

Yoichi Sumi, Takeshi Koseki

Pour une édition critique informatisée de l'*Encyclopédie* : quelques précisions sur les métadonnées

« Et le chemin est long du projet à la chose ».
Molière

Le travail que nous présentons ici est le résultat de recherches dans le domaine de la lexicographie des dictionnaires, plus particulièrement du premier tirage de la première édition de l'*Encyclopédie*, recherches qui ont été menées notamment grâce à la subvention allouée par le Ministère de l'Education nationale du Japon.

La question des métadonnées de l'*Encyclopédie* a pris une importance majeure depuis quelques années, en raison du développement de ce qu'on appelle aujourd'hui l'« informatisation ». Pour expliquer les enjeux du problème, nous prendrons l'exemple des grands travaux de Schwab et de Rex réalisés au cours des années 1970 portant sur l'inventaire du dictionnaire parisien, dont le mérite principal a été de mettre en lumière le fait que le dictionnaire n'est pas seulement un témoin de la langue, mais aussi un objet de recherche qu'on peut analyser à partir de données concrètes et quantifiables. Simplement, l'inventaire des deux Américains se présente sous forme de gigantesques imprimés en plusieurs volumes, ce qui cantonne trop souvent repérages ou vérifications à des manipulations élémentaires et limitées.

Nous avons relancé cette question des métadonnées en en faisant un thème majeur dans un projet de recherche expliqué ci-après par notre collègue KOSEKI Takeshi. La réflexion présentée ici s'inscrit dans le cadre d'une démarche scientifique que nous appelons, au Japon, « la *Bunri-Yogô* », à savoir la fusion des approches littéraires et scientifiques. Ce domaine de recherche combine les apports de la philologie et de la technologie en privilégiant deux axes méthodologiques. Le premier part d'une lecture personnelle des articles encyclopédiques en décrivant leur contenu, suivant un certain nombre de critères, de « métadonnées », afin d'établir une base de données qui sera, espérons-le, un nouvel inventaire des articles et des documents relatifs à la genèse du grand dictionnaire.

Le rôle des scientifiques, en second lieu, est d'offrir aux littéraires, pour leur lecture analytique, un texte encyclopédique ayant déjà subi un traitement informatique.

Le dernier point sur lequel nous voudrions nous arrêter concerne le caractère fondamentalement collectif de notre travail. Autour de huit chercheurs qui participent activement à ce projet se réunissent plus de trente dix-huitiémistes bénévoles qui apportent leur collaboration à cette gigantesque œuvre, et nous ne saurions trop insister sur la générosité avec laquelle ils y participent.

L'enregistrement des métadonnées, effectué collectivement, devient une tâche incontestablement difficile, qui risque facilement de se limiter à un simple jeu fantaisiste. Il nous a donc semblé opportun d'unir nos efforts pour qu'un véritable accord de lecture s'établisse entre les chercheurs. Une formation quasi professionnelle s'impose pour réaliser la meilleure harmonisation possible.

Liste des co-auteurs :

