

**L'EXTENSION DU *GRAND CORPUS DES GRAMMAIRES*
FRANÇAISES, DES REMARQUES ET DES TRAITÉS SUR LA LANGUE :
QUESTIONS THÉORIQUES ET MÉTHODOLOGIQUES**

Wendy Ayres-Bennett & Bernard Colombat

Université de Cambridge & Université Paris Diderot (UMR 7597)

Résumé

Cet article envisage l'extension du *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue* (2011) à un ensemble plus vaste de textes couvrant les xvii^e et xviii^e siècles. Après une présentation du corpus existant, il évoque les principes mis en œuvre pour sa continuation, en analysant notamment les critères de sélection pour les choix des textes, sans oublier les difficultés que l'entreprise soulève (découpage temporel, hétérogénéité des textes, mise en œuvre de l'outillage).

Mots-clés

corpus, grammaire, remarques sur la langue, français, xvii^e s., xviii^e s.

Abstract

In this article we discuss how the *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue* (2011) will be expanded by the addition of a large number of seventeenth- and eighteenth-century texts. Following a brief presentation of the existing corpus, the principles governing the second phase are outlined, notably the criteria used for selecting the new texts, as well as the difficulties associated with such an enterprise (choice of temporal limits; heterogeneity of the texts; selection of search tools, etc.).

Keywords

corpus, grammar, remarks on the French language, French, seventeenth century, eighteenth century

1. INTRODUCTION

Les critères pour définir un corpus ont été examinés pendant plus de vingt ans (voir, par ex., Biber 1993 ; Sinclair 1996 ; Habert, Nazarenko & Salem 1997). Citons la définition de Sinclair (2004) :

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.

Dans cette définition figurent deux termes clés : sélection (“selected”) et représentativité (“to represent”). Autrement dit, « c’est l’opération de choix raisonné parmi les composants disponibles qui crée un corpus » (Habert 2000, p. 13). À ces critères Sinclair (2004) ajoute ceux d’« homogénéité » et d’« équilibre ». Jusqu’ici la plupart des études se sont concentrées sur les critères pour la construction d’un corpus de données linguistiques qui représenterait les usages d’une langue ou variété linguistique. Dans cet article nous présentons plutôt les critères adoptés pour la construction d’un corpus de textes métalinguistiques et en particulier de grammaires françaises et de volumes de remarques sur la langue française.

En 2011 est parue chez Classiques Garnier Numérique la première tranche du *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue (XIV^e-XVII^e siècles)* ; elle comprend trois volets : Grammaires françaises de la Renaissance, Grammaires françaises du XVII^e s., et Remarques sur la langue française (XVII^e s.). Grâce à une licence nationale, le corpus est disponible (depuis avril 2014) dans toutes les bibliothèques publiques et universitaires en France. Cette première tranche contient 48 textes, soit 3,36 millions de mots ou environ 15000 pages (voir le Tableau 1).

TABLEAU 1
La première tranche du *Grand Corpus*

	GRAMMAIRES RENAISSANCE	GRAMMAIRES XVII ^e S.	REMARQUES	TOTAL
Auteurs	17	12	11	40
Textes	21	12	15	48
Mots	961.289	884.644	1.513.389	3.359.322
Pourcentage du total	28.6	26.3	45	
Signes (total)	10.518.439	10.306.165	14.878.421	35.703.025
Signes (hors balises)	6.073.013	6.079.591	10.151.461	22.304.065
Balises	659.537	627.581	701.805	1.988.923
Pages	3383	3927	7741	15.051
Moyenne	161 p./ouvrage	327 p./ouvrage	516 p./ouvrage	313 p./ouvrage

Fortement outillée et balisée selon des critères spécifiques, elle permet des recherches complexes dans des domaines comme l'histoire des théories linguistiques, l'histoire de la langue ou, plus généralement, l'histoire culturelle. La « recherche simple » permet d'interroger le corpus en plein texte, en filtrant éventuellement la recherche avec des critères relevant des données bibliographiques ; l'emploi d'opérateurs, de la fonction « recherche floue » et des jokers permet de combiner les critères de recherche et de trouver les variantes. La « recherche avancée » mobilise en revanche des critères complexes qui permettent de recueillir les données en fonction de trois axes : « Auteurs et titres », « Exemples », « Citations ».

Une deuxième tranche de 49 textes est actuellement en préparation par les deux auteurs de cet article. Elle ajoutera les grammaires du XVIII^e siècle et étoffera le corpus des volumes de remarques en y joignant d'autres volumes publiés au XVII^e siècle et en étendant le corpus au XVIII^e siècle (voir l'Annexe). Pour illustrer l'intérêt apporté par les nouveaux textes, on retiendra trois points.

1.1. Le renouvellement du traitement des parties du discours dans les grammaires

La division en parties du discours est une constante, on peut même dire une ossature de la tradition occidentale : on la retrouve dans la plupart des grammaires françaises, mais, à l'époque considérée, elle connaît des variations considérables dans :

- leur nombre : des trois de Buffier (noms, verbes, modificatifs) aux neuf traditionnelles, à laquelle vient s'ajouter l'adjectif, s'il est séparé du nom ; Wailly parle des particules *ou* interjections et Girard subordonne les secondes aux premières, constituant spécifiquement le nombre comme septième partie d'oraison ; Beauzée place l'article dans la catégorie de l'adjectif et regroupe prépositions et adverbes sous la catégorie commune des mots supplétifs ;
- leur ordre de traitement : l'article peut précéder le nom (Régner-Desmarais) – ou le substantif (Girard) –, mais aussi le suivre (Restaut) ; rangé dans la catégorie des adjectifs, il peut même suivre à la fois le nom et le pronom (Beauzée) ; l'ordre peut faire l'objet d'une numérotation explicite, comme chez Girard ;
- la forme de leur traitement : les parties du discours peuvent faire l'objet d'un chapitre spécifique avec leurs sous-catégorie qui constitue alors une section ; mais les catégories peuvent faire l'objet d'un chapitre spécifique qui vient s'insérer dans l'économie générale de l'ouvrage (ainsi chez Condillac 1775 et 1782) ;

- leur dénomination : Beauzée distingue explicitement les « éléments de l’oraison » (nos parties du discours) des éléments de la syntaxe, préférant dire dans l’*Encyclopédie méthodique* que « les parties du discours sont les mêmes que celles de la pensée » ; s’il choisit de faire de l’article une variété d’adjectif, il renonce à bouleverser la tradition en s’abstenant de le renommer ; le plus obscur Fauleau divise les parties du discours en parties *nominales* (nom, article, pronom), en parties *verbales* (verbe, préposition, adverbe), et en parties *succursales* (conjonction, particule, interjection).

