)

Plusieurs possibilités s'offrent au lexicographe pour rendre compte du performatif, de ce que l'on appelle le « pouvoir des mots » : il peut fournir une marque qui indique l'appartenance d'une lexie à la catégorie des insultes, si tant est qu'elle fait partie des étiquettes prévues dans sa microstructure ; il peut avoir recours à une définition de type métalinguistique qui met au second plan le sens dénotatif d'un mot par rapport à son appartenance à un univers de discours particulier, ce qui est le plus souvent le cas dans le contexte du discours injurieux ; ou bien il peut ajouter une note relative à son utilisation et à sa réprobation par rapport aux normes éthique et sociale.

21. *Marie-couche-toi-là* — *Wiktionnaire, le dictionnaire libre (wiktionary.org)* [page consultée le 13 août 2024]

22. *traînée* — *Wiktionnaire, le dictionnaire libre (wiktionary.org)* [page consultée le 13 août 2024]

Et pourtant, comme nous le verrons par l'analyse d'un échantillon d'insultes sexistes dans les différents dictionnaires tant professionnels que profanes, il est rare que ces insultes soient repérées comme telles. Tout comme le contenu des définitions qui les concerne qui ne permet pas toujours de mesurer leur violence symbolique, l'absence de marquage adéquat tend à minimiser voire à éliminer leur contenu insultant.

6. Le marquage des insultes (marques et indicateurs pragmatiques)

Comme nous venons de l'indiquer, l'insertion d'informations pragmatiques dans les dictionnaires revêt plusieurs formes qui dépendent du système de marquage des données lexicographiques défini par la direction éditoriale de chaque dictionnaire. Nous avons choisi trois dictionnaires de professionnels pour notre étude en raison de leurs particularités par rapport à ce type d'informations : les équipes du Robert peuvent être vues comme des précurseurs par rapport à la vision de la variation linguistique et on peut considérer que l'évolution du rôle des marques à l'intérieur du *Petit Robert* dans les dernières décennies en témoigne (Farina, 2014 et 2020). Sur cette base et dans une optique comparative inédite, la typologie des marques proposée par le dictionnaire *Usito* peut aussi être considérée comme novatrice (Farina, 2020 : 112-113). Si, par contre, on peut observer que le marquage proposé par le TLFi est peu systématique et qu'il ne procède pas de décisions éditoriales clairement définies, ce dictionnaire possède une catégorie particulière dans la structure de ses articles, les crochets, définie dans le paratexte du dictionnaire comme décrivant les « conditions d'emploi de la vedette ». Ces conditions d'emploi correspondent souvent au contexte pragmatique d'utilisation des mots, permettant une description plus précise de ce contexte que ne le font de simples étiquettes. Ces crochets contiennent ainsi souvent des informations détaillées que l'on ne trouve dans aucun autre dictionnaire (Farina, 2009). Sur la base des observations de Michela Murano selon qui, dans les « dictionnaires collaboratifs sur internet qui recensent les mots d'argots et les mots du français familier et populaire » se révèlent des « connaissances des lexicographes profanes [qui] s'avèrent précieuses et peuvent dépasser celles des professionnels » (2017), nous avons sélectionné deux des dictionnaires objets de son étude, le *Bob* et le *Dictionnaire de la Zone*. Pour le premier, elle le repère comme le seul à posséder un système de marquage diastratique et diaphasique, et nous avons pu remarquer qu'il introduit, de plus, des gradations dans la force perlocutoire des injures. Pour le *Dictionnaire de la Zone*, ce n'est pas tant l'apposition de marques mais plutôt celle de notes d'usage, issues parfois de commentaires postés par ses lecteurs, qui nous a semblé intéressante. Le *Wiktionnaire*, comme représentant par excellence de la lexicographie collaborative en ligne, ne pouvait qu'apparaître aussi dans cette comparaison.

Pour le repérage des injures à caractère sexiste comme telles, nous nous basons sur les définitions juridiques tant de l'injure, telle que déjà citée dans la loi française, que du sexisme, pour lequel nous nous appuyons cette fois sur l'énoncé de la « Loi tendant à lutter contre le sexisme dans l'espace public et modifiant la loi du 10 mai 2007 » du gouvernement belge :

Art. 2. Pour l'application de la présente loi, le sexisme s'entend de tout geste ou comportement qui, dans les circonstances visées à l'article 444 du Code pénal, a manifestement pour objet d'exprimer un mépris à l'égard d'une personne, en raison de son appartenance sexuelle, ou de la considérer, pour la même raison, comme inférieure ou comme réduite essentiellement à sa dimension sexuelle et qui entraîne une atteinte grave à sa dignité.²³

À l'instar des juristes qui ont proposé ces définitions, et d'autres chercheurs qui ont analysé ce type d'insultes avant nous, nous considérons que l'insulte sexiste « n'est pas seulement une désignation de l'autre en fonction de son sexe, mais une manifestation de l'inégalité des rapports de sexe dans la société et un moyen de la maintenir » (Lebugle Mojdehi, 2018 : 174). Pour éviter que l'on remette en question le fait de considérer comme telles les insultes sexistes que nous avons choisi d'analyser, nous ne présentons ici qu'une poignée de ces insultes pour lesquelles ce caractère ne semble pas discutable. Les trois premières insultes sont selon le Rapport du Haut Conseil à l'égalité entre les femmes et les hommes celles qui sont le plus fréquemment citées par les femmes françaises qui ont subi une injure sexiste : dans 64 % des cas, en effet, l'insulte dénoncée contient les mots *salope* (27 %), *pute* (21 %) ou *connasse* (16 %) (Haut Conseil à l'égalité, 2019 : 7). Pour pouvoir évaluer la prise en compte d'autres réalités que le français de France, en particulier dans les ressources collaboratives, j'ai ajouté à cette série les mots *guidoune* et *bitch* recensés par Laforest et Vincent parmi les formes « considérées par l'ensemble de la communauté québécoise comme intrinsèquement insultantes (en dehors de tout contexte) » (2004 : 65). Enfin, j'ai ajouté *mal-baisée* dont le contenu sexiste semble lui aussi évident, illustré par Chiflet et Deveaux comme suit :

La pauvre, elle n'a pas eu la chance d'être comblée sexuellement et c'est forcément sa faute si ça se passe mal puisqu'il n'y a pas de « mal-baiseurs ». Quoique, en cherchant bien, on peut toujours trouver un vulgaire « mou du dard » ou un « impuissant », dont *Le Robert* nous explique qu'il est « incapable physiquement d'accomplir l'acte sexuel », nous confirmant qu'un homme ne peut pas « mal baiser ». Il peut ou il ne peut pas, point final. (2019 : 13)

Dans les microstructures des dictionnaires traditionnels, la marque « injurieux » n'est apparue que très récemment. L'augmentation progressive de l'information sur l'utilisation comme injure dans les indications pragmatiques fournies par le TLFi nous permet de dater à la fin des années 1980 une plus grande sensibilité lexicographique à ce phénomène, plus de la moitié des indications relevées se trouvant dans les cinq derniers volumes (sur seize), publiés à partir de 1986. La comparaison des articles *bougnoule* publié en 1975 et indiqué simplement comme « péjoratif » et de *youtre* et *youpin* publiés en 1994 qui sont catégorisés comme « injure à caractère raciste » nous semble significative dans ce sens. Pour ce qui est du *Petit Robert* , un changement est intervenu à la fin de l'année 2023 ou au début 2024 par l'apparition d'une note « Terme d'injure pour une femme » à la fin de la plupart des articles injurieux pour les femmes que nous avons relevés sans notes en juin 2023. C'est la raison pour laquelle on ne la trouve dans la table des abréviations ni du *Petit Robert* , ni du TLFi, mais qu'on la trouve dans celle d' *Usito* . Cela n'empêche ni le *Petit Robert* , ni le TLFi d'étiqueter certains mots en les

23. https://etaamb.openjustice.be/fr/loi-du-22-mai-2014_n2014000586

accompagnant de l'indication « injure » ou « injure raciste » à côté des marques de connotation pour le premier ou dans des crochets destinés aux indications pragmatiques pour le second. Pour aucun des trois cependant il n'a été possible de déduire de la comparaison de mots qui la portent et qui ne la portent pas les critères qui ont été utilisés pour cette apposition, et l'on peut facilement en déduire qu'il n'y a aucune systématisme dans ce marquage. De plus, très peu de mots sont indiqués comme injurieux dans ces dictionnaires : nous n'avons pu chiffrer le nombre d'injures indiquées comme telles dans le dictionnaire *Usito* qui ne permet pas de recherches plein texte mais pour les deux autres dictionnaires professionnels consultés, on recense huit injures désignant des femmes qui sont indiquées comme telles dans le *Petit Robert* dont une seule est définie comme « injure sexiste » - *pisseuse*²⁴ -, dix dans le TLFi (qui a quand même une nomenclature plus de deux fois supérieure à celle du *Petit Robert*). Neuf insultes sont indiquées comme « racistes » dans le *Petit Robert*, deux dans le TLFi.

Certains dictionnaires de profanes, par contre, font montre de bien plus de systématisme dans cet étiquetage. C'est le cas surtout du *Wiktionnaire* où l'étiquette « insulte » est apposée systématiquement sur la plupart des 551 lexies classifiées comme telles dans la catégorie « insultes en français »²⁵, mais aussi du dictionnaire d'argot *Bob* en ligne qui appose très fréquemment l'indicateur « insulte » parfois accompagné d'une catégorisation comme « méprisante » ou « à caractère moral, idéologique ou politique » et a aussi une marque « terme de mépris » qui admet des gradations (« terme de profond mépris », « de mépris moral »). Le *Dictionnaire de la Zone* a aussi une marque « injure » apposée systématiquement sur les lexies injurieuses.

Le tableau ci-dessous nous permettra de comparer le marquage du petit échantillon d'injures sexistes que nous avons présentées plus haut dans les différents dictionnaires cités²⁶ pour montrer combien leur force connotative semble amoindrie par leur traitement lexicographique :

24. Nous ne comprenons pas le principe qui a porté les rédacteurs du *Petit Robert* à le distinguer des autres insultes envers les femmes non caractérisées comme sexistes pour leur part. Dans l'article *pisseuse* du *Dictionnaire de la Zone*, nous trouvons l'information suivante : « Ce terme vient de l'idée sexiste qui sous-entend que les jeunes femmes seraient plus sujettes à l'envie d'uriner que les hommes » (<https://www.dictionnairedelazone.fr/dictionary/lexical/p/pisseuse> [page consultée le 13 août 2024]). Sur cette base, nous avançons une interprétation : contrairement aux autres mots utilisés pour désigner les femmes, celui-ci partirait d'un stéréotype qui ne reposerait pas sur une réalité. En conclure que la comparaison des femmes à des prostituées ne reposerait pas sur une idée tout à fait erronée de leur nature, comme le mot *pétasse* indiqué seulement comme « vulgaire » dans le *Petit Robert* pourrait nous le faire penser, n'est probablement pas une aberration.

25. https://fr.wiktionary.org/w/index.php?title=Cat%C3%A9gorie:Insultes_en_fran%C3%A7ais&pageuntil=fiotte#mw-pages

26. Toutes les citations des articles se réfèrent à une consultation des pages en 2024.

	<i>Petit Robert</i>	<i>TLFi</i>	<i>Usito</i>	<i>Wiktionnaire</i>	<i>Bob</i>	<i>DicoZone</i>
 salope 	fam. vulg. T. d'injure	Trivial/ [Injure forte, à l'adresse d'une femme, avec ou sans idée précise de débauche]	vulg. et péj.	vulg. / péj. / injurieux. Note intégrée pour un des sens ²⁷	Mot injurieux /insulte contre femme méchante, insulte contre femme (avec forte connotation sexuelle, ou mépris)	Injure
 pute 	péj. vulg. T. d'injure	Trivial / en manière d'injure	ABSENT	vulg. et péj. / vulg. Injurieux	péj. / insulte c/	Injure
 connasse 	vulg. et méprisant	vulg.	injur. et vulg.	vulg. Injurieux	insulte générale et méprisante contre femme / terme de mépris	Injure
 bitch 	ABSENT	ABSENT	ABSENT	Anglicisme / Vulgaire	Pas de marque	ABSENT
 guidoune 	Canada péj. vulg. terme injurieux à l'adresse d'une femme (sans connotation sexuelle)	ABSENT	vulg. et péj.	Québec / Québec désuet	ABSENT	ABSENT

27. « Le mot peut être employé de façon ludique entre partenaires sexuels, dans un contexte où l'entente suppose qu'il perde son caractère blessant. »

mal-baisée	Pas de marque	ABSENT	ABSENT	fam.	Formule dépréciative, humiliante	Pas de marque mais une indication dans l'étymologie ²⁸
-------------------	---------------	--------	--------	------	----------------------------------	---

Ce tableau nous permet de confirmer la non-systématicité quant au repérage de la valeur offensive de mots utilisés comme des insultes dans la plupart des ressources. Seuls les mots *pute* et *salope* font l'unanimité sur leur valeur insultante. Pour les autres, il nous paraît évident que l'on peut trouver un modèle de marquage dans les ressources de la langue « non conventionnelle », et en particulier dans les dictionnaires d'argot *Bob* et le *Dictionnaire de la Zone*. Se basant sur des attestations de ces mots dans des discours où ils sont foison et où la violence qu'on peut leur attribuer est sans aucun doute déductible de leur contexte, les auteurs de ce type de dictionnaires ont probablement une meilleure base pour mesurer leur valeur performative.

Josiane Boutet décrit ainsi les effets de la violence verbale véhiculée par l'insulte :

Selon la nature des insultes, mais aussi selon les situations sociales, selon les relations intersubjectives entre les personnes, selon les relations de pouvoir et d'autorité entre elles, les effets sur soi sont très variables : on peut avoir mal, avoir peur, être humilié, vexé, contrarié, désespéré, en colère, blessé, outragé, se sentir calomnié, menacé, etc. (2016 : 201)

Le pouvoir perlocutoire des insultes n'est ainsi pas seulement relié à une intention, mais elle dépend aussi d'une relation entre un locuteur et son interlocuteur et, surtout, de sa réception par cet interlocuteur. Pourtant, si l'on observe d'une manière plus détaillée les informations connotatives et pragmatiques fournies par les dictionnaires, il semble évident que c'est l'intention de celui qui profère l'injure qui est toujours prise en considération. De même que la marque « péjoratif » indique un point de vue de l'interlocuteur exprimé dans l'usage d'un mot, la catégorisation du contexte d'énonciation des insultes indique la cible visée par cet interlocuteur : l'insulte est « à l'adresse d'une femme » ou « contre une femme ». Seule l'indication « humiliante » proposée par le *Bob* dans l'article « mal-baisée » fait état de ce qui peut être ressenti par l'interlocutrice.

Confirmant sa grande hospitalité envers la « sagesse des foules », le *Wiktionnaire* accueille pour sa part une acception particulière de *salope* classifiée comme appartenant strictement à la « sexualité » qui contient une note que l'on pourrait considérer comme assumant aussi le point de vue de la réceptrice : la *salope* serait dans ce sens un « partenaire masculin ou féminin, salace, lubrique ou soumis. » et la note indique que « le mot peut être employé de façon ludique entre partenaires

28. « Exprime l'idée machiste qu'une femme qui n'obtient pas de satisfaction sexuelle ne peut qu'être aigrie. » <https://www.dictionnairedelazone.fr/dictionary/lexical/m/baiser> [page consultée le 13 août 2024] (Information étymologique provenant du *Dictionnaire de l'argot* de Larousse)

sexuels, dans un contexte où l'entente suppose qu'il perde son caractère blessant.²⁹ » Ne reste-t-on pas ici dans une interprétation orientée, tout comme dans le cas des autres marques, par un rédacteur qui n'assume qu'un point de vue, celui d'un locuteur qui a le rôle de meneur de jeu et qui définit la réaction de son interlocuteur (en particulier, de « l'autre sexe », clairement féminin) en fonction d'intentions qu'il dépouille de toute malignité ? Le traitement lexicographique illustré ici participe selon nous à consolider ce que le Conseil appelle « l'entre-soi "masculin" » et par là même à « renforcer les stéréotypes de sexe » et « légitimer les inégalités » (Haut Conseil à l'Égalité entre les femmes et les hommes, 2019 : 6).

7. Conclusion

Nadine Vincent souligne la responsabilité importante qui repose sur le lexicographe par rapport à son public : son discours, discours d'autorité reconnu comme tel, est de ce fait un « outil idéologique très puissant » (2022 : 125). Elle analyse aussi, à partir de l'exemple des usages « polémiques » de *woke* la complexité d'une approche neutre de la part du dictionnaire qui, tout en étant « ouvert à l'évolution de la langue », ne devrait pas être « influençable par les groupes de pression qui composent la société ». Pour éviter cet écueil, et pouvoir refléter les différents points de vue d'une société plurielle, il devrait selon elle se faire porteur de la « polyphonie » qui émerge dans l'espace public.

Cette affirmation nous semble tout à fait pertinente pour ce qui concerne la néologie néo-féministe qui prolifère dans les dictionnaires collaboratifs, mais elle ne s'applique pas vraiment au problème de la sanction de la discrimination d'origine sexiste. Dans le cas des violences sexistes telles qu'elles s'expriment à travers notre langue, et de la diffusion de stéréotypes qui découlent de cette vision inégalitaire entre hommes et femmes dans nos sociétés, en effet, on peut considérer qu'elle n'est plus source de polémiques sociétales aujourd'hui. Le Haut Conseil à l'égalité entre les femmes et les hommes observe aujourd'hui encore, dans son rapport annuel de 2024, une forte présence du sexisme en France, voire une augmentation des discours et des violences sexistes, mais il n'en est pas de même des jugements par rapport à ce phénomène. Ce rapport souligne en effet que, paradoxalement, s'il y a une recrudescence des pratiques et des discours sexistes, « la population est de plus en plus consciente et tolère de moins en moins les violences sexistes et sexuelles ». Ce qui illustrerait « le décalage entre cette prise de conscience et le maintien des stéréotypes qui continuent de forger les mentalités et les comportements³⁰ ».

Aussi la trace d'un discours d'un autre temps, d'une idéologie dépassée, où la femme, pour des lexicographes exclusivement de sexe masculin, jouait le rôle de l'autre infériorisée et dépréciée, subsiste-t-elle dans ces ouvrages.

29. [salope — Wiktionnaire, le dictionnaire libre \(wiktionary.org\)](#) [page consultée le 13 août 2024]

30. <https://www.haut-conseil-egalite.gouv.fr/stereotypes-et-roles-sociaux/travaux-du-hce/article/6eme-etat-des-lieux-du-sexisme-en-france-s-attaquer-aux-racines-du-sexisme>

L'observation du Haut Conseil sur le fait que les manifestations du sexisme sont « perçues comme banales », et qu'elles bénéficient d'une « tolérance sociale » rend difficile la possibilité de recul critique par rapport à ces manifestations pour ceux qui travaillent à la rédaction des dictionnaires. La plus grande participation de rédacteurs, de provenances plus variées que celle des lexicographes professionnels, à la création et modification des articles de dictionnaires dans les ressources profanes semble pouvoir aider la lexicographie à bénéficier d'un sens critique élargi par rapport à un phénomène comme le sexisme. C'est ce qui semble ressortir d'un usage plus systématique de l'indication des insultes comme telles dans un dictionnaire collaboratif comme le *Wiktionnaire*.

C'est ce que nous avons tenté de démontrer par notre analyse des marques et des indicateurs pragmatiques relatifs aux injures sexistes, mais nous avons vu qu'on pourrait aussi l'analyser dans le maintien de renvois à l'intérieur des articles ou de la présentation des éléments définitionnels.

Toutes ces observations nous poussent à conclure que le traitement des mots qui relèvent du sexisme ordinaire devraient faire l'objet d'un travail critique approfondi. À l'instar de Vincent, nous avons aussi tenté de souligner des éléments qui nous semblaient potentiellement enrichissants pour les différents types de ressources dans la perspective d'une influence réciproque. Comme elle, nous souhaiterions en effet que « la lexicographie professionnelle et la lexicographie collaborative s'influencent et se contaminent, l'une transmettant sa méthode, et l'autre son ouverture sur l'ensemble des emplois qui composent la langue française » (2017).