Ao Yasuyoshi (U. du Kyushu), ARIGA Nobumichi (U. de Kyoto), BABA Akira (U. préfectorale de jeunes filles de Gunma), FUCHIDA Masashi (U. de Hitotsubashi), FUKUI Satoshi (ancien doctorant à l'U. SOPHIA), HARADA Yuri (U. Paris 4), HASHIMOTO Itaru (U. Hosei), HEMMI Tatsuo (U. de Niigata), IDA Hisashi (U. Aoyama-gakuin), INOUE Sakurako (U. Keio), KANAZAWA Fumio (U. de Tokyo), KITAGAKI Kiyoshi (Institut universitaire Caritas), KOJIMA Ryuji (U. Paris 10), KONNO Kiwahito (U. de Shizuoka), KOSEKI Takeshi (U. de Hitotsubashi), KOSHI Morihiko (U. de Shirayuri), KUMAMOTO Tetsuya (U. préfectorale d'Iwate), MANABE Kiyotaka (U. Keio), MASUDA Makoto (U. de Kyoto), Mitsushima Naoko (U. Médicale et Dentaire de Tokyo), NAKAMURA Hidetoshi (U. Paris 3), NISHIKAWA Junko (U. de Tokyo), OBA Shizue (U. Waseda), ODEISHI Atsuko (U. Waseda), OHASHI Kantaro (U. de Tokyo), OKUBO Naru (ancien assistant à l'U. Keio), ORIKATA Nozomi (U. Meiji), OZAWA Shinji (U. de technologie d'Aichi), SAITO Yamato (U. de Tokyo), SAKAKURA Yuji (U. Rikkyo), SEKITANI Kazuhiko (U. Kwansai), SHINOHARA Hiroharu (U. Keio), SHIRAKAWA Rie (U. Sophia), SHIRASU Takashi (U. Sophia), SUMI Yoichi (U. Chubu), SUZUKI Tadafumi (U. Waseda), SUZUKI Yuko (U. de Tokyo), TAGUCHI Takuomi (U. d'Utsunomiya), TAHARA Takahide (U. Mejiro), TAMADA Atsuko (U. Chubu), TERADA Motoichi (U. de la ville de Nagoya), TOKUNAGA Satoko (U. Keio), TOZUKA Mariko (U. Dokkyo), TSUJIBE Daisuke (U. Fukuoka), UCHIMURA Rina (Institut universitaire de jeunes filles d'Aoyama Gakuin), Watanabe Yoshihito (U. de technologie d'Aichi), YOSHINARI Yu (U. de Tokyo)

Yoichi Sumi
U. Chubu Nagoya

Notre équipe se propose d'établir, à partir des diverses données contenues dans les 17 volumes de « discours » de l'*Encyclopédie*, une base de données (ou plutôt de « métadonnées ») susceptible de servir d'outil de recherches à tous ceux qui abordent ce dictionnaire monumental et quelque peu chaotique.

Quels sont dès lors les éléments dignes d'y être relevés ? Tel a été le point de départ des discussions de notre équipe. Elles nous ont amenés à répondre que cette base de données devrait aboutir à une sorte d'édition critique de l'*Encyclopédie*. Pour donner une description suffisamment scientifique de chaque article, sont à relever les éléments suivants :

- A. Vedette de l'entrée
- B. Localisation de l'entrée (page, colonne, ligne)
- C. Numérotation
- D. Longueur de l'entrée
- E. Partie du discours et genre
- F. Désignant
- G. Attribution
- H. Renvois aux articles du texte
- I. Références aux planches
- J. Références bibliographiques explicites
- K. Citations
- L. Tableaux
- M. Caractères d'imprimerie spéciaux
- N. Remarques Schwab

Yoichi Sumi, chef de notre équipe, dont on vient de lire un avant-propos, a déjà présenté notre plan à l'occasion du XII^e Congrès International des Lumières (Montpellier), et en a donné un compte rendu dans un article¹. Nous apportons ici, en guise de rapport d'étape de notre projet, quelques précisions accompagnées d'exemples.

A. Vedette de l'entrée

Comme on le sait, Schwab avait remis en cause la notion d'« article » dans son *Inventory of Diderot's Encyclopédie*². Il arrive parfois qu'un article soit rédigé par plus de deux auteurs ou complété par une intervention de l'éditeur ; dans ces cas-là, chaque partie considérée comme de

1. Yoichi Sumi, « Sur l'importance des métadonnées de l'*Encyclopédie* – annonce d'un projet japonais », in Muriel Brot et Sante A. Viselli (éd.), *Lectures de Jacques Proust*, Presses Universitaires de la Méditerranée, 2008, p. 93-102.

2. Schwab, Richard N. With the Collaboration of Walter E. Rex, *Inventory of Diderot's Encyclopédie*, t. 1, in *SVEC*, 80, 1971, p. 19-20.

nature différente constitue une entrée (*entry* en anglais) indépendante. Le premier exemple d'intervention apparaît à la page 8 du tome premier ; l'article ABAISSEMENT *ou* ABATTEMENT se termine par un paragraphe commençant par « *En un mot³ ». Schwab avait donc divisé cet article en deux entrées, en retenant ces trois mots comme la vedette de la seconde. Cette méthode nous semble raisonnable, et nous l'avons donc suivie. En plus, ces interventions sont signalées par la vignette « IN » dans notre base de données, afin d'être distinguées des articles ordinaires.

Mais nous ne sommes pas entièrement d'accord avec Schwab, dont l'*Inventory* