

PIERRE CORBIN
UMR 8163 (STL), Université de Lille

Les dictionnaires monolingues généraux du français actuel gratuits en ligne : évolutions récentes (2020)

Free Online General Monolingual Dictionaries of Current French: Recent Developments (2020)

Abstract

This paper presents a selection of free online general dictionaries of current French in the middle of 2020, focusing on several topics: the metadata, sometimes abundant, explicit and detailed, sometimes brief or lacking; the initiators or authors of dictionaries, who may be academies (e.g. *Dictionnaire de l'Académie française*), university departments (e.g. *Usito*, produced at Sherbrooke University, Quebec), publishers (e.g. *Dico en ligne* by Le Robert), software developers (e.g. *Dictionnaire Cordial* by Synapse Développement), collaborative organizations (e.g. *Wiktionnaire*), individuals (e.g. *Interactive Language Toolbox*, written and maintained by Flemish teachers of French as a foreign language) or unknown persons (e.g. *Dictionnaire français* on L'Internaute website); the hosts of these dictionaries, who may also be news websites (e.g. *sensAgent* on Le Parisien website); the user interface, more or less friendly and interactive; the update of headwords and data and the quality of linguistic descriptions, which can vary significantly from one dictionary to another.

Keywords: online dictionary, free dictionary, French dictionary

Si des repérages ont déjà été proposés pour des répertoires collaboratifs du lexique “non conventionnel” du français (Dolar 2017, Murano 2017), il n’existait pas jusqu’à présent de vue d’ensemble de l’offre gratuite de dictionnaires généraux de cette langue, qui est d’une autre ampleur et puise à diverses sources. En cumulant plusieurs procédures, non détaillables ici, on pouvait en effet recenser, début 2019, environ 25 dictionnaires généraux en ligne gratuits du français moderne (pas toujours le plus récent), répartis sur une trentaine de sites (Corbin, Gasiglia 2020). Dans cet ensemble un peu modifié un an et demi plus tard et disparate à tous égards (origine, contenus, affichage, métadonnées), les répertoires qui vont être présentés – environ la moitié de l’offre globale – se signalent par une ou plusieurs de trois propriétés

combinables incitant à les consulter préférentiellement (ce à quoi contribue aussi leur référencement, la plupart sortant parmi les premières réponses à des requêtes lexicales effectuées avec Google) : l'évolutivité de leur interface, l'enrichissement de leurs contenus et leur qualité lexicographique.

1. Des dictionnaires institutionnels

Deux dictionnaires, différents à plusieurs égards mais dotés l'un et l'autre de paratextes éclairants, émanent d'institutions, partagent à ce titre une vocation de service public et sont consultables sur des sites sans publicité aux interfaces récentes.

1.1. La 9^e édition du Dictionnaire de l'Académie française

À l'achèvement, pas encore imminent, de sa rédaction, la 9^e édition du *Dictionnaire de l'Académie française* [DAF9] sera celle à la fois dont la publication sera la plus distante de celle de la précédente, la 8^e [DAF8] datant de 1935, et dont l'élaboration aura duré le plus longtemps, les premiers fascicules de sa version imprimée remontant à 1986.

Publié en volumes, dont les trois premiers parurent en 1992, 2000 et 2011, le DAF9 est aussi consultable en ligne depuis 2000 sur le site *atilf.fr*¹, qui permet des recherches transversales dans l'intégralité du texte, dont le visuel a été modernisé en 2019 et où la nomenclature, fin juillet 2020, reste arrêtée à *sabéisme*. On peut par contre, depuis février 2019, suivre l'avancée de la rédaction, qui, un an plus tard, a atteint *sérénissime*, sur le site *dictionnaire-academie.fr*², sans fonctions de recherche mais avec un affichage adapté aux smartphones (« Dossier de presse »³, p. 15), ce qui ancre dans la modernité un emblème de la tradition.

La durée de l'élaboration du DAF9, outre la lenteur propre à l'institution, tient au changement d'échelle et de mode rédactionnel par rapport à ses prédécesseurs. Presque double de celle du DAF8, sa nomenclature avoisinera 60 000 items⁴ (l'ordre de grandeur de celle du *Petit Robert* [PR] ou de la composante de langue du *Petit Larousse illustré* [PLI]) et les articles, préparés pour les académiciens par un service lexicographique et désormais dotés d'une structure arborescente combinant symboles alphanumériques et diacritiques, offrent des descriptions linguistiques étoffées et illustrées de nombreux exemples (forgés, selon la coutume).

Par les « Exemples de mots nouveaux » et les « Exemples de mots supprimés » que présente le site de l'Académie (boutons « Le Dictionnaire » puis « La 9^e édition »⁵), on mesure l'ampleur de la modernisation de la nomenclature, qui, avec le recul conforme au projet du dictionnaire, s'enrichit du lexique du XX^e siècle le mieux consolidé dans l'usage en couvrant largement sa deuxième moitié, dernière

1 <https://academie.atilf.fr/9/>.

2 <https://dictionnaire-academie.fr/> et, spécifiquement, <https://www.dictionnaire-academie.fr/nouveautes> pour la publication au *Journal officiel* du plus récent fascicule en février 2020.

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

4 <https://dictionnaire-academie.fr/presentation>, § « Les ressources éditoriales ».

5 <http://www.academie-francaise.fr/le-dictionnaire/la-9e-edition>.

décennie incluse (*négationnisme* (1990⁶), *euro* (1995), *exoplanète* (1998)). Une rubrique d'« Addenda » aux volumes parus, accessible par un bouton voisin⁷ et devant figurer à la fin du quatrième tome, atteste aussi cette visée d'actualisation. L'accueil de mots apparus au XXI^e siècle reste pourtant exceptionnel (*darmstadtium* (2003), *copernicium* (2010)), le *DAF* n'ayant pas vocation à capter les usages les plus récents, éventuellement éphémères. Depuis février 2019, néanmoins, ceux qui ont le statut de recommandations officielles sont évoqués indirectement via des liens de fin d'articles (« Voir aussi ») dirigeant vers ceux correspondants de la base *FranceTerme* (par exemple, s.v. **fixe** adjectif, « (Motocycle et cycle) **vélo à pignon fixe** » mène à l'article **vélo à pignon fixe** de celle-ci, qui date de 2013 la recommandation de cet item comme terme).

Par ailleurs, depuis les articles appropriés, d'autres liens ouvrent sur des préconisations de l'Académie elle-même (« Dire, ne pas dire »⁸, par exemple s.v. **espèce** (quatre occurrences)) ou sur l'espace francophone via la *Base de données lexicographiques panfrancophone* (par exemple s.v. **ambiance**).

Enfin, depuis décembre 2019, les neuf éditions du *DAF* sont consultables dans la même interface et accessibles, selon pertinence, depuis n'importe quel motif de requête, ce qui offre un espace inédit d'investigation d'environ 250 000 articles⁹.

1.2. Usito

Proposé en ligne sur abonnement payant depuis 2012, le dictionnaire *Usito* est venu enrichir l'offre gratuite en octobre 2019 sur le site *usito.usherbrooke.ca*¹⁰.

Élaboré à l'Université de Sherbrooke entre 2004 et 2011, *Usito*, dont la nomenclature, indexée dans la rubrique « Tous les mots du dictionnaire »¹¹, dépasse 73 000 items en juillet 2020, est un dictionnaire électronique natif¹² partageant avec le *DAF9*, outre une mission de service public (présentée comme didactique pour le répertoire québécois), un positionnement normatif revendiqué dans son paratexte¹³, qui affirme sa vocation à décrire le lexique français actuel neutre et soutenu du Québec en s'appuyant sur un corpus écrit de plus de 50 millions de mots et en le contrastant avec le lexique hexagonal de mêmes registres. Dans les articles, cette option se concrétise par des marques indiquant l'aire de pertinence des usages et par la spécification d'emplois critiqués ou recommandés au Québec ou en France. L'horizon évoqué d'une ouverture panfrancophone plus large¹⁴ n'est encore qu'un projet, puisque même les acadianismes ne sont traités que dans une liste indépendante de la nomenclature générale¹⁵, à l'instar

6 Les dates de première attestation mentionnées dans ce texte proviennent du *PR* en ligne et d'*Usito* (§ 1.2.).

7 <http://www.academie-francaise.fr/le-dictionnaire-la-9e-edition/addenda>.

8 <http://www.academie-francaise.fr/dire-ne-pas-dire>.

9 <https://www.dictionnaire-academie.fr/presentation>, § « Les neuf éditions du *Dictionnaire de l'Académie* ».

10 <https://usito.usherbrooke.ca/>.

11 <https://usito.usherbrooke.ca/index/mots#a>.

12 <https://www.usherbrooke.ca/usito/usito-cest/numerique-et-gratuit/>.

13 <https://www.usherbrooke.ca/usito/> et <https://www.usherbrooke.ca/usito/lorientation-normative/>.

14 <https://www.usherbrooke.ca/usito/usito-cest/louverture-a-la-francophonie/>.

15 <https://usito.usherbrooke.ca/index/lexiques/acadianismes#a>.

des belgicisms¹⁶ et des helvétismes¹⁷, qui sont confiés à des spécialistes européens extérieurs à l'équipe rédactionnelle.

68

La normativité d'*Usito*, qui vise les « emplois qui sont le plus négativement perçus dans la société québécoise (certains emprunts critiqués à l'anglais, sacres, expressions vulgaires, étiquettes discriminatoires ou racistes, etc.) »¹⁸, se manifeste par l'omission, ou par le traitement dans des annexes thématiques (par exemple « Les sacres en français québécois »¹⁹) vers lesquelles redirigent certains articles (**calice**, **tabernacle**...), ou, pour les « anglicismes critiqués »²⁰, par l'illustration exclusive, dans les articles dont ils sont les adresses, de l'emploi de leurs équivalents québécois recommandés (*s.v.* **big data**, exemplification de *mégadonnées*).

L'ancrage québécois d'*Usito* n'hypothèque pas son usage comme dictionnaire du français de France – dans les limites fixées à sa couverture lexicale, qui exclut « les emplois les plus marginaux (trop vieux ou trop récents ; trop locaux ; trop spécialisés) »²¹ –, puisque, une fois défalqués les québécismes et les items référant à des réalités canadiennes ou québécoises, et prises en compte les recommandations officielles propres au Québec (spécialement vis-à-vis d'anglicismes), l'ordre de grandeur des unités lexicales simples ou complexes et des emplois spécifiés dans l'index de la nomenclature est de 65 000 unités propres à l'Hexagone ou partagées par les deux communautés, ce qui est celui de dictionnaires imprimés de milieu de gamme.

Diverses dates de première attestation postérieures à 2000 témoignent de l'actualisation continue d'*Usito* (*tweet* (2008), *cougar* nom féminin et *hashtag* (2009), *bitcoin* et *vapoter* (2010), *cryptomonnaie* et *selfie* (2013), *maladie à coronavirus 2019* et *COVID-19* (2020)), qui s'effectue cependant pour partie dans des directions différentes de celle des dictionnaires français, en corrélation avec des écarts d'usage et de norme entre les deux communautés : *baladodiffuser* (2004) ou *microbillet* (2009), absents du *PR*, sont dans *Usito* ; *égoportrait* (2013), adresse dans celui-ci, n'est que mentionné comme privilégié au Québec *s.v.* **selfie** dans le *PR*.

Les descriptions linguistiques détaillées et illustrées de contextualisations forgées ou attestées (surtout dans des textes québécois) s'affichent dans une interface donnant à voir fonctionnellement la structure des articles par l'association de symboles alphanumériques, d'indentations et de contrastes de couleurs et par la dissociation du traitement des emplois des unités lexicales et de l'indication de leurs réseaux, ces derniers étant isolés, en plein écran, dans sa partie droite, où apparaît aussi la liste des items traités qui incluent le motif de requête.

2. Des dictionnaires d'éditeurs

Trois dictionnaires produits par les deux principaux éditeurs français de ce type d'ouvrages, ancrés dans l'imprimé et peu ou pas dotés de métadonnées, font aussi partie de l'offre gratuite en ligne.

16 <https://usito.usherbrooke.ca/index/lexiques/belgicisms#a>.

17 <https://usito.usherbrooke.ca/index/lexiques/helv%C3%A9tismes#a>.

18 <https://www.usherbrooke.ca/usito/orientation-normative/selection-des-emplois-decrits/>.

19 https://usito.usherbrooke.ca/articles/th%C3%A9matiques/vincent_1.

20 https://usito.usherbrooke.ca/index/anglicismes_critiqu%C3%A9s#a.

21 <https://www.usherbrooke.ca/usito/orientation-normative/les-mots-nouveaux-dans-usito/>.

2.1. Le Dictionnaire de français de Larousse

Parmi les ressources lexicographiques, encyclopédiques, pratiques, ludiques et documentaires gratuites du site *larousse.fr*²², le bouton « Langue française » donne accès à un *Dictionnaire de français* [*DdF-Lar*], ouvert aux publicités et liens commerciaux, concernant lequel les informations fournies sur sa page d'accueil depuis au moins début 2020 se limitent à « 135 000 définitions, 90 000 articles²³, 92 000 synonymes... »²⁴. Rien n'indique que la source du cœur de ce dictionnaire est le *Grand usuel Larousse* [*GUL*] de 1997, lui-même issu du *Grand Larousse en 5 volumes* de 1987, ni que ses actualisations répercutent celles du *PLI*, avec des variations de formulation ou de mise en forme, notamment pour des sigles (*RFID*, *RSA*), et des décalages dans le temps (par exemple pour le sens internautique de *troll* (2005), encore absent du *DdF-Lar* début août 2020), et ne sont donc pas à l'échelle du dictionnaire matrice.

Comme pour *Usito*, deux colonnes asymétriques scindent un plein écran, celle de droite listant les items correspondant au motif saisi tandis que celle de gauche, plus large, affiche les articles, dont les contenus sont distribuables sur six rubriques ordonnées (« Définitions », « Expressions », « Synonymes », « Homonymes », « Difficultés », « Citations ») enchaînées ou accessibles par des boutons²⁵. Cet ensemble est composite, les trois premières rubriques résultant d'une modularisation des articles du *GUL* avec suppression des indicateurs numériques de leurs subdivisions, alors que les trois autres proviennent de l'agglomération de ressources indépendantes, ce qui n'est pas non plus spécifié.

2.2. Le dictionnaire Robert de Google

Plus masquée encore est la source des articles de dictionnaire que Google est susceptible de proposer, depuis 2018, comme première réponse à des requêtes de forme *x définition*, dont rien n'indique qu'ils procèdent, avec des simplifications et des déstructurations, de ceux d'une évolution de la composante de langue du *Robert illustré* [*RI*] intégrée au logiciel de correction *Robert correcteur* [*RCorr*] (Gasiglia à paraître et 2020). La puissance de réponse potentielle du Robert de Google [*R-G*] devrait donc être sensiblement inférieure aux 160 000 définitions annoncées pour le *RI* jusqu'à son millésime 2017, ce qui lui laisse encore un bon avantage sur les dictionnaires en ligne les mieux référencés dont les liens vers les articles pertinents s'affichent à la suite. Encore faut-il, toutefois, que les requêtes s'ajustent à de possibles dégroupements de réponses : la saisie de *permettre définition* ne mène qu'aux emplois transitifs du verbe, l'accès aux emplois pronominaux requérant celle de *se* avant l'infinitif (alors que le *RI* incorporé au *RCorr* regroupe les deux emplois). Par ailleurs, à la différence des versions électroniques du *RI* en ligne ou en applications, le *R-G* ne mène pas aux articles à partir de formes fléchies.

22 <https://www.larousse.fr/>.

23 La nomenclature, consultable via un index accessible depuis le bas du site, est en fait d'environ 83 000 items.

24 Début 2019, elle évoquait en outre « 34 000 expressions », « 29 000 contraires », « 15 000 homonymes » et « 6 000 articles » éclairant des difficultés linguistiques.

25 L'activation conjointe des six rubriques s'observe, par exemple, s.v. **main**.

2.3. *Le Dico en ligne du Robert*

70

Depuis mi-avril 2020, Le Robert propose aussi gratuitement, sous son propre label et sur un site spécifique²⁶ animé par des blogs et des divertissements lexicaux, un autre avatar de la composante lexicale du RI, le *Dico en ligne* [DI], très proche de celle-ci et similairement actualisé (cf. article **covid**), mais avec une mise en forme différente et des écarts de sélection des items. Selon Jannièrre (2020), il comporterait « 170 000 mots » (ce qui inclut nécessairement une masse d'unités polylexicales) et « quelque 200 000 synonymes » (décompte d'occurrences et non de mots différents).

Les données sont distribuées en cinq rubriques enchaînées et accessibles sélectivement par les boutons « DÉFINITIONS », « CONJUGAISON », « SYNONYMES », « EXEMPLES » et « 17^e SIÈCLE », dont la mise en œuvre varie selon les items : effective pour toutes *s.v.* **courir**, mais seulement pour « DÉFINITIONS » *s.v.* **oganesson**. Les « SYNONYMES » émanent d'une ressource autre que le RI, les « EXEMPLES » sont des attestations « non révisées par Le Robert » qui proviennent de textes du domaine public ou de sites libres, et la reproduction d'articles du *Dictionnaire universel* de Furetière documente le « 17^e SIÈCLE ».

3. Des dictionnaires issus de l'ingénierie linguistique

Pour prendre pied sur le marché des logiciels de correction (cf. § 2.2.), Le Robert s'est associé avec les Éditions Diagonal, qui commercialisent en propre *Prolexis*, utilisé notamment dans la presse et l'édition, et le *Dictionnaire Myriade*²⁷. Deux autres acteurs français de l'ingénierie linguistique proposant des versions en ligne gratuites de dictionnaires qu'ils développent (sans donner accès à leurs nomenclatures) retiennent l'attention tant par les spécificités de ceux-ci que par leur dissémination sur des sites hébergeurs.

3.1. Le dictionnaire Cordial

Celui qui intègre le plus de mots récents est, sur *cordial.fr*, *Le dictionnaire Cordial*²⁸ [DCor], composante lexicographique du logiciel de correction *Cordial*, lui aussi bien diffusé, que produit Synapse Développement. À en juger par les écarts entre les valeurs affichées sur la page d'accueil du site début 2020 et en octobre 2018, ce dictionnaire a connu entretemps une actualisation importante, avec « 216 000 entrées lexicales » et « 310 000 définitions » contre « 150 000 » et « 190 000 ». Même s'il faut minorer ces valeurs du fait de dédoublements d'articles, par exemple pour des noms variant en genre (**youtubeur** n. m. et **youtubeuse** n. f.), la nomenclature s'est enrichie d'items entrés dans l'usage chaque année au moins jusqu'en 2013 (*selfie*) qui puisent dans divers registres, de *boloss* à *zythologie* en passant par *aquabiking* ou *mégadonnées*.

Les articles sont des séquences linéaires de rubriques titrées, qui, selon les cas, peuvent adjoindre aux définitions, numérotées même en cas de monosémie, des indicateurs d'origine et d'usage (certains étant

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

27 <https://www.prolexis.com/>. Diagonal a aussi développé l'interface du DAF9 (§ 1.1.) pour le site *dictionnaire-academie.fr* (<https://www.dictionnaire-academie.fr/credits>).

28 <https://www.cordial.fr/dictionnaire/>.

redoublés par des icônes), des expressions, des difficultés, des synonymes (distribués en sous-ensembles ni numérotés, ni typés), des contraires (*idem*), des attestations d'emplois (illustrant des cooccurrences) et des mots proches sémantiquement (malheureusement pas tous référés au même lemme pour des formes graphiques ambiguës : par exemple, *s.v. vente* nom, mots du champ sémantique du verbe *venter*).

Des versions antérieures du *DCor*, moins riches en entrées comme en données, sont hébergées, sous le nom de *Cordial dico*, par trois sites accueillant des publicités :

*dictionnaire.reverso.net*²⁹ : sans métadonnées ; actualisation ne paraissant pas postérieure à 2009 (*vuvuzela*) ; données limitées à des définitions et des synonymes et antonymes, affichés successivement dans des rectangles titrés indépendants (« *Dictionnaire Français Définition* » et « *Dictionnaire Français Synonyme* ») et assortis de marques d'usage ; numérotation des emplois seulement pour les items polysémiques et sans corrélation avec celle des synonymes et antonymes ;³⁰

*la-conjugaison.nouvelobs.com*³¹ : « **définitions** de plus de 150 000 mots de la langue française, leurs **synonymes** » ; actualisation arrêtée autour de 2005 (*churro*, *sudoku*) ; données de même nature (avec cependant de possibles listes d'expressions après les définitions, par exemple *s.v. fil*) et de même codification que sur *dictionnaire.reverso.net*, mais accessibles alternativement par les boutons « DÉFINITION » et « SYNONYME » ;

*universalis.fr*³² : « plus de **122 000 entrées** »³³ ; actualisation moindre que sur le site précédent, bien que l'horizon de datation soit similaire (*scroll* (2004) et dérivés, *churro*, mais pas *sudoku*) ; données limitées aux définitions (non numérotées) et expressions (numérotées), enchaînées et sans contraste typographique (par exemple *s.v. cœur*).

3.2. Le dictionnaire sensAgent

Si des versions du dictionnaire *sensAgent* [*sA*] sont également accessibles sur quatre sites, leur offre de consultation diffère de celle du *DCor* à plusieurs égards.

La plus ancienne, intitulée *Définitions et traductions avec Alexandria* [*sA-AI*], enfouie dans les profondeurs du site *tv5monde.com* et difficilement accessible si l'on ignore son URL³⁴, remonte à l'époque où son entreprise conceptrice s'appelait Memodata³⁵, avec un copyright 2003.

Aujourd'hui nommée Sensegates, la société, spécialisée dans le traitement automatique des langues et la sémantique lexicale³⁶, propose la consultation du *sA* sur un site propre doté d'un copyright

29 <https://dictionnaire.reverso.net/>.

30 L'ensemble de ces dispositions se rencontre, par exemple, *s.v. fixer*.

31 <https://la-conjugaison.nouvelobs.com/definition/>.

32 <https://www.universalis.fr/accueil/#dictionnaire-fr>.

33 Fenêtre surgissante à partir de <https://www.universalis.fr/accueil/#dictionnaire-fr>. L'accès alphabétique à la nomenclature se fait depuis le lien « Consulter le dictionnaire de l'Encyclopædia Universalis » en bas de l'écran.

34 <http://www.tv5monde.com/cms/chaine-francophone/outils/p-7550-Traducteur-Alexandria.htm?sl=fr&terme=accorderie-&tl=fr&ok.x=17&ok.y=10>.

35 <http://www.memodata.com/>.

36 <http://www.sensagent.com/common/online-2008/popup/PopUp-informationcompany.fr.html>.

2013³⁷, qui partage avec le *sA-Al* de faire accéder aux données unilingues par une interface de traduction plurilingue dans laquelle il faut sélectionner le français comme langue source et comme langue cible.

L'avatar unilingue de cette interface est hébergé par le journal *Le Parisien*³⁸ [*sA-Par*] avec un copyright 2000-2016. Cette version apparaît de façon privilégiée dans les réponses à des requêtes lexicales effectuées avec Google.

C'est en revanche une interface unilingue différente que propose à nouveau *tv5monde.com*, avec cette fois une meilleure visibilité, sous l'intitulé « Rechercher un mot dans le dictionnaire » au sein de la rubrique « Découvrir le français »³⁹ [*sA-tv5*], mais sans métadonnées identificatrices.

Plus dépouillée que les autres versions, *sA-tv5* limite ses données à des définitions (numérotées pour les items polysémiques) et des synonymes et est aussi moins actualisée, même si toutes ont en commun de ne pas l'être à la même échelle que le *DCor* (*selfie*, attesté depuis 2013, est une des entrées les plus récentes). En revanche, divers manques de *sA-Al*, *sA* et *sA-Par*, dont les contenus sont similaires⁴⁰ et s'affichent par défilement linéaire ou par sélection de boutons et dont les deux derniers ont la même feuille de styles, sont compensés par le recours aux mêmes ressources libres externes (le dictionnaire de Littré pour l'ancien, *Wikipédia* pour le récent).

Peu remarquable en tant que répertoire conventionnel, le *sA* a pour singularité, sur ces trois portails, d'être accompagné d'une ontologie lexicale dite « dictionnaire analogique », charpente de l'outillage taliste de Memodata / Sensegates au titre de laquelle était crédité *WordNet* en 2003 pour le *sA-Al*, qui place les emplois des items sur les nœuds d'arborescences complexes et non explicitées, et qui peut accueillir des mots récents ignorés du dictionnaire classique (*locavorisme* (pas antérieur à 2007), *youtubeur* (2011)).

4. Un dictionnaire d'origine indéterminée

D'autres sites d'information que *nouvelobs.com* (§ 3.1.) intègrent des dictionnaires, mais, à la différence de la version du *DCor* qu'héberge celui-ci et faute de métadonnées, leur origine n'a pas été identifiée. L'un d'eux, le *Dictionnaire français sur linternaute.com*⁴¹ [*DF-LInt*], accessible aussi depuis *lefigaro.fr*⁴² du fait de l'appartenance de L'Internaute au Groupe Figaro via le CCM Benchmark Group (cf. Piquard 2015), se signale par son enrichissement continu et certaines possibilités d'exploration transversale.

Évaluable à environ 83 000 unités à partir de l'index de chaque lettre ou du bouton « Abécédaire » sous la fenêtre de saisie, sa nomenclature accueille régulièrement de nouveaux items de diverses natures, certains des plus récents (2016) relevant du lexique chimique (*moscovium*, *nihonium*, *oganesson*, *tennessine*).

37 <http://traduction.sensagent.com/>.

38 <http://dictionnaire.sensagent.leparisien.fr/index.html>.

39 <https://langue-francaise.tv5monde.com/decouvrir/dictionnaire>.

40 Une des plus récentes illustrations en est l'intégration commune de *covid-19* (2020).

41 <https://www.linternaute.fr/dictionnaire/fr/>.

42 <https://www.lefigaro.fr/culture>, menu « Langue française », bouton « Dictionnaire ».

Une consultation thématique est permise par trois regroupements de sous-ensembles d'adresses accessibles, en haut de la page d'accueil et en bas des pages consultées, à partir des boutons « Les thèmes » (pour 87 domaines d'activité), « Les usages » (pour 8 registres et les emplois figurés) et « Types de mots » (pour 51 catégories grammaticales).

La consultation par saisie de motifs est, elle, compliquée par l'interclassement, dans les résultats des requêtes, des liens vers les articles avec d'autres qui mènent à des pages non pertinentes du site et avec des annonces de Google.

Dans les articles, qui s'affichent sous le bouton « Définitions », les données se succèdent au gré de l'activation variable des rubriques « Féminin » et « Pluriel » (pour les flexions), « Sens » (toujours renseignée, avec numérotation des emplois même en cas de monosémie), « Exemple » (pas systématique), « Synonymes » et « Contraires » (contingentes), « Étymologie » (rare) et « Traduction anglais » (occasionnelle)⁴³. Trois autres boutons optionnels, « Synonymes », « Conjugaisons » et « Expressions », donnent accès à des informations indépendantes, les synonymes étant simplement listés alphabétiquement, tandis que les expressions sont décrites dans les champs « Sens » et « Origine ».

La rédaction ne suit pas un protocole strict, ce qui, concernant les « Sens », se traduit par des oscillations entre définition des mots et caractérisation des référents (**nihonium** : « Élément chimique ... » vs **livermorium** : « Le Livermorium est un élément chimique ... ») et par des associations libres des deux types de discours (**chikungunya** : « Maladie infectieuse tropicale qui se transmet à l'être humain par la piqûre de moustiques que l'on retrouve surtout en Afrique et en Asie. Les douleurs articulaires sont les premiers symptômes de la maladie. Dans certains cas, le chikungunya peut entraîner la mort, particulièrement chez les personnes faibles, personnes âgées et nouveau-nés. »). On observe par ailleurs divers indices d'un déficit de contrôle rédactionnel (**moscovium** : « Moscovium est un élément chimique... » ; **slim** adjectif : « Ce dit d'un pantalon... » ; **youtubeur** : « Pluriel youtubeuse » ; **antépisode** : « L'histoire de l'antépisode précède donc à l'autre. »).

5. Des dictionnaires collaboratifs

Internet a fait naître des dictionnaires reposant sur l'accumulation d'articles de rédacteurs spontanés. Selon le statut des productions individuelles, ils peuvent être dits « collaboratifs », « contributifs » ou « participatifs ». Tous n'ont pas la même dynamique : certains stagnent sur leur compilation initiale de données lexicographiques libres, donc anciennes ; deux émergent, avec des profils différents.

5.1. Le Dictionnaire collaboratif de Reverso

Outre la version du *DCor* décrite au § 3.1., le site *dictionnaire.reverso.net* accueille un *Dictionnaire collaboratif*⁴⁴ [*DColl-Rev*], sans autres métadonnées que l'indication du nombre de données cumulées pour les diverses langues documentées⁴⁵ et qui, pour le lexique le plus récent, se révèle plus riche, avec, année après année, des propositions de définitions, sonorisées à l'intention des déficients visuels et

43 La consultation des articles **cheval** et **gentrification** donne une vue d'ensemble du dispositif microstructurel.

44 <https://dictionnaire.reverso.net/CollabDict.aspx?srcLang=-1&targLang=-1&lang=fr>.

45 « 2 728 569 Résultats » le 1^{er} août 2020.

accompagnées le cas échéant de marques d'usage, pour des items apparus jusqu'en 2016 (*nihonium*, *oganesson*).

74 En associant ainsi deux dictionnaires de genèse différente, *dictionnaire.reverso.net*, bien placé parmi les réponses aux requêtes de définitions sur Google, offre, dans les limites d'articles minimalistes, une couverture lexicale importante pour les usages contemporains.

5.2. Le Wiktionnaire

De facture très différente, le *Wiktionnaire* [*Wk*] sur *fr.wiktionary.org*, branche française du projet *Wiktionary* de la floridienne Wikimedia Foundation⁴⁶, est d'une autre ampleur à tous égards, à commencer par la documentation accessible à partir de sa page d'accueil⁴⁷, en y lisant la « Présentation du projet » ou en activant les liens « À propos du Wiktionnaire » en bas de l'écran (pour aller à la page « Wiktionnaire:À propos ») ou « **3 852 100 pages** »⁴⁸ en haut à droite (pour consulter « Wiktionnaire:Statistiques »).

Cette dernière page, qui pointe l'avancement des articles pour l'ensemble des langues prises en compte (potentiellement toutes), fait état, pour le français, de 385 968 lemmes (incluant locutions et noms propres) le 1^{er} août 2020⁴⁹, ce qui se traduit notamment par l'accueil de mots récents de divers registres le plus ouvert de l'ensemble des dictionnaires généraux de français gratuits en ligne.

L'organisation des articles, classique, peut mobiliser une vingtaine de rubriques étiquetées (cf. n. 46). La rédaction suit des protocoles réglés, spécialement pour les définitions, et les emplois décrits sont, autant que possible, documentés par des attestations prises à diverses sources et dûment référencées.

Le *Wk* a un avatar revendiqué sis en Finlande et apparaissant peu dans les réponses aux requêtes, le *Webtionnaire* [*Wb*] sur *webtionnaire.fr*⁵⁰, interface traductionnelle plurilingue qui, pour le français, avec un décalage d'actualisation⁵¹ et une mise en forme différente, reprend avec des adaptations, notamment pour les contextualisations, le cœur des articles de sa source (définitions, exemples, flexions, catégorisations diverses, synonymes, prononciation, dans cet ordre).

6. Un dictionnaire de didacticiens

Également ignoré des réponses aux requêtes de définitions, parce que produit à la KU Leuven pour des apprenants néerlandophones, le *Dictionnaire d'apprentissage du français langue étrangère et seconde* [*DAFLES*], absorbé dans l'*Interactive Language Toolbox* sur *ilt.kuleuven.be/inlato*⁵², a plusieurs caractéristiques originales :

46 <https://fr.wikipedia.org/wiki/Wiktionnaire>.

47 https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil.

