

# Recherches sur Diderot et sur l'Encyclopédie

Numéro 31-32 (Avril 2002)

L'Encyclopédie en ses nouveaux atours électroniques: vices et vertus du virtuel

Alain Cernuschi

## La question d'un découpage par matières : l'exemple du corpus musicographique de l'*Encyclopédie*.

### Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

**revues.org**

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

### Référence électronique

Alain Cernuschi, « La question d'un découpage par matières : l'exemple du corpus musicographique de l'*Encyclopédie*. », *Recherches sur Diderot et sur l'Encyclopédie* [En ligne], 31-32 | Avril 2002, mis en ligne le 16 mars 2008. URL : <http://rde.revues.org/index3063.html>

DOI : en cours d'attribution

Éditeur : Société Diderot

<http://rde.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://rde.revues.org/index3063.html>

Ce document est le fac-similé de l'édition papier.

Tous droits réservés

Alain CERNUSCHI

## La question d'un découpage par matières : l'exemple du corpus musicographique de l'*Encyclopédie*

Le programme annonçant des « Paroles d'utilisateurs », il convient de préciser, en préambule, dans quelle perspective particulière j'ai été amené à tester l'*Encyclopédie* numérisée, à mesurer ses apports, ses limites actuelles, mais aussi ses promesses. Cette évaluation très modeste et ponctuelle de l'instrument informatique s'est faite à la lumière d'une expérience antérieure, d'une recherche menée il y a quelques années sur la place de la musique dans l'*Encyclopédie*<sup>1</sup> — dont je vais rapidement présenter les principes puisqu'elle oriente mon propos d'aujourd'hui.

L'enjeu était non seulement de comprendre comment les questions musicales sont traitées dans les 17 volumes de texte et dans les *Planches*, mais aussi — et surtout — de repérer et d'analyser la manière dont ce domaine se trouve articulé à d'autres domaines de connaissances (acoustique, médecine, grammaire, poétique, etc.). Il s'agissait là de prendre au sérieux le programme épistémologique des éditeurs, selon lequel l'*Encyclopédie* « doit exposer autant qu'il est possible l'ordre et l'enchaînement des connaissances humaines »<sup>2</sup>. Plus largement, j'ai essayé de comprendre à travers cette étude pourquoi la musique a représenté un centre d'intérêt constant pour les philosophes. L'*Encyclopédie*, par son ambition cartographique, permet, je crois, de saisir le rôle carrefour que la musicologie des Lumières tient dans la topographie des savoirs propre à l'époque. Mais ce travail m'a aussi conduit — c'est ce qui nous intéresse ici — à mener une réflexion sur le texte concret de l'*Encyclopédie* et sur sa lisibilité, et en particulier à définir un mode d'investigation qui puisse

1. Ce travail a paru sous le titre *Penser la musique dans l'Encyclopédie*, Paris, Champion, 2000 [Coll. « Les Dix-huitièmes siècles », n° 47].

2. Début du « Discours préliminaire » (*Enc.*, I, i).

prendre en compte les innombrables disparates et dysfonctionnements de l'œuvre. On le sait, l'*Encyclopédie* n'accomplit pas parfaitement le modèle textuel théorisé par ses éditeurs. Comme le dit Diderot dans son article ENCYCLOPÉDIE, elle n'est qu'une première réalisation du projet. Mais là est sa grandeur ! « [La] produire une fois, quelque informe qu'[elle] fût », voilà l'exploit dont il s'honore et qu'il désigne comme ce qui « ne pouvait être que la tentative d'un siècle philosophe » (*Enc.*, V, 644b). En d'autres termes, les défauts du texte, qui peuvent irriter l'utilisateur moderne, font partie intégrante de son sens parce qu'ils renvoient, au même titre que ses plus grandes réussites, à sa condition historique de production, aux conditions réelles, épistémologiques et discursives, dans lesquelles le projet encyclopédique des Lumières a affronté l'épreuve de sa concrétisation.

J'insiste sur ce point parce qu'il impose, à mon sens, un principe concernant l'informatisation du texte. Certes, l'édition électronique n'a de sens que si elle facilite la circulation dans les textes ; aussi, pour l'*Encyclopédie*, il paraît souhaitable que les dysfonctionnements, en particulier dans le classement et les renvois, soient autant que possible contournés. Toutefois, ce que je viens de souligner sur la valeur signifiante de ces imperfections impose une règle d'or, (qui est d'ailleurs celle de toute édition scientifique !) : l'*Encyclopédie* numérisée doit absolument donner accès au texte tel qu'il figure dans l'édition originale, avec tous ses ratés *qu'elle doit permettre de repérer comme tels*. Autrement dit, les améliorations dans les systèmes de repérage et de circulation ne doivent être apportées que dans une interface qui signale clairement son statut : celui d'un appareil critique.

Venons-en maintenant à la question plus précise des études sectorielles dans l'*Encyclopédie*, c'est-à-dire des études conçues selon un découpage par matière. Mon travail sur la musique me servira de référence pour soulever des questions qui concernent en fait toute recherche du même type, et pour mettre en valeur ce que l'instrument informatique peut apporter de ce point de vue. Car il est à souhaiter que l'*Encyclopédie* numérisée suscitera la multiplication de telles études.

Je militerai d'abord pour une perspective non pas monographique — qui isolerait une matière du reste de l'œuvre — mais cartographique — c'est-à-dire qui envisage l'articulation d'un domaine donné à tous les autres — perspective fidèle au point de vue encyclopédique voulu par Diderot et D'Alembert<sup>3</sup>. L'incidence pratique d'un tel parti pris est que le corpus à constituer ne doit pas se limiter aux articles spécifiquement consacrés au champ de connaissance visé ; il doit aussi inclure tous les articles ou extraits d'articles appartenant à d'autres domaines mais traitant de, ou se

3. C'est la perspective esquissée ici même par Gilles Blanchard — mais sur un secteur déterminé et en entrant de façon détaillée dans les textes.

référant à ce champ de connaissance. C'est ce qu'on peut nommer un *corpus sectoriel extensif*. Pour mesurer l'importance d'un tel principe, je signalerai simplement que le corpus musicographique que j'ai réuni (un peu plus d'un millier d'articles) comprend un tiers d'entrées qui relèvent, explicitement ou implicitement, d'autres secteurs de connaissance.

Du coup, de tels corpus se configurent en cercles concentriques. Ainsi, pour la musique, le noyau est constitué par les articles explicitement rattachés à ce domaine (par ce que j'appelle le *désignant* : *Musique*, *Lutherie*, *Chant*, etc.) ; et il est complété par les entrées sans désignant mais qui appartiennent visiblement au même ensemble (mêmes sujets, même auteur, réseau interne des renvois). Les franges du corpus comprennent d'abord des articles entièrement consacrés à des questions musicales mais rangés sous d'autres désignants (p. ex., si l'article TROMPETTE est classé sous *Lutherie*, l'article COR, lui, est rangé sous *Chauder. & Chasse*) ; plus périphériquement, viennent enfin les articles inscrits dans un autre domaine et dont seul un extrait concerne la musique (comme l'article ÉCOLE de D'Alembert, classé sous *beaux Arts*).

Lorsque j'ai réuni ces textes, il m'était apparu que tout corpus de ce type restait ouvert, se caractérisait par des limites floues — pour des raisons de fait et de principe. De fait : parce qu'il me semblait alors impossible de prétendre pouvoir réunir exhaustivement tous les articles hors domaine contenant des références musicales. De principe : parce que, pour les franges les plus extérieures d'un tel corpus, les arbitrages sont nombreux et dépendent des partis pris de la recherche ; la pertinence d'une référence périphérique n'apparaît souvent qu'à la lumière de l'étude fouillée d'un dossier.

Le nouvel instrument informatique lève l'obstacle pratique mais il amplifie l'obstacle de principe. En effet, une recherche plein texte par mots-clés permet d'atteindre à un repérage exhaustif : les tests que j'ai menés sur le texte électronique, dans les deux versions disponibles, m'ont en tout cas permis de repérer un certain nombre d'articles qui m'avaient échappé (j'y reviendrai) ; en revanche, ils ont multiplié les cas limites, où le chercheur doit se demander si une simple allusion trouve assez de pertinence pour être incluse dans son corpus. Tout intégrer ce que la machine donne risque de produire un corpus illisible. Peut-être n'est-il pas inutile de rappeler ici cette évidence que l'instrument informatique ne permet en rien de faire l'économie d'une démarche de construction des données (sélection, délimitation, structuration).

Le corpus une fois établi, la recherche principale consiste bien sûr dans l'analyse des textes. Mais une part intéressante de la réflexion peut aussi se dégager de l'étude du corpus dans son ensemble. Il s'agit en quelque sorte de déterminer son « profil encyclopédique » : tel domaine a-t-il été traité par un seul encyclopédiste ou par plusieurs — et dans ce cas

note-t-on une répartition, stable ou évolutive, ou des chevauchements (articles composites) ; combien ce domaine suscite-t-il d'entrées dans le dictionnaire, et de quelle longueur ; ces articles sont-ils isolés, articles de tête ou articles insérés dans des ensembles ; sont-ils munis de désignants uniformes ou variés ; quelle est la proportion d'articles sans désignants, d'articles non signés ; quelle est la proportion entre les franges et le noyau, autrement dit quel est le taux d'empiètement du domaine considéré sur les autres ; quel degré d'interconnexion par renvois explicites ; quelle quantité de liens avec les planches ? De telles caractérisations statistiques ne deviennent parlantes que dans une démarche comparative. Ainsi, pour les trois secteurs principaux de mon corpus, il s'est révélé que les modes de traitement encyclopédique réservés à la musique, à la lutherie et à l'acoustique étaient extrêmement distincts, comme si la « consistance » épistémologique des objets visés n'était pas la même. Ce type d'approche globale d'un corpus donne des indications précieuses pour orienter l'analyse proprement dite des textes. Mais, élargie à l'ensemble des domaines traités dans l'*Encyclopédie*, la comparaison des « profils encyclopédiques » apporterait aussi, je crois, une connaissance profondément renouvelée de l'œuvre.

Or, la numérisation du texte rend tout à fait envisageable un tel programme. C'est en fonction de cet horizon que j'ai essayé d'évaluer les deux versions électroniques de l'*Encyclopédie* actuellement disponibles.

Techniquement, le mode d'approche statistique que je viens d'esquisser suppose la constitution du corpus sectoriel en base de données. Par conséquent, l'attente principale vis-à-vis de l'*Encyclopédie* numérisée, de ce point de vue, est qu'elle permette de construire facilement de telles bases de données. En particulier, il me paraît très opportun que les résultats de requêtes, les listes d'intitulés sélectionnés, puissent être exploitées directement et qu'elles ne soient pas de simples tables d'accès au texte. Or — mais j'avoue ici mes compétences très restreintes en la matière — il m'a semblé que dans une version comme dans l'autre, il est impossible de travailler avec ces listes (par exemple d'en retrancher les items qui se révèlent parasites) et impossible de les exporter dans une base de données. En fait, l'idéal serait que le texte numérisé soit déjà couplé avec une base de données globale, un inventaire informatisé des entrées. Voilà un souhait très général.

Venons-en à un examen plus précis. Dès qu'on parle de bases de données, deux questions se posent : celle de l'unité pertinente à adopter et celle de l'uniformisation des paramètres pris en compte.

Question de l'unité, d'abord. Pour les études par domaine, l'unité évidente est l'article, c'est-à-dire chaque entrée signalée par un intitulé en capitales (grandes ou petites), et complété ou non d'un désignant. C'est le choix de la version ARTFL, d'autant plus précieuse qu'elle distingue les

deux types d'articles. Le découpage plus analytique qu'avaient adopté Schwab et Rex, dans leur *Inventory*<sup>4</sup>, subdivisant les articles lorsqu'ils comprennent plus d'une signature, convenait à un inventoriage qui cherchait avant tout à repérer toutes les interventions d'un collaborateur, mais serait trop parcellisant pour une recherche sectorielle. Inversement, le regroupement des articles par paquets, qui caractérise la version Redon, est peu approprié à la constitution de listes pertinentes, d'autant plus lorsque les critères de regroupement sont arbitraires (ce qui m'a quelquefois semblé le cas dans cette version...).

Sur la question des paramètres, il serait à souhaiter, de façon générale, que les listes d'articles qui résultent d'une requête précisent chaque fois, s'il y a lieu, le désignant et le ou les auteurs (c'est encore un avantage de la version ARTFL sur le CD Rom Redon), mais qu'elle précise aussi la longueur de l'article et sa localisation dans les volumes originaux. L'inventaire de Schwab pourrait servir de référence.

J'aimerais surtout m'arrêter aux difficultés que soulèvent les désignants. Il y en a trois, bien connues et rappelées ici même par Gilles Blanchard. L'hétérogénéité de leur forme typographique ; la diversité de leur formulation ; leur absence fréquente.

L'hétérogénéité de leur forme de présentation a entraîné ce qui m'apparaît comme l'un des défauts les plus importants des deux versions. N'ont été pris en compte, selon mes enquêtes partielles, que les désignants entre parenthèses. Or, il arrive très fréquemment que la désignation du domaine encyclopédique figure en italique mais sans parenthèses à la suite immédiate de l'intitulé ou au début de la première phrase de l'article. Exemple : « SON, *en Musique* ;... », ou « DIASTÈME, sub. m. dans la *Musique ancienne*, signifie... ». Toutes ces entrées-là ont été considérées, à tort, comme démunies de classement encyclopédique. Cette saisie très lacunaire des désignants fait qu'une recherche par domaine produit des listes extrêmement déficientes. Pour le corpus musical, j'ai pu quantifier ce déficit : ce sont plus de 60 % des articles munis de désignant qui manquent à l'appel...

La diversité des abréviations et des formulations est en revanche un obstacle que les astuces classiques de la recherche informatique, avec les caractères « jokers », permet en grande partie de contourner. Mais dans la perspective que je développe ici, il serait à souhaiter qu'une uniformisation soit apportée, non pas bien sûr au niveau du texte mais dans cette interface éditoriale qui devrait en faciliter l'interrogation. Ici encore, le travail accompli par Schwab et Rex pourrait servir de référence. C'est au même niveau de l'inventaire informatisé que pourrait être introduite une

4. SVEC 80, 83, 85, 91-93 (Genève/Oxford, 1971-1972).

indexation matière pour les articles démunis de désignant. Signalons en passant que créer des classes de désignants ne présente pas seulement un intérêt pratique (faciliter la recherche), mais révèle aussi de significatives variations d'un domaine de connaissance à l'autre : certains se caractérisent par la régularité et l'économie des formulations ; d'autres donnent lieu à une prolifération de solutions (le recoupement avec les auteurs semble aussi parfois pertinent : Jaucourt, notamment, multiplie les désignants de manière très libre !).

Je n'aimerais pas terminer par cette énumération de ce qu'on ne peut pas encore faire avec les versions actuelles de l'*Encyclopédie* numérisée. Il faut aussi souligner tout ce qu'elles apportent déjà. J'insisterai pour ma part sur l'exhaustivité des repérages plein texte par mots clés (avec un bémol : cela suppose une saisie aussi exacte que possible du texte original, idéal auquel les versions disponibles ne tendent que très imparfaitement ; ici, le CD Rom Redon me paraît toutefois présenter un net avantage par rapport à son concurrent...).

L'extraordinaire rapidité de telles requêtes représente bien sûr l'apport technique le plus incontestable. Mais au-delà, il me paraît important de chercher à mesurer leur portée dans la procédure de réunion de données. Je m'en tiens ici à un pointage, le résultat de deux interrogations avec les mots « musique » et « chant ». Ce sont trente-six articles supplémentaires (v. liste annexée) que j'ai pu finalement ajouter à mon corpus musicographique, après un tri entre près de 350 entrées nouvelles. 36 sur 1132, ce n'est sans doute pas très lourd (3,2 %). La répartition de ces articles dans le corpus se révèle toutefois significative : un seul concerne le noyau (une entrée-renvoi, qui avait été omise par Schwab) ; six s'ajoutent dans la première extension (celle des articles totalement centrés sur des questions liées à la musique ou à l'acoustique mais rangés sous des désignants extérieurs). C'est donc surtout dans la seconde extension du corpus que l'apport de la recherche informatisée se révèle prépondérant : 29 articles (les quatre cinquièmes) retenus pour un extrait pertinent. Comme le corpus initial contenait 133 extraits d'articles, le gain s'élève ici à 22 %.

Par cette irremplaçable puissance de repérage des passages relatifs à un domaine mais insérés dans des articles d'autres domaines, l'*Encyclopédie* numérisée est l'instrument idoine pour les recherches sectorielles, attentives aux questions d'articulation entre domaines, que j'appelle de mes vœux. Mais je souhaite aussi avoir montré que le bon usage de cet instrument ne va pas sans des réflexions méthodologiques et épistémologiques, et que celles-ci, en retour, sont porteuses de propositions pour perfectionner l'instrument.