1.2. La richesse terminologique

La richesse terminologique est une constante des grammaires de cette période. Pensons par ex. aux « adjectifs physiques / métaphysiques » de Du Marsais réinterprétés par Beauzée, encore plus à Girard et à ses fonctions exprimées le plus souvent par des termes en *-ifs* : *subjectif*, *attributif*, *objectif*, *terminatif*, *circonstanciel*, *conjonctif*, *adjonctif*. Mais il y a d’autres exemples, comme celui, moins connu, de Vallange :

Prénoms pour articles ; *prénoms apostrophés* pour articles élidés ; *gradation grammaticale* pour comparatif et superlatif ; *noms appellatifs* ; modes *absolu* et *relatif* ; *modaux* (i.e. adverbess de manière) ; *postpositif* pour génitif ; *modes invocatif*, *imprécatif*, *plaintif* ; *suprême* (différentes formes pour l’expression du haut degré). Une terminologie phonétique et graphique : *solissone* réfère aux voyelles, et *bissone*, *trissone* portent respectivement sur les groupes de deux ou de trois « solissons » (voyelles graphiques) qui ne font ensemble qu’un son. (CTLF, notice 2403 : J.-M. Fournier, rév. M. Lecolle)

La base va nous fournir une quantité de métatermes linguistiques spécifiques. Pensons cette diversité non comme une hétérogénéité, mais comme une richesse partagée, qu’une base de donnée permet d’exploiter... à condition de savoir comment chercher (cf. *infra* pour les outils).

1.3. Un nouvel éclairage du point de vue des remarques

Comme nous le verrons (2.2), l’élargissement du corpus nous oblige de repenser la définition de différents genres, et on remarque surtout que les distinctions entre grammaires / remarques / dictionnaires / journaux linguistiques ne sont pas étanches. Cet élargissement entraîne également la considération de nouveaux publics, qu’il s’agisse d’étrangers ou de francophones hors de France.

L’ajout d’une deuxième tranche de textes soulève des questions méthodologiques et théoriques importantes. Dans la section suivante nous traiterons de la question du choix des textes, et dans la section 3 nous considérons quelques conséquences de la création d’un tel corpus.

2. CRITÈRES POUR LE CHOIX DES TEXTES

À mesure que l'on avance dans le temps, le choix des textes devient plus complexe pour plusieurs raisons – le découpage temporel (2.1), l'hétérogénéité de la matière, de la taille et des supports (2.2), le nombre d'ouvrages à retenir (2.3) et l'hétérogénéité dans la notoriété des auteurs et la diffusion des ouvrages (2.4).

2.1. *Le découpage temporel*

On sait que tout découpage chronologique est nécessairement arbitraire. Mais il est commode. Il y a d'abord les périodes définies par rapport à une notion évaluative, « âge classique », « Renaissance », « Lumières », pour ne pas parler des « siècles obscurs », etc. L'intitulé « Grammaires françaises de la Renaissance » (cf. la base existante) est commode, la Renaissance étant une période malléable qui obéit à des temporalités différentes (elle commence par ex. en Italie bien avant de s'installer en France), en tout cas elle est assez vaste pour enfermer un avant et un après *xvi*^e siècle.

A priori la division en « siècle » semble plus simple, même si elle est la plus arbitraire. En fait, elle ne pose guère de problème pour la grammaire du *xvii*^e siècle et pour les remarqueurs ; notons, par exemple, que le *Journal* de Domergue continue à être publié de 1784 à 1792, malgré le commencement de la Révolution française. La situation est néanmoins différente pour les grammaires du *xviii*^e siècle : les conséquences des événements politiques sur le plan social, institutionnel et intellectuel (création des Écoles centrales, institution de la grammaire générale comme discipline d'enseignement, mouvement des Idéologues, etc.) enjambent la fin du *xviii*^e et le début du *xix*^e siècle. Il fallait choisir. La décision a été prise : on s'arrêtera avant. S'arrêter après, dans les années 1810, était possible aussi, peut-être meilleur, mais cela nous posait des problèmes à la fois théoriques (selon quels principes sélectionner les ouvrages du *xix*^e siècle, époque à laquelle la production grammaticale explose ?) et (bêtement) techniques, du fait du traitement d'un corpus devenu trop gros, sachant que l'élaboration de la phase précédente avait nécessité une douzaine d'années. Nous nous arrêtons donc à 1789 (l'ouvrage de Bauchaint date précisément de cette année), à l'exception de deux ouvrages au statut un peu particulier : ceux de Harris et une des œuvres de Domergue (cf. *infra*).

2.2. *Définition du genre vs hétérogénéités*

L'hétérogénéité de la matière

En ce qui concerne les volumes de remarques et d'observations sur la langue française, pour la première tranche nous nous sommes concentrés sur les textes les plus 'classiques', tels les *Remarques sur la langue françoise* de Vaugelas (1647) ou les *Remarques Nouvelles* de Bouhours (1692 [1675]), les commentaires sur

Vaugelas, de nature plutôt positive ou négative, et quelques textes d'un intérêt particulier, tel le volume de Macé (1651) qui rassemble une « méthode générale et raisonnée », un traité sur les difficultés orthographiques et une compilation des *Remarques* de Vaugelas avec les commentaires de La Mothe Le Vayer et d'un auteur anonyme. L'extension de la base soulève nécessairement une réflexion sur la définition du genre. Dans nos publications précédentes (voir, par ex., Ayres-Bennett et Seijido 2011, p. 47-54), nous avons identifié quatre caractéristiques des *Remarques* de Vaugelas :

- Du point de vue du *contenu*, le point de départ est un doute sur une question de bon usage ou l'examen de l'usage douteux.
- Du point de vue de la *forme*, les remarques sont le plus souvent assez courtes : elles s'enchaînent pêle-mêle sans ordre logique.
- Le *public* implicitement désigné est constitué d'honnêtes gens français qui n'aiment pas le pédantisme.
- *L'idéologie* sous-jacente du genre tend à être normative, mais la variation sociolinguistique est également prise en compte.

Il est pourtant évident que même au xvii^e siècle les remarqueurs ultérieurs n'adoptent pas tous ces caractéristiques, et la question de l'homogénéité du genre devient plus épineuse plus on avance dans le temps. En ce qui concerne le xvii^e siècle, nous pourrions regrouper les quinze textes supplémentaires en quatre grands types : (a) des textes qui représentent ce qu'on pourrait appeler la « préhistoire » du genre : Wapy 1634, La Mothe Le Vayer¹ 1638, Dupleix 1645 ; (b) six volumes de type plus ou moins classique¹ : Aisy 1685, Bérain 1675, Bouhours 1674, Cassagne 1690, Leven de Templéry 1698, Morvan de Bellegarde 1695 ; (c) des textes destinés à un public particulier : Poullain de la Barre 1691 pour les Genevois, Saint-Maurice 1672 principalement pour les étrangers ; (d) des textes où l'on voit les mêmes préoccupations mais qui se distinguent des textes classiques de sorte qu'ils mettent en question la définition du genre : anon. 1657, Callières 1693a, 1693b, Renaud 1697.

Prenons le cas du texte anonyme publié en 1657. Il s'agit d'une grammaire destinée seulement aux Français « qui ont quelque première teinture de la pureté, & de la netteté de notre langue », qui intègre les remarques et reprend les grands débats des remarqueurs précédents, et adopte notamment les « belles Remarques de M. de Vaugelas » (Au lecteur). Le chapitre intitulé « Explication de certains Noms, plus difficiles dans le divers usage qu'ils ont en notre langue » (Image 1) illustre bien que la distinction entre grammaires et remarques n'est pas étanche à cette époque.

1 Notons, pourtant, que Bouhours (1674) et Leven de Templery (1698) organisent leur contenu en sections thématiques, que Jean d'Aisy (1685) commence son ouvrage par un abrégé de grammaire et organise les remarques en onze sections, dont huit correspondent aux parties du discours.

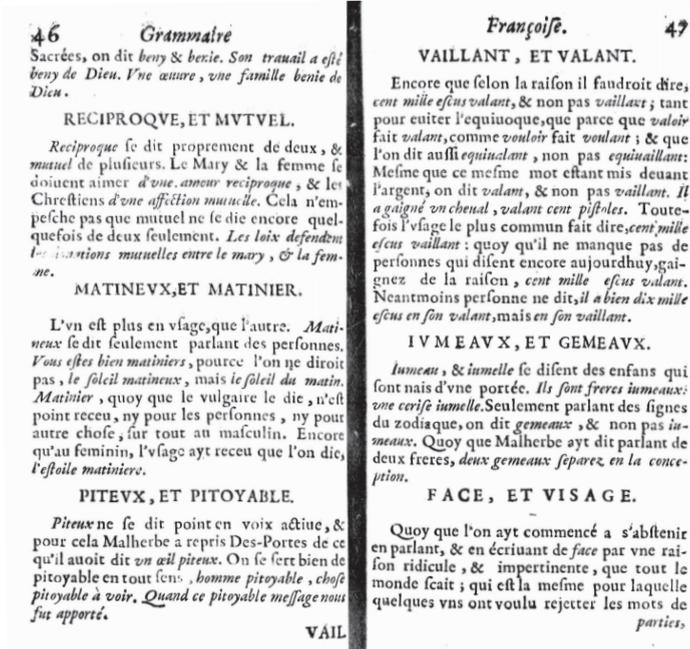


Image 1. Anonyme de 1657

Autre exemple, les deux volumes de Callières qui illustrent l'intégration des remarques sur la langue français sous forme de conversation; le deuxième volume (1693b) est surtout intéressant car il inclut la discussion de façons de parler bourgeoises (Image 2).

Quant au XVIII^e siècle, la sélection de textes devient plus compliquée car il y a moins de textes « classiques » tels que les *Remarques* d'Olivet (1767)² ou les commentaires de Patru et de Corneille sur les *Remarques* de Vaugelas (1738). Nous étions donc obligés de choisir des textes représentatifs de l'héritage des remarqueurs :

- textes qui s'intéressent aux régionalismes dont le texte clé est *Les Gasconismes corrigés* de Desgrouais (1766), mais aussi Mauvillon (1747), Dubois de Launay (1775), et Prémontval (1759-1761)³ ;

2 Ce volume comporte pourtant trois « opuscules » : une « Prosodie Française » (p. 15-142), des « Essais de Grammaire » (p. 143-250) qui traitent des parties du discours et, la partie la plus longue (p. 251-389), cent courtes « Remarques sur Racine », regroupées en deux parties, la première traitant des expressions surannées et la seconde des phrases qui ont « quelque sorte d'irrégularité ».

3 Pour citer Hassler (2004, p. 174), Prémontval « veut surtout moderniser la méthode des remarqueurs, tout en perpétuant la forme des observations trouvées chez Vaugelas, Corneille et Bouhours qu'il consulte à tout moment et qui lui permettent de garder le sentiment d'un français pur et clair et d'échapper à la contagion des façons de parler vicieuses, courantes à Berlin »

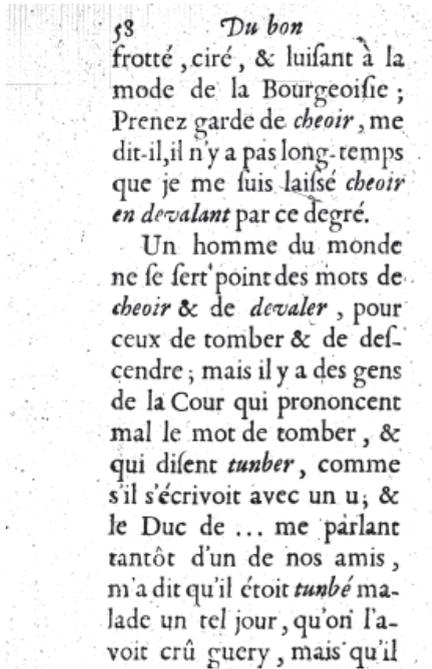


Image 2. Callières 1693b

- textes qui ont pour titre « Dictionnaire... » (Desfontaines 1727 [¹1726], Demandre 1802 [¹1769]) qui traitent souvent des néologismes et des difficultés linguistiques et qui sont donc en quelque sorte à cheval entre remarques et dictionnaires. Par exemple, le texte de Desfontaines ressemble beaucoup aux volumes de remarques organisés selon l'ordre alphabétique (Image 3) ;
- le *Journal de la langue françoise, soit exacte, soit ornée* de Domergue (1784-1792), dont le but est indiqué dans la préface du premier numéro : « Journal rédigé par une société de Grammairiens, ouvert à toutes les questions que le public lui adressera et présentant constamment toutes les solutions demandées ». En traitant ainsi de l'usage « douteux », Domergue se situe dans la tradition des remarqueurs et préfigure le genre des chroniques de langage (Ayres-Bennett 2015) ;
- le texte d'Arnauld (1707) qui, dans sa dernière partie (p. 107-152), examine l'attitude des remarqueurs envers l'usage ; l'ouvrage se situe donc dans la tradition des commentaires sur les *Remarques*.

10 A.
heur, p. 607. de la nouvelle Edition en
3. vol. tom. 1.
 AVANCEMENT D'HOIRIE. L'Auteur
 des *nouvelle Fables*, parlant du fils d'un
 Peintre habile, qui peignoit aussi-bien
 que son pere encore vivant, dit que
 ce fils reveillant la sublime industrie
 de son pere, s'est fait donner en avan-
 cement d'hoirie une part de son pin-
 ceau.
 Coppel digne héritier d'un Apelle nouveau,
 Qui reveillant sa sublime industrie,
 T'es fait donner la part de son Pinceau
 En pur avancement d'hoirie.
 [*Fab. 16. l. 1.*]
 AVANT DE est mieux dit qu'*avant*
 que de. Ex. » Astrée *avant* de se reti-
 » rer au Ciel, avoit choisi les campa-
 » gnes pour son dernier azyle. [*Georg.*
 » p. 217.] La pierre retomboit *avant*
 » d'être arrivée à la cime. » [*Ibid.*
 p. 256.] & *passim*.
 AVANTAGE. » Que ces vaisseaux
 » tirent l'*avantage* d'avoir été con-
 » struits sur une montagne qui m'est
 » dédiée. » *Profit nostris in montibus*
ortas. C'est-à-dire, *tirent avantage*.
 [*En. l. 5. p. 297.*]
 AVANTAGEUX (*Adj.*) Ce mot

Image 3. Desfontaines 1727

L'hétérogénéité de la taille

Pour les grammaires, les 24 ouvrages ou recueils totalisent env. 10 300 pages, soit une moyenne de 430 p. par texte. Mais sous cette moyenne, se cachent d'énormes disparités : de 33 pages (ouvrage inachevé, et non publié, de Mme du Châtelet) à 783 p. (Régnier-Desmarais), obligeant parfois une publication en 2 volumes, 910 p. (pour Girard) et 1336 p. (pour Beauzée), en passant par les 90 p. des *Elemens de la grammaire française* de Lhomond. L'œuvre de Du Marsais, telle qu'elle sera proposée, se présente comme un ensemble de 152 textes de taille très variable : ainsi, à l'intérieur des articles de l'*Encyclopédie*, de quelques lignes (*Absolument*, 4 li. ; *Alternative*, 6 li.) à 20 pages in-folio en deux colonnes (*figure*, 7 p. ; *article*, 18 p. ; *construction*, 20 p.), soit 22, 62 et 71 pages dans l'éd. de F. Douay (1987). Il a donc fallu s'extraire de la contrainte des entités « livre » ou « volume », en leur préférant celle de « texte », quelle que soit la taille de celui-ci.

La même hétérogénéité est évidente pour le volet « remarqueurs ». D'une part nous avons les douze remarques de Cassagne, publiées à la fin des *Essais de lettres familières* de Furetière (1690), qui consistent en 46 pages (p. 225-270). D'autre part, en ne prenant que la rubrique « langue exacte » du Journal de Domergue, nous arrivons à plus de 3000 pages.

L'hétérogénéité des supports

Dans plusieurs cas, le corpus créé provient de supports inhabituels : ainsi, qui veut rendre compte de l'œuvre de Du Marsais ne peut s'appuyer que sur quelques textes le plus souvent très courts, certains publiés à titre posthume, et sur les articles de l'*Encyclopédie*, qui ne constituent pas un ouvrage indépendant. De même, on a retenu un texte incomplet : le fragment de la *Grammaire raisonnée* de Madame du Châtelet, réduite à trois chapitres (cf. notice 2426 du *CTLF*), restée sous forme manuscrite, et éditée seulement en 1947.

2.3. Exhaustivité et/ou représentativité

La quantité d'ouvrages à retenir est une question importante : il ne peut être question d'exhaustivité à l'époque considérée, car la liste est ouverte : on dénichera *toujours* dans une quelconque bibliothèque européenne un nouvel ouvrage relevant du domaine, ouvrage dont le découvreur nous prouvera l'importance. Cela ne veut pas dire qu'il ne faut pas continuer à explorer les bibliothèques. Mais ce qu'on visera en priorité, c'est la représentativité, qui peut être fournie par l'histoire sérielle : on établit des séries sur une base chronologique et sur un faisceau de critères parmi lesquels la notoriété, de l'auteur ou de l'ouvrage, ne peut être le seul.

2.4. L'hétérogénéité dans la notoriété et la diffusion des ouvrages

Certains auteurs sont des grands noms de la linguistique (Girard, l'abbé d'Olivet, ou les deux rédacteurs des articles pour l'*Encyclopédie*, Du Marsais et Beauzée), voire de la philosophie (Condillac). D'autres sont au contraire des inconnus ou des quasi inconnus, tels Jean de Wapy (ou Wepy), Malherbe (il ne s'agit évidemment pas de François de Malherbe), Sauvage de Villaire, Fauleau, Bauchaint, à tel point qu'on a parfois de la peine à établir une notice bio-bibliographique, voire à connaître le prénom de l'auteur (ainsi pour Fauleau), ou les dates de sa vie (ainsi pour Wapy, Malherbe, Fauleau, Bonté, Bauchaint, etc. ; pour Dubois de Launay ou Sauvage de Villaire, on connaît seulement la date de leur mort).

Parfois il n'existe qu'une seule édition, dont un exemplaire est particulièrement difficile à trouver : ce fut le cas pour Bauchaint et Bonté ; les différents numéros du *Journal de Domergue* sont éparpillés dans plusieurs bibliothèques. Parfois au contraire, l'ouvrage a connu un grand succès éditorial, avec une pléthore d'éditions : ainsi, selon le principe de retenir la dernière du vivant de l'auteur, nous avons choisi pour Restaut, la huitième édition ; pour Wailly, la dixième (l'ouvrage ayant reçu des remaniements constants) ; pour Lhomond, la septième.

Cette hétérogénéité constitue-t-elle un défaut ? Pas forcément, si l'on veut établir un échantillonnage, et on peut la voir comme révélatrice de la diversité des

approches (ce qui était aussi le cas pour le XVI^e siècle, mais pour des raisons différentes). Consacrer un nouveau corpus sur la base de textes archi-connus n'aurait pas grand sens.

3. EFFETS DE LA CRÉATION D'UN TEL CORPUS

3.1. *Les avantages*

Les avantages de la création d'un corpus pour un domaine donné sont assez connus pour nous éviter de nous étendre sur les points généraux. Insistons sur quelques spécificités : les textes numérisés et outillés sont disponibles sous trois formes (fac-similé, version diplomatique, version légèrement modernisée). L'outillage fin permet des recherches quantitatives plus fiables et plus exhaustives qu'auparavant. Prenons le cas de Cicéron dans la base existante : grâce aux thesaurus, nous pouvons non seulement identifier 369 références à cette autorité latine, mais aussi préciser s'il est cité comme : source d'un exemple (codage 1), source d'une citation ou d'une opinion (codage 3), autorité linguistique sans exemple (codage 2) ou sans citation (codage 4), voire comme personnage politique (codage 5). En cherchant « verbe neutre » on identifie 43 occurrences de la locution, mais en nous servant de la fonction «verbe “et 3” neutre »⁴, on identifie encore 39 contextes pertinents.

3.2. *Le risque du canon*

Notons pourtant deux dangers potentiels associés à la création d'un tel corpus, et en premier lieu la question du canon. D'une part on risque de créer une sorte de canon de textes qui seront beaucoup étudiés à cause des outils de recherche disponibles. D'autre part on risque l'éviction ou la marginalisation de textes qui seront peu ou pas étudiés, malgré leur intérêt, parce qu'ils ne sont pas inclus dans le corpus de référence⁵. Un exemple typique est représenté par l'étude récente de Pagani-Naudet qui décrit son choix de textes pour son étude de la façon suivante :

Pour ce qui est de l'ensemble des textes sur lesquels portera notre investigation, nous suivrons la sélection que propose le corpus Garnier. C'est une sage facilité que de nous en remettre aux spécialistes qui ont sélectionné ces textes pour leur représentativité. (2014)

4 Deux termes éloignés l'un de l'autre par un nombre défini de mots, en l'occurrence 3.

5 On connaît les effets du canon alexandrin, dû à Aristophane de Byzance et à Aristarque, sur la transmission des textes : ce classement par ordre de mérite des œuvres littéraires grecques considérées comme modèles dans leur genre a certes permis la sauvegarde de quelques-unes ; mais il a entraîné aussi la perte irrémédiable de beaucoup d'autres. L'informatisation des textes et la constitution de bases de données créent un danger similaire, et ce de façon tout à fait paradoxale puisque ces nouveaux outils permettent de traiter des masses considérables de textes. Mais les bases de données (si vastes soient-elles) supposent toujours une sélection, et sélection (si large soit-elle) signifie aussi élimination. Sur cette question en littérature, voir par ex. Bernard 2011 ; pour l'histoire des idées linguistiques, Sanders 2000.

3.3. La question de l'outillage

La base utilise le moteur Babel développé par Classiques Garnier Numérique et permet les recherches spécifiques suivantes :

- repérage des auteurs en cinq catégories (cf. *supra* l'ex. de Cicéron) ;
- repérage des œuvres : (1) ouvrages sources d'exemples, (2) titres des ouvrages sources de citations ou d'opinions, (3) titres des ouvrages cités qui ne sont ni source d'exemple, ni source de citation ;
- repérage des exemples, en fonction de la langue (actuellement 62 langues ou dialectes répertoriés), avec indication, le cas échéant, de l'auteur et du titre texte ;
- repérage des citations, en indiquant la langue, l'auteur, le titre du texte ;
- repérage fin du métalangage utilisé dans l'ouvrage, par un thesaurus métalinguistique (termes isolés) complété par une liste des collocations intéressantes données dans l'environnement documentaire du texte.

Sont possibles des recherches sans accent (adaptées au clavier QWERTY) et, comme attendu, une utilisation étendue d'opérateurs de recherche. Les possibilités sont si nombreuses que plusieurs voies sont possibles pour aboutir à un même résultat et que chacun utilise une stratégie de recherche spécifique. Une aide en ligne relativement détaillée a été développée, mais l'expérience prouve que peu de gens explorent complètement les outils de recherche proposés.

Le traitement pose un certain nombre de difficultés dont on peut donner comme exemples :

1. la constitution des thesaurus des noms d'auteurs et des titres, qui doivent nécessairement être harmonisés ; pour certains textes, les variantes peuvent être considérables⁶ ;
2. le traitement des caractères spéciaux.

On prendra le cas du grec, dont certains auteurs (surtout Beauzée [cf. image 4] et Harris) donnent d'assez larges extraits, souvent difficiles à lire à cause des ligatures. Le grec, traité initialement par une police grecque « maison », difficile à utiliser, le sera dorénavant par une police Unicode⁷.

3. la différence entre les opérateurs de recherche du *Grand Corpus* et les expressions régulières communément admises (tableau) en linguistique de corpus.

6 Par ex., dans la base des grammaires de la Renaissance, une recherche pour Octovien de Saint-Gelais engendre 29 résultats, mais 15 formes différentes, toutes répertoriés chez Palsgrave, dont : « Leuesque d'Angoulesme, l'Evcsque Anguellesme, d'Anguillesme, the bisshope of Anguillesme, the bysshoppe of Angolesme, etc. ».

7 Dans la base existante, il y a près de 1000 exemples grecs, contre environ 10 500 pour le latin, et 150 000 pour le français.

<p>{r*chi} génère <i>reflechi</i>, mais aussi <i>ra-fraichi</i> {r*chi*} génère <i>rafraischir</i>, <i>rafreschir</i>, <i>rechigne(r)</i>, <i>retranchions</i>, <i>richissime</i>, <i>ro- chiers</i>, [de] <i>rechief</i>, etc. à côté de <i>refléchie</i></p>	<p>{r.chi} génère <i>brachio</i>, <i>enrichir</i>, <i>enrichit</i>, etc. {^r.chi} génère <i>Richi</i>, <i>rechigner</i>, <i>rachistisme</i>, etc. {r*chi} génère n'importe quoi (<i>architectus</i>, <i>Chio</i>, <i>Colchis</i>, etc.) {r.*chi} génère 6860 résultats : toujours un {r} suivi, plus ou moins loin, de {chi} {^r.*chi} génère 950 résultats ({^} impose que le {r} soit début de mot)</p>
---	---

À la lecture de ce tableau, on comprendra que celui qui se sert conjointement de ces deux bases, au contenu très fortement complémentaire, risque d'être gêné par la différence de valeur des caractères utilisés. Une harmonisation serait à désirer.

4. CONCLUSION

Celui qui conçoit un corpus a toujours une idée plus ou moins précise de ce qu'il pense pouvoir découvrir grâce aux résultats espérés suite à son exploration. En l'occurrence, les concepteurs avaient pour objectifs à la fois une meilleure connaissance des théories linguistiques mises en œuvre dans la période considérée et une meilleure connaissance de l'histoire de la langue française. En fait, dans la première étape de leur travail, ils ont été parfois surpris par des résultats qui n'étaient pas conformes à leurs attentes. Prenons un exemple. Le concept de « substantif » a été associé au verbe dans l'Antiquité, chez un auteur important pour la suite, Priscien (*uerbum substantiuum*), puis beaucoup plus régulièrement au nom à partir du moment (xii^e siècle) où les médiévaux ont subdivisé ce dernier en substantif et adjectif (*nomen substantiuum* vs *nomen adiectiuum*) et, surtout, ont stabilisé l'opposition. Dans la période xv^e-xvii^e siècle, bien éloignée de Priscien, on s'attendrait donc à trouver la seconde association bien plus souvent que la première. Or si l'on trouve la séquence {nom* substanti*} 237 fois, celle de {verb* substanti*} est encore attestée 202 fois, et nous disposons d'indices forts sur la survie de ce couple dans la suite de la tradition linguistique : ainsi Frantext-CTLF, pour des textes s'étalant du xviii^e au xx^e siècle, nous donne 61 résultats pour {verbe substantif} contre 6 seulement pour {nom substantif}. L'exploration fine des ouvrages du xviii^e siècle nous permettra peut-être de donner des explications à ce (re)basculement.

Le corpus échappe donc à son concepteur, puisqu'il lui fournit des données inattendues. Faut-il s'en inquiéter ? Certainement pas. C'est même sans doute l'intérêt majeur d'un corpus, que d'échapper aux intentions du créateur, car il est susceptible de lui fournir, à lui, mais aussi et surtout à l'utilisateur quelques surprises, de renouveler quelques idées préconçues. Une fonction essentielle des corpus est de générer des potentialités infinies, d'avoir leur vie propre, et donc, une fois la dynamique lancée, d'être justement... incontrôlables !

CORPUS UTILISÉS

- Grand Corpus des grammaires françaises, des remarques et des traités sur la langue (XIV^e-XVII^e s.)* (dir. B. Colombat, J.-M. Fournier ; dir. W. Ayres-Bennett pour le volet Corpus des remarques sur la langue française), 2011, Paris, Classiques Garnier Numérique.
- Corpus de textes linguistiques fondamentaux (CTLF)* (dir. B. Colombat, A. Pelfrène), <http://ctlf.ens-lyon.fr/default.htm>
- Frantext-CTLF* (dir. B. Colombat, A. Pelfrène, G. Souvay), <http://www.frantext.fr/ctlf/>

RÉFÉRENCES

- Ayres-Bennett, Wendy, 2015. « La Persistance de l'idéologie linguistique des Remarqueurs dans les chroniques de langage de 1925 à nos jours », *Circula*, 1 : *La Médiation des idéologies linguistiques : tradition et continuité dans la presse écrite*, 44-68.
- Ayres-Bennett, Wendy & Sejjido, Magali, 2011. *Remarques et observations sur la langue française : Histoire et évolution d'un genre*, Paris, Classiques Garnier.
- Bernard, Michel, 2011. *L'Histoire littéraire au risque de l'informatique. La Question du canon littéraire*, Paris, Presses Sorbonne Nouvelle.
- Biber, Douglas, 1993. "Representativeness in corpus design", *Literary and Linguistic Computing*, 8(4), 243-257.
- Bouhours, Dominique, 1692 [1675]. *Remarques nouvelles sur la langue française*, 3^e éd., Paris, G. et L. Josse.
- Habert, Benoît, 2000. « Des corpus représentatifs : de quoi, pour quoi, comment ? », Mireille Bilger (éd.), *Linguistique sur corpus. Études et réflexions*, Perpignan, Presses Universitaires de Perpignan, 11-58.
- Habert, Benoît, Nazarenko, Adeline & Salem, André, 1997. *Les Linguistiques de corpus*, Paris, Colin.
- Hassler, Gerda, 2004. « Bon usage et langue parfaite. La présence des remarqueurs dans le débat sur l'universalité de la langue française au XVIII^e siècle », Philippe Caron (éd.), *Les Remarqueurs sur la langue française du XV^e siècle à nos jours*, Rennes, Presses Universitaires de Rennes, 167-183.
- Macé, Jean, 1651. *Méthode universelle pour apprendre facilement les langues, pour parler purement et écrire nettement en français*, 2^e éd., Paris, J. Jost.
- Pagani-Naudet, Cendrine, 2014. « L'amplification dans le discours des grammairiens (XVI^e-XVII^e siècles », *Exercices de rhétorique* [En ligne], 4 | 2014, mis en ligne le 05/12/2014, consulté le 5/12/2015.
URL : <http://rhetorique.revues.org/362> ; DOI : 10.4000/rhetorique.362
- Sanders, Carol, 2000. "Linguistic Historiography: A Survey with particular reference to French linguistics at the turn of the century", *Journal of French Language Studies*, 10(2), 273-292.
- Sinclair, John, 2004. "Developing Linguistic Corpora: A Guide to Good Practice", version en ligne URL : <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>, consulté le 5/12/2015.
- Vaugelas, Claude Favre de, 1647. *Remarques sur la langue française utiles à ceux qui veulent bien parler et bien écrire*, Paris, Veuve J. Camusat et P. le Petit.

ANNEXE : LISTE DES TEXTES

1. Grammaires françaises et grammaire générale d'expression française : XVIII^e s.

- RÉGNIER-DESMARAIS, Abbé François-Séraphin, 1706. *Traité de la Grammaire française*, Paris, Coignard.

- BUFFIER, le Père Claude, 1709. *Grammaire française sur un plan nouveau*, Paris, N. Le Clerc *et al.*
- VALLANGE, DE, 1721. *Grammaire française raisonnée*, Paris, C. Jombert, A. Cailleau, F. Barois.
- MALHERBE, V., 1725. *La langue française expliquée dans un ordre nouveau*, Paris, Nicolas Le Breton.
- DU MARSAIS, César Chesneau :
 – *Les véritables principes de la grammaire ou Nouvelle grammaire raisonnée pour apprendre la langue latine*, 1729, 10 p. (sans lieu, sans éditeur [Brocas]).
 – 4 textes publiés dans les *Œuvres complètes* de Du Marsais rassemblées par Duchosal et Milon, Paris, Pougin, an V, 1797 (tome 3) : *Dissertation sur la Prononciation et sur l'Orthographe de la langue française* ; *Lettre d'une jeune Demoiselle à l'Auteur des Vrais Principes de la langue française* ; *Inversion* ; *Fragment sur les causes de la parole* ;
 – 147 articles de l'*Encyclopédie* rédigés par Du Marsais (tomes 1-7, 1751-1757).
- RESTAUT, Pierre, 1758 [1730]. *Principes généraux et raisonnés de la Grammaire française*, 8^e éd., Paris, Vve Lottin, J. Desaint et C. Saillant.
- DU CHÂTELET, Émilie, entre 1736-1749. *Grammaire raisonnée*. Manuscrit éd. par Ira O. Wade, 1947.
- VAL(L)ART, Abbé Joseph, 1744. *Grammaire Française*, Paris, Desaint et Saillant.
- GIRARD, Abbé Gabriel, 1747. *Les vrais Principes de la Langue française*, 2 vol., Paris, Le Breton.
- SAUVAGE DE VILLAIRE, 1749. *Abrégé de la grammaire française pour ceux qui n'ont point étudié*, Paris, G. Desprez & P. G. Cavalier.
- WAILLY, Noël-François DE, 1786 [1754]. *Principes généraux et particuliers de la langue française, Dixième éd. revue et considérablement augm.*, Paris, J. Barbou.
- D'ACARQ, Jean-Pierre, 1760. *Grammaire française philosophique*, Genève, Moreau & Lambert.
- BEAUZÉE, Nicolas, 1767. *Grammaire Générale ou exposition raisonnée des éléments nécessaires au langage pour servir de fondement à l'étude de toutes les langues*, 2 vol., Paris, J. Barbou.
- CONDILLAC, Étienne BONNOT DE, 1775. *Table des matières, Discours préliminaire, Grammaire* [Tome I du *Cours d'étude pour l'instruction du Prince de Parme*], Parme, Imprimerie Royale.
- CONDILLAC, Étienne BONNOT DE, 1782. *Cours d'étude*, vol. I : *Grammaire dans laquelle on considère le langage comme un instrument qui n'est pas moins nécessaire pour penser que pour communiquer nos pensées*, Aux deux ponts.
- DOMERGUE, Urbain, 1791 [1778]. *Grammaire française simplifiée, élémentaire*, 4^e édition, Paris, L.-M. Guillaume.
- LHOMOND, Charles-François, 1790 [1780]. *Elémens de grammaire française*, 7^e éd., Paris, Colas.
- FAULEAU, 1786. *Métaphysique de la langue française*, Paris, impr. de Quillau.
- D'ACARQ, Jean-Pierre, 1787. *Remarques sur la Grammaire française de M. de Wailly*, Saint-Omer, H. F. Boubers.
- FABRE, Abbé, 1787. *Syntaxe française, ou nouvelle grammaire simplifiée*, Paris, Périsse.
- BONTÉ, André [en religion, le Père François-Xavier, Capucin de Rouen], 1788. *Grammaire française raisonnée*, Paris, Le Barrois.
- BAUCHAINT, 1789. *Principes de la langue française... à l'Usage des Demoiselles*, Saint-Malo, l'auteur.
- HARRIS, James, 1796. *Hermès, ou Recherches philosophiques sur la grammaire universelle, ouvrage traduit de l'anglais de Jacques Harris, avec des remarques et des additions par François Thurot*, Paris, Impr. de la République.
- DOMERGUE, Urbain, 1798-1799. *Grammaire générale analytique distribuée en différents mémoires lus et discutés à l'Institut national de France*. T. 1 : *Premier mémoire. Sur la proposition grammaticale*, Paris, Charles Houel.

