



CLASSIQUES
GARNIER

MULLER (Ch.), « Comptes rendus », *Cahiers de lexicologie*, n° 7, 1965 – 2,
p. 119-124

DOI : [10.15122/isbn.978-2-8124-4260-5.p.0121](https://doi.org/10.15122/isbn.978-2-8124-4260-5.p.0121)

La diffusion ou la divulgation de ce document et de son contenu via Internet ou tout autre moyen de communication ne sont pas autorisées hormis dans un cadre privé.

© 2012. Classiques Garnier, Paris.
Reproduction et traduction, même partielles, interdites.
Tous droits réservés pour tous les pays.

COMPTES RENDUS

BAUDELAIRE, *Les Fleurs du Mal*. Concordances, index et relevés statistiques établis d'après l'édition Crépet-Blin par le Centre d'Etude du Vocabulaire français de la Faculté des Lettres de Besançon, avec la collaboration de K. Menemencioglu, 1 vol. br., 21 × 27, 246 p., 28 francs. Paris, s. d. (1965) Ed. Larousse.

On pourrait se contenter de quelques lignes pour signaler cette publication ; il suffirait d'attirer l'attention de ceux qui s'intéressent à Baudelaire, ou plus généralement au langage poétique, sur ce relevé lexical complet du texte et de ses variantes : chacun des mots qui y figurent étant remis à sa place alphabétique et suivi de tous les vers qui le contiennent, un coup d'œil donne une première mesure de ses emplois, de ses valeurs et des associations qu'il peut créer ; on citerait aisément quelques exemples suggestifs ; on signalerait les tableaux récapitulatifs, qui à la fin du volume regroupent certaines données numériques ; et il ne resterait qu'à se féliciter une fois de plus de l'heureuse collaboration des machines, qui permet au linguiste d'épargner son temps, et à souhaiter que la collection ainsi inaugurée nous propose bientôt d'autres concordances semblables.

Ce serait traiter à la légère l'énorme travail qu'a exigé cet ouvrage, et les nombreux problèmes de méthode qu'il a posés (1), justement parce qu'il ouvre une série et qu'il engage l'avenir. Les options les plus futiles en apparence, dans ce cas, prennent leur importance si elles doivent être reportées sur d'autres œuvres ; car l'unité et la stabilité des normes est la condition première des comparaisons et des synthèses futures. C'est à ce titre que l'on doit soumettre l'ouvrage à un examen scrupuleux, et même vétilleux. Le meilleur moyen de reconnaître le travail des auteurs n'est-il pas de mettre en lumière les difficultés qu'ils ont rencontrées, même si quelques-unes leur ont résisté, et d'aider à corriger les quelques imperfections de ce premier essai ?

Notons d'abord la réussite technique, obtenue après de nombreux tâtonnements, qui permet maintenant de passer directement des listes de la tabulatrice à l'impression, par reproduction photographique ; économie considérable, qui rend plus accessibles de telles publications. Remarquons aussi que ce classement intégral des formes a été obtenu par les seuls moyens de la mécanographie, dont la lourdeur relative est compensée par des avantages appréciables dans l'examen et le tri des formes homographes qui s'ensuit.

(1) On trouvera un reflet des travaux préparatoires effectués à Besançon sous la direction de B. Quemada dans les numéros I, II et III du *Bulletin d'Information du Laboratoire d'analyse lexicologique* (Besançon, 1960).

Les opérations commencent, on le sait (2), par l'établissement d'une bande perforée, d'où l'on tire mécaniquement un double fichier : cartes-mots et cartes-contextes. On a pris comme unité de contexte le vers, même court, solution économique qui se révèle presque toujours très suffisante pour les identifications grammaticales, et même pour l'intelligence des sens.

Le résultat de cette première opération est un catalogue des formes (3), accompagnées de leurs références ; ce qui revient (comme dans les Index de Besançon) à réunir (*le*) bois et (*je*) bois, mais à séparer cette dernière forme de (*nous*) buvons, comme l'on sépare yeux de œil. Certes, on peut s'en tenir là ; c'est ce que fit un précurseur, William T. Bandy, auteur du *Word Index to Baudelaire's poems* (ronéotypé, Madison, 1939) ; une telle liste de formes est déjà un outil précieux ; mais elle n'est guère utilisable sans un constant recours au texte ; car celui qui s'intéresse à l'*or* chez Baudelaire devra d'abord se débarrasser des références qui le mènent à la conjonction homonyme ; et l'image du *ver* ne sera constituée qu'après avoir rassemblé singuliers et pluriels :

O vers ! noirs compagnons sans oreille et sans yeux !

et après avoir écarté :

Je te donne ces vers...

et :

Crispe ses poings vers Dieu...

L'index des formes ne devient donc un index de mots qu'après la séparation des homographes et le regroupement des formes fléchies. Lourde tâche, en français plus qu'ailleurs, et à laquelle les machines participent assez peu ; elles ne peuvent que séparer les formes homographes quand celles-ci ont été préalablement codées par une intervention humaine.

Ce codage peut se limiter aux formes ambiguës ; la plupart du temps, la catégorie grammaticale suffit pour rattacher correctement à un lemme, et un code numérique d'un signe (de 0 à 9) convient, si l'on ne tient pas à distinguer nombres, genres, personnes, temps et modes des formes qui dépendent d'un même lemme ; on a ainsi bois 0 (= substantif), singulier ou pluriel, et bois 2 (= verbe), 1^{re} ou 2^e personne du singulier ; ou bien or 0 et or 5 (= conjonction). Reste le cas des homographes de même catégorie : (*le*) voile et (*la*) voile, air (= aspect) et air (= atmosphère), les deux voler, (*il*) convient et (*ils*) convient, (*je*) vis (de voir) et (*je*) vis (de vivre), etc. Il faut ici introduire des discriminants non grammaticaux.

Mais on peut aussi coder tous les mots du texte, ce qui prépare des comptages par catégorie grammaticale. C'est la solution adoptée pour cette concordance. Elle néglige toutefois les homographes sans différence de catégorie, qui ont dû être ensuite séparés « à la main ». A ce stade du travail interviennent les décisions, parfois embarrassantes ; ainsi on a réuni fonds singulier (« dans mon fonds le plus ténébreux ») et fonds, pluriel de fond (« les gouffres sans fonds ») ; en revanche, on a codé différemment mort dans « le mort joyeux »

(2) Voir à ce sujet le n° I du *Bulletin* déjà cité, et l'exposé de B. Quemada au Colloque de Strasbourg en 1957 (*Lexicologie et lexicographie françaises et romanes*, C.N.R.S., 1961, p. 53 sqq.).

(3) Sur les index de mots et les index de formes, voir Ch. Muller. *Les Index de vocabulaire*, in *Bulletin des Jeunes Romanistes*, n° 4, p. 9 sqq. (1961) et n° 8, p. 44 sqq (1963).

et « un mort libre et joyeux » (4) ; on a traité comme substantif (5) la forme *belle* dans :

A te voir marcher en cadence
Belle d'abandon...

On a réuni tous les sens de *air* sous une seule vedette, ce qui est défendable, mais contraire à la tradition lexicographique ; on n'a pas distingué *Dieu* et *dieu*, ce qui du reste est souvent difficile et qui a la caution de plusieurs dictionnaires (dont le *Petit Larousse*, où *Dieu* et *Satan* figurent le premier dans la partie générale, le second dans la partie historique et géographique). Mais on a distingué *diane* et *Diane*.

C'est après ce codage grammatical de toutes les formes que l'on obtient l'index statistique qui occupe les pages 203 à 225 : chaque forme y est accompagnée de sa fréquence et de sa « répartition », c'est-à-dire du nombre de poèmes où elle figure. Ce recensement exclut les variantes, traitées dans un index particulier (pp. 237-244). Notons une défaillance de la machine à la page 205 (*brisés* 4 scindé en deux) et une faute de frappe à la page 223 (*syllables*) (6).

Pour passer de cet index statistique à la concordance, il a fallu éliminer les mots à fréquence très élevée, regrouper les formes fléchies, faire suivre chaque mot-vedette des contextes et des références.

Si la concordance était faite sans aucune élimination, elle comporterait autant de contextes, donc de vers, qu'il y a de mots dans le recueil, soit un peu plus de 30 000 ; la seule entrée « *le*, art. déf. », serait suivie de quelque 2 500 citations, une bonne moitié des *Fleurs du Mal*. On pouvait, pour éviter cet encombrement bien inutile, écarter de la concordance tous les mots qui dépassaient une certaine fréquence : mais dans la concordance d'une autre œuvre, la liste des éliminations serait différente ; on a donc dressé une liste de mots grammaticaux, une cinquantaine en tout, choisis à la fois d'après leur fréquence et leur faible intérêt lexicologique, et formant des séries cohérentes : ainsi on écarte *eux*, qui n'apparaît que 12 fois, parce qu'il appartient aux pronoms personnels, mais on conserve *comme*, qui a 331 occurrences et dont les contextes occupent plus de deux pages : quel secours pour celui qui s'intéresse aux comparaisons !

On a regroupé les formes fléchies de façon diverse : pour les verbes, toutes les formes, y compris les participes adjectivés, sont réunies ; au contraire les formes d'un adjectif ou d'un substantif restent distinctes : *plein*, *pleine*, *pleines* et *pleins* fournissent quatre entrées, que l'on se contente de rapprocher si elles ne se suivent pas dans l'ordre alphabétique ; ainsi *yeux* vient se placer après *œil*, et *rêves* n'est plus séparé, de *rêve*, comme dans l'index statistique, par *réveil*, *revenir*, *rêver*, etc. Quelques oublis cependant dans ce regroupement : *travaux* est séparé de *travail* par *travailler* et *travailleur* ; *fol* et *folles* n'ont pas rejoint *fou* ; *vieux*, *vieil*, *vieille*, etc. sont dispersés. D'autre part on a réuni, sans doute pour gagner de la place, des formes codées différemment comme *mort* masc. et *mort* fém.

(4) Les deux *pêche*, tous deux substantifs féminins, ont été correctement distingués, ainsi que *tour* masculin et *tour* féminin.

(5) J'hésite aussi à considérer comme une forme du verbe *suivre* : « ... un beau vaisseau qui prend le large, chargé de toile, et va roulant *suivant* un rythme doux, et paresseux, et lent ».

(6) Autre indication pour un erratum : le vers « Que d'ignobles bourreaux plantaient dans tes chairs vives » comporte une faute de frappe (*qui* pour *que*) qui n'a été rectifiée que dans une seule des citations.

Le point le plus contestable me semble être le traitement des participes. Je sais que c'est là le chapitre le plus désespérant de la casuistique lexicographique, mais j'ai peine à approuver la solution adoptée ici. Lors du codage, on avait distingué trois cas :

- forme impersonnelle de verbe code 3
- adjectif qualificatif ou participe passé employé sans
auxiliaire code 4
- substantif code 0

Le classement des formes en *-ant* dans ces trois catégories ne crée guère de difficultés ; elles sont codées 3 quand elles sont précédées de *en* (gérondif), quand elles sont senties comme invariables, quand elles sont transitives ; elles sont chiffrées 4 quand elles sont variables ou senties comme variables. Cela ne soulève aucune objection. Mais dans l'établissement de la concordance, on a rangé le tout à la suite du verbe, alors qu'en général on considère que *charmant*, *plaisant*, *languissant*, *trionphant*, etc. sont des unités lexicales devenues indépendantes. Mais le plus déroutant est que lorsque le verbe correspondant n'apparaît pas sous d'autres formes dans le recueil, ces mots retombent au niveau des adjectifs. Ainsi *trionphant*, *trionphante* et *trionphants* donnent trois entrées, alors que *languissants* et *languissante* sont réunis sous *languir* ; il est évident que dans la concordance d'une œuvre où *trionpher* serait présent et où manquerait *languir*, le classement de ces adjectifs serait différent. Or, c'est là un des effets d'une norme constante : chaque forme doit venir se classer dans une case lexicale préétablie ; idéal difficile à concilier avec une linguistique rigoureuse, certes ; mais dans un état de langue donné, la norme ne devrait pas dépendre des contingences du texte analysé.

Pour le participe présent, le problème est du reste simple, et les contextes ambigus sont très rares. On n'en dira pas autant du participe passé, qui offre des difficultés quasi insolubles. Le critère de l'auxiliaire me paraît sans grande valeur. Je ne vois pas de différence *lexicale* entre :

Il est *tombé* dans cette geôle

et :

Andromaque, des bras d'un grand époux *tombée*

ni entre :

La Cloche *fêlée* (titre)

et :

Moi, mon âme est *fêlée*

Pas plus, à mon compte, qu'on ne serait fondé à classer différemment « le ciel *pur* » et « l'air est *pur* ».

Il y a certes des participes que leur adjectivation a détachés du verbe parent : tels *sacré*, *interdit*, *étourdi*, *parfait*..., etc. Mais le fait est révélé par des critères sémantiques, forcément subjectifs, et de force très inégale suivant les contextes ; il est assez rare qu'ils soient corroborés par des faits de position (un *parfait* amant). On n'a donc aucune peine à reconnaître qu'il existe un participe *adoré* et un adjectif *adoré* ; mais dans une bonne partie des contextes, il est difficile ou impossible de dire auquel des deux on a à faire ; or, si le lexicographe peut sélectionner ses citations et écarter les cas douteux, le statisticien doit classer tous les cas. Aussi ne peut-on reprocher aux auteurs de la concordance d'avoir indifféremment reclassé sous les verbes toutes les formes de participes passés, à l'exception de quelques cas du type *sacré*. Tout au plus jugera-t-on imprudent le classement d'*interdit*, codé comme verbe dans :

... des péchés qui t'ont *interdit* le tombeau

et comme adjectif dans :

... d'un gouffre *interdit* à nos sondes.

Mais dont les deux occurrences sont ensuite réunies sous une vedette d'adjectif.

Ces détails de classement, du reste, ne nuisent aucunement à l'usage normal de la concordance, qui est la lecture synthétique de contextes ayant un mot en commun. L'essentiel est que les regroupements soient correctement faits lors des relevés récapitulatifs, ce qui est le cas (sauf peut-être pour *fol*, oublié p. 230). Ils ne prendraient quelque importance que si l'on devait établir des concordances plus étendues ou très nombreuses et comparer entre eux les données de divers textes. Mais la chose aurait-elle un sens si ces textes n'appartiennent pas à un même état de langue ?

Le volume contient également un index statistique des 500 mots les plus fréquents. Cette fois il ne s'agit plus de formes, mais bien d'unités lexicales englobant leurs variantes morphologiques, avec indication, pour chacune de ces variantes, de sa fréquence et de sa répartition. Seuls entrent dans ce relevé les mots de la concordance, à l'exclusion des mots grammaticaux éliminés de celle-ci. Son point de départ n'est donc créé ni par une fréquence déterminée, ni par un rang dans l'échelle des fréquences, ni (comme chez Pierre Guiraud) par une limite plus ou moins conventionnelle entre mots pleins et mots-outils. Son point final est assez arbitraire ; à fréquence égale, les mots sont classés alphabétiquement ; or le 500^e est l'adjectif *tendre*, qui a la fréquence 7 ; il laisse donc passer devant lui tous les mots de fréquence 7 qui le précèdent dans l'alphabet, mais non ceux qui le suivent : *trionphant*, *ville* et bien d'autres sont ainsi écartés de ce palmarès pour une raison toute formelle. Il fallait évidemment ici appliquer la règle des *ex-æquo* dans les concours (7), et pousser la liste jusqu'à la fin de la fréquence 7.

Cette partie, d'ailleurs, cache quelques autres erreurs. Il semble qu'on ait voulu ici procéder à une autre distribution des formes nominales du verbe. Ainsi les participes présents adjectivés sont traités à part du verbe : *vivre* d'une part, *vivant* de l'autre ; les participes passés, adjectivés ou non, y sont au contraire joints (cf. par exemple *connu*, *damné*, *tombé*, etc.) ; cela semble parfaitement justifié. Mais par ailleurs on a généralement joint les adjectifs substantivés aux adjectifs : *vide* (« le vide », emploi dit neutre) aussi bien que *mourant* (« un mourant »), ce qui est déjà plus contestable, surtout pour le premier de ces deux cas. Enfin on a intégré au verbe des infinitifs substantivés (*coucher* de soleil, par exemple), mais pas tous (cf. *baiser*, *souvenir*, etc.). Cette redistribution, acceptable dans l'ensemble, a causé quelques flottements : *maudit*, tant adjectif que substantif, *condamné* (*id.*), rejoignent leurs verbes ; mais *pendu* n'a pas été compté avec *pendre*, ni *brûlé* avec *brûler* (au passage, notons un lapsus qui a rangé *marchand* sous *marcher*) ; quant au cas de *mort*, substantif masculin, il a déjà été relevé.

Il est donc prudent de ne pas utiliser sans vérification les fréquences fournies par cette partie de l'ouvrage, et de consulter de préférence la concordance proprement dite, où les contextes permettent un classement immédiat.

Puisque le mot de statistique est employé à plusieurs reprises dans l'ouvrage, exprimons un regret : bien que les auteurs ne se soient proposé que de livrer des matériaux bruts, n'eût-il pas été indiqué de donner un tableau récapitulatif des fréquences et des effectifs correspondants ? La répar-

(7) Procédé employé dans le *Frequency Dictionary of Spanish Words* de Juilland et Rodriguez (Mouton, 1964).

tition par catégories grammaticales, qui est établie, fournit une indication stylistique importante. Mais la structure quantitative du vocabulaire, classé par fréquence, est aussi un élément essentiel, trop souvent passé sous silence. On peut certes l'extraire de l'un des index, mais au prix d'un travail assez long. Ce vœu nous ramène à la question des normes du comptage. Question irritante, certes, mais qui doit être résolue au prix de quelques conventions simples pour ouvrir la voie aux études comparatives. Malgré les imperfections dont il ne faudrait pas exagérer l'importance, ce dépouillement intégral et attentif d'un texte littéraire très significatif mérite d'être exploité par de nombreux chercheurs, et retiendra l'attention de tous ceux qui se sont déjà livrés aux tâches harassantes de l'analyse lexicale.

Ch. MULLER.