Alain CERNUSCHI  
*Université de Lausanne (Suisse)*

## Annexe

Liste des trente-six articles complétant le corpus musical de l'*Encyclopédie* établi dans : Alain Cernuschi, *Penser la musique dans l'Encyclopédie*, Paris, Champion, pp. 673-696 (mêmes conventions).

- ANTITHÈSE [extr.], *Bell. Lett.*, (G = Mallet), I, 517a.  
 CABALE [extr.], *Philos.*, (C = Pestré), II, 475b.  
 CATALOGUE [extr.], *Littérat. & Librair.* (ns), II, 759b.  
 CERVEAU [extr.], *Anatom.*, (L = Tarin), II, 862a.  
 COMIRIS [extr.], *Littér.*, (ns), III, 683a.  
 CREUX [extr.], (ns), IV, 460b.  
 CRIN [extr.], *Corderie*, IV, 471a.  
 ÉCRITURE-SAINTE [extr.], *Théol.*, (G = Mallet), V, 361b.  
 ENSEMBLE [extr.], *Peint.*, (Watelet), V, 713a.  
 EPHEMERIUS [extr.], *Hist. anc.*, (\* = Diderot), V, 772b.  
 FIL D'ARGENT [extr.], (ns/# *Chambers*), VI, 790b.  
 HEMI, *en Musique. Voyez* (SEMI.), (ns), VIII, 112a.  
 HÉMISTICHE [extr.], *Littérature*, (de Voltaire), VIII, 113b.  
 JETTER LE PLOMB SUR TOILE [extr.], *Plombier*, (ns), VIII, 529b.  
 IMPRIMEUR [extr. dans la section *Imprimeur-Libraire ordinaire du Roi (Hist. litt.)*, 629b], (ns), VIII, 624a.  
 LACEDEMONNE, *république de* [extr.], *Hist. de Grèce*, (D.J. = Jaucourt), IX, 152b.  
 LIVRE [extr.], *Littér.*, (ns), IX, 601a.  
 MACHICOT, [extr.], *Hist. eccles.*, (ns), IX, 793b.  
 MARSYAS, *Mythol.*, (D.J. = Jaucourt), X, 159b [1,0].  
 MENUS PLAISIRS ou *simplement* MENUS, *Hist. mod.*, (ns), X, 345b.  
 MYSIE [extr.], *Géog. anc.*, (D.J. = Jaucourt), X, 920b.  
 ORGE, *grain d'*, [extr.], *Tisserander. Imprim.*, (ns), XI, 633a.  
 POÉTIQUE, ART [extr.], *Poésie*, (D.J. = Jaucourt), XII, 847b.  
 POÉTIQUE HARMONIE [extr.], *Poésie*, (D.J. = Jaucourt), XII, 848a.  
 POIRIER [extr.], *Commerce de bois*, (D.J. = Jaucourt/# *Dict. du commerce*), XII, 883b.  
 PRÉCENTEUR, *Jurispr.*, (A = Boucher d'Argis), XIII, 268a [0,2].  
 REDITE [extr.], *Gram.*, (ns), XIII, 876a.  
 REMPLISSAGE [extr.], *Gramm.*, (ns), XIV, 100b.  
 SENS EXTERNES [extr.], *Physiol.*, (Jaucourt), XV, 29b.  
 SUPERFLU [extr.], *Gram.*, (ns), XV, 661b.  
 THÉÂTRE DES GRECS [extr.], *Archit. greq.*, (D.J. = Jaucourt), XVI, 234b  
 TRAIT, *terme de rubrique*, (D.J. = Jaucourt/# *Du Cange*), XVI, 531b [0,1].  
 TROCHOÏDE [extr.], *en Géométrie*, (O = D'Alembert), XVI, 683a.  
 VASES *de théâtre, Antiq. greq. & rom.*, (D.J. = Jaucourt), XVI, 854a [1,0].  
 ZIL, *Hist. nat.* [sic ! il s'agit d'un instrument de musique...], (ns), XVII, 714a [0].  
 SERRURERIE [extr.: section *Fer en fil*, ou *fil de fer*, 813a], *Art méchan.*, (Lucotte), XVII, 812a (in « Articles omis »).

