

# Recherches sur Diderot et sur l'Encyclopédie

25 (1998)  
Varia

Robert Morrissey, John Iverson et Mark Olsen

## L'ENCYCLOPÉDIE DE DIDEROT SUR INTERNET\*

### 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

Robert Morrissey, John Iverson et Mark Olsen, « L'ENCYCLOPÉDIE DE DIDEROT SUR INTERNET\* », *Recherches sur Diderot et sur l'Encyclopédie* [En ligne], 25 | 1998, mis en ligne le 19 août 2007. URL : <http://rde.revues.org/index1911.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/index1911.html>

Document généré automatiquement le 27 janvier 2012. La pagination ne correspond pas à la pagination de l'édition papier.

Tous droits réservés

**Robert Morrissey, John Iverson et Mark Olsen**

## L'ENCYCLOPÉDIE DE DIDEROT SUR INTERNET\*

1 Le ARTEL Project de l'University of Chicago annonce la mise sur l'Internet d'une version électronique de l'Encyclopédie, ou Dictionnaire raisonné des Sciences, des Métiers et des Arts de Denis Diderot, Jean le Rond D'Alembert et al.1.

Mère, en quelque sorte, de toutes les encyclopédies modernes l'Encyclopédie, cet ouvrage collectif d'une « société de gens de lettres », reste comme un monument colossal de l'Âge des Lumières. Diderot lui-même la compare tantôt à une ville avec ses rues et ses quartiers, tantôt à un labyrinthe où l'on risque de se perdre, tantôt à un arbre avec ses racines, ses branches et ses ramifications, tantôt enfin à un monde dont l'ordre encyclopédique servirait de carte ou mappemonde. Autant d'images qui traduisent à la fois l'énormité de l'entreprise et le souci d'ordre mais aussi la complexité voire la menace de désordre qui hante la volonté démesurée de devenir le lieu de rencontre de tous les savoirs. Comme on le sait, l'Encyclopédie cherche, on dirait parfois désespérément, à lier tout à tout ; symboliquement par l'arbre des connaissances, d'une manière plus concrète par les classifications en tête des articles et surtout par les renvois. On a beaucoup dit sur le prétendu système des renvois, sur son efficacité, sur sa nature subversive mais aussi sur ses incohérences et ses insuffisances. Diderot lui-même n'a-t-il pas affirmé qu'il aurait fallu que l'éditeur ait en main tous les articles avant la publication pour assurer la cohésion et la cohérence de l'ensemble ?

Mais que les renvois ainsi que les autres liens et classements lui donnent une vraie cohérence ou qu'ils n'en fournissent que l'illusion, l'Encyclopédie a toujours

\* Ce projet a été présenté à la journée «L'Encyclopédie. Du réseau au livre et du livre au réseau » organisée le 5 juin 1988, à l'École des Hautes Études en Sciences Sociales (EHESS, Paris), par Robert Morrissey et Philippe Roger (Actes à paraître).

1. L'ARTFL (Project for American and French Research on the Treasury of the French Language) est une collaboration entre le CNRS et l'University of Chicago. L'ARTFL est rattaché à l'INaLF (Institut National de la Langue Française) dont il est l'antenne américaine présenté, aux lecteurs un défi d'interprétation qui tenait à l'énormité de ses dimensions. Elle attire notre regard parce qu'elle est un monument. Elle incarne la volonté des philosophes de redéfinir le monde ; elle porte en elle les traces d'une lutte pour changer le sens des mots afin de faire voir autrement les choses. Mais elle s'érige aussi comme vaste témoignage d'un ensemble de connaissances, de pratiques et de techniques. Se proposant de représenter tous les savoirs du monde, de les fixer pour les conserver, pour assurer qu'ils survivent aux bouleversements — aux révolutions, disait Diderot —, qui pourraient conduire au retour des ténèbres, elle nous livre une sorte de vaste inventaire du monde de cette époque ou plutôt un résumé qui prétend à un ordre qui ne cesse de lui échapper.

L'initiative de créer une édition électronique de l'Encyclopédie relève donc d'une part de la place stratégique que celle-ci occupe dans les productions des Lumières et d'autre part de sa forme et de ses dimensions mêmes. Œuvre emblématique, œuvre problématique, œuvre monstrueuse et insaisissable, elle semble faire appel en quelque sorte aux moyens informatiques jusque dans sa conception. Pour la première fois, il sera facile de faire ce à quoi les éditeurs nous invitaient avec beaucoup d'insistance : suivre les renvois, naviguer entre articles, planches et légendes, consulter les différentes entrées classées sous une même rubrique. Ainsi l'espoir que formulaient les encyclopédistes de voir s'établir une relation dialogique entre les articles aussi bien qu'entre l'œuvre et les lecteurs se verra sinon réalisé, du moins porté dans le domaine du possible sur le plan pratique.

Dans un sens et toute proportion gardée, nous sommes confrontés à des problèmes semblables à ceux auxquels se heurtaient Diderot et D'Alembert. Ou bien ils pouvaient faire comme ils ont fait, c'est-à-dire publier l'Encyclopédie volume par volume, avec pour conséquence des incohérences, des oublis et des imperfections. Ou bien ils pouvaient viser une plus grande

perfection, attendre d'avoir le tout, de le corriger, de l'améliorer, et alors seulement le publier. Face aux exigences pratiques et économiques mais aussi face à la menace d'un perfectionnisme paralysant, ils ont pris le parti et le risque de passer à l'action, d'accepter un compromis entre l'idéal et l'expédient: ils ont commencé la publication lettre par lettre, quitte ensuite à préconiser une nouvelle édition revue et corrigée, voire une refonte de l'ensemble. Car ils ne savaient que trop bien « que la première édition d'une Encyclopédie, ne peut être qu'une compilation très-informe & très-incomplète » (V, 645). Dans le meilleur des mondes, pour arriver à l'Encyclopédie idéale, il faudrait, comme l'envisageait Diderot, établir des mécanismes et une pratique de correspondance générale permettant d'accueillir et de rediffuser les informations envoyées de toute la République des Lettres<sup>2</sup>. On dirait presque qu'il rêvait des possibilités de l'Internet : une correspondance rapide et générale, la possibilité d'établir des liens entre différents savoirs, une structure dynamique permettant de tenir compte des nouvelles hypothèses et des dernières découvertes.

Bien sûr, il ne s'agit pas pour nous de faire une encyclopédie, mais d'en faire une édition électronique aussi fidèle, aussi correcte que possible. Mais nous aussi avons des contraintes économiques et le désir de mettre l'Encyclopédie à la disposition des chercheurs aussitôt que possible. Et nous aussi, nous devons accepter un compromis entre l'idéal et l'expédient. Pourtant si, comme nous l'avons dit, la nature statique ou la fixité du livre imprimé a été très vite ressentie<sup>2</sup>. « Il ne seroit pas inutile d'établir des correspondances dans les lieux principaux du monde lettré, & je ne doute point qu'on n'y réussît. On s'instruira des usages, des coutumes, des productions, des travaux, des machines, &c. si on ne néglige personne, & si l'on a pour tous ce degré de considération que l'on doit à l'homme désintéressé qui veut se rendre utile. » (V, 645) comme une contrainte et une limitation par Diderot, l'Internet offre une souplesse et une ouverture qui sont tout à fait intéressantes dans le contexte d'un travail de cette envergure. Car transformée en base de données, la version électronique de l'Encyclopédie devra être considérée comme une chose dynamique dans le sens qu'il sera toujours possible et relativement facile d'y apporter des corrections, des amendements, des commentaires éditoriaux ainsi que des nouveaux outils d'accès et d'analyse. Grâce au courrier électronique, les observations et les suggestions des utilisateurs et des spécialistes pourront être cueillies, accumulées et intégrées périodiquement à l'Encyclopédie électronique qui devient de la sorte une édition dynamique constamment en voie d'amélioration.

Nous avons témoigné au nom de nos Collègues & au nôtre, & nous témoignons encore notre reconnaissance à tous ceux qui voudront bien nous faire appercevoir nos fautes. [...] De plus la reconnaissance dont nous parlons doit s'étendre, comme il est juste, sur ceux qui nous adresseront directement & immédiatement leurs remarques. Un tel procédé ne peut avoir pour objet que le bien public & celui de l'ouvrage. (« Avertissement », t. III, p. x)

État actuel des travaux : la saisie

Le travail de saisie a déjà été réalisé à partir d'un exemplaire microfiché de la première édition de Paris. Le résultat de cette saisie effectuée manuellement pourrait être comparé à une transcription dactylographiée de l'ensemble du texte de l'Encyclopédie. Nous avons pris cette version et nous avons procédé à un travail de repérage. Il a fallu distinguer toute une gamme d'« objets » tels que les articles, les classifications, les renvois, etc. Tout ce travail de classification et d'identification que nous avons effectué a été fait par repérage automatique sur des critères de positionnement, de typographie etc. sans vérification manuelle.

Quelques chiffres sont très parlants pour donner une idée de l'envergure du projet. L'Encyclopédie contient environ 20,5 millions d'occurrences de mots et 391 000 mots différents. Sur l'ensemble des dix-sept tomes d'articles et les onze tomes de planches, nos logiciels ont identifié quelque 76 242 parties distinctes du texte (principalement les articles et les légendes). Nous avons classé cet ensemble en 44 632 articles principaux, 28 366 sous-articles et 2 575 légendes. Ainsi l'article principal HERBE, par exemple, est suivi d'un ensemble de sous-articles parmi lesquels on trouvera HERBE AUX ÂNES, HERBE AUX HÉMORROÏDES, HERBE AUX MITES, HERBE AUX NOMBRILS, HERBE À ROBERT et HERBE AUX VERRUES. En plus de ces catégories, il reste 669 « objets » divers tels que les pages de titre, le Discours préliminaire, les préfaces, les avertissements, les sous-ensembles de

planches, etc. Il y a approximativement 64 000 renvois entre articles (dont un certain nombre à des articles inexistantes !) et 12 000 renvois aux planches.

Nous estimons avoir pu identifier par procédés automatiques plus de 99 % des articles et sous-articles, un peu moins en ce qui concerne les indications d'auteurs, de classification et de fonction grammaticale. À cause de la diversité des formes de renvois, le niveau d'identification est légèrement inférieur, de l'ordre de 95 %, et pour les planches il est quelque peu plus bas encore. Il faut aussi se souvenir du fait que, pour le moment, nous n'avons rien ajouté au texte. Ainsi si (ce qui est souvent le cas) les encyclopédistes ont négligé d'indiquer la fonction grammaticale d'un terme (par exemple, « verbe ») l'ordinateur ne pourra pas l'identifier par sa fonction grammaticale. Il sera toujours possible par la suite d'introduire des corrections et d'ajouter des informations supplémentaires. Il serait souhaitable, par exemple, d'intégrer les attributions d'articles faites par des chercheurs au fil des années, notamment celles établies par Jacques Proust et John Lough, et les renseignements intégrés par Richard Schwab dans son *Inventory of Diderot's Encyclopédie*.

Dès le début nous avons décidé qu'il faudrait saisir à la fois le texte dans son ensemble et une image numérisée de chaque page. La transcription du texte permettra aux utilisateurs de chercher des occurrences de mot ou des phrases et de balayer le texte de plusieurs manières. Mais il fallait aussi l'image numérisée — c'est-à-dire un fac-similé de chaque page — parce que nous savions que la version « dactylographiée » comporterait certaines lacunes et insuffisances. En seront omis, par exemple, les tables, les partitions musicales et autres images qui émaillent le texte des articles. Et puis, la saisie comporte des erreurs. D'après quelques sondages préliminaires, nous estimons que pour les mots français et sans aucun travail de correction, le taux d'erreur est de l'ordre d'un sur mille caractères. Certaines confusions se reproduisent plus souvent que d'autres : celle entre « f » et « s » tel que l'on l'imprimait au XVIIIe siècle [f] est la plus fréquente (il y aurait une erreur sur cent « [f] » ou « f »). Une partie de ces erreurs pourra être corrigée automatiquement. Mais il nous est clair que pour arriver à un texte qui soit conforme aux normes de correction en vigueur dans les éditions de nos jours, il faudrait un investissement financier qui dépasse nos moyens. L'image numérisée de chaque page permettra aux chercheurs d'y remonter pour vérifier la transcription dactylographiée ainsi que de voir les éléments graphiques qui n'ont pas été saisis. Pour remonter de la transcription à l'image, on n'aura qu'à cliquer sur l'indication de page toujours présente à l'écran.

Développement d'un prototype de logiciel d'interrogation:

Nous avons mis en place un prototype de logiciel qui permet aux utilisateurs d'effectuer plusieurs types de recherches dans l'Encyclopédie et de spécifier la manière dont ils voudraient que ces résultats soient affichés sur l'écran. Le lecteur peut donc faire des demandes par:

Entrée. Exemple: « accélération ». Dans ce cas s'affiche sur l'écran une liste d'articles et sous-articles consacrés à ce sujet. Mais on pourra aussi engager une recherche plus large, contenant par exemple la série de lettres « nation » (ainsi entre autres : « aliénation », « carnation » et, bien sûr « nation »). Pour consulter l'article « nation », il suffit de cliquer sur ce mot dans la liste.

Auteur. Exemple : «Rousseau ». Apparaît sur l'écran une liste de tous les articles signés de Rousseau. En cliquant sur le titre d'un article, on arrive directement au texte.

Classification. Exemple : « chasse ». On obtient une liste de tous les termes que les encyclopédistes avaient classés dans cette catégorie (par exemple : « abois », « acharner », « amener », « braconnier », etc.). On pourrait aussi demander de voir tous les articles de « Rousseau » classés sous la rubrique « musique » ou tous les articles de Diderot portant sur la « morale ».

Fonction grammaticale. Exemple : « verbe ». Très souvent cette catégorie est utilisée conjointement avec une autre, par exemple avec classification. Ainsi si l'on demandait tous les « verbes » ayant été classés sous la rubrique « chasse », on obtiendrait parmi d'autres : « acharner », « affriander », « enlamer », « fouger » etc.

En plus des recherches par les différentes catégories indiquées ci-dessus, l'utilisateur pourra balayer l'ensemble de l'Encyclopédie pour trouver:

Des occurrences de mots. Exemples : « navire », « persécuter ». On pourra également rechercher toutes les formes d'un mot (navire, navires ; persécute, persécutent, persécutoit,

persécutoient, etc.), ou les radicaux d'une même racine (« navigable », « navigateur », « navigation »).

Des co-occurrences de mots. Exemple : « folie » et « religion » dans une même phrase, dans un même paragraphe ou bien dans un même article ou sous-article. Des locutions, des séries de mots. Exemple: « honnête homme ».

Ces résultats peuvent être affichés dans un contexte variable (une ligne ou bien environ dix lignes) et l'on peut toujours choisir de remonter des concordances directement au paragraphe, à la page, ou à l'article. Les opérateurs booléens permettent aux utilisateurs de raffiner leurs demandes (par exemple, au lieu de chercher la co-occurrence de « folie » et « religion », on pourrait, par exemple chercher « folie » ou « religion »).

Souvent pourtant la recherche d'un mot à travers l'ensemble de l'Encyclopédie livre une quantité d'occurrences quelque peu paralysantes. Si l'on demande toutes les occurrences de « religion » par exemple, on se trouve confronté à 3 484 occurrences ! Pour s'orienter dans cette masse d'information, le chercheur peut demander d'en voir la fréquence d'occurrence par article. Dans le cas de « religion », il verrait s'afficher une liste lui fournissant cette information :

113 RELIGION  
 71 CHRISTIANISME 64 ATHÉES  
 49 FANATISME  
 49 FOI  
 39 JUIF  
 37 POLYTHÉISME 31 OECONOMIE  
 29 ÉCLECTISME 27 LOI  
 26 ARISTOTÉLISME...  
 1 ZENDICISME 1 ZENDIK  
 1 ZIRICZÉE  
 1 ZONE  
 1 ZWOL

Il pourrait ensuite commencer par consulter les articles où le terme apparaît le plus souvent. Le même type de recherche permet de situer les références explicites à des auteurs ou personnages historiques. Quand on demande de voir les références à Newton, par exemple, on s'étonne de trouver le plus grand nombre d'occurrences non pas à « Newtonianisme » mais dans l'article « Wolstrobe ». En fait, c'est là que le chevalier de Jaucourt a choisi de raconter la vie du grand penseur anglais, parce que c'est sa ville natale !

Les Planches:

Dans la mesure du possible, nous relierons aux images correspondantes les renvois aux planches. Ainsi l'article AURORE BORÉALE renvoie à une belle planche qui « représente la fameuse aurore boréale de 1726, telle qu'elle parut à Paris le 19 octobre 1726 à huit heures du soir » (I, 888). Mais souvent ces renvois sont très imprécis. De plus, comme, lors de l'interruption de la publication du texte de l'Encyclopédie, la publication des images est devenue une fin en soi, il y a beaucoup d'images qui ne sont explicitement rattachées à aucun article. Pour cette raison, nous avons pris le parti de traiter les planches comme les articles ; elles sont divisées en « groupes » et « sous-groupes » d'après la « Table alphabétique des matières contenues dans les onze volumes du Recueil de Planches ». Ainsi trouvera-t-on dans le groupe « Agriculture et économie rustique » un ensemble de sous-groupes tels que « labourage », « semoirs », « façon des foins », « batteurs en grange ». Comme pour les articles, il suffira de cliquer sur ces indications pour consulter les planches et les légendes<sup>3</sup>.

Accès:

En plus du site de Chicago, nous avons l'intention de créer un site miroir en France en collaboration avec nos collègues à l'Institut National de la Langue Française, laboratoire du CNRS<sup>4</sup>. Nous espérons mettre l'Encyclopédie électronique à la disposition des chercheurs et du public au mois de septembre 1998. Les conditions d'accès restent à déterminer.

Remarques en conclusion:

Nous pensons que l'Encyclopédie électronique ouvrira de nouvelles pistes de recherches importantes. Pour la première fois, il sera possible de comparer systématiquement les correspondances et les écarts entre les définitions proposées des mots et l'utilisation de ceux-ci à travers l'ensemble de l'ouvrage. Dans peu de temps, nous espérons avoir des mécanismes permettant de suivre et d'afficher pour un article ou un groupe d'articles les ramifications des renvois sur plusieurs générations. Nous comptons aussi donner la possibilité d'examiner les renvois entre les différentes classifications de savoir. Face à l'idéal de l'arbre généalogique des connaissances proposé au début de l'Encyclopédie se dressera l'arbre bien plus tortueux des renvois et des références qui se font réellement dans les articles.

Si l'on poussait la logique de Diderot jusqu'au bout, il aurait fallu qu'il attendît d'avoir tout en main pour que chaque auteur pût se corriger en fonction du travail de définition que l'Encyclopédie se proposait de réaliser. Alors seulement le cercle aurait-il été bouclé. Mais, en fait, Diderot était plus que conscient qu'il poursuivait un idéal irréalisable.

Nous avons vû, à mesure que nous travaillions, la matière s'étendre, la nomenclature s'obscurcir, des substances ramenées sous une multitude de noms différents, les instruments, les machines & les manœuvres se multiplier sans mesure, & les détours nombreux d'un labyrinthe inextricable se compliquer de plus en plus. Nous avons vû combien il en coûtait pour s'assurer que les mêmes choses étoient les mêmes, & combien, pour s'assurer que d'autres qui paroissent très-différentes, n'étoient pas différentes. [...] Mais nous avons vû que de toutes les difficultés, une des plus considérables, c'étoit de le produire une fois, quelque informe qu'il fût, & qu'on ne nous raviroit pas l'honneur d'avoir surmonté cet obstacle. (V, 644)

L'Encyclopédie électronique aura sans doute ses défauts, mais elle aura le mérite d'exister et de se concevoir comme une édition dynamique orientée vers un idéal de perfection qui sera, peut-être, toujours à venir.

---

### ***Pour citer cet article***

Référence électronique

Robert Morrissey, John Iverson et Mark Olsen, « L'ENCYCLOPÉDIE DE DIDEROT SUR INTERNET\* », *Recherches sur Diderot et sur l'Encyclopédie* [En ligne], 25 | 1998, mis en ligne le 19 août 2007. URL : <http://rde.revues.org/index1911.html>

---

### ***Droits d'auteur***

Propriété intellectuelle

---

***Licence portant sur le document*** : Propriété intellectuelle