48 Valeur du 1^{er} août 2020.

49 Pour comparaison, il y en avait 371 019 le 20 mars 2019 et 380 448 le 20 février 2020.

50 <http://webtionnaire.fr/>.

51 À la différence du *Wk*, le *Wb* n'a pas encore intégré *Covid* et *Covid-19* le 1^{er} août 2020.

52 <https://ilt.kuleuven.be/inlato/>.

une nomenclature limitée aux 5 000 lemmes et 6 731 vocables français jugés les plus utiles, au début des années 2000, sur la base de l'exploration d'un corpus de presse franco-belge d'environ 50 millions de mots⁵³ ;

une rédaction des articles inspirée par des modèles linguistiques (lexique-grammaire, théorie sens-texte) et lexicographiques (*Collins COBUILD*) contemporains ;

une interface de consultation qui permet d'accéder, via sa base de données sous-jacente, à des articles intégraux ou à toute une gamme d'informations particulières, et qui, pour des compléments, oriente vers divers sites ;

et des métadonnées dispersées dans les publications savantes de ses concepteurs, dont le site ne fait pas état.

7. Quelles structurations pour les dictionnaires à l'ère électronique ?

Le tour d'horizon qui vient d'être présenté permet de prendre la mesure des incidences de la bascule de l'imprimé vers l'informatique, pour l'élaboration de dictionnaires comme pour leur diffusion, sur les structurations de leurs nomenclatures et de leurs contenus informationnels.

7.1. Structuration des nomenclatures

La composition des nomenclatures des dictionnaires est la résultante de deux forces : les moyens dont on dispose pour élaborer ces répertoires et les finalités qu'on leur assigne.

Comme toute industrie, la lexicographie imprimée commerciale cherche à générer des profits en diversifiant ses produits à destination de publics plus ou moins précisément cernés, et les catalogues des éditeurs se structurent de ce fait en segments variablement concurrentiels (dictionnaires de référence, pratiques de formats échelonnés, scolaires pour différents niveaux...) dont les nomenclatures sont calibrées, avec des variations, en fonction de principes partagés : pour l'essentiel, sélection des items par ordre décroissant d'usualité (apprécié intuitivement et/ou sur corpus), pondération des technicisms entre les différents domaines d'activité et de connaissance, limitation des diaphasmes, diastratismes, diatopismes et diachronismes dans les répertoires à visée didactique, infléchissement vers le lexique des disciplines enseignées dans ceux pour l'école et le collègue. Certes, les équilibres instaurés à partir de ces préceptes lors de l'élaboration des dictionnaires peuvent être marginalement altérés, pour certains d'entre eux, par les petites doses d'actualisations annuelles surfant sur les néologismes et par les ouvertures à des francophonismes de diverses aires, pas assez nombreux pour les locuteurs de celles-ci tout en étant trop pour ceux de France (cf. Corbin, Gasiglia 2010 : 47-49), mais on peut considérer que, foncièrement, les nomenclatures des répertoires contenus dans les limites physiques de volumes imprimés répondent à des critères de structuration.

Il en va autrement pour les dictionnaires gratuits en ligne. Les notions de catalogue, gamme ou segment n'y sont pas pertinentes, aucun fournisseur de répertoire n'en proposant plus d'un et aucun destinataire n'étant, sauf pour le *DAFLES*, expressément désigné. Seulement moins de la moitié de ceux qui ont été décrits ont une nomenclature structurée : le *R-G* et le *DL*, qui émanent, avec des évolutions,

⁵³ Cf. Verlinde, Selva & Binon (2009 : 219).

du *RI*, dictionnaire encyclopédique imprimé de référence du Robert (§§ 2.2. et 2.3.) ; le *DAF9*, corrélat électronique d'un dictionnaire imprimé en cours d'achèvement qui s'inscrit dans la tradition académique d'enregistrement d'un français consolidé "de bon aloi" (§ 1.1.) ; *Usito*, dictionnaire électronique natif qui, pour le français du Québec, partage la visée prescriptive des *DAF* et prend appui sur un corpus (§ 1.2.) ; et le *DAFLES*, également pur répertoire électronique informé par un corpus, limité au lexique français le plus usuel (§ 6.). Et tous les autres, sur des bases et avec des moyens différents, proposent des nomenclatures cumulatives plus ou moins hétérogènes, enrichies à des vitesses variables : le *DdF-Lar* actualisé à partir du *PLI*, d'échelle différente (§ 2.1.) ; *DCor*, *sA* et *DF-LInt*, dont les sources sont inconnues (§§ 3. et 4.) ; et *DColl-Rev* et *Wk*, collaboratifs (pour ce dernier après absorption de dictionnaires anciens libres), donc aléatoires (§ 5.).

7.2. Structuration des contenus

Les contenus dictionnaires qui s'affichent sur Internet reçoivent une codification, qui peut s'ajouter à des balisages plus spécifiques des documents sources et qui se matérialise en partie dans les styles adoptés pour différencier des types de données. Pour autant, ils n'ont pas nécessairement une structuration lexicographique cohérente. Si celle-ci est avérée pour le *DAF9* et *Usito* (§ 1.), des dérives s'observent sur d'autres sites, qu'il s'agisse, sans exhaustivité, de la perte de la cohérence linguistique d'articles dans le *R-G* par la compression de la profondeur structurelle de leurs sources (§ 2.2.), ou, dans le *DCor*, d'amalgames de réponses pour des formes lexicales sans discrimination de leurs lemmes (§ 3.1.), ou de l'agglutination de données provenant de sources hétérogènes, comme dans le *DdF-Lar* et le *DI* (§§ 2.1. et 2.3.), ou encore de l'association de traductions aux contenus métalinguistiques unilingues, comme dans le *DF-LInt* (§ 4.). De proche en proche, sur Internet, c'est toute la conception classique du dictionnaire comme texte structuré cohérent qui vacille.⁵⁴

Références bibliographiques

Dictionnaires, encyclopédies, bases et logiciels

Base de données lexicographiques panfrancophone, <http://www.bdlp.org/>.