Bibliographie

- BOUTET, Josiane (2016), *Le pouvoir des mots*, Paris : La Dispute.
- CAMPBELL, Elisabeth (2006), « Le Traitement des insultes sexistes et racistes dans les dictionnaires bilingues », *The French Review*, Vol. 80, No. 1, p. 112-137 URL : <https://www.jstor.org/stable/25480589>
- COTTER, Colleen et John Damaso (2007), « Online Dictionaries as Emergent Archives of Contemporary Usage and Collaborative Codification », *QMOPAL - Queen Mary's Occasional Papers Advancing Linguistics*, Londres : Queen Mary University of London, p. 1-11.
- DE SCHRUYVER, Gilles-Maurice (2003), « Lexicographers' dreams in the electronic-dictionary age », *International Journal of Lexicography*, 16(2), p. 143-199.
- DOLAR, Kaja (2017), « Les dictionnaires collaboratifs non institutionnels dans l'espace francophone : éléments de typologie et bilan », dans C. Molinari et N. Vincent (dir.), *Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, Repères DoRiF, 14. URL : <https://www.dorif.it/reperes/kaja-dolar-les-dictionnaires-collaboratifs-non-institutionnels-dans-lespace-francophone-elements-de-typologie-et-bilan/>
- DUBOIS, Jean et Claude Dubois (1971), *Introduction à la lexicographie : le dictionnaire*, Paris, Larousse.
- FARINA, Annick (2005), « Lexicographie et discrimination », dans Raus, R. (éd.), *Linguaggi e discriminazione*, Centro internazionale di Ricerca sulle Donne, Université de Turin. URL: <https://www.cirsde.unito.it/it/formazione/corso-line-introduzione-agli-studi-di-genere/moduli-di-secondo-livello/linguaggi-e>
- FARINA, Annick (2009), « Problèmes de traitement des "pragmatèmes" dans les dictionnaires bilingues », dans Heinz, M. (dir.), *Le Dictionnaire maître de langue*, Berlin : Frank & Timme, coll. Metalexikographie, p. 245-264
- FARINA, Annick (2014), « Les marques d'usage dans les dictionnaires monolingues français : l'exemple du *Petit Robert* » dans Heinz, M. (dir.), *Les sémiotiques des dictionnaires*, Berlin : Frank & Timme, coll. Metalexikographie, p. 111-132.
- FARINA, Annick (2020), « Les marques dans les dictionnaires bilingues généraux : l'exemple des marques de domaine dans les dictionnaires français-italien », dans G. Armianov (éd.), *Marques de registre dans les dictionnaires bilingues*, Paris : Inalco Presses, p. 109-130
- FARINA, Annick (2022), « Dizionari e stereotipi di genere », Podcast *Forlilpsi per l'inclusione*. URL: <https://sites.google.com/forlilpsi.unifi.it/public-engagement/forlilpsi-per-linclusione/6-inclusione-di-genere-e-lgbtq/dizionari-e-stereotipi-di-genere?authuser=0>
- FARINA, Annick (2024 à paraître), « Dirigismo e inclusione di genere: i diversi tipi di norme nell'insegnamento delle lingue straniere », dans E. Jafrancesco et al. (éds), *Educazione all'uguaglianza di genere ed educazione linguistica*, Florence : Firenze University Press.
- FLORY, Julienne (2016), *Injuriez-vous. Du bon usage de l'insulte*, Paris : La Découverte.

- FOUBERT, Océane (2023), « Neologisms in contemporary feminisms: For a redefinition of feminist linguistic activism », *GLAD!* 15. URL: <http://journals.openedition.org/glad/7722>
- FRAISSE, Geneviève (1996), *La différence des sexes*, Paris : PUF.
- Haut Conseil À l'Égalité entre les femmes et les hommes (2019), *1er état des lieux du sexisme en France*, Rapport no 2018-01-07 STER 038, Paris.
- HUSTON, Nancy (2002), *Dire et interdire. Eléments de jurologie*, Paris : Payot et Rivages.
- KOTTELAT, Patricia (2009), « Lexicographie et clichés doxiques : la permanence de stéréotypes sexistes dans un dictionnaire d'apprentissage contemporain, le *Robert Junior 99* », *Studi e Ricerche, Quaderni del Dipartimento di Scienze del Linguaggio e Letterature Moderne e Comparate dell'Università di Torino*, 4, p. 165-185.
- LAFOREST, Marty et Diane Vincent (2004), « La qualification péjorative dans tous ses états », *Langue française*, n° 144, p. 59-81. URL : <https://www.cairn.info/revue-langue-francaise-2004-4-page-59.htm>
- LEBUGLE MOJDEHI, Amandine (2018), « Stéréotypes de genre et sexisme : principaux registres d'insultes dans les espaces publics », *Cahiers du Genre*, n° 65, p. 169-191. URL : <https://www.cairn.info/revue-cahiers-du-genre-2018-2-page-169.htm>
- Le Petit Robert de la langue française 2025* (2024), Dictionnaires Le Robert, édition numérique.
- MEYER, Christian M. et Iryna Gurevych (2012), « Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography », dans Granger, S. et Magali Paquot (éds.), *Electronic Lexicography*, Oxford : Oxford University Press, p. 259-291.
- MOLINARI, Chiara (2021), « La lexicographie numérique : un nouvel outil pour l'enseignement du FLE ? », *Études de linguistique appliquée*, 201, 27-47. URL : <https://doi.org/10.3917/ela.201.0028>
- MURANO, Michela (2014), « La lexicographie 2.0 : nous sommes tous lexicographes ? », *Cahiers de recherche de l'école doctorale en linguistique française*, n° 8, Trieste : Edizioni Università di Trieste, p. 147-162.
- MURANO, Michela (2017), « Une lexicographie deux fois populaire : les dictionnaires collaboratifs du français 'non conventionnel' », *Repères DoRiF, n. 14 – Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, URL : <https://www.dorif.it/reperes/michela-murano-une-lexicographie-deux-fois-populaire-les-dictionnaires-collaboratifs-du-francais-non-conventionnel/>
- MURANO, Michela (2019), « Les dictionnaires collaboratifs et la phraséologie : de la description du patrimoine existant à l'invention de nouvelles séquences figées », *Éla. Études de linguistique appliquée*, 2019/2 (N° 194) : 193-209.
- PRUVOST, Jean (2008), « Dictionnaires d'hommes et/ou de femmes : parcours historique, bibliographique et heuristique », dans Farina, A. et Raus, R. (dir.), *Des mots et des femmes : rencontres linguistiques*, Florence : FUP, p. 41-68.
- REY, Alain (1983), « Norme et dictionnaire (domaine du français) », dans É. Bédard et J. Maurais (dir.), *La norme linguistique*, Québec – Paris : Conseil de la langue française - Le Robert, p. 541-569.

- REY-DEBOVE, Josette (1967), « La définition lexicographique : bases d'une typologie formelle », *Travaux de Linguistique et de Littérature*, n. V, 1, p. 141-159.
- REY-DEBOVE, Josette (2003), « La philosophie des dictionnaires Le Robert ou les chemins de l'intelligible » dans A. Francoeur et al., *Les dictionnaires Le Robert : Genèse et évolution*, Montréal : Presses de l'Université de Montréal, p. 100-109.
- REY-DEBOVE, Josette (2004), « Les femmes et les dictionnaires », *Revue d'aménagement linguistique*, 107, p. 227-234.
- SAJOUS, Franck, Amélie Josselin-Leray et Nabil Hathout (2018), « The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology », *Lexis*, 12, URL: <http://journals.openedition.org/lexis/2322>
- SAJOUS, Franck et John Humbley (2022), « Mesures d'isolement sanitaire dans *Wiktionnaire* et *Wikipédia* : néologie et lexicographie ou néonymie et terminographie ? », *Estudios Románicos*, 31 : 175-201.
- SAJOUS, Franck (2023), « Quantité et qualité dans le *Wiktionnaire* : de la diversité... à la rigueur ? », dans Nadine Vincent (dir.), *La lexicographie en ligne contribue-t-elle à une meilleure description du français ?*, *Linx*, n° 86, (2023-1). URL : <http://journals.openedition.org/linx/9835>
- STEFFENS, Marie (2017), « Lexicographie collaborative, variation et norme : le projet 10-nous », *Repères DoRiF, n. 14 – Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, URL : <https://www.dorif.it/reperes/marie-steffens-lexicographie-collaborative-variation-et-norme-le-projet-10-nous/>
- VINCENT, Nadine (2017), « Présence et description d'emplois québécois dans des dictionnaires disponibles gratuitement en ligne », *Repères DoRiF, n. 14 – Dictionnaires, culture numérique et décentralisation de la norme dans l'espace francophone*, URL : <https://www.dorif.it/reperes/nadine-vincent-presence-et-description-demplois-quebecois-dans-des-dictionnaires-disponibles-gratuitement-en-ligne>
- VINCENT, Nadine (2022), « Faut-il adapter les dictionnaires à l'air du temps ? Proposition d'un traitement polyphonique du mot woke. Regards linguistiques sur des mots polémiques », *Circula*, 15, p. 122-145.

TITRE: *FÉMINICIDE* : UN MOT AMBIGU POUR UNE SOCIÉTÉ DIVISÉE

AUTEURE: LAETITIA CHICOINE, UNIVERSITÉ DE SHERBROOKE

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 296-322

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22409](http://hdl.handle.net/11143/22409)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22409](https://doi.org/10.17118/11143/22409)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Féminicide : un mot ambigu pour une société divisée

Laetitia Chicoine¹, Université de Sherbrooke

Résumé :

Le sémantisme de *féminicide* ne fait pas consensus dans la francophonie. Certains parlent de meurtre misogynne, de crime de possession ou de crime contre une conjointe alors que d'autres y voient un homicide comme un autre. Ces divergences d'opinions ne sont toutefois pas reflétées dans les définitions qu'offrent actuellement les dictionnaires. Le débat témoigne de la lourde charge sociale que le mot porte en lui. L'équipe du Projet porc-épic désire éclaircir cette controverse et redéfinir l'emploi sensible qu'est *féminicide*. Dans l'objectif de décrire les emplois polémiques de l'actualité, nous proposons un traitement lexicographique qui tire avantage de l'espace numérique et qui ne se restreint pas aux limites imposées aux dictionnaires traditionnels.

Mots-clés :

Féminicide, dictionnaires en ligne, mots polémiques, lexicographie, Projet porc-épic

Abstract :

There is no consensus on the semantics of *feminicide* in the French-speaking world. Some call it misogynistic murder, a crime of possession or a crime against a spouse, while others see it as a homicide like any other. However, these differences of opinion are not reflected in current dictionary definitions. This debate bears witness to the heavy social burden that the word carries. That's why the Projet porc-épic team wants to shed light on this controversy and redefine the sensitive term *feminicide*. Projet porc-épic, whose mandate is to describe polemical uses of current events, proposes a lexicographic treatment that takes advantage of digital space.

Keywords :

Feminicide, online dictionaries, polemical words, lexicography, Projet porc-épic

1. L'auteure remercie la professeure Nadine Vincent pour ses commentaires pertinents et constructifs ainsi que pour la relecture attentive de la première version de cet article.

1. Introduction

Selon l'Office des Nations unies contre les drogues et le crime (ONUDC), 87 000 femmes ont été tuées intentionnellement en 2017. Parmi celles-ci, 58 % ont perdu la vie en raison d'un ex-conjoint, d'un conjoint ou d'un membre de la famille (*Global Study on Homicide*, 2019 :10). Ce phénomène de violence envers les femmes est de plus en plus connu sous le nom de *féminicide*. Le terme, qui rappelle entre autres le terme *génocide* ou *infanticide*, est un mot qui frappe, qui attire l'attention. Les médias l'ont employé et diffusé notamment pour désigner les meurtres de femmes commis par un conjoint ou un ancien conjoint.

Si la fréquence alarmante des meurtres de femmes explique leur présence importante dans les médias au cours des dernières années, le terme *féminicide* a aussi retenu l'attention en raison d'un non-consensus sur ce qu'il désigne. Alors qu'au Québec, *féminicide* a intégré l'usage sans trop de vague, les journaux européens, principalement de France, ont servi de vitrine au débat public. En effet, pour certains, le meurtre d'une femme par un conjoint ou un ex-conjoint ne correspond pas à un féminicide alors que pour d'autres, oui. Cette divergence sémantique, observable dans l'usage, transparait aussi dans les statistiques officielles des États qui ne possèdent pas tous la même définition du féminicide (lorsqu'il figure dans le code de lois), ce qui rend difficile la comparaison des données (*Global Study on Homicide*, 2019 : 24).

En 2012, avant que l'usage de *féminicide* explose dans les médias et que s'ensuive une fluctuation des sens qui y sont associés, l'Organisation mondiale de la Santé (OMS), l'UNODC ainsi que le *Wiktionnaire* proposent une définition du mot. Dans les années qui suivent, *féminicide* sera attesté dans plusieurs autres dictionnaires (voir section 6).

Encore aujourd'hui, l'on se rend compte que le mot *féminicide* est utilisé pour renvoyer à différentes réalités et lorsqu'on consulte les ouvrages de référence, la réponse qu'on y trouve ne semble pas convenir à l'emploi qui en est fait dans la société. C'est pourquoi le Projet porc-épic, qui étudie les emplois polémiques de l'actualité, désire éclaircir ce cas. L'objectif est d'observer l'emploi qu'il est fait de *féminicide* au Québec comme en Europe francophone afin de proposer une description lexicographique sans ambiguïté de ce mot controversé.

Cet article s'attardera, dans un premier temps, à l'entrée de mot *féminicide* dans le domaine de la recherche; à la manière dont il est décrit par les organismes mondiaux et dans la loi; à son usage dans les journaux du Québec et de l'Europe francophone et au débat qu'il a suscité dans les journaux. Dans un second temps, nous observerons la manière dont il a été décrit dans différents dictionnaires et nous en proposerons un traitement dans le Projet porc-épic.

2. Le mot *féminicide* dans différents milieux

Cet article se concentre sur le traitement qu'en ont fait les dictionnaires, mais il faut savoir que le féminicide a été décrit dans d'autres contextes hors de la discipline lexicographique, notamment dans le milieu de la recherche, dans les codes de loi et par certains organismes internationaux. La pluridisciplinarité du concept implique nécessairement une certaine fluctuation dans les définitions.

2.1. L'émergence du mot dans le milieu de la recherche

Le terme a d'abord émergé sous la forme de *femicide* en anglais. C'est la sociologue américaine Diana E. H. Russell qui l'utilise pour la première fois en 1976. Elle ne le définit toutefois qu'en 1990, à la suite du drame de l'école Polytechnique à Montréal ayant eu lieu l'année précédente où 14 étudiantes en génie ont été tuées (Dawson et Carrigan, 2020 : 3 ; Fregoso, 2022 : 318). À ce moment, Russell définit le terme comme « the murder of women by men motivated by hatred, contempt, pleasure, or a sense of ownership of women² » (Caputi et Russel, 1990 : 34 dans Dawson et Carrigan, 2020 : 3). Le choix du mot *femicide*, construit sur la même forme que *génocide* (à l'aide du suffixe *-cide*, « tuer »), avait comme objectif de mettre en lumière que les violences spécifiquement dirigées envers les femmes résultent d'une idéologie et d'un système de pouvoir, ce que n'illustre pas le terme *homicide* (Von Joeden-Forgey, 2022 : 583).

Le néologisme est repris quelques années plus tard, en 1994, par l'anthropologue mexicaine Marcela Lagarde alors que la découverte de centaines de meurtres de femmes au Mexique fait la couverture dans les médias. Lagarde est l'une des premières en Amérique latine à qualifier ce type d'assassinats de *féminicide* (Fregoso, 2022 : 319). En espagnol, c'est la forme *feminicidio* qui est préférée à *femicidio* pour conserver la racine latine *femini* et pour éviter que l'homicide soit perçu comme le meurtre d'un homme et le féminicide, comme le meurtre d'une femme sans considérer le motif du genre comme déclencheur de la violence (Lapalus et Mora, 2020 : 156). Néanmoins, la forme *femicidio* est également attestée dans le milieu académique et politique (Bellami, 2018 : 138). Pour ce qui est de la francophonie, la graphie *féminicide*, choisie selon le même raisonnement que *feminicidio*, constitue la forme la plus répandue, mais coexiste avec la variante *fémicide*. D'ailleurs, cette seconde forme est considérée comme un anglicisme dans certains ouvrages de référence³, car elle est construite à partir de *fem-* qui n'est pas une forme productive en français.

2. ... « le meurtre de femmes par des hommes motivés par la haine, le mépris, le plaisir ou un sentiment de propriété à l'égard des femmes » (traduction libre)

3. Le *Grand dictionnaire terminologique* (GDT) et *Le Petit Robert 2024* indiquent que la variante *fémicide* est un anglicisme.

Dans les années 2000, la définition évolue et le féminicide devient « le meurtre de femmes commis par des hommes parce qu’elles sont femmes » (Russel, 2001 dans Dawson et Carrigan, 2020 : 3). À l’époque, le terme était surtout associé aux meurtres des femmes en Amérique latine puisqu’il a été très tôt utilisé dans les travaux des chercheuses et dans les lois de cette région. Ces nombreuses études soulignent d’ailleurs la passivité des gouvernements et leur responsabilité dans la perpétuation des féminicides (Dawson et Carrigan, 2020 : 5). Le terme est désormais reconnu à l’échelle mondiale. En 2012, le Conseil académique des Nations unies organise d’ailleurs la première conférence mondiale, en Autriche, à propos de l’enjeu du féminicide.

Le féminicide continue d’être un sujet de recherche d’actualité, avec de nouvelles études récentes. Par exemple, Dawson et Carrigan (2021) ont examiné différents motifs et indicateurs liés au sexe ou au genre (« sex/gender-related motives and indicators » ou SGRMI) dans le but de faire du féminicide un crime distinct, c’est-à-dire qui possède des caractéristiques qui lui sont propres. Les résultats témoignent que certains de ces facteurs sont significativement plus fréquents dans les meurtres perpétrés par des hommes sur des femmes que dans les autres types d’homicides. L’étude relève plusieurs caractéristiques précédant l’incident ayant une forte corrélation avec des meurtres de femmes commis par des hommes comme des contacts antérieurs avec la police, une séparation récente ou des menaces antérieures (Dawson et Carrigan, 2021 : 9). D’autres auteurs ont proposé une définition avec l’intention que le féminicide soit défini autrement que par les sexes respectifs des personnes impliquées. C’est le cas de Cecchi *et al.* qui proposent que le féminicide soit défini « as a murder perpetrated because of a failure to recognize the victim’s right to self-determination⁴ » (2022 : 5). Ici, les auteurs mettent de l’avant que la perpétuation du féminicide est étroitement liée à la violation des droits de la femme, dont l’absence de reconnaissance du droit, pour les femmes, de prendre leurs propres décisions et de vivre leur vie selon les conditions qui leur conviennent.

2.2. La récupération du mot dans les organisations internationales et nationales

En 2012, l’OMS propose quatre catégories distinctes pour décrire le féminicide. La première catégorie, le « féminicide intime », qui est « commis par un époux ou par un petit ami, actuel ou ancien » (OMS, 2012), correspond au sens qui ne fait justement pas l’unanimité au sein de la population. La deuxième catégorie est le « féminicide au nom de l’honneur » qui renvoie au meurtre d’une femme sous prétexte que celle-ci aurait déshonoré sa famille par un comportement dit inacceptable comme une relation sexuelle hors mariage ou une tenue vestimentaire jugée inappropriée. La troisième catégorie est le « féminicide lié à la dot », c’est-à-dire lorsque de jeunes femmes sont tuées par leur belle-famille ou harcelées jusqu’au suicide pour avoir apporté une somme d’argent insuffisante lors du mariage. La quatrième catégorie est le « féminicide non intime » qui correspond au meurtre d’une femme commis par un individu avec lequel la victime n’a pas été en relation (OMS, 2012). L’ONU

4. ... « comme un meurtre perpétré en raison de la non-reconnaissance du droit de la victime à l’autodétermination » (traduction libre)

définit également le féminicide, mais pour sa part sous onze formes différentes. Celles-ci reprennent entre autres le féminicide lié à la dot, le crime dit d'honneur et le meurtre commis par un partenaire tout comme l'OMS, mais aussi d'autres situations telles que le meurtre de femmes en contexte de conflits armés, à la suite de mutilations génitales, après accusation de sorcellerie ou encore un fœticide ou un infanticide (où le fœtus ou l'enfant était de sexe féminin) (*Global Study on Homicide*, 2019).

L'ONUUDC en collaboration avec l'Entité des Nations unies pour l'égalité des sexes et l'autonomisation des femmes (ONU Femmes) a également créé un cadre statistique où dix caractéristiques et contextes ont été identifiés afin de mieux baliser les féminicides et d'avoir des points de référence communs à l'international. Parmi ces éléments, on retrouve une femme ou une fille tuée par son partenaire intime ou par un membre de la famille, une femme qui a subi de la violence précédemment à son meurtre, une femme qui a été exploitée illégalement avant de mourir, etc. (OCFJR, 2024). Au Canada, l'Observatoire canadien du féminicide pour la justice et la responsabilisation (OCFJR), créé en 2017, s'inspire des avancées de l'ONUUDC et de l'ONU Femmes pour ses propres recherches et notamment pour sa conceptualisation du féminicide. Par exemple, l'OCFJR identifie des indicateurs pouvant précéder un féminicide tels que les antécédents d'actes violents, la misogynie, une grossesse, une séparation ou des comportements de contrôle (OCFJR, 2019 : 45-59).

2.3. Le féminicide dans les textes de loi

Il est intéressant d'observer que le féminicide est reconnu par plusieurs législations à travers le monde et de voir comment ce crime est décrit dans les différents articles de loi.

En date de 2022, 19 pays, tous de l'Amérique du Sud et de l'Amérique centrale, avaient intégré le féminicide à leur code de lois. Depuis, quatre pays de l'Union européenne lui ont donné un statut juridique : la République de Chypre, la République de Malte, la Croatie et la Belgique.

Tableau 1 : Législations portant sur le féminicide/fémicide (source : OCFJR, 2022)

Tableau 2.1: Résumé du type et de l'année de la législation sur les fémicides ou féminicides par pays

Pays	Année	Loi or Droit?	Législation	Terme utilisé
Argentine	2012	Loi n° 26791 – Modifications du code pénal sur le fémicide	Non	Fémicide
Belgique	2022	La Belgique va légiférer contre le femicide	En cours	--
Bolivie	2013	Loi n° 348 – Loi globale visant à garantir aux femmes une vie sans violence	Oui	Féminicide
Brésil	2015	Loi n° 13.104 – Code pénal prévoyant que le féminicide constitue une circonstance qualifiante pour le crime d'homicide	Non	Féminicide
Chili	2010	Loi n° 20.480 – Modifie le code pénal et la loi 20.066, établissant le "féminicide"	Non	Fémicide
Colombie	2008/ 2015	Réforme du Code pénal/Loi Rosa Elvira Cely	Non	Féminicide
Costa Rica	2007	Loi n° 8589 sur la pénalisation de la violence à l'égard des femmes	Non	Fémicide
République dominicaine	2014	Loi n° 550-14 – Établit le Code pénal de la République dominicaine	Non	Féminicide
Équateur	2014	Code pénal intégral organique	Non	Fémicide
El Salvador	2010	Loi spéciale globale pour une vie sans violence pour les femmes	Oui	Féminicide
Guatemala	2008	Loi contre le fémicide et autres formes de violence à l'égard des femmes	Oui	Fémicide
Honduras	2013	Décret 23-2013 portant réforme du code pénal	Non	Fémicide
Malte	2022	Le féminicide comme circonstance aggravante de l'homicide	Oui	Féminicide
Mexique	2012	Réforme du code pénal Droit général de l'accès des femmes à une vie sans violence	Oui	Féminicide
Nicaragua	2012	Loi n° 779 - Loi intégrale contre la violence à l'égard des femmes et réformes de la loi n° 641, "Code pénal".	Oui	Fémicide
Panama	2013	Adopter des mesures pour prévenir la violence à l'égard des femmes et réformer le code pénal pour criminaliser le féminicide et punir les actes de violence à l'égard des femmes	Oui	Fémicide
Paraguay	2016	Art. 50 de la loi n° 5777 - De la protection intégrale des femmes, contre toutes les formes de violence	Oui	Féminicide
Pérou	2013	N° 29819 - Loi qui modifie l'article 107 du code pénal en y intégrant le féminicide	Oui	Féminicide
Uruguay	2017	No. 19538 - Loi modifiant l'article 312 du Code pénal	Non	Fémicide
Venezuela	2007/ 2014	Loi organique sur le droit des femmes à une vie sans violence	Oui	Fémicide

D'autres traités politiques ont été mis en place contre les violences de genre. Parmi ceux-ci, on retrouve la Convention interaméricaine pour la prévention, la répression et l'éradication de violence contre les femmes (Congrès Belém do Pará), premier traité international (OCFJR, 2019 : 20) ainsi que la Convention d'Istanbul, instaurée en 2011 et aujourd'hui signée par près de 40 États. De la même manière, quelques pays comme l'Espagne et l'Italie ont également instauré une loi contre la violence de genre, sans aborder spécifiquement le féminicide toutefois.

En Amérique centrale et en Amérique du Sud, le féminicide a été inclus de différentes façons dans le code juridique. Selon les États, le féminicide peut correspondre, par exemple, à une circonstance aggravante d'un homicide, à un crime autonome intégré dans le Code pénal ou à un crime en dehors du Code pénal, jugé dans une loi autonome (Bellami, 2018 : 135-136 ; OCFJR, 2019 : 18). Cette dernière méthode, en présentant le féminicide comme un crime distinct, offre une reconnaissance légale et sociale plus importante.

Ce qui retient notre attention est la manière dont est défini le féminicide dans ces lois. Même si la description varie d'un État à l'autre, un élément revient dans la majorité des pays : le féminicide est restreint à la sphère intime, ce qui signifie le plus souvent que le crime doit être commis par un membre de la famille ou par le conjoint pour être reconnu comme tel. Parfois, même le crime commis par un ancien conjoint n'est pas considéré comme un féminicide (Bellami, 2018 : 138). Néanmoins, certains pays, dont le Costa Rica et le Pérou, incriminent les féminicides non intimes, c'est-à-dire ayant lieu dans la sphère publique (*Global Study on Homicide*, 2019 : 58-61).

La divergence des définitions va plus loin que la considération de la sphère intime ou publique. La perspective adoptée varie selon les pays. Par exemple, la Bolivie a désiré mettre de l'avant que le féminicide est souvent le point culminant d'une série de violences exercée par le même individu, alors que le Guatemala a plutôt défini le féminicide du point de vue d'un rapport de domination ayant lieu « dans le contexte des relations inégales de pouvoir entre les hommes et les femmes » (article 3 E, dans Bellami, 2018 : 139). Pour sa part, la Colombie définit le féminicide comme le meurtre d'une femme « en raison de sa condition de femme ou en raison de son identité de genre », ce qui inclut les victimes transsexuelles (Bellami, 2018 : 138-139).

En juin 2023, la Belgique, avec sa *Loi sur la prévention et la lutte contre les féminicides, les homicides fondés sur le genre et les violences*, se distingue de la majorité des pays latino-américains en englobant une multitude de contextes sous le terme *féminicide*. En effet, la nouvelle loi belge définit quatre catégories de féminicide : le féminicide intime, le féminicide non intime, le féminicide indirect (qui résulte d'une violence comme un avortement forcé ou la mutilation génitale féminine) et l'homicide fondé sur le genre, ce qui inclut tous les membres de la communauté LGBTQI+ (*Loi sur la lutte et prévention des féminicides*).

Tout compte fait, les définitions légales du féminicide, comme les emplois dans l'usage, sont diverses et plus ou moins restreintes selon l'État. Il faut garder en tête qu'il s'agit de définitions qui ont été bâties selon différents systèmes de loi préexistants et dans différentes langues. Le cadre juridique impose des éléments dont les chercheurs et les lexicographes, par exemple, n'ont pas à tenir compte dans leur description. L'absence d'une définition consensuelle et universelle n'a donc rien d'étonnant. Néanmoins, le survol des définitions juridiques permet de constater que le féminicide acquiert un statut de plus en plus important à l'échelle mondiale au fil du temps.

3. L'usage dans les médias

L'emploi de *féminicide* dans les articles de journaux du Québec et de l'Europe francophone a explosé en 2019. C'est toutefois en 2021 que le nombre d'occurrences atteint son paroxysme. Puisque le terme était peu employé auparavant et que la polémique a enflammé les médias francophones environ au même moment, la majorité des occurrences analysées pour la rédaction de cet article ont été publiées entre 2018 et 2022. Les articles de journaux ont été collectés à partir de la banque de

données *Eureka*. Les articles analysés font partie soit du groupe de sources « Canada fr », soit de celui « Europe fr ». Sans prétendre avoir analysé l'ensemble des occurrences se situant dans cet intervalle de temps, nous avons dépouillé les contextes de plusieurs centaines d'articles afin d'en dégager le sens octroyé à *féminicide* et de noter s'il y avait prise de position dans le débat sémantique concernant le terme. Les informations ont été consignées dans une grille d'annotation.

Avant de s'attarder à ce qui ressort dans l'usage journalistique, il faut savoir que quelques événements constituent des références en matière de féminicide au sein de la francophonie. Ces archétypes, qui ont forgé l'image collective liée à ce type de meurtre, influencent nécessairement l'emploi du mot. D'abord, en contexte québécois, la tragédie de Polytechnique en 1989, à Montréal, est l'un des événements étroitement associés au féminicide. Dans cette tuerie, un homme a divisé une classe d'étudiants en deux groupes, les hommes d'un côté et les femmes de l'autre, avant de tirer sur ces dernières. Cette tragédie dans laquelle 14 femmes ont été tuées et 13 blessées est décrite à répétition dans les médias comme « le premier féminicide de masse de l'histoire du Québec » (Michaud, 2021), même s'il a fallu plusieurs années avant que l'évènement soit nommé comme tel : « Il y avait un "avant Poly" et un "après Poly". Il y avait aussi des conditions sociales qui avaient permis que cette tuerie se produise. Et il aura fallu près de 30 ans pour utiliser les mots "attentat antiféministe" et "féminicide". » (Elkouri, 2019)

De même, les femmes autochtones disparues et assassinées au Canada au cours des dernières décennies sont associées au féminicide. Le haut taux de violence vécu par les femmes autochtones a été mentionné à répétition dans les médias : « Certains considèrent également la mort ou la disparition des 1181 femmes autochtones entre 1980 et 2012 (3 femmes par mois) comme autant de féminicides canadiens passés sous silence. » (Lord, 2016)

En Europe, ce sont surtout les meurtres de femmes commis par un conjoint ou un ex-conjoint qui ont accaparé le paysage médiatique. Parmi ceux-ci, l'assassinat d'Alexia Daval, tuée et brûlée par son mari, a profondément marqué la société française et a contribué à lancer les débats sur la pertinence du mot *féminicide* dans les journaux.

En France, il faut attendre la fin 2017 pour que le mot [féminicide] surgisse dans le débat public. Tout bascule avec l'affaire Alexia Daval, cette femme de 29 ans retrouvée carbonisée dans un bois de Haute-Saône en octobre 2017. Son mari, Jonathann Daval, participera aux marches en sa mémoire, avant d'avouer son meurtre en janvier 2018. (Perrin, 2019)

Ces archétypes, bien qu'ils aient fortement influencé la conceptualisation du féminicide, ne couvrent pas l'entièreté des usages de la presse écrite. Lorsqu'on observe les occurrences de *féminicide* dans les journaux québécois et de l'Europe francophone, on retrouve d'abord les quatre sens définis par l'OMS. En effet, le féminicide lié à la dot et le féminicide au nom de l'honneur, deux types de féminicides moins ancrés dans la réalité occidentale, sont rapportés dans les journaux québécois et franco-européens. C'est le cas dans ces articles parus dans *Le Monde* et dans *L'Express*.

L'OMS a toutefois évalué qu'au moins 5 000 crimes d'honneur étaient perpétrés chaque année, surtout au Moyen-Orient et en Asie du Sud, et que le féminicide lié à la dot concerne «entre 7600 et 25 000 jeunes mariées» (Durand, 2018).

Ce genre de féminicide n'est pas un cas isolé. En 2019, 30 % des meurtres en Iran relevaient de la même catégorie : celle des «crimes d'honneur», c'est-à-dire l'assassinat d'une femme par son tuteur masculin, sous prétexte de préserver la réputation de la famille. (Müller, 2020)

Pour sa part, le féminicide intime est le seul sens contesté mais également le plus attesté dans les journaux, et ce, autant au Québec qu'en Europe.

Atroce féminicide à Saint-Étienne, dans la Loire. Un trentenaire a tué sa conjointe devant leur garçonnet de deux ans. (Robert, 2021)

Le meurtre d'une femme de 32 ans, qui aurait été poignardée sauvagement à mort par son mari hier dans un quartier familial de Montréal, a créé une véritable commotion, étant le deuxième féminicide à survenir depuis le début de l'année au Québec. (Deland et Aubin, 2024)

Le mot *féminicide* est employé si fréquemment pour désigner le meurtre d'une femme par son conjoint ou son ex-conjoint que le terme semble restreint à cet unique sens dans certains articles européens.

En 2020, les féminicides (ou crimes conjugaux) étaient beaucoup moins nombreux que l'année précédente a annoncé Éric Dupond-Moretti sur les réseaux sociaux... Que cachent ces chiffres historiques ? (Gauthier, 2021)

Le féminicide est un fléau qui prend de plus en plus d'ampleur en Europe, la France comme la Grande-Bretagne voient des femmes mourir chaque jour sous les coups de leur compagnon ou d'un membre de leur famille. (Laurence O, 2021)

Le suicide forcé, par exemple une femme qui met fin à ses jours pour échapper à la violence que lui inflige son conjoint, est également catégorisé comme un féminicide dans plusieurs articles de journaux, même si ces cas ne sont pas comptabilisés dans les statistiques officielles comme un féminicide.

C'est le point invisibilisé des violences faites aux femmes : le « suicide forcé », celui qu'une femme sous emprise finit par commettre pour être enfin délivrée de son bourreau, la seule porte de sortie qu'elle soit capable d'entrevoir à ce moment précis. (Schwartzbrod, 2022)

[...] elles ont élargi leur mode de calcul en 2022 en adoptant « la nouvelle définition d'Inter Orga Féminicides : tout meurtre ou suicide forcé d'une femme en raison de son genre, quels que soient son âge ou les circonstances. Les féminicides s'inscrivent dans un contexte de violence patriarcale et systémique et/ou au croisement d'autres systèmes d'oppression ». (Babey, 2023)

Pour ce qui est du féminicide non intime, il est présent dans les journaux mais dans une moindre mesure. Ce sont surtout les événements ayant causé de multiples victimes qui sont décrits. C'est le cas de la tuerie de Polytechnique mais aussi des assassinats incessants de la ville mexicaine de Ciudad Juárez, dite « capitale du meurtre de femmes », où plusieurs milliers de meurtres et disparitions de femmes ont été dénombrés (Bodiou et Chauvaud, 2019).

Disparues. Dépecées. Autopsiées. Car, à Ciudad Juarez, depuis plus de 20 ans, les femmes meurent. À un rythme si élevé et dans de telles proportions que l'on y parle de « féminicide ». (Vallet, 2017)

L'appellation *féminicide de masse* est parfois employée pour ce type de meurtre quand il fait plusieurs victimes en même temps.

Le 6 décembre 1989, quatorze jeunes femmes sont assassinées à l'École polytechnique de Montréal. Leur tueur a laissé un message expliquant son geste par la haine des féministes. Il faudra attendre de nombreuses années pour que le crime soit nommé comme tel : le premier féminicide de masse revendiqué. (Jouan, 2019)

En dehors de tels événements, le meurtre d'une femme unique, qui ne fait pas partie des trois autres catégories ci-dessus, est rarement nommé comme un féminicide dans les journaux.

Par ailleurs, le terme *féminicide* est aussi parfois employé pour parler de l'avortement d'un fœtus féminin.

Pour avoir voté en faveur de cette motion, on a remis en doute les convictions féministes de Rona Ambrose. Elle l'a fait parce qu'elle est préoccupée par le féminicide, cette pratique répandue dans certaines communautés culturelles d'avorter les fœtus féminins. (Bellemare, 2012)

L'ONU classe d'ailleurs les foeticides et les infanticides sur des personnes de sexe féminin comme des féminicides (*Global Study on Homicide*, 2019). Dans certaines régions du monde, apprendre que le sexe du futur enfant est féminin mène les parents à procéder à un avortement, la femme étant considérée économiquement comme un fardeau pour la famille.

Dans les journaux, *féminicide* est également employé pour référer au phénomène social. Cette acception montre le caractère collectif et systématique des violences envers les femmes qui persiste parce que les organisations sociétales le permettent.

[...] l'anthropologue mexicaine Marcela Lagarde [...] a défini le féminicide comme étant « l'une des formes extrêmes de la violence de genre », un « génocide contre les femmes » et un « crime d'État ». (l'historienne Christelle Taraud citée dans Lemonnier, 2022)

Le dernier sens de *féminicide* mentionné dans les journaux est celui de « meurtre de femme en raison de son sexe ». Cet emploi correspond en fait à la définition qu'on retrouve dans tous les dictionnaires actuellement et est justement utilisé quand on cite un dictionnaire.

Commençons par définir ce qu'est le féminicide. Selon le dictionnaire français Larousse, le féminicide est un meurtre d'une femme ou d'une jeune fille, en raison de son appartenance au sexe féminin. (Kisombe, 2021)

Le mot «féminicide» a mis longtemps à s'imposer. D'abord concurrencé par «fémicide», utilisé par l'OMS, il est adopté par l'ONU en 2012, avant de faire son entrée dans *Le Petit Robert* en 2015. Si le mot est récent, «le meurtre de femmes parce que femmes», lui, ne l'est pas. » (Ripa, 2022)

Sans surprise, cela laisse supposer qu'une partie des journalistes s'appuient sur les ouvrages de référence pour définir le féminicide et que ceux-ci exercent une influence non négligeable sur son utilisation. C'est pourquoi, dans cet article, nous recensons les emplois de l'usage afin de proposer un article lexicographique fidèle à celui-ci.

4. Quand le sémantisme de *féminicide* suscite la controverse

Alors que *féminicide* apparaît abondamment dans les médias pour nommer les drames du quotidien, le mot provoque simultanément un débat dans les journaux quant à sa désignation. Des gens de différents milieux, entre autres des linguistes, des avocats et des procureurs, des associations et des individus spécialisés en violence de genre, débattent quant à savoir si le meurtre d'une femme par son partenaire ou un de ses anciens partenaires devrait être considéré comme un féminicide.

Certains défendent que le féminicide consiste exclusivement en un meurtre ayant eu lieu en raison du sexe féminin de la victime. Dans le contexte où le meurtre est commis par le conjoint ou l'ex-conjoint, est-ce que le meurtrier tue la femme parce qu'elle était une femme ou parce qu'elle était *sa* femme, *sa* partenaire ? Cette relation de proximité et d'intimité entretenue entre les deux individus laisse penser que d'autres motifs que le sexe féminin de la victime, tels que la jalousie, le sentiment d'abandon ou la vengeance envers sa partenaire ou ex-partenaire (*Global Study on Homicide*, 2019 : 37), mèneraient à l'acte violent et que, pour cette raison, il ne faudrait pas considérer ce type de meurtre comme un féminicide. C'est ce que nomme d'ailleurs l'essayiste et philosophe française Bérénice Levet dans un article du *Figaro*.

L'homme qui tue sa compagne ou son ex-compagne ne tue pas une femme, il tue sa femme, la femme avec laquelle il vit ou avec laquelle il a vécu, avec laquelle il a peut-être eu des enfants. Féminicide il y aurait si quelque homme ou quelques hommes réunis s'emparaient d'un groupe de jeunes filles ou de femmes et les vouaient à la mort, les exterminaient pour la seule raison d'être nées femmes. Ce serait la seule acception rigoureuse. (Levet, 2021)

Des motifs autres peuvent alors intervenir et pousser l'homme à commettre l'acte (*Global Study, On Homicide*, 2019 : 37). De plus, cette situation pourrait aussi bien survenir dans un couple formé par deux femmes que par deux hommes ce qui exclurait que le meurtre soit fondé sur le sexe, comme le mentionne la procureure Dominique Laurens : « On ne tue pas cette femme parce que c'est une femme, mais parce qu'elle est la compagne. Les ressorts sont les mêmes dans tous les meurtres de

conjoints qu'ils soient hommes ou femmes » (citée dans AFP, 2019). Ainsi, selon cette conception, seuls les meurtres commis par un homme qui n'a pas été un conjoint de la victime consisteraient en des féminicides, ce qui explique pourquoi plusieurs préfèrent employer un autre mot pour décrire ces tragédies. C'est ce qu'en conclut le spécialiste de psychopathologie et de psychiatrie légale Daniel Zagury : « Je ne sais pas s'il tue sa femme parce que c'est une femme. C'est un meurtre, un assassinat, un homicide » (cité dans Van Kasteel, 2018).

À cette opinion se confrontent ceux qui croient que *féminicide* devrait désigner tout meurtre de femmes, perpétré par des hommes, témoignant d'un système de violence dirigé envers celles-ci. Effectivement, parmi les meurtres au sein de couples, les femmes sont statistiquement les principales touchées. Selon les données de l'OMS, les hommes sont ceux qui commettent le plus ce crime, toutes cultures confondues (Deilbreil, 2019). C'est pourquoi, selon cette perspective, le terme *féminicide* devrait mettre de l'avant l'existence d'une violence spécifiquement dirigée vers les femmes, et ce, peu importe que cette violence soit causée par un conjoint ou par un inconnu.

Un féminicide, c'est lorsqu'une femme est assassinée parce qu'elle est une femme. On ne parle pas d'une personne qui est au mauvais endroit au mauvais moment. Quand un homme tue sa partenaire, par exemple, c'est un féminicide. Quand un homme fonce dans une foule pour tuer des femmes, c'est un féminicide. (Bergeron, M., 2021)

Le féminicide pourrait aussi se traduire comme un drame qui n'aurait pas eu lieu si la victime avait été du sexe opposé. C'est la vision que rapporte le journaliste Steve Bergeron de *La Tribune*.

Vous me direz alors que toute femme assassinée par un homme dans un contexte conjugal ou familial est automatiquement victime d'un féminicide ? C'est le constat auquel en arrivent de plus en plus de gens. Leur raisonnement est de se dire : « Si ces femmes avaient été des hommes, elles seraient probablement encore vivantes aujourd'hui. » C'est-à-dire qu'elles n'auraient pas été perçues comme les possessions de leur conjoint, comme des êtres humains que l'on peut contrôler, violenter, et qui ne méritent pas, dans l'esprit de leur meurtrier, de vivre en dehors de la relation. (Bergeron, S., 2021)

En résumé, cette conception plus large du féminicide sous-tend l'objectif d'une prise de conscience que les meurtres de femmes, dont ceux commis par un conjoint ou un ex-conjoint, ne sont pas des cas isolés mais qu'ils s'inscrivent dans la problématique plus grande des violences dirigées envers le sexe féminin. C'est d'ailleurs l'opinion de la professeure de psychologie sociale Patrizia Romito : « Ces violences, multiples et répétées [...] sont structurelles; partout, elles sont le produit d'un système patriarcal et représentent un des moyens qui lui permettent de perdurer et de continuer à opprimer les femmes » (2022 : 477-478).

Céline Caudron, coordonnatrice de l'association *Vie féminine*, croit aussi qu'il est nécessaire de nommer la réalité telle qu'elle est si l'on désire du changement : « ce n'est plus un fait divers, mais un phénomène de société, et on peut mettre un mot sur ce phénomène : féminicide. Et quand on nomme les choses, elles existent, et on peut en parler, pour améliorer la situation. » (Van Kasteel,

2018). La linguiste Julie Abbou abonde dans le même sens et mentionne que « [q]uand on peut parler de féminicide, on fait apparaître une réalité massive de meurtres, qui n'étaient jusque-là que d'anecdotiques crimes passionnels » (cité dans Huon, 2022).

Si le mot *féminicide* octroie une visibilité à la violence envers les femmes, à l'inverse, les expressions *crime passionnel* et *drame conjugal*, qui ont été très populaires à une époque, minimisent cette violence. Sur cette idée, Marlène Schiappa, secrétaire d'État chargée de l'Égalité en France entre les femmes et les hommes de 2017 à 2020, énonce que ces formulations ont comme conséquence « de faire du meurtrier non plus un criminel mais un passionné qui tuerait dans un élan non contrôlé » (citée dans D'Haenens, 2018). Du même avis, la professeure de droit Stéphanie Wattier mentionne que « parler de crime passionnel invisibilise le problème et "justifie" en quelque sorte le crime qui a eu lieu » (citée dans Royen, 2021).

5. Féminicide dans les dictionnaires

Nous avons choisi d'analyser les articles de six références lexicographiques. Parmi celles-ci, nous retrouvons deux dictionnaires conçus en France, soit le *Larousse* (version en ligne) et *Le Petit Robert* (PR2024), un dictionnaire québécois, *Usito*, et un dictionnaire collaboratif, le *Wiktionnaire*. Les définitions de deux organismes linguistiques ont aussi été prises en compte, soit le *Grand dictionnaire terminologique* (GDT) et *FranceTerme* qui diffusent respectivement la norme en vigueur au Québec et en France. Nous avons choisi ces sources lexicographiques, car nous les considérons comme les sources les plus sérieuses et les plus à jour pour nous informer du traitement lexicographique du cas de *féminicide*, mais aussi d'autres mots polémiques traités dans le cadre du Projet porc-épic.

Le *Wiktionnaire* atteste *féminicide* dès 2012. En juin 2014, *féminicide* entre au *Petit Robert* (PR2015) ; il devient un terme officialisé en France la même année et est attesté dans *FranceTerme* en septembre. Quelques années plus tard, d'autres dictionnaires suivent le mouvement : en 2020, *féminicide* entre dans *Usito* et dans *Le Petit Larousse illustré* (PLI2021), puis dans le *Grand dictionnaire terminologique* en 2021. Durant cette période, la fréquence du terme augmente de manière significative dans les médias au point où *Le Petit Robert* en fait son mot de l'année en 2019.

Lorsqu'on s'attarde à ce qu'en disent ces différentes sources, c'est du pareil au même. Les dictionnaires, tout comme le GDT et *FranceTerme*, proposent sensiblement la même définition, soit que le féminicide est le meurtre d'une femme parce qu'elle est une femme. Malgré la polysémie présente dans l'usage, dont un sens controversé, les différents ouvrages ne proposent pas davantage d'informations. Quelques nuances départagent les différents articles lexicographiques, mais ce sont des variations mineures qui mènent toutes à la même idée.

Dans cette section, nous examinerons pour chacun des articles divers éléments : la classe de mot (adjectif, nom), l'absence ou la présence de marque d'usage ainsi que les définitions en fonction des différents sens attestés dans les milieux experts et dans l'usage. Attardons-nous d'abord au *Petit Robert 2024* qui décrit *féminicide* à la fois comme adjectif, comme nom épïcène (*un* ou *une féminicide*) pour désigner le meurtrier ou la meurtrière, et comme nom masculin pour le meurtre lui-même.

féminicide [feminisid] **adjectif et nom**

ÉTYM. 1855 ◊ du radical du latin *femina* « femme » et *-cide*
Famille étymologique ⇨ FEMME.

FORMES
singulier : féminicide
pluriel : féminicides
singulier : fémicide
pluriel : fémicides

PRONONCIATION
◀ féminicide [feminisid]

■ DIDACT.
1 RARE Qui tue une femme.
= N. Un, une féminicide.
2 N. m. (1986) Meurtre d'une femme, d'une fille en raison de son sexe.
« permettre le prononcé de peines plus lourdes en introduisant la notion de « féminicide » dans le Code pénal » (Houellebecq). Reconnaissance juridique du féminicide par plusieurs pays d'Amérique latine.
= On rencontre aussi l'anglicisme *fémicide*.

Figure 1 : Article *féminicide* du *Petit Robert 2024*

Il faut d'abord mentionner la marque « didactique » qui précède les définitions. Selon le dictionnaire, cette marque désigne un « mot ou emploi propre à la langue savante et qui n'est pas employé dans le français standard » (*Le Petit Robert*). Si cette indication était appropriée il y a une dizaine d'années, elle ne l'est assurément plus aujourd'hui : *féminicide* est désormais bel et bien répandu dans l'usage. Quant aux définitions, *Le Petit Robert* atteste deux sens. Le premier décrit *féminicide* lorsqu'il désigne la personne qui a commis l'acte. Cette acception est peu fréquente dans l'usage, ce qu'indique bien la marque « rare ». Le second sens correspond au sens générique décrit dans l'ensemble des dictionnaires : « Meurtre d'une femme, d'une fille en raison de son sexe. ». On mentionne également dans un exemple que le terme est reconnu au sens de la loi en Amérique latine et que la variante *fémicide* est considérée comme un anglicisme. Il s'avère nécessaire de souligner l'apparence engagée de cet article. Déjà que *Le Petit Robert* a été, en 2014, dans les premiers dictionnaires à intégrer le mot à sa nomenclature, la citation, qui plaide pour une sanction plus lourde, et l'exemple construit qui rappelle que le *féminicide* figure dans les textes de loi de l'Amérique latine, laissent présumer une prise de position de la part du *Petit Robert*, soit que le *féminicide* devrait être intégré au Code pénal français. Or, il ne s'agit pas du rôle d'un dictionnaire qui se veut plutôt un reflet objectif de l'usage : « le dictionnaire général n'a pas comme fonction d'orienter l'usage, mais de le décrire. Il est le témoin de la langue de son époque » (Vincent, 2022 : 124).

Du côté du *Larousse*, *fémicide* n'est décrit qu'en tant que nom et une seule définition, « Meurtre d'une femme ou d'une jeune fille, en raison de son appartenance au sexe féminin », est indiquée. Ce qui est intéressant est que le terme est qualifié de « crime sexiste », mettant ainsi en évidence la dimension discriminatoire du fémicide envers les femmes. On mentionne également que ce crime ne fait pas partie du Code pénal français.

fémicide

nom féminin

(du latin *femina*, femme, et *cædere*, tuer)

Meurtre d'une femme ou d'une jeune fille, en raison de son appartenance au sexe féminin. Crime sexiste, le fémicide n'est pas reconnu en tant que tel par le Code pénal français.

Figure 2 : Article *fémicide* dans le *Larousse* (version en ligne)

FranceTerme, qui constitue une banque de données où sont diffusées les recommandations publiées au *Journal officiel de la République française*, indique une définition équivalente à celle des autres sources : « Homicide d'une femme, d'une jeune fille ou d'une enfant en raison de son sexe ».

fémicide n.m.

Journal officiel du 16/09/2014

Domaine

DROIT - SCIENCES HUMAINES

Définition

Homicide d'une femme, d'une jeune fille ou d'une enfant en raison de son sexe.

Note

Le fémicide peut désigner un meurtre à caractère individuel ou systématique.

Voir aussi

[gynophobie](#)

[homicide sexiste](#)

Équivalent étranger

femicide (en) , femicidio (es) , fémicide (en) , femicidio (es) , femminicidio (it)

Figure 3 : Article *fémicide* dans *FranceTerme*

Cette définition se démarque par l'ajout que ce type de meurtre touche les femmes et les jeunes filles, mais également les enfants de sexe féminin. L'article de *FranceTerme* est particulièrement intéressant pour les deux renvois proposés, soit *gynophobie* et *homicide sexiste*. Ceux-ci mettent bien de l'avant la dimension sexiste du crime, c'est-à-dire discriminante à l'égard des femmes. Par ailleurs, des équivalents dans d'autres langues (anglais, espagnol et italien) sont indiqués. Finalement, *FranceTerme* spécifie que ce type de meurtre peut correspondre à un acte individuel mais aussi à un meurtre systématique, ancré dans un mouvement de violence envers la femme.

Tout comme *FranceTerme*, le GDT présente cette distinction entre les féminicides individuels et systématiques.

féminicide

— anglais : femicide

Domaines :
criminologie
sociologie › genre

Auteur :
Office québécois de la langue française

Dernière mise à jour : 2021

Définition

Fait pour une personne de causer la mort d'une femme ou d'une fille en raison de sa condition féminine.

Note

Un féminicide peut constituer un meurtre individuel ou systématique.

Terme privilégié

féminicide n. m.

En France, le terme *féminicide* est recommandé officiellement par la Commission d'enrichissement de la langue française, depuis 2014.

Le terme *féminicide* est composé de *fémin-* (du latin *femina*, « femme ») et du suffixe *-icide*, qui signifie « suppression, destruction, meurtre ». Il est formé sur le même modèle que les termes *parricide* et *régicide*, par exemple.

Le terme *féminicide* peut aussi désigner la personne qui cause la mort.

Terme déconseillé

fémicide

L'emprunt à l'anglais *fémicide* est déconseillé parce qu'il n'est pas légitimé en français au Québec et qu'il est mal formé. En effet, *fém-* n'est pas un élément formant usité en français.

Figure 4 : Article *féminicide* dans le *Grand dictionnaire terminologique*

Un élément distinctif de la définition du GDT est le segment « en raison de sa condition féminine » qui, davantage que « en raison de son sexe », rend visible tout un système de discrimination envers la femme. La description du GDT est aussi intéressante par rapport à la section sur le terme privilégié. On y indique des informations étymologiques, que le terme est officialisé en France et qu'il peut désigner la personne coupable du meurtre. La variante *fémicide* figure comme terme déconseillé.

Pour ce qui est du dictionnaire *Usito*, quelques éléments méritent d'être soulignés. D'abord, comme dans le *Larousse*, *féminicide* n'y est traité que comme un nom. Ensuite, l'article n'atteste pas la variante *fémicide*. Enfin, la définition comporte une partie distincte des autres dictionnaires. En effet, elle mentionne que la femme ou la fille est tuée *en particulier* en raison de son sexe. La présence de *en particulier* sous-entend que le meurtre est fondé sur le sexe féminin de la victime, mais pas exclu-

sivement. Cela rejoint le débat sur les meurtres de femmes commis par un conjoint ou un ex-conjoint où certains distinguent ce type de meurtre du féminicide sous prétexte que les motifs du crime ne seraient pas seulement liés au sexe de la victime. Ici, par la présence de *en particulier*, la définition d'*Usito* englobe le meurtre commis par un conjoint ou un ex-conjoint.

fémicide [feminisid] n. m.

Action de tuer une femme, une fille, en particulier en raison de son sexe.

↑ [homicide](#).

Commettre un fémicide.

Être coupable de fémicide.

Des victimes de fémicides.

« parler de fémicides et, plus largement, de prévention des violences faites aux femmes » (La Presse, 2020).

REM. L'emploi de *fémicide* a fait l'objet d'une recommandation officielle en France.

ÉTYMOLOGIE

1986 (*in* NPR); du latin *femina* « femme » et *-(i)cide*.

Figure 5 : Article *fémicide* dans *Usito*

Finalement, le *Wiktionnaire* propose trois sens pour l'adjectif et trois sens pour le nom *fémicide*.

⇒ **Adjectif** [[modifier le wikicode](#)]

fémicide \fe.mi.ni.sid\ masculin et féminin identiques

- Qui a **tué** une ou plusieurs **filles** ou **femmes**.
 - Au contraire le geste homicide, ou plutôt fémicide, doit être accompli sans colère et sans haine.* — (Pierre Brunel, Charles Baudelaire, *Les fleurs du mal* – “Les fleurs du mal” entre “fleurir” et “défleurer”, 1998)
 - C'est le 45e fémicide perpétré depuis janvier 2020, selon un décompte d'organisations féministes.* — (Agence QMI, Algérie: une femme égorgée de sang-froid, Le Journal de Montréal, 17 novembre 2020)
 - C'est le second fémicide depuis janvier lorsqu'une journaliste de la télévision publique avait été tuée par son mari.* — (Agence QMI, Algérie : un père décapite sa fille et jette son corps dans la forêt, journaldemontreal.com, 19 février 2021)
- Qui **blesse** les femmes.
 - Je me demande si, en présence de tels précédents humains, nous sommes bien venus, nous autres hommes, créatures censées raisonnables, à faire le procès à une bête [...] qui n'a été entraînée à mal faire que par la dangereuse contagion de la monomanie fémicide.* — (Alphonse Toussenel, *Le monde des oiseaux : ornithologie passionnelle*, volume 2, page 9, 1855)
 - Depuis quelques années, on ne se serre plus. Le relâchement du lacet fémicide a été, comme jadis lors des paniers, la conséquence du développement extravagant des hanches par la crinoline. Mais il paraît que ces pauvres dames n'ont fait que changer de torture...* — (Jules Lecomte dans *Le monde illustré*, volume 1-7, page 99, 1863)
 - Quoi qu'il en soit, le temps des tailles de guêpes est revenu, et nous verrons la belle génération qui sortira de ces corsets fémicides. On ne comprime pas impunément l'appareil organique indispensable à la vie ;* — (*Les femmes du monde*, Bachaumont (pseud.), 1876)
- Qui **détruit** la **féminité**.
 - Plutôt que misogynie (une femme honnête y est victime d'un homme sans scrupules), l'intrigue est « fémicide », elle travaille à la destruction de la féminité, atteint les femmes dans leurs traits spécifiques, voire leur dénie le droit.* — (*Bulletin de la Société J.-K. Huysmans*, n° 97, page 42, 2004, Durtal)

	Singulier	Pluriel
Masculin	fémicide	fémicides
et féminin	\fe.mi.ni.sid\	

Notes [[modifier le wikicode](#)]

Le terme « fémicide » spécifie la victime, mais pas l'agresseur qui peut dans sa stricte définition être une femme ^[2].

Figure 6 : Article *fémicide* (adjectif) dans le *Wiktionnaire*

Malgré l'apparente richesse de l'article du *Wiktionnaire*, beaucoup d'éléments ne répondent pas aux critères lexicographiques. *Féminicide* en tant qu'adjectif est rare dans l'usage. Le *Wiktionnaire* indique tout de même trois emplois différents pour l'adjectif, dont « Qui a tué une ou plusieurs filles ou femmes ». Il faut toutefois noter que deux des trois citations du premier sens ne sont pas appropriées. En effet, dans « le 45^e féminicide » et « le second féminicide », le mot *féminicide* n'est pas employé comme adjectif mais plutôt comme nom. Pour ce qui est du deuxième et du troisième sens, « Qui blesse les femmes » et « Qui détruit la féminité », ils n'ont pas été relevés dans les articles analysés dans *Eureka*.

➔ Nom commun [modifier le wikicode]

féminicide \fe.mi.ni.sid\ masculin

- | Singulier | Pluriel |
|----------------|-------------|
| féminicide | féminicides |
| \fe.mi.ni.sid\ | |
- (Droit) Meurtre d'une femme, d'une jeune fille ou d'un enfant en raison de son sexe.
 - Dans ce contexte est survenu le « **féminicide** » de Ciudad Juárez : la mort par viol, mutilations, étranglement, supplices, découpes et incinération... — (Michel Wieviorka, Jean-Paul Brodeur, *L'empire américain* ?, 2004)
 - Dans le cadre du couple, le **féminicide** est qualifié de « conjugal » ou « intime » et désigne le meurtre de femmes par leur conjoint ou ex-conjoint. Si les **féminicides** recouvrent une réalité plus variée, les **féminicides conjugaux** sont ceux qui retiennent le plus l'attention en France. — (Buisson, Charlotte, et Jeanne Wetzels. « Chapitre II. Des violences omniprésentes et continues », Charlotte Buisson éd., *Les Violences sexistes et sexuelles*. Presses Universitaires de France, 2022, pp. 28-48.)
 - Ajouter un exemple
 - (Néologisme) Violence faite à une femme fondée sur sa condition féminine.
 - Le terme "**féminicide**" se base sur la définition juridique de la violence faite à la femme, précisée à l'article 1^{er} de la Convention de Belém do Pará : "on entend par violence contre la femme tout acte ou comportement fondé sur la condition féminine qui cause la mort, des torts ou des souffrances physiques, sexuelles ou psychiques à la femme, aussi bien dans sa vie publique que dans sa vie privée". — (Parlement européen, *Rapport sur les meurtres de femmes (féminicides) en Amérique centrale et au Mexique et le rôle de l'Union européenne dans la lutte contre ce phénomène*, 20 septembre 2007)
 - Meurtrier d'une femme.
 - On se demande ce que peut bien représenter ce monstre, ce **féminicide** qu'est Barbe-Bleue ? Il tue sa huitième femme, Eveline, après avoir « étranglé » les sept premières. — (Luc Badesco, *La Génération poétique de 1860*, volume 1, Nizet, 1971, page 481)



➔ Dérivés [modifier le wikicode]

- féminicidaire
 - continuum féminicidaire
- féminicide de masse

Figure 7 : Article *féminicide* (nom) dans le *Wiktionnaire*

Pour ce qui est du nom, le *Wiktionnaire* propose également trois sens dont « Meurtre d'une femme, d'une jeune fille ou d'un enfant en raison de son sexe » et « Meurtrier d'une femme » qui ont été relevés dans les dictionnaires abordés plus haut. Quant au sens « Violence faite à une femme fondée sur la condition féminine », il nous apparaît inusité, n'ayant pas été repéré dans l'usage journalistique également. La marque « droit » qui précède le premier sens nous semble aussi incongrue étant donné que *féminicide* ne figure toujours pas dans le Code pénal français ni dans celui de la majorité des États francophones.

À la suite à ce tour d'horizon, il apparaît qu'un écart non négligeable existe entre la façon dont *fémicide* est employé dans l'usage et la description faite par les dictionnaires et les organismes officiels. En effet, ceux-ci s'en tiennent à une définition générique, « Meurtre d'une femme en raison de son sexe », alors que l'usage est plus diversifié, se rapprochant des catégories définies par l'OMS. L'usage journalistique témoigne ainsi d'une réalité bien plus complexe où les violences sont ancrées dans un système de domination masculine et dans des milieux culturels où les femmes ne sont pas reconnues et ont parfois peu ou pas de droits.

Afin de mieux refléter l'usage actuel, mais aussi de mettre en lumière les enjeux sociétaux qu'implique le fémicide, nous croyons que les ouvrages lexicographiques devraient davantage nuancer leur description, notamment par rapport au non-consensus sur le meurtre d'une femme par un conjoint ou ex-conjoint. De plus, plusieurs sens spécifiques ne sont pas du tout mentionnés, autant dans la définition que dans les citations ou les exemples. Devant le constat que l'usage est beaucoup plus vaste que ce que proposent les dictionnaires et qu'un sens fait polémique, l'équipe du Projet porc-épic propose son article lexicographique qui se veut en accord avec les emplois présents dans l'actualité.

6. La proposition du Projet porc-épic

Le Projet porc-épic, créé en 2021, est dirigé par la professeure Nadine Vincent (Université de Sherbrooke) et est rattaché au laboratoire LexiQcorpus du Centre de recherche interuniversitaire sur le français en usage au Québec (CRIFUQ). Ce projet de recherche a comme mandat d'étudier les emplois polémiques afin d'informer les usagers sur leurs contextes d'utilisation, leurs connotations et leurs dénotations.

Selon la tradition lexicographique, les dictionnaires de langue ont pour fonction de décrire les usages les plus populaires, et ce, de manière tout à fait objective (Vincent, 2022 : 124). Pour notre part, nous avons une approche distincte. Notre objectif est plutôt de proposer un article lexicographique qui prenne en compte l'ensemble des sens répertoriés du mot en question. Nous visons à ce que notre description soit la plus complète possible à l'aide de nombreuses mises en contexte, d'exemples construits et de citations. Un autre point qui distingue nos descriptions lexicographiques est que le Projet porc-épic n'a aucune restriction en termes d'espace. En tirant profit de l'espace numérique, il est possible de fournir des articles lexicographiques complets, même si les termes polémiques peuvent nécessiter davantage d'indications. Par exemple, nous avons la possibilité de développer les remarques, sans que la longueur de celles-ci soit un enjeu. Des liens hypertextes peuvent également être intégrés à nos entrées afin de favoriser l'accès à des sources fondamentales et à des informations complémentaires sur le sujet.

Nous avons fait des choix qui diffèrent des articles des dictionnaires examinés dans les pages précédentes. D'abord, nous offrons une description polyphonique qui comprend tous les sens répertoriés dans l'usage afin de désambigüiser au maximum l'utilisation du terme. C'est pourquoi nous considérons le meurtre d'une femme commis par un conjoint ou un ex-conjoint comme un féminicide dans notre article, tout en marquant une distinction avec les autres types de féminicides. Cela s'explique par la relation particulière qu'entretient l'homme avec la victime, qui a nécessairement un effet sur la motivation derrière le meurtre. Selon le rapport de l'ONUDC, les motifs qui peuvent pousser les hommes à tuer leur partenaire sont la possessivité, la jalousie, la peur de l'abandon ou encore des problèmes mentaux sévères (*Global Study on Homicide*, 2019 : 37). À l'exception de l'instabilité mentale, les autres motifs ne s'appliquent pas au féminicide non intime.

Ensuite, en plus de la victime, nous avons pris la décision d'identifier le meurtrier dans la définition : il s'agit donc du meurtre d'une femme commis par un homme. Ce choix distingue notre article de celui des autres dictionnaires, mais il est fondé sur de nombreuses sources. Effectivement, tous les cas rapportés dans les journaux décrivent le meurtre de femmes par des hommes. C'est aussi ce que les travaux sur la question mentionnent. Selon l'OMS, les meurtres au sein des couples, toutes cultures confondues, sont presque toujours commis par les hommes (Deilbreil, 2019) et la situation contraire (des femmes qui tuent leur partenaire masculin) est très rare et découle souvent de la légitime défense de la femme (*Global Study on Homicide*, 2019 : 37). Nous n'avons répertorié aucun contexte dans l'usage où le mot *féminicide* désignerait le meurtre d'une femme par une autre femme et les données sont déficientes à ce propos : la plupart des études ne précisent pas si les meurtres de femmes commis par des femmes sont inclus et ces chiffres ne sont pas recensés, ou du moins connus, pour l'instant (OCFJR, 2022 : 84). Il est reconnu toutefois que des femmes, souvent des membres de la famille de la victime et des femmes sous l'emprise de l'autorité d'un homme, sont parfois impliquées dans certains types de féminicides (OMS, 2012 : 1), mais les données suggèrent que le phénomène correspond généralement à des violences envers des femmes, réalisées par des hommes.

Finalement, nous définissons le féminicide en cinq acceptions. L'ordre d'apparition de ces sens a été choisi en fonction de leur fréquence dans l'usage journalistique au Québec et en Europe francophone. Nous n'avons pas retenu *féminicide* en tant qu'adjectif (ex. *un mari féminicide*) ni le nom sous le sens d'une personne qui a assassiné ou tenté d'assassiner une ou plusieurs femmes (ex : *un féminicide récidiviste*), car ces deux emplois sont pratiquement absents de notre corpus de données et ne sont pas polémiques : ils n'ont pas suscité de questionnements quant à leur sémantisme.

Nous rappelons que notre objectif n'est pas de reproduire le travail des autres dictionnaires ni de compétitionner avec eux. L'objectif du Projet porc-épic est de pointer les tensions et les ambiguïtés suscitées par un mot pour ensuite apporter des précisions et les nuances nécessaires afin de décloisonner la langue.

Voici notre proposition d'article lexicographique, reflétant à la fois les tendances de l'usage et les sources scientifiques sur le sujet.

FÉMINICIDE, NOM MASCULIN

1. Meurtre d'une femme par son conjoint ou son ex-conjoint, motivé par le contrôle (par exemple par jalousie ou à la suite de l'annonce d'une rupture).

SYNONYMES : *féminicide intime* (appellation de l'OMS), *féminicide conjugal*. Les termes *drame conjugal* et *crime passionnel* sont contestés parce qu'ils minimisent la gravité du crime et la responsabilité du meurtrier.

« *Féminicide : elle aurait été tuée puis brûlée par son ex-conjoint* » (Le Journal de Québec, 2022)

« *Tout porte à croire qu'il s'agit de l'ex-conjoint de la victime, ce qui fait de l'affaire un possible féminicide.* » (La Presse+, 2022)

« *Un féminicide de plus qui aurait sans doute pu être évité. Car [la victime] aurait déjà déposé plusieurs mains courantes contre son mari.* » (L'Humanité, 2020)

REMARQUES : 1) La pertinence du mot *féminicide* est parfois contestée en ce sens, notamment parce que l'homicide volontaire constitue déjà le crime le plus lourdement sanctionné et que la reconnaissance juridique du féminicide n'entraînerait pas une peine supplémentaire. De plus, la définition de ce mot n'étant pas stable, les critères pour établir qu'il s'agit bien d'un féminicide seraient difficiles à démontrer. 2) Dans les contextes consultés, on ne trouve qu'un seul cas d'un homme ayant tué une femme qu'il connaissait, sans avoir été son conjoint. Il la harcelait cependant pour sortir avec elle.

2. Meurtre motivée par la haine des femmes, commis par un homme qui ne connaissait pas la ou les victimes.

SYNONYME : *féminicide non intime* (appellation de l'OMS)

Féminicide de masse.

Le féminicide de Polytechnique en 1989.

« *Plus d'une vingtaine de personnes ont participé, mardi soir, à une marche silencieuse à la chandelle en mémoire de toutes les femmes victimes de féminicide, dont les 14 victimes de Polytechnique et les femmes autochtones disparues et assassinées.* » (La Nouvelle Union, 2018)

3. Meurtre d'une femme ou d'une adolescente, le plus souvent par un membre masculin de la famille, sous prétexte qu'elle aurait déshonoré les siens (par un choix vestimentaire considéré comme inapproprié, une relation sexuelle avant le mariage, un soupçon d'infidélité, la demande d'un divorce, etc.).

SYNONYME : *crime commis au nom de « l'honneur »* (appellation de l'OMS)

« *Ce crime abject, tuer ses propres enfants pour une question d'honneur, ne tient pas la route. C'est plutôt un cas de « féminicide », rien de moins.* » (La Presse, 2012)

« *La Cour suprême brésilienne a établi qu'il était inconstitutionnel d'invoquer « un crime d'honneur » comme circonstance atténuante dans les affaires de féminicides.* » (AFP Infos Françaises, 2021)

« *Le meurtre en mai de [la victime], infanticide et féminicide aux relents de « crime d'honneur », a provoqué un tollé en Iran. Des médias ont dénoncé une « violence institutionnelle » dans une République islamique « patriarcale ».* » (Le Figaro, 2020)

REMARQUE : Le terme *crime d'honneur* est souvent contesté, car il implique que la victime méritait de mourir pour avoir terni la dignité familiale alors que ce type de meurtre constitue une violation de droits reconnus internationalement. On préfère la formulation *crime dit d'honneur* ou *crime commis au nom de l'honneur* pour indiquer que le manquement à « l'honneur » n'est pas une justification valide.

4. Meurtre d'une femme nouvellement mariée, le plus souvent par un membre masculin de la famille, par exemple parce que la dot que l'épouse doit offrir à la famille du marié est insuffisante; PAR EXTENSION suicide d'une femme harcelée par sa belle-famille pour les mêmes raisons.

SYNONYME : *féminicide lié à la dot* (appellation de l'OMS)

« Pour les enquêteurs, un mobile du crime n'a pu être identifié mais la piste du féminicide en Inde n'est pas exclue. Et pour cause, il s'agit d'un véritable phénomène social dans le pays. Ce phénomène meurtrier est particulièrement courant chez les familles les plus pauvres car les filles risquent moins d'être des sources de revenus et que dans cette culture, c'est la famille de la femme qui doit payer une dot en plus du coût du mariage. » (Santé+ Magazine, 2020)

5. Avortement sélectif en raison du sexe féminin du fœtus; meurtre, violence physique directe ou négligence d'un enfant en bas âge menant à sa mort, souvent parce que les femmes rapportent moins à la famille que les hommes.

Elle « est préoccupée par le féminicide, cette pratique répandue dans certaines communautés culturelles d'avorter les fœtus féminins. » (Le Journal de Québec, 2012)

« En Amérique centrale, face au nombre croissant de jeunes femmes assassinées, on invente le mot « féminicide ». Dans le monde arabe, si on n'enterre plus la fille vivante à sa naissance, on a 1001 façons subtiles de s'en débarrasser. » (Le Matin, 2016)

REMARQUE : Le sens de *féminicide* varie selon le contexte culturel et géographique. Certains types de féminicides sont plus fréquents dans une région du monde et sont rares ailleurs. Le féminicide fait d'ailleurs partie du code pénal de nombreux pays d'Amérique latine, et est restreint à la sphère conjugale dans la plupart des États occidentaux. Le féminicide peut correspondre à un acte individuel, mais pour plusieurs, le terme illustre un problème social collectif qui est perpétué à travers le monde et qui découle du maintien des inégalités envers le sexe féminin.

Conclusion

Bien qu'il soit quotidiennement dans les médias, le terme *féminicide* ne fait pas consensus quant à son sémantisme. Le féminicide non intime, comme la tragédie de Polytechnique, n'est pas remis en cause. Le débat disputé dans les journaux de la francophonie porte plutôt sur l'inclusion ou la non-inclusion des meurtres de femmes commis par un conjoint.

Cet article a mis de l'avant que la conceptualisation du féminicide dans les différentes sphères de la société fluctue et que ce flottement sémantique alimente une confusion dans la population. Dans les mêmes journaux qui utilisent *féminicide* pour parler des meurtres par conjoint, plusieurs spécialistes de champs divers se confrontent et remettent en question la légitimité de ce sens. Si l'on se penche sur les définitions légales, la majorité des codes de loi, à l'inverse, le restreignent à la sphère conjugale et excluent les autres contextes. Du côté des dictionnaires, les descriptions très larges, qui n'écartent pas les meurtres commis par un conjoint mais ne les intègrent pas clairement non plus, n'aident pas à y voir clair. Au final, l'hétérogénéité des définitions des chercheurs, des organismes et des dictionnaires rend difficile, non seulement la compilation des données, mais aussi la reconnaissance du féminicide dans la sphère publique.

Le Projet porc-épic a comme mission de dissiper la confusion entourant cet emploi sensible. Il va sans dire que l'exercice comporte son lot de défis. Après avoir examiné les définitions de dictionnaires, d'organismes internationaux, d'articles scientifiques et l'usage actuel dans les journaux, nous avons choisi d'indiquer dans notre article que le féminicide correspond autant aux meurtres de femmes tuées par un partenaire que les autres meurtres de femmes : dans les deux cas, les femmes sont la cible d'une violence spécifiquement dirigée envers le sexe féminin, et ce, à l'échelle planétaire. Dans cette optique où « le féminicide est un combat à la fois intime et international » (Bodiou et Chauvaud, 2019), l'article spécifie que ce crime est commis par un homme et qu'il peut se produire dans différents contextes et milieux culturels, à l'aide de toutes les nuances et remarques jugées nécessaires.

Le cas de *féminicide* témoigne de la complexité de décrire les mots polémiques. Beaucoup d'informations sont nécessaires pour que l'utilisateur connaisse le contexte qui entoure le terme. Cela nécessite de l'espace, ce que les dictionnaires papier traditionnels ne peuvent souvent pas se permettre, mais l'exploitation du format numérique rend possible la description de ces emplois sensibles. Par ailleurs, comme tous les mots, mais particulièrement pour les emplois polémiques de l'actualité, il faudra rester à l'affût de l'évolution sémantique de *féminicide*. Il ne serait d'ailleurs pas étonnant que la question de l'inclusion des victimes de genre féminin (personne qui s'identifie aux caractéristiques féminines de la société) et les femmes transgenres soit prochainement soulevée alors que la société fait face présentement à la reconnaissance de ces identités, notamment avec l'émergence du pronom *iel* et plus largement avec l'écriture inclusive.

Bibliographie

- « 13 juillet 2023. – Loi sur la prévention et la lutte contre les féminicides, les homicides fondés sur le genre et les violences », Strada lex. [en ligne : https://www.stradalex.com/fr/sl_src_publ_leg_be_moniteur/toc/leg_be_moniteur_fr_31082023_1/doc/mb2023044133]
- AFP (2019), « Féminicide : un mot qui fait école, mais aussi débat », *Challenge*, 22 novembre.
- BABEY, Sophie (2023), « Paris : huit rues rebaptisées aux noms de victimes de féminicide tuées dans les Pyrénées-Orientales », *L'Indépendant*, 7 novembre.
- BELLAMI, Victoria (2018), « Intégrer, définir, réprimer et prévenir le «féminicide/féminicide» en Amérique latine », *Autrepart*, vol. 1, n° 85, p. 133-148.
- BELLEMARE, Michel (2012), « Courriers », *Le Journal de Québec*, 1^{er} octobre.
- BERGERON, Mickaël (2021), « Des mortes sur notre conscience », *La Tribune*, 23 mars.
- BERGERON, Steve (2021), « Élargir la notion de féminicide », *La Tribune*, 3 juillet.
- BODIOU, Lydie et Frédéric Chauvaud (2019), « Féminicides, fémicides et violences de genre », dans Lydie Bodiou, Frédéric Chauvaud, Ludovic Gaussot, Laurie Laufer et Beatriz Santos (dir.), *On tue une femme. Le féminicide. Histoire et actualité*, Paris, Hermann, p. 11-64.
- CECCHI, Rossana, Matilde Sassani, Giulia Agugiario, Emanuele Caroppo, Pietro De Lellis, Alessandra Sannella, Marianna Mazza, Tomoya Ikeda, Toshikazu Kondo, et Vittoria Masotti (2022), « A medico-legal definition of femicide », *Legal medicine*, vol. 59
- DAWSON, Myrna et Michelle Carrigan (2021), « Identifying femicide locally and globally : Understanding the utility and accessibility of sex/gender-related movies and indicators », *Current Sociology*, vol. 69, n° 5 , p. 682-704.
- DEILBREIL, Alexia (2019), « Quand l'homicide conjugal ne se résume pas au féminicide », dans Lydie Bodiou, Frédéric Chauvaud, Ludovic Gaussot, Laurie Laufer et Beatriz Santos (dir.), *On tue une femme. Le féminicide. Histoire et actualité*, Paris, Hermann, p. 361-378.
- DELAND, Maxime et Erika Aubin (2024), « Poignardée une vingtaine de fois », *Le Journal de Montréal*, 27 janvier.
- D'HAENENS, Delphine (2018), « Drame conjugal, féminicide, meurtre : les mots pour le dire », *La Voix du Nord*, 2 février.
- DURAND, Anne-Aël (2018), « Qu'est-ce que le "féminicide" », *Le Monde*, 2 février.
- ELKOURI, Rima (2019), « Attentat de Polytechnique : "Il a tué mes filles" », *La Presse*, 24 novembre.
- FREGOSO, Rosa-Linda (2022), « La "guerre la plus longue" : féminicide sexuel systémique en Amérique », dans Christelle Taraud, *Féminicide : une histoire mondiale*, Paris, La Découverte, p. 317-333.

- GAUTHIER, Victor (2021), « Meurtre de Sabina Nessa : un crime qui choque Kate et William », *Le Journal des Femmes*, 29 septembre.
- HUON, Julie (2022), « Ni prescrire ni proscrire », *Le Soir*, 7 octobre.
- JOUAN, Hélène (2019), « Il y a trente ans à Montréal, le premier féminicide de masse », *Le Monde*, 15 novembre.
- KISOMBE, Ardevie (2021), « Oui, nous pouvons prévenir le féminicide! », *Le Nouvelliste*, 3 octobre.
- LAPALUS, Marylène et Mariana R. Mora (2020), « Fémicide/féminicide. Les enjeux politiques d'une catégorie juridique et militante », *Travail, genre et sociétés*, vol. 1, n° 43, p. 155-160.
- LEMONNIER, Marie (2022), « "Il nous faut, toujours et encore, dénoncer la pandémie mondiale de féminicides" », *L'OBS*, 13 mars.
- LEVET, Bérénice (2021), « "Féminicides": une sémantique militante pour une réalité tragique et complexe », *Figaro*, 14 juin.
- LORD, Jean-Marc (2016), « "Épidémie" de féminicides au Honduras », *Le Nouvelliste*, 8 juillet.
- MICHAUD, Yanick (2021), « Poser des actions concrètes pour contrer la violence faite aux femmes », *Le Journal Saint-François*, 24 novembre.
- MÜLLER, Quentin (2020), « Le terrible assassinat au nom de "l'honneur" de Romina, 14 ans, bouleverse l'Iran », *L'Express*, 23 juin.
- OCFJR (2019), *#Cestunfémicide : Comprendre les meurtres des femmes et filles basés sur le genre au Canada en 2019*, Observatoire canadien du fémicide pour la justice et la responsabilisation, Conseil de recherches en sciences humaines du Canada.
- OCFJR (2022), *#Cestunfémicide : Comprendre les meurtres de femmes et filles liés au sexe et au genre au Canada de 2018 à 2022*, Observatoire canadien du fémicide pour la justice et la responsabilisation, Conseil de recherches en sciences humaines du Canada.
- OCFJR (2024), « Définir et mesurer le fémicide », Observatoire canadien du fémicide pour la justice et la responsabilisation, [en ligne : <https://femicideincanada.ca/fr/quest-ce-que-le-feminicide/definir-et-mesurer-le-femicide/>]
- O., Laurence (2020), « Mon petit ami m'a frappé [sic], m'a violé [sic] et a menacé de violer ma sœur de 10 ans », *Santé+ Magazine*, 6 janvier.
- OMS (2012), *Comprendre et lutter contre la violence à l'égard des femmes*, Organisation mondiale de la Santé et Organisation Panaméricaine de la Santé.
- PERRIN, Dominique (2019), « Féminicide : mot masculin qui tue », *Le Monde*, 22 novembre.
- RIPA, Yannick (2022), « Tuées par des femmes », *L'Histoire*, 1^{er} novembre.
- ROMITO, Patrizia (2022), « Masculinisme, féminicides, fémicides », dans Christelle Taraud, *Féminicide : une*

histoire mondiale, Paris, La Découverte, p. 477-487.

ROBERT, Aurélie (2021), « Il poignarde sa femme devant leur enfant à Saint-Etienne », *Le Journal des Femmes*, 4 janvier.

ROYEN, Marie-Cécile (2021), « Le crime passionnel invisibilise », *Le Vif Express*, 26 août.

SCHWARTZBROD, Alexandra (2022), « Insidieux », *Libération*, 3 novembre.

UNODC, *Global Study on Homicide : Gender-related killing of women and girls* (2019), [En ligne : https://www.unodc.org/documents/data-and-analysis/gsh/Booklet_5.pdf]

VALLET, Élisabeth (2017), « États-Unis - Les femmes au pied du mur », *Le Devoir*, 11 mars.

VAN KASTEEL, Jean (2018), « Meurtre de Maria à Charleroi : "le féminicide n'est pas un fait divers" », *DHnet*, 8 mars.

VINCENT, Nadine (2022), « Faut-il adapter les dictionnaires à l'air du temps ? Proposition d'un traitement polyphonique du mot *woke* », dans Geneviève Bernard Barbeau et Nadine Vincent (dir.), *Regards linguistiques sur des mots polémiques*, numéro thématique de *Circula : revue d'idéologies linguistiques*, n° 15, p. 122-145.

VON JOEDEN-FORGEY, Elisa (2022), « Féminicide en génocide, féminicide et génocide », dans Christelle Taraud, *Féminicide : une histoire mondiale*, Paris, La Découverte, p. 577-591.

TITRE: LE QUÉBEC EST-IL UNE PROVINCE? REGARD LEXICOGRAPHIQUE SUR UN EMPLOI POLÉMIQUE

AUTEUR: NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE, CRIFUQ, LABORATOIRE LEXIQCORPUS

PUBLICATION: LES DICTIONNAIRES NUMÉRIQUES DANS L'ESPACE FRANCOPHONE, DES RESSOURCES PORTEUSES DE CULTURE ET D'IDÉOLOGIES

DIRIGÉE PAR : NICOLAS SORBA, UNIVERSITÀ DI CORSICA PASQUALE PAOLI, ET NADINE VINCENT, UNIVERSITÉ DE SHERBROOKE

ÉDITEUR: LES ÉDITIONS DE L'UNIVERSITÉ DE SHERBROOKE

COLLECTION : LEXIQCORPUS

ANNÉE: 2024

PAGES: 323-345

ISBN: 978-2-7622-0368-4

URI: [HTTP://HDL.HANDLE.NET/11143/22410](http://hdl.handle.net/11143/22410)

DOI: [HTTPS://DOI.ORG/10.17118/11143/22410](https://doi.org/10.17118/11143/22410)

CET OUVRAGE EST MIS À DISPOSITION SELON LES TERMES DE LA LICENCE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL.

Le Québec est-il une province? Regard lexicographique sur un emploi polémique

Nadine Vincent¹, Université de Sherbrooke, CRIFUQ, laboratoire LexiQcorpus

« Nous ne voulons pas être une province "pas comme les autres",
nous voulons être un pays comme les autres »

- Pierre Bourgault, journaliste, homme politique et intellectuel québécois

Résumé :

Depuis les années 1960, l'utilisation du mot *province* pour désigner le Québec est contestée par certains nationalistes qui préfèrent lui donner le titre de nation ou d'État. Après une accalmie à la suite du référendum de 1995, la polémique autour de la façon de désigner le Québec réapparaît de plus en plus souvent dans les médias depuis 2015 sans que les dictionnaires n'en tiennent compte. Nous proposons ici une mise à jour sur la question et une description polyphonique du mot *province* au Québec de façon à rendre compte des différentes idéologies portées par ce mot tout en explorant une nouvelle lexicographie rendue possible par le recours au numérique.

Mot-clés :

Lexicographie, État, nation, province de Québec, Québec

Abstract :

Since the 1960s, the use of the word *province* to designate the Quebec has been contested by certain nationalists who prefer to give it the title of nation or state. After a lull following the 1995 referendum, the controversy surrounding how to designate the Quebec has reappeared more and more often in the media since 2015 without dictionaries taking it into account. Here we offer an update on the question and a polyphonic description of the word *province* in Quebec in order to account for the different ideologies carried by this word while exploring a new lexicography made possible by the use of digital technology.

Keywords :

Lexicography, state, nation, province of Quebec, Québec

1. L'auteure remercie Daniella Coderre Porras et Serge D'Amico de la relecture attentive qu'ils ont faite de la première version de cet article.

1. Introduction

Les mots peuvent être porteurs de connotations différentes selon les personnes qui les utilisent. Nous nous intéressons ici au mot *province*, subdivision administrative et territoriale du Canada qui permet notamment de désigner le Québec depuis 1867. La collocation *province de Québec* a été remise en question – voire dénoncée – au cours des dernières décennies par certains nationalistes qui attribuent plutôt au mot *province* son sens premier de « territoire conquis » ou « territoire des vaincus », du latin *provincia* (de *pro-*, « pour », et de la racine de *vincere* « vaincre »). Notons que ce débat entre initiés est généralement ignoré par le grand public, surtout dans les plus jeunes générations.

Nous nous concentrerons sur l'évolution de cette contestation dans un corpus journalistique québécois et proposerons une nouvelle description du mot *province* en fonction des critères du Projet porc-épic (voir section 6) portant sur le traitement lexicographique d'emplois polémiques.

2. Quelques balises historiques

À la suite de la conquête du Canada par la couronne britannique (1759-1760), et de la cession du territoire par la France lors de la signature du traité de Paris (1763), la Grande-Bretagne nomme la nouvelle colonie *Province of Quebec*.

[L]e 14 novembre 1763, George III désignait James Murray « capitaine général et gouverneur en chef de la province de Québec », cette commission étant enregistrée le 28 novembre 1763. L'expression « province de Québec » faisait son entrée dans l'histoire. Pour les historiens de l'école de Montréal, c'était la marque du conquérant.²

Le Québec gardera cette appellation jusqu'à l'adoption de l'Acte constitutionnel, « [l]oi britannique adoptée en 1791 qui divise la Province de Québec en deux entités politiques, soit le Bas-Canada et le Haut-Canada³ ». En 1841, lors de la proclamation de l'Acte d'Union, les deux colonies sont de nouveau réunies pour former le Canada-Uni, constitué du Canada-Ouest (ancêtre de l'Ontario actuel) et du Canada-Est (ancêtre du Québec). Enfin, quelques décennies plus tard, le Dominion du Canada est créé par l'Acte de l'Amérique du Nord britannique, loi adoptée à la Chambre des Communes de Londres.

En 1867, deux colonies britanniques de l'Atlantique (Nouvelle-Écosse et Nouveau-Brunswick) et le Canada-Uni (Québec et Ontario) sont réunis pour former un seul grand « Dominion ». Les délégués du nouveau Dominion et les élus britanniques optent pour une monarchie constitutionnelle et une fédé-

2. <https://www.septentrion.qc.ca/blogue/du-slogan-la-belle-province-a-la-devise-je-me-souviens-sur-les-plaques-dimmatriculation> [page consultée le 2 mars 2024]

3. <https://www.assnat.qc.ca/fr/patrimoine/lexique/acte-constitutionnel-de-1791.html> [page consultée le 16 décembre 2023]

ration centralisée. Les responsabilités politiques sont distribuées entre deux paliers de gouvernements, celui du « fédéral » et ceux des « provinces ». (Bédard, 2017 : 8)

C'est à cette seconde incarnation du mot *province* attribuée au Québec depuis 1867 que nous nous intéressons dans le présent article.

3. Naissance de la diplomatie québécoise

Le 22 juin 1960, le gouvernement libéral de Jean Lesage prend le pouvoir au Québec, donnant naissance à la Révolution tranquille. En France, Charles de Gaulle est président de la République depuis le 8 janvier 1959. Rapidement, les deux gouvernements vont tisser des liens qui mèneront à l'inauguration de la Délégation du Québec à Paris, le 5 octobre 1961. C'est l'aboutissement d'un projet que certains portaient depuis des décennies⁴ et la reprise d'une relation France-Québec mise à mal depuis la Conquête de 1763⁵.

Lors de l'inauguration de la Délégation, le premier ministre Lesage est reçu à Paris comme un chef d'État et est « à la hauteur de la situation. Le discours qu'il prononça lors de l'inauguration montre que la Révolution tranquille qui s'amorçait, allait de pair avec l'ouverture sur le monde du Québec francophone⁶ » (Sarra-Bournet, 2012 : 52).

L'affirmation du rôle international du Québec se poursuit au cours de la décennie 1960 et trouve sa plus claire expression avec ce qu'il est convenu d'appeler la doctrine Gérin-Lajoie. Alors ministre de l'Éducation du Québec, Paul Gérin-Lajoie présente, le 12 avril 1965,

la thèse selon laquelle le Québec pouvait agir à l'étranger dans ses domaines de compétence sans « surveillance » ni « contrôle » d'Ottawa. [Le premier ministre Jean Lesage] appuya totalement la position de son ministre et donna ainsi naissance à la doctrine qui fonde toujours la politique étrangère du Québec. (Rioux, 5 octobre 2011)

Cette autonomisation du Québec touche peu à peu la façon dont l'élite québécoise choisit de se présenter au monde. Par exemple, pour souligner le centenaire de la création du Canada, la ville de Montréal est désignée pour devenir l'hôte de l'exposition universelle de 1967. Le diplomate québécois André Patry est nommé chef du protocole.

4. « Dans le portrait un peu sombre que l'on a tracé des années 1950, on a souvent oublié de rappeler que René Chaloult avait réclamé la création d'une maison du Québec à Paris dès les années 1940. Peu avant sa mort subite, Duplessis envisagea d'ailleurs de nommer Jean Désy, ambassadeur du Canada à Paris, délégué général pour l'Europe. » (Rioux, 4 octobre 2011)

5. Au retour d'une visite au Québec en avril 1960, le président de Gaulle recommandera à son ministre de la Culture, André Malraux, de s'occuper du Québec. « Malraux exprima alors le malaise qu'il éprouvait à la pensée que la France avait négligé le Québec pendant 200 ans et son souhait de mieux l'y faire connaître. » (Sarra-Bournet, 2012 : 51)

6. En 2024, en plus de Paris, le Québec dispose de délégations générales à Bruxelles, Dakar, Londres, Los Angeles, Mexico, Munich, New York et Tokyo et de délégations à Atlanta, Boston, Chicago, Houston et Rome.

[A]lors que le Québec s'apprête à accueillir une cinquantaine de chefs d'État et de gouvernement, il [...] en profite pour éliminer discrètement certains symboles coloniaux et affirmer la personnalité internationale du Québec. La « province » disparaît des documents officiels et l'on parle alors de l'État du Québec, comme disait souvent Jean Lesage. (Rioux, 5 octobre 2011 : A1)

Questionné à ce sujet, Claude Morin, un des principaux rédacteurs de discours [de Lesage] le confirme : « Ça fait des années qu'on utilise la formule Gouvernement du Québec. Province et provincial sont devenus des archaïsmes. On n'utilise plus Gouvernement de la Province de Québec. Province de Québec a rapidement cédé la place à État québécois ou État du Québec.⁷ »

Cette désignation d'*État*, en vigueur au Québec depuis ce temps, a même parfois été remplacée par le mot *pays* au grand dam de certaines fédéralistes, notamment chez les linguistes (voir Nemni, 2000).

Les nationalistes québécois aiment beaucoup employer le mot Québec sur le même plan que des mots désignant des états indépendants, et même – ô ironie ! – que des mots désignant des fédérations : « la Belgique, la Suisse, le Québec » est un type d'énumération fréquent dans des discours apparemment neutres, notamment sous la plume de linguistes québécois qui étudient la francophonie [...]. (Paquot, 2000 : 142)

4. Les institutions du Québec

Le mot *province* pour désigner le Québec étant remis en question au cours des années 1960, il disparaît peu à peu du nom de plusieurs institutions où le Québec est plutôt associé à un État ou à une nation (par l'utilisation de l'adjectif *national*). Est-ce un mouvement volontaire ou inconscient? Il n'est pas facile de trouver des textes témoignant clairement de l'objectif de faire disparaître le mot *province* de l'espace public, mais les cas de transformation sont trop nombreux pour tenir du hasard.

4.1. Les sociétés d'État

Partout au Canada, il existe des organismes publics gérés par des conseils d'administration dont la majorité des membres sont nommés par le gouvernement duquel ils relèvent. Ainsi, CBC/Radio-Canada, Postes Canada ou VIA Rail sont des sociétés d'État relevant du gouvernement fédéral. Quand ces organismes relèvent des gouvernements provinciaux, on parle généralement de *sociétés de la Couronne*. Ainsi, Hydro One (Ontario), Énergie NB Power (Nouveau-Brunswick) ou BC Hydro (Colombie-Britannique) sont des sociétés de la Couronne. Au Québec, Hydro-Québec, la Société des alcools du Québec (SAQ) ou Héma-Québec sont des sociétés d'État.

7. <https://www.septentrion.qc.ca/blogue/du-slogan-la-belle-province-a-la-devise-je-me-souviens-sur-les-plaques-dimmatriculation> [page consultée le 2 mars 2024]

Parmi les provinces et territoires, seuls trois gouvernements font exceptions et nomment leurs organismes publics *sociétés d'État*. Il s'agit du Québec, de la Saskatchewan et du Nunavut. Quant au Yukon, il fait cavalier seul en nommant ces organismes *sociétés publiques*.

Les entreprises appartenant à l'État ont toujours eu une importance particulière au Québec et en Saskatchewan. Après la victoire électorale historique de la Cooperative Commonwealth Federation en Saskatchewan en 1944, les sociétés d'État sont considérées comme un moyen important de bâtir une économie diversifiée. Depuis le début des années 1960, les gouvernements successifs du Québec ont aussi recours aux sociétés d'État pour diversifier l'économie provinciale, créer et conserver des emplois, et favoriser l'émergence de gestionnaires francophones. Hydro-Québec en est un bon exemple. Des sociétés d'État sont aussi créées au Québec dans les industries forestières, gazières et pétrolières, ainsi que dans les secteurs de l'acier et de l'amiante. (*Encyclopédie canadienne*⁸)

Même si elles sont pareillement nommées *société d'État*, les entreprises publiques du Québec et de la Saskatchewan diffèrent. En Saskatchewan, elles relèvent toutes de la Couronne, directement, par le biais de la Société des investissements de la Couronne (CIC) ou indirectement, par le biais du Conseil du Trésor de la province. Selon la définition donnée par le gouvernement de la Saskatchewan : « [u]ne société d'État est une entreprise appartenant à la Couronne ou à la Reine du chef d'une province. Elles sont constituées par une loi et rendent compte au gouvernement par l'intermédiaire d'un ministre.⁹ »

Au Québec, les sociétés d'État relèvent exclusivement du gouvernement québécois et le mot *Couronne* est absent de la terminologie de cette administration gouvernementale.

4.2. L'Assemblée nationale du Québec

De 1867 à 1968, le Parlement du Québec compte deux chambres, le Conseil législatif, constitué de 24 conseillers nommés par le lieutenant-gouverneur (représentant de la Couronne britannique) et la Chambre élue, « l'Assemblée législative de la province de Québec¹⁰ ». En 1968, le Conseil législatif est aboli, et la chambre unique est rebaptisée *Assemblée nationale du Québec*.

Ailleurs au Canada, les chambres élues des provinces et territoires portent le nom de *Chambre législative* (de la Colombie-Britannique, de l'Ontario, du Manitoba, du Yukon, du Nouveau-Brunswick, etc.), à l'exception de Terre-Neuve, dont la chambre élue est appelée *Chambre d'assemblée de Terre-Neuve-et-Labrador*.

8. <https://www.thecanadianencyclopedia.ca/fr/article/crown-corporation> [page consultée le 25 février 2024]

9. <https://www.saskatchewan.ca/bonjour/government/crown-corporations> [page consultée le 25 février 2024]

10. <https://www.assnat.qc.ca/fr/patrimoine/lexique/assemblee-legislative-de-la-province-de-quebec.html> [page consultée le 28 décembre 2023]

4.3. La fête nationale

En 1834, le journaliste et homme politique patriote Ludger Duvernay dote ceux qu'on appelle alors « les Canadiens français d'une fête nationale annuelle qui serait célébrée le 24 juin. Depuis des temps immémoriaux, le solstice d'été était une occasion de réjouissances. L'Église catholique avait récupéré cette fête païenne en la mettant sous le patronage de saint Jean-Baptiste. » (Lebel, 1985)

La Saint-Jean-Baptiste, d'abord fête de l'affirmation politique des Canadiens français du Québec, devient peu à peu la fête de tous les francophones du Canada, et même de la diaspora, ainsi que l'officialise le pape Pie X en 1908.

Dans un bref, il déclare, pour perpétuelle mémoire : « Nous établissons, Nous constituons et Nous proclamons saint Jean-Baptiste patron spécial auprès de Dieu des fidèles franco-canadiens, tant de ceux qui sont au Canada que ceux qui vivent sur une terre étrangère » (Drouin, 1991 : 19).

Il faudra attendre près de 70 ans pour que le gouvernement du Québec, sous René Lévesque, adopte une loi¹¹ qui fasse du 24 juin une fête à première vue plus territoriale qu'identitaire.

Le 24 juin, jour de la fête du saint patron des Canadiens français, est officiellement devenu la Fête nationale du Québec [...]. Ainsi en a décidé hier le Conseil des ministres, qui du même coup a relégué encore un peu plus loin dans l'histoire les caractères ethniques et religieux de la Saint-Jean-Baptiste. (s.a., *Le Devoir*, 19 mai 1977)

Il s'est donc produit une évacuation des Canadiens français hors Québec de la Saint-Jean-Baptiste au Québec. Pourtant, la fête de la nativité de saint Jean le Baptiste est célébrée dans les communautés francophones partout au Canada. (Drouin, 2016)

Aujourd'hui, pour parler du 24 juin au Québec, on utilise fréquemment la double dénomination : *Saint-Jean-Baptiste* et *fête nationale du Québec*. Pourtant, les publics cibles ne sont pas les mêmes : la Saint-Jean-Baptiste est la fête des Canadiens français de tout le Canada et la fête nationale, la fête de tous les Québécois, quelle que soit leur langue maternelle.

Ailleurs au Canada, quand on utilise l'adjectif *national*, c'est pour qualifier une activité pancanadienne, par exemple la *Journal nationale de la vérité et de la réconciliation*, congé férié fédéral qui se célèbre le 30 septembre. Seule exception : la *Fête nationale de l'Acadie*¹², qui se célèbre le 15 août. Un même adjectif réfère donc à des nations distinctes : le Québec, le Canada, l'Acadie. L'adjectif *national* n'est pas utilisé dans les autres provinces ou territoires pour désigner une fête identitaire. L'Ontario,

11. « Loi sur la fête nationale ». <https://www.legisquebec.gouv.qc.ca/fr/document/lc/f-1.1>

12. L'Acadie ne désigne pas un territoire défini, mais une nation qui regroupe les francophones des provinces maritimes, notamment ceux dont les ancêtres proviennent de l'Ouest de la France. Le français acadien est une variété de français distincte du français québécois.

par exemple, célèbre le 1^{er} juin le *Jour de l'Ontario* et la Colombie-Britannique souligne le premier lundi du mois d'août la *Journée de la Colombie-Britannique*.

4.4. La capitale nationale

Dans le même esprit d'affirmation du statut de nation pour le Québec, la désignation de la ville de Québec est passée de *vieille capitale* à *capitale nationale*.

Depuis les années 1990, on parle de « capitale nationale », surtout depuis la création de la Commission de la capitale nationale, en 1995. À cette époque, l'adjectif « national » pouvait encore sembler audacieux et pouvait heurter certaines sensibilités. Depuis, même les milieux les plus autorisés à Ottawa l'ont reconnue comme légitime et d'emblée elle s'impose à présent dans la désignation de la ville. (Gallichan, 2008 : 22)

Notons qu'à moins de 500 kilomètres à l'ouest de la ville de Québec, la ville d'Ottawa, capitale du Canada, se fait aussi appeler *capitale nationale*.

4.5. La Belle Province

Le slogan du Québec, *La Belle Province*, qui désigne le Québec, a été inscrit sur les plaques d'immatriculation des véhicules routiers de 1963 à 1978. Depuis 1978, il est remplacé par la devise du Québec : *Je me souviens*. Qu'est-ce qui a mené à cette modification? Selon Denis Vaugois, historien et ancien ministre du gouvernement Lévesque élu en 1976, c'est à lui que l'on doit ce changement.

Il raconte que lors du premier caucus du Parti québécois, après la victoire de novembre 1976, les candidats élus ont reçu leur premier chèque de député. Il a alors un choc.

[C]e chèque est émis par le Gouvernement de la Province de Québec. Je tombe de haut, incrédule. « Province », ce mot n'est-il pas à proscrire? Oui et non. Ça dépend! Il faut en effet bien en comprendre le sens. Les fédéraux aiment utiliser ce mot pour bien marquer le statut d'infériorité politique du Québec. [...]

Étymologiquement et historiquement, pour les péquistes bon teint, le mot province (*pro vincere*) est associée à la subdivision, la subordination, voire à la défaite. Assez curieusement, les services de la trésorerie de l'Assemblée nationale n'avaient pas vu passer le train. Leur libellé, « gouvernement de la province », avait des allures d'artefact. [...] La modification ne tarda pas.¹³

Après la correction sur les chèques, Vaugois aurait proposé de remplacer *La Belle Province* par *Je me souviens* sur les plaques d'immatriculation, appuyé notamment par Lucien Lessard et Lise Payette.

13. <https://www.septentrion.qc.ca/blogue/du-slogan-la-belle-province-a-la-devise-je-me-souviens-sur-les-plaques-dimmatriculation> [page consultée le 2 mars 2024]

Peu après, alors que l'on discutait d'assurance automobile, j'ai avoué à mes collègues que ce qui me fatiguait, c'était d'avoir une plaque automobile arborant les mots *La Belle Province*. En nous affichant ainsi, nous admettions être en tutelle; étions-nous donc si fiers d'être une province? (Vaugeois cité dans Savard, 2019 : 200)

C'est le ministre des Transports, Lucien Lessard, qui officialisera la décision quelques jours plus tard.

Dans un communiqué émis le 15 décembre 1976 et cité par au moins quatre quotidiens le lendemain (*Soleil, Presse, Nouvelliste, Sherbrooke Record*), le ministre Lessard annonçait [que les plaques d'immatriculation] de 1978 présenteraient une modification importante, la devise Je me souviens remplaçant le slogan *La belle province*. (Deschênes, 2018)

Bien que l'expression *La Belle Province* soit encore régulièrement utilisée dans les médias québécois pour désigner le Québec, certains justifient son remplacement par sa faible fréquence au Québec. Ainsi, l'auteur jeunesse Pierre-Alexandre Bonin écrit sur une page web destinée à un jeune public : « de 1963 à 1977, les plaques d'immatriculation portaient une autre devise! Eh oui, à cette époque, on pouvait plutôt lire «La Belle Province». Cette expression était pourtant utilisée davantage au Canada anglais et en France qu'au Québec.¹⁴ »

Malgré toutes ces transformations dans les désignations administratives, le remplacement de *province* par *État* ou *nation* dans l'usage n'a jamais été dominant. C'était une pratique tacite chez les nationalistes, mais le mot *province* a toujours été utilisé couramment dans l'espace public.

5. Le « retour » de province

Bien que le mot *province* ait toujours été très présent dans l'usage, sa fréquence a semblé croître à partir de la fin du 20^e siècle dans des contextes où depuis quelques décennies il avait été remplacé par *État* ou *nation* ou où le mot *Québec* était utilisé seul. Cette guerre des mots, plus implicite qu'affirmée, et menée dans certains milieux spécifiques, a sans aucun doute échappé à la majorité de la population québécoise et ne faisait pas partie des débats publics. Puis, peu à peu, on l'a vue apparaître dans les médias.

5.1. Réactions dans les journaux

Jusqu'en 2015, il était rare de lire dans les journaux des plaintes au sujet de l'utilisation du mot *province* pour désigner le Québec. Citons en contre-exemple cette réponse faite en 1991 par l'éditorialiste de *La Presse*, Guy Cormier, ouvertement fédéraliste, à un lecteur. Nous sommes alors à quelques années du référendum de 1995 sur la souveraineté du Québec.

14. <https://www.histoirecanada.ca/getmedia/2c6777df-2c47-4401-838d-e18affaebad5/Kay2022MarLaBelleProvinceFR.pdf.aspx> [page consultée le 31 août 2024]

Voici donc une lettre d'un lecteur anonyme, qui semble me reprocher d'endosser avec enthousiasme la désignation de « province de Québec » pour nommer le Québec. Je n'ai jamais écrit rien de tel. [...] « Province de Québec » est une expression traduite de l'anglais « Province of Quebec ». C'est une appellation anglaise pour une création de la Couronne anglaise. Il n'y eut jamais de « province de Québec » sous le régime français. Ce que le régime français connaissait, c'est la Nouvelle-France. (Cormier, 1991 : B2)

Après deux référendums perdus par le camp du oui (1980 et 1995) sur la question de l'indépendance du Québec, les nouvelles générations n'ont sans doute plus été sensibilisées aux connotations possibles du mot *province* et y sont revenues, généralement inconscientes de sa charge politique pour certaines personnes des générations précédentes.

Or, depuis quelques années, des revendications surgissent à l'occasion dans les médias traditionnels et les médias sociaux numériques demandant qu'on condamne le mot *province* et qu'on emploie les mots *État* ou *nation* pour désigner le Québec. Vingt ans et plus après le dernier référendum, ces revendications indiqueraient-elles le retour d'un sentiment national? Cette affirmation serait périlleuse à démontrer. Nous nous concentrerons plutôt sur cette ambiguïté portée par le mot *province*, qui semble neutre pour certains et insultant pour d'autres. Force est de constater qu'une sensibilité que l'on croyait perdue n'était en fait qu'en dormance, et que peu à peu s'élèvent des voix demandant de ne plus associer le Québec à une province.

Le mot province vient du latin *pro victis*, qui signifie « territoire des vaincus ». [...] Qu'on soit fédéraliste ou souverainiste, du moment qu'on est de bonne foi, force est d'admettre que le Québec est un État et qu'en ce moment, celui-ci est membre à part entière d'une fédération. Plutôt que d'arborer le titre de province, ne serait-il pas mieux que le Québec se nomme l'État fédéré du Québec ? (Jacques, 2015 : A8)

En fait, la question qui se pose, c'est si ce sont les revendications nationalistes qui reprennent du gallon, ou si c'est le mot *province* qui augmente en fréquence et provoque des réactions. Nous remarquons en effet que ceux qui en dénoncent l'utilisation, notamment le journaliste Antoine Robitaille, évoquent souvent son présumé « retour ».

Parce qu'il renferme cette majorité francophone, le Québec n'est pas une « province » comme les autres dans ce Dominion. Il est foyer de la nation québécoise. Conséquemment, depuis la Révolution tranquille, Jean Lesage et les premiers ministres qui suivirent ont remis l'expression « province de Québec », propre aux époques précédentes. Ils parlèrent « du Québec », de « l'État du Québec ». (Le retour du vocable « province » dans le discours public, notamment médiatique, pour désigner l'État ou le gouvernement du Québec, illustre la tentation de « reprovincialisation » de l'idée du Québec.) (Robitaille, 2015 : A6)

Budget « provincial », politique « provinciale », la « province de Québec », la « belle province », la province par-ci, la province par-là. On assiste actuellement à un retour à la dénomination du Québec, son gouvernement, son État, son territoire, son statut, comme « province »; même de la part de ceux qui, il y a 25 ou 30 ans, n'auraient pas utilisé ce mot. [L]a re-provincialisation des esprits semble s'accélérer.

On aperçoit le mot province dans des communiqués. De jeunes souverainistes l'utilisent même. [...] Difficile de ne pas conclure que c'est une certaine idée du Québec qui se perd. (Robitaille, 2018 : 12)

À la mort de l'ancien premier ministre péquiste Bernard Landry, la chroniqueuse Josée Legault raconte le rapport difficile qu'il avait avec les mots *province* et *provincial* et parle elle aussi d'un « retour ».

Bernard Landry était profondément allergique aux mots « province » et « provincial ». C'était connu et quiconque l'oubliait s'en mordait rapidement les doigts. Lorsqu'il les prononçait lui-même, c'était pour mieux en faire ressortir l'archaïsme méprisant que ces mots évoquent par leur définition même. [...] Connaissant bien le latin, il savait pertinemment que le mot « province » vient de *pro victis*. Ce qui veut dire « territoire des vaincus ». Qui voudrait donc affubler un État moderne comme le Québec, le sien propre en plus, d'une épithète aussi dégradante? Et pourtant... Depuis quelques années, les vocables « province » et « provincial » opèrent un retour en force dans l'espace public, politique et médiatique. C'est comme si nous ne savions plus le sens des mots ni le pouvoir qu'ils ont de nous élever dans notre perception de nous-mêmes. Ou, au contraire, de nous rabaisser. (Legault, 2018)

De même, quelques années plus tard, quand deux juristes proposent de modifier la Constitution du Québec contenue dans la Loi constitutionnelle de 1867, l'éditorialiste Robert Dutrisac rappelle des notions importantes et en profite pour souligner que « les jeunes générations ont recommencé à employer libéralement » le mot *province* pour parler du Québec.

Des notions comme peuple du Québec, capitale nationale, souveraineté partagée pourraient faire partie de la loi fondamentale. Le mot province, que les jeunes générations ont recommencé à employer libéralement, pourrait disparaître au profit du terme État du Québec. Après tout, depuis la Révolution tranquille, le gouvernement du Québec, que ce soit dans ses lois ou ses communications, a cessé d'employer ce premier vocable, qui - il n'est pas inutile de le rappeler - viendrait du latin *pro victis*, ou territoire des vaincus. (Dutrisac, 2021, p. A6)

Dans les journaux, les prises de position se multiplient. Parfois on condamne un politicien parce qu'il a malencontreusement utilisé le mot *province*, d'autres fois on s'apitoie sur le sort du Québec, en constatant qu'il s'abaisse au rang de province. Dans l'exemple suivant, un lecteur dénonce l'humiliation que fait subir le fédéral au Québec, et évoque les deux premiers ministres Trudeau, fils et père, sans même sembler s'en rendre compte.

Le gouvernement de Justin Trudeau, au moyen de sa grosse Banque du Canada qui lui fabrique de l'argent à volonté, a signifié au Québec sans même le vouloir qu'il est redevenu plus que jamais une simple province, une province sans trop de moyens.

[...]

[U]ne déclaration comme celle de Pierre Trudeau ne peut aujourd'hui qu'être lue avec plus d'amertume qu'au moment où elle fut prononcée en avril 1994 à l'Université de Montréal : « Les professeurs enseignent le nationalisme dans les écoles. Ils disent à leurs élèves que le Québec est un État et c'est faux. Le Québec est une province, un canton, un district ». (Martin, 2020)

5.2. Réactions dans les médias sociaux numériques

Le débat sur le mot *province* apparaît aussi dans les médias sociaux numériques. Nous ne signalons ici que quelques cas apparus sur X (anciennement Twitter) dans des comptes signés (non anonymes).

Par exemple, en juillet 2021, le compte de L'Action nationale (revue aujourd'hui indépendantiste, publiée depuis 1917) fait un parallèle avec le mot *indiens*, que l'on a remplacé dans la langue courante par *Premières Nations*.

On ne parle plus d'« indiens » même si une loi fondamentale les désigne comme tels, pourquoi insister sur le terme « province » pour désigner le Québec? (@actionnationale, 13 juillet 2021)

En mars 2022, réagissant à un titre de TVA Nouvelles et à un texte du journal gratuit *24 heures*, le chroniqueur du *Journal de Québec*¹⁵, Antoine Robitaille, corrige les emplois de *province* qu'il juge abusifs et crée le mot-clic #déprovincialisons. Il l'utilisera deux fois dans la même journée et à quelques reprises dans les jours suivants. Le mot-clic (ou le mot seul) est parfois repris par d'autres internautes.

TVA Nouvelles : La moitié des cégeps de la province débordent.¹⁶

Précision : « La moitié des cégeps ***du Québec*** débordent. » #déprovincialisons (@Ant_Robitaille, 25 mars 2022)

#déprovincialisons (@TetroAlexis, 25 mars 2022)

24 heures : Peu importe que vous soyez piéton, cycliste ou automobiliste, chaque famille de quatre personnes au Québec assume 6900 \$ par année en dépenses publiques pour le transport automobile dans la province.¹⁷

« Peu importe que vous soyez piéton, cycliste ou automobiliste, chaque famille de quatre personnes au Québec assume 6900 \$ par année en dépenses publiques pour le transport automobile [...]. » #déprovincialisons (@Ant_Robitaille, 25 mars 2022)

L'ancienne mairesse de Magog se lance en politique provinciale (@tvanouvelles, 27 mars 2022)

« Provinciale » #déprovincialisons (@nacroix)

Pourquoi pas « en politique québécoise »? Déprovincialisons le Québec! (@DavChampagne, 27 mars 2022)

Le 5 avril 2022, la ministre caquiste des relations internationales utilise le mot *province* pour désigner le Québec et reçoit quelques critiques, notamment de l'essayiste Frédéric Lacroix, de l'ancien

15. TVA Nouvelles, *24 heures* et *Le Journal de Québec* sont tous trois propriétés de Québecor Média.

16. <https://www.tvanouvelles.ca/2022/03/25/la-moitie-des-cegeps-de-la-province-debordent-1-1> [page consultée le 3 mars 2024] Ce titre a aussi été repris la même journée dans *Le Journal de Québec*.

17. <https://www.24heures.ca/2023/08/17/ya-t-il-vraiment-trop-dautos-au-quebec-comme-le-pense-le-ministre-de-leconomie-pierre-fitzgibbon> [page consultée le 3 mars 2024]

député péquiste André Boulerice, de la députée libérale Christine St-Pierre et bien sûr du journaliste Antoine Robitaille.

Notre province est tellement belle (@NadineGirault, 5 avril 2022)

La ministre des relations internationales devrait évidemment savoir qu'il s'agit de l'État du Québec. Les mots comptent. (@nacroix, 7 avril 2022)

Elle adore probablement être une ministre « provinciale ». (@BoulericeAndre, 7 avril 2022)

Étonnant de la part de celle qui dirige la diplomatie québécoise. Le Québec est un état fédéré! (@st-pierre_ch, 7 avril 2022)

Ah oui, une « prôvince »? #déprovincialisons! (@Ant_Robitaille, 7 avril 2022)

Le débat ne se résume bien sûr pas à ces quelques messages. Certains rappellent « les faits » : le Québec est une province. D'autres sont prêts à l'appeler dès maintenant un pays. Chose certaine, le mot *province* fait réagir. Quant au mot-clic #déprovincialisons, il a été réutilisé à quelques reprises par son auteur, mais ne s'est pas généralisé.

6. Traitement du mot *province* dans les dictionnaires

Soyons réaliste : le débat autour de l'utilisation du mot *province* au Québec existe bel et bien, mais il reste relativement marginal. Est-il assez important pour apparaître dans les dictionnaires de langue générale? Non. Qu'ils soient faits en France ou au Québec, les dictionnaires n'en disent rien. Et même, la description qu'il font du mot *province* rend le débat caduc.

Prenons, par exemple, le dictionnaire français *Le Petit Robert* et le dictionnaire québécois *Usito*. Tous deux définissent le sens historique de *province* (marqué « Histoire romaine » dans *Le Petit Robert* et « Antiquité romaine » dans *Usito*) qui correspond à un territoire conquis par Rome hors de l'Italie. Et tous deux attestent également un sens propre à la politique canadienne où la province correspond à un « État fédéré¹⁸ ».

3 (1867 [Acte de l'Amérique du Nord britannique]) MOD. Au Canada, État fédéré doté d'un gouvernement propre, souverain dans le domaine de ses compétences. Les dix provinces canadiennes. Les provinces de l'Ouest, les Provinces maritimes. VIEILLI La Belle Province : le Québec.

Figure 1 : Sens de *province* en contexte canadien dans *Le Petit Robert 2024*

18. Cette quasi-synonymie entre *province* et *État fédéré* est aussi présente dans le *Trésor de la langue française informatisé* (1972-1994) et dans le *Grand dictionnaire terminologique* de l'Office québécois de la langue française (fiche de 2003).

◆ SPÉCIALT (AU CANADA) Subdivision administrative et territoriale correspondant à chacun des dix États fédérés ayant son gouvernement propre.

Les provinces canadiennes.

La province de Québec.

La province de l'Ontario, de la Colombie-Britannique.

« À une assemblée libérale, il appuie une résolution exigeant une réforme en profondeur de la Constitution de 1867 afin [...] d'assurer que les provinces disposent à l'avenir de ressources financières suffisantes » (J.-P. Warren, 2015).

- *Les provinces maritimes.*

- *Les provinces atlantiques, formées des provinces maritimes et de Terre-Neuve-et-Labrador.*

- (AVEC DES MAJUSC.) *La Belle Province* : le Québec.

Figure 2 : Acception de *province* en contexte canadien dans *Usito*

Les dictionnaires traitent bien sûr *province* comme un mot polysémique, faisant deux sens très différents de la province romaine et de la province canadienne. À la lecture des définitions données, la *province de Québec* est synonyme de l'*État fédéré du Québec*, ce qui devrait satisfaire tous les bel-ligérants. Ce serait exact et suffisant si on faisait abstraction de la perception des usagers, du regard subjectif qu'ils peuvent porter sur un emploi, et qui constitue sa connotation. Ici, les dictionnaires n'en tiennent pas compte.

Parmi les sources consultées, seul le *Wiktionnaire*, dictionnaire collaboratif, fait état du débat. Dans une note suivant la définition de *province* en contexte canadien, il indique que le choix de ce mot a été délibérément fait par Londres pour éviter de donner aux différentes constituantes du Canada le nom *État*, porté par les territoires américains. On peut supposer que cette distinction avait pour but d'éviter des tentations d'annexion aux États-Unis, présentes notamment au Québec au 19^e siècle, mais aussi de pouvoir définir de façon autonome les obligations et prérogatives d'une province, sans influence de la structure adoptée par le voisin américain.

11. (Canada) (Administration) (Depuis 1867) État doté d'un gouvernement élu et faisant partie de la fédération canadienne. **Note** : Cette appellation a été choisie (plutôt qu'État) par le gouvernement britannique pour éviter toute analogie avec les États des États-Unis d'Amérique. L'État québécois l'a fait disparaître de son vocabulaire, pour se désigner soi-même, dans les années 1970 (voir citation ci-dessous).

- Le Canada compte dix **provinces** et trois territoires.
- Le Québec et l'Ontario sont les deux **provinces** centrales du Canada.
- Je me rappelle qu'une fois élu [en 1976], à l'une de nos premières réunions du caucus, quand nous avons reçu notre premier chèque de député, j'ai constaté qu'il était émis par le gouvernement de la « **Province** de Québec ». J'étais insulté d'avoir un chèque de la « **province** », parce que ma bataille comme fonctionnaire, c'était de faire sauter le mot **province** pour le remplacer par État. On trouvait d'ailleurs ce mot dans les discours de Jean Lesage dans les années 1960, et je m'étais battu à l'Agence de coopération culturelle et technique pour que le Québec ait un statut d'État, pas de **province**. Ce mot est synonyme de colonie. — (Denis Vaugeois et Stéphane Savard, *Entretiens*, Boréal, Montréal, 2019, pages 199-200)

Figure 3 : Sens canadien de *province* dans le *Wiktionnaire*¹⁹

La note précise aussi que le Québec a fait disparaître le mot *province* « de son vocabulaire, pour se désigner soi-même », mais situe ce changement dans les années 1970 plutôt que dans les années 1960, ainsi que nous l'avons établi (voir sections 2 et 3) et ainsi que le rappelle la citation du ministre péquiste Denis Vaugeois qui mentionne le rôle de précurseur de Jean Lesage dans les années 1960.

Dans le *Wiktionnaire*, la locution *Belle Province* est traitée dans un article autonome (voir figure 4). L'exemplification est constituée de trois citations de 2006, ce qui témoigne qu'elle est encore en usage. Aucune allusion n'est faite dans cet article à la connotation du mot *province* au Québec, sinon un commentaire dans une note qui indique que la locution est « souvent utilisée de manière plaisante, étant considérée comme un cliché par une bonne partie de la population québécoise. » Le mot *cliché* semble ici faire référence au côté pittoresque, voire folklorique de cette expression plutôt qu'à la polémique politique dont le mot *province* est l'objet.

19. Page consultée le 31 août 2024.

Belle Province \bel pɔv.vɛs\ féminin singulier

1. (Géographie) (Avec l'article *la*) (Surnom) Province de Québec au Canada.

- Lors de la crise du verglas, le *Globe and Mail* avait également jugé que la **Belle Province** s'agitait beaucoup pour bien peu. — (*Le Devoir*, 21 septembre 2006)
- C'était en 2004, au Festival d'été de Québec. *Bérurier Noir*, après 15 ans d'absence au Québec, venait rallier pour un soir sur les plaines d'Abraham toute « la jeunesse itinérante » de la **Belle Province**, voire du Canada. — (*Le Devoir*, 25 août 2006)
- Le recours à des migrants à fort potentiel démographique se rencontre même dans des États fédéraux comme le Canada où la « **Belle province** », le Québec, en panne démographique depuis quelques décennies mène une politique d'immigration francophone active pour maintenir son identité. — (Christian Pradeau et Jean-François Malterre, *Migrations et territoires*, dans *Les cahiers d'Outre-Mer* n^o 234/vol. 59, Presses Universitaires de Bordeaux, 2006)

Notes [modifier le wikicode]

Cette expression a figuré sur les plaques d'immatriculation québécoises de 1963 à 1978. Elle est encore assez courante aujourd'hui, quoi qu'elle soit souvent utilisée de manière plaisante, étant considérée comme un cliché par une bonne partie de la population québécoise.

Figure 4 : Article *Belle Province* dans le *Wiktionnaire*²⁰

7. Traitement du mot *province* dans le Projet porc-épic

Le Projet porc-épic (PPÉ), créé à l'Université de Sherbrooke en juin 2021, s'inscrit dans les travaux du laboratoire LexiQcorpus du CRIFUQ (Centre de recherche interuniversitaire sur le français en usage au Québec). Il a comme objet le traitement lexicographique d'emplois polémiques. Le PPÉ ne décrit pas les acceptions consensuelles et se concentre sur celles qui font débat, ce qui lui permet de mettre l'accent sur des cas trop pointus pour être soulignés par les dictionnaires généraux. De plus, tout en respectant la méthodologie lexicographique, il se permet des innovations inconcevables pour un dictionnaire général, comme de multiplier les descriptions d'une même acception en fonction de ses différentes interprétations, de préciser les sèmes qu'un mot ne contient pas plutôt que de se limiter à ceux qu'il contient, de décrire un nom propre, etc. Le mot *province* en contexte québécois est donc tout désigné pour intégrer le Projet porc-épic.

Deux options s'offrent à nous ici : soit de décrire le mot *Québec* et de spécifier s'il s'agit d'une province, d'un État, d'une nation, d'une société distincte ou autre, en précisant qui utilise quel mot et dans quels contextes; soit de décrire *province* en contexte québécois, et de présenter ses deux versants en fonction des connotations qui lui sont ou non attribuées. Nous opterons ici pour la seconde option, mais la première sera sans aucun doute à réaliser aussi dans le cadre du PPÉ.

20. Page consultée le 31 août 2024.

Comme il est de mise dans le PPÉ, l'article lexicographique lui-même sera précédé d'un préambule expliquant l'objet de la polémique. Dans le cas de *province*, le préambule résumera en environ une page les sections 1 à 4 du présent texte.

Voici ce que nous proposons pour l'article lexicographique lui-même. Il est divisé en deux macrosens. Le premier décrit brièvement le sens de *province* en contexte canadien. Le deuxième traite *province de Québec* (ou *province* pour nommer le Québec), ce qui inclut la locution *La Belle Province*, encore très courante dans l'usage. L'emploi majoritaire et non marqué est d'abord présenté, suivi de l'emploi péjoratif de *province*, plus marginal, mais tout de même bien implanté dans le discours public au Québec.

PROVINCE, NOM FÉMININ

I. Au Canada, division administrative et territoriale ayant son gouvernement propre et exerçant des pouvoirs constitutionnels.

Le Canada compte dix provinces et trois territoires.

L'éducation et la santé sont des champs de compétence des provinces.

« Alors que les provinces exercent des pouvoirs constitutionnels de plein droit, les territoires ont des pouvoirs délégués sous l'autorité du Parlement du Canada. » (Gouvernement du Canada²¹²²)

II. Spécialement PROVINCE DE QUÉBEC ou employé seul PROVINCE : État de la fédération canadienne, arrosé par le fleuve St-Laurent, et dont la métropole est Montréal.

→ ÉTAT, NATION, SOCIÉTÉ DISTINCTE

◇ **SANS CONNOTATION.** En général, le mot *province* est employé de façon neutre et descriptive pour désigner le Québec.

L'entreprise « a sondé un échantillon aléatoire de 1122 personnes âgées de 18 ans ou plus et vivant dans la province de Québec. » (Le Devoir, 2024²³)

« Si les nuages et la pluie doivent faire leur apparition en début de journée dans le sud de la province, les régions de Québec et de la Beauce auront droit à une journée relativement ensoleillée avant que le ciel ne se couvre en début de soirée. » (Le Journal de Montréal, 2024²⁴)

21. Les citations à l'intérieur de l'article proposé respectent les règles de présentation des articles lexicographiques et non d'un article scientifique. Pour cette raison, nous ne mettons pas les références des articles journalistiques cités dans la bibliographie. Nous les indiquons exceptionnellement en note de bas de page.

22. <https://www.canada.ca/fr/affaires-intergouvernementales/services/provinces-territoires.html>

23. <https://www.ledevoir.com/environnement/808073/majorite-quebecois-rejettent-approche-caquiste-dossier-northvolt>

24. <https://www.journaldemontreal.com/2024/03/02/pluie-nuages-et-temps-doux-une-fin-de-semaine-aux-allures-de-printemps>

— **La Belle Province** : Le Québec.

« Près d'une vingtaine de productions de la Belle Province (dont neuf longs métrages) ont [...] été sélectionnées au 48e Festival international du film de Toronto (TIFF), qui débute demain soir. » (Le Journal de Montréal, 2023²⁵)

« Au décompte final, [la compagnie] aura réduit ses effectifs de quelque 1700 employés au Québec. C'est tout de même significatif, étant donné que l'entreprise compte 9000 employés dans la Belle Province. » (Le Nouvelliste, 2023²⁶)

◊ **AVEC CONNOTATION NÉGATIVE, DÉPRÉCIATIVE.** Dans certains cercles politiques et médiatiques nationalistes, on évite d'associer le Québec à une province depuis les années 1960. Le mot *province* référerait à un territoire conquis, à une colonie. Se limiter au statut de province serait pour le Québec de renoncer à son histoire, plus ancienne que celle du Canada, à sa culture distincte, à sa responsabilité en tant que seul État majoritairement de langue française en Amérique et à sa capacité à se représenter lui-même sur la scène internationale. Dans les discours prospectif, on distingue parfois le « Québec(-)province » du « Québec(-)pays ».

« Réassignés à notre statut minoritaire à la suite du dernier référendum, recalés au niveau d'une simple province au sein de la fédération canadienne, une province qui d'ailleurs en impose de moins en moins, nous constatons que la question nationale sombre aujourd'hui dans l'insignifiance, dans d'incessantes chicanes fédérales-provinciales. » (Francine Pelletier, citée dans L'Actualité, 2023²⁷)

« Cantonné dans un statut d'État semi-libre, le Québec province ne pèse pas lourd dans les affaires qui comptent. Veut-il être laïque à sa façon qu'il ne le peut pas si Ottawa n'aime pas sa manière. Ses lois pour protéger le français sont soumises à l'approbation d'une cour issue d'un autre État dont le programme entre en contradiction avec le sien. [...] Quant à l'idée d'être vert dans un État pétrolier comme le Canada, bonne chance ! Sur la scène politique internationale, de plus, le Québec n'existe pas. Ses idées seraient-elles les meilleures du monde qu'on ne les entendrait pas. » (Le Devoir, 2021²⁸)

Les « auteurs [de l'étude] soutiennent notamment que le déficit d'un Québec-pays serait moins élevé que celui d'un Québec-province » (Radio-Canada, 2023²⁹)

25. <https://www.journaldemontreal.com/2023/09/05/le-festival-de-toronto-deroule-son-tapis-rouge-au-cinema-quebecois>

26. <https://www.lenouvelliste.ca/chroniques/yvon-laprade/2023/09/17/quest-ce-qui-se-passe-chez-olympel-3NO2AFYOTJHTFIF6SJ6NLUYRWQ/>

27. <https://lactualite.com/societe/nationalisme-quebecois-et-diversite-le-point-de-rupture/>

28. <https://www.ledevoir.com/opinion/chroniques/762063/chronique-peuple-chambreur>

29. <https://ici.radio-canada.ca/nouvelle/2020401/budget-an-1-parti-quebecois>

— **La Belle Province** : Le Québec inféodé au Canada.

« L'appellation "La belle province" qu'on utilisait autrefois pour désigner le Québec revient en force et nous ramène dans un passé qui nous réduit au rang de "La belle soumise". On redevient colonisés comme avant, alors que le Québec est reconnu depuis longtemps comme un état dans le Canada. » (Le Journal de Montréal, 2021³⁰)

« Pour Ottawa, l'objectif ultime consiste à ce que le Québec devienne un reliquat folklorique muselé, édenté, dégriffé et stérilisé. Aux yeux des fédéraux, la "Belle Province" ne peut être tolérée que dans deux positions subalternes : à plat ventre ou à genoux. » (La Tribune, 2022³¹)

« Aux Jeux olympiques d'hiver, le Canada exploite sa Belle Province. Elle travaille docilement pour lui. Elle est obligée de renoncer à sa couleur bleue pour se draper intégralement dans l'unifolié rouge et blanc. » (Le Journal de Montréal, 2022³²)

REMARQUE 1. Chez les personnes réfractaires au mot *province*, on utilise plutôt le mot *Québec* seul, ou on dira le *gouvernement*, *l'État*, *la nation*, *le territoire du Québec* (ou le *gouvernement*, *l'État*, *la nation*, *le territoire québécois*).

« Le Québec est à solidifier ses liens avec ses partenaires américains, européens et africains, alors qu'il revoit sa stratégie dans la région indopacifique en évoquant des changements potentiels en Chine. » (PC, Radio-Canada, 2023³³)

« En somme, on veut que l'État du Québec ait pleins pouvoirs dans ces champs de compétence. » (La Presse, 2022³⁴)

« Le Québec forme une nation. Le français est la langue officielle et commune de cette nation. » (Le Journal de Montréal, 2021³⁵)

30. <https://www.journaldemontreal.com/2021/09/26/le-quebec-nest-pas-une-province-comme-une-autre>

31. <https://www.latribune.ca/2022/10/22/le-reveil-du-geant-9f83215cf080ff9e5181b00f32d01a9f/>

32. <https://www.journaldemontreal.com/2022/02/24/olympiques--le-canada-sapproprie-notre-succes>

33. <https://ici.radio-canada.ca/nouvelle/1950087/quebec-international-diplomatie-chine-afrique>

34. <https://www.lapresse.ca/contexte/2022-06-26/extrait-la-revolution-tranquille-60-ans-apres/qu-est-ce-que-le-modele-quebecois.php>

35. <https://www.journaldemontreal.com/2021/05/14/voir-un-quebec-se-tenir-debout-fait-du-bien>

REMARQUE 2. Tout comme *province*, l'adjectif *provincial* est parfois neutre et parfois connoté négativement. Tous les autres dérivés de *province*, appliqués au Québec, sont toujours connotés, soit positivement (*déprovincialisation*, *déprovincialiser*, etc.), soit négativement (*provincialisation*, *provincialiser*, *provincialiste*, *reprovincialisation*, etc.)

« Présentement, on ne rêve plus. On s'acharne plutôt à reprovincialiser jusqu'au trognon ce peuple qui menaçait gravement de se prendre pour lui-même... » (René Lévesque dans *Le Journal de Montréal* du 27 août 1970, cité dans *Le Journal de Montréal*, 2023³⁶)

« Devenu premier ministre en avril 2003, huit ans après le référendum, il s'est donné une mission : reprovincialiser le Québec. » (*Le Journal de Montréal*, 2023³⁷)

« Sans rapport de force devant le gouvernement canadien, les beaux discours nationalistes provinciaux du gouvernement [...] sonnent creux. » (*Le Devoir*, 2022³⁸)

« Si la Révolution tranquille avait marqué une tentative de "déprovincialisation" du Québec, les quinze années de gouvernement libéral entamées en 2003 allaient marquer une ère de "reprovincialisation", en abandonnant la question du régime politique au profit d'un discours dépolitisé chantant la réussite individuelle. » (*Le Devoir*, 2020³⁹)

« S'il y avait beaucoup à espérer d'un Québec en voie de se déprovincialiser, il n'y avait plus rien à attendre du Québec reprovincialisé d'après le 20 mai 1980. » (*Le Devoir*, 2001⁴⁰)

Figure 5 : Article *province* dans le Projet porc-épic

8. Conclusion

Parce qu'ils doivent réalistement s'en tenir à décrire l'usage dominant, en termes notamment de fréquence et de dispersion, les dictionnaires généraux laissent bien souvent dans l'ombre des emplois plus marginaux, mais néanmoins représentatifs de courants idéologiques qui circulent dans la société.

La polyvalence de la lexicographie numérique permet de multiplier les projets de description de la langue et de se concentrer au besoin sur des emplois spécifiques, les emplois polémiques en ce qui concerne le Projet porc-épic.

36. <https://www.journaldemontreal.com/2023/06/24/rene-levesque-lincarnation-de-lamour-du-quebec>

37. <https://www.journaldemontreal.com/2023/04/14/une-nostalgie-nommee-jean-charest>

38. <https://www.ledevoir.com/opinion/idees/749757/elections-quebec-2022-pour-une-election-sur-l-avenir-de-notre-nation?>

39. <https://www.ledevoir.com/opinion/chroniques/584909/un-destin-provincial?>

40. Bouthillier, Guy (2001) « Le dit et le non-dit dans le rapport Larose », *Le Devoir*, 6 septembre, p. A 7.

Pour le cas du mot *province* au Québec, ce traitement hors des sentiers battus permet d'extraire des livres d'histoire et des essais de spécialistes une utilisation de la langue inconnue du grand public⁴¹, parce que vivant en marge de l'usage courant, et de la placer sous les projecteurs de la lexicographie. Il contribuera aussi, nous l'espérons, à donner des clés aux nouvelles générations pour leur permettre de comprendre la place de la langue dans une lutte d'affirmation nationale qui se mène aussi bien dans les débats politiques que dans le choix des mots.

41. Comme toute généralisation, celle-ci connaît des exceptions. Ainsi, l'auteur jeunesse Pierre-Alexandre Bonin écrit sur une page web présentant le Québec à un jeune public : « Tu as peut-être remarqué que nous parlons beaucoup d'emblèmes et de symboles "nationaux" plutôt que "provinciaux" comme c'est le cas ailleurs au Canada. C'est parce que le Québec est considéré comme une nation distincte, la seule province unilingue francophone au sein du Canada. [...] Il y a aussi l'Assemblée nationale, où siège le gouvernement du Québec, et notre fête nationale, la Saint-Jean-Baptiste, célébrée le 24 juin. » <https://www.histoirecanada.ca/getmedia/2c6777df-2c47-4401-838d-e18affaebad5/Kay2022MarLaBelleProvinceFR.pdf.aspx> [page consultée le 31 août 2024]

Bibliographie

- BÉDARD, Éric (2017), « 1867, la première fondation du Canada », dans *Histoire, économie et société*, n° 4, p. 8-27.
- CORMIER, Guy (1991), « Comédie musicale », *La Presse*, 24 mars, p. B2.
- DESCHÊNES, Gaston (2018), « La devise sur les plaques d'immatriculation: à chacun ses mérites », *Le Journal de Québec*, 29 octobre. [En ligne : <https://www.journaldequebec.com/2018/10/29/la-devise-sur-les-plaques-dimmatriculation-a-chacun-ses-merites>]
- DROUIN, François (1991), « Pourquoi la Saint-Jean-Baptiste », *Cap-aux-Diamants*, n° 26, p. 18-19.
- DROUIN, Mathieu (2016), « La Saint-Jean-Baptiste : une fête connue, mais méconnue », *Histoire Canada*. [En ligne : <https://www.histoirecanada.ca/consulter/arts-culture-et-societe/la-saint-jean-baptiste-une-fete-connue-mais-meconnue>]
- DUTRISAC, Robert (2021), « Faire avancer le Québec », *Le Devoir*, 26 janvier, p. A6.
- GALLICHAN, Gilles (2008), « Pour en finir avec l'expression «Vieille Capitale» », *Cap-aux-Diamants*, n° 94, p. 22-24.
- JACQUES, Mathieu (2015), « Pourquoi s'appeler encore une province? », *Le Devoir*, 5 novembre, p. A 8.
- LEBEL, Jean-Marie (1985), « DUVERNAY, LUDGER (baptisé Joseph-Ludger) », dans *Dictionnaire biographique du Canada*, vol. 8, Université Laval/University of Toronto. Fiche révisée en 2021. [En ligne : http://www.biographi.ca/fr/bio/duvernay_ludger_8F.html]
- LEGAULT, Josée (2018), « Bernard Landry et la nation », *Le Journal de Montréal*, 8 novembre. [En ligne : <https://www.journaldemontreal.com/2018/11/08/bernard-landry-et-la-nation>]
- MARTIN, Réjean (2020) « Une province », *Le Nouvelliste*, 19 juin. [En ligne : <https://www.lenouvelliste.ca/2020/06/19/une-province-21061381045544c805d7d332ceec0d28/>]
- NEMNI, Monique (2000), « Lettre à Larousse » dans le dossier « Cité libre corrige le Petit Larousse illustré », *Cité libre*, 28(4), 146-152.
- PAQUOT, Annette (2000), « La victoire du discours nationaliste au Québec », *Cité libre*, 28(4), 141-145.
- RIOUX, Christian (2011), « Il y a 50 ans, le Québec à Paris - 2 – Un grand élan diplomatique », *Le Devoir*, 4 octobre, p. A 1.
- RIOUX, Christian (2011), « Il y a 50 ans, le Québec à Paris - 3 - Une passion disparue », *Le Devoir*, 5 octobre, p. A 1.
- ROBITAILLE, Antoine (2015), « Philippe Couillard à Queen's Park - Les "vraies" suites », *Le Devoir*, 13 mai, p. A6.
- ROBITAILLE, Antoine (2018), « Non au budget "provincial" », *Le Journal de Québec*, 27 mars, p.12.

SARRA-BOURNET, Michel (2012) « La naissance de la Délégation Générale du Québec à Paris », *Études canadiennes / Canadian Studies*, n° 73, p. 47-60.

SAVARD, Stéphane (2019), *Denis Vaugeois, entretiens*, Boréal, 381 p.