2. Remarques sur la langue française : XVII^e s. (complément)

- WAPY [ou WEPY], Jean DE, 1634. *Les Remarques sur la langue françoise Pour parler à la Mode & selon l'air du Temps*, Pont à Mousson, P. Vincent.
- LA MOTHE LE VAYER, François DE, 1669 [1638]. *Considerations sur l'Eloquence françoise de ce temps*, in *Œuvres de François de La Mothe Le Vayer...* Nouvelle édition, augmentée de plusieurs nouveaux traittez..., tome 4, Paris, L. Billaine.
- DUPLEIX, Scipion, 1645. *Les Lumières de Mathieu de Morgues...*, Condom, A. Manas.
- ANONYME, 1657. *Grammaire françoise avec quelques remarques sur cette langue, selon l'usage de ce temps*, Lyon, M. Duhan.
- SAINT-AURICE, Alcide de Bonnacese, sieur de, 1672. *Remarques sur les principales difficultés que les estrangers ont en la langue françoise*, Paris, E. Loyson.
- BOUHOURS, Dominique, 1674. *Doutes sur la langue françoise proposez à Messieurs de l'Académie Françoise par un gentilhomme de province*, Paris, S. Mabre-Cramoisy.
- BÉRAIN, Nicolas, 1675. *Nouvelles Remarques sur la langue française*, Rouen, E. Viret.
- AISY, Jean D', 1685. *Le Genie de la langue françoise. Suite du Genie de la langue française*, 2 vol., Paris, L. d'Houry.
- CASSAGNE, Jacques, 1690. « Nouvelles remarques sur la langue française » publiées après les *Essais de lettres familières sur toutes sortes de sujets*, Paris, J. Le Febvre.
- POULLAIN DE LA BARRE, François, 1691. *Essai de remarques particulières, Sur la Langue Française*, Genève, s.n.
- CALLIÈRES, François DE, 1693a [1692]. *Des Mots à la mode, et des nouvelles façons de parler*, Troisième Edition Augmentée [...], Paris, C. Barbin.
- CALLIÈRES, François DE, 1693b. *Du bon, et du mauvais usage, dans les manières de s'exprimer [...]*, Paris, C. Barbin.
- MORVAN DE BELLEGARDE, Jean-Baptiste, 1695. *Réflexions sur l'Elegance et la politesse du style*, Paris, A. Pralard.
- RENAUD, André, 1697. *Manière de parler la langue française selon ses differens styles [...]*, Lyon, C. Rey.
- LEVEN DE TEMPLERY, Joseph, 1698. *Nouvelles remarques sur la langue française*, Paris, M. & G. Jouvenel.

3. Remarques sur la langue française : XVIII^e s.

- ARNAULD, Antoine, 1707. *Regles pour discerner les bonnes et les mauvaises critiques des traductions de l'Ecriture-Sainte en Française [...]*, Paris, C. Hugier.
- DEFONTAINES, Pierre-François Guyot, 1727 [1726]. *Dictionnaire néologique à l'usage des beaux esprits du siècle*. Seconde édition corrigée & augmentée, [Paris, P.-N. Lottin ?].
- CORNEILLE, Thomas et PATRU, Olivier, 1738. *Remarques de M. de Vaugelas sur la langue française avec des notes de MM. Patru et T. Corneille*, 3 vol., Paris, Didot.
- [MAUVILLON, Éléazar DE], 1747. *Remarques sur les germanismes. Ouvrage utile aux Allemands, aux François, et aux Hollandois, &c.*, Amsterdam, P. Mortier.
- PRÉMONTVAL, André Pierre Le Guay DE, 1759-1761. *Préservatif contre la corruption de la langue française*, 2 vol., Berlin, s.n.
- DESGROUAIS, 1766. *Les Gasconismes corrigés*, Toulouse, impr. de J.-J. Robert.
- OLIVET, Pierre-Joseph Thoulier D', 1767. *Remarques sur la langue française*, Paris, J. Barbou.
- DEMANDRE, A., 1802 [1769]. *Dictionnaire de l'élocution française, contenant les principes de grammaire, logique, rhétorique, versification, syntaxe... [...]*, nouv. éd. rev., corr. et [...] augm. par l'A. Fontenai, 2 vol., Paris, Delalain fils.
- [DUBOIS DE LAUNAY, Henri], 1775. *Remarques sur la langue française à l'usage de la jeunesse de Lorraine*, Paris, Libraires Associés.
- DOMERGUE, Urbain, 1784-1792. *Journal de la langue française, soit exacte, soit ornée*, Lyon, Domergue.