Collins COBUILD English Language Dictionary (1987). London / Glasgow: Collins.

Cordial 21 Pro / Référence (2016). Logiciels téléchargeables pour PC. Toulouse: Synapse Développement.

DAF8: Dictionnaire de l'Académie française. 8^e éd. 2 vol. (1932–1935). Paris: Librairie Hachette.

DAF9: Dictionnaire de l'Académie française. 9^e éd. 3 vol. (1992, 2000, 2011). Paris: Imprimerie Nationale Éditions (et Librairie Arthème Fayard pour les tt. 2 et 3).

Dictionnaire Myriade langue française. Gamme de dictionnaires pour Mac et PC. Sophia Antipolis: Éditions Diagonal.

FranceTerme. <http://www.culture.fr/franceterme>.

Furetière, Antoine (1690) *Dictionnaire universel* [...]. 2 vol. La Haye / Rotterdam: Arnout & Reinier Leers.

Grand Larousse en 5 volumes (1987). Paris: Larousse.

⁵⁴ Merci à Nathalie Gasiglia pour son assistance technique et sa relecture critique.

- GUL*: *Grand usuel Larousse* (1997). 5 vol. Paris: Larousse.
- Littré, Émile (1863–1873) *Dictionnaire de la langue française*. 4 vol. Paris: Librairie Hachette et C^{ie}.
- PLI*: *Le Petit Larousse illustré 2021* (2020). Paris: Larousse.
- PR*: *Le Petit Robert de la langue française*. Éd. 2021 (2020). Paris: Le Robert. – En ligne sur abonnement <https://www.lerobert.com/dictionnaires/francais/langue/dictionnaire-le-petit-robert-de-la-langue-francaise-abonnement-annuel-3133099010272.html>.
- Prolexis 7*. Gamme de logiciels de correction pour Mac et PC. Sophia Antipolis: Éditions Diagonal.
- RCorr*: *Le Robert correcteur*. Logiciel téléchargeable pour Mac et PC. Nouv. Version (2016). Paris: Le Robert.
- RI*: *Le Robert illustré 2017* (2016). Paris: Le Robert. – Abonnement en ligne gratuit avec l'achat de l'imprimé.
– Application *Dixel mobile* (2009–2014). Paris: Le Robert / Sophia Antipolis: Éditions Diagonal.
– Application *Le Robert mobile* (2015–2019). Paris: Le Robert / Sophia Antipolis: Éditions Diagonal.
- Wikipédia*. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal.
- WordNet*. <https://wordnet.princeton.edu/>.

Autres références

- Corbin, Pierre, Gasiglia Nathalie (2010) “Représentations lexicographiques actuelles du lexique français: aperçus sur les choix de nomenclatures.” [In:] *Cahiers de lexicologie*. Vol. 96; 35–65.
- Corbin, Pierre, Gasiglia Nathalie (2020) “Les dictionnaires monolingues généraux du français ‘actuel’ gratuits en ligne (début 2019).” [In:] *Congrès mondial de linguistique française - CMLF 2020. SHS Web of Conferences*. Vol. 78; 20 p., https://www.shsconferences.org/articles/shsconf/pdf/2020/06/shsconf_cmlf2020_05008.pdf (consulté le 10/09/2020).
- Dolar, Kaja (2017) “Les dictionnaires collaboratifs non institutionnels dans l’espace francophone: éléments de typologie et bilan.” [In:] *Repères-Dorif*. Vol. 14, http://www.dorif.it/ezone/ezone_articles.php?art_id=380 (consulté le 31/07/2020).
- Gasiglia, Nathalie (à paraître) “Identification des articles dictionnaires non référencés fournis par Google.” [In:] *Études de linguistique appliquée*.
- Gasiglia, Nathalie (2020) “Étude généalogique d’articles dictionnaires affichés par Google.” [In:] *Academic Journal of Modern Philology*. Vol. 9; 91-105.
- Jannière, Virginie (2020) “Le Robert lance son dictionnaire en ligne gratuit.” [In:] *CNews*. 17 avril 2020, <https://www.cnews.fr/culture/2020-04-17/le-robert-lance-son-dictionnaire-en-ligne-gratuit-947936> (consulté le 31/07/2020).
- Murano, Michaela (2017) “Une lexicographie deux fois populaire: les dictionnaires collaboratifs du français ‘non conventionnel’.” [In:] *Repères-Dorif*. Vol. 14, http://www.dorif.it/ezone/show_issue.php?iss_id=25 (consulté le 31/07/2020).
- Piquard, Alexandre (2015) “Le Figaro frappe un grand coup dans le numérique.” [In:] *Le Monde*. 1^{er} octobre 2015. https://www.lemonde.fr/entreprises/article/2015/10/01/le-figaro-rachete-le-groupe-internet-ccm-benchmark_4780530_1656994.html (consulté le 31/07/2020).
- Verlinde, Serge, Selva Thierry & Binon Jean (2009) “Les bases de données au service d’un dictionnaire d’(auto-)apprentissage pour allophones.” [In:] *Lexique*. Vol. 19; 217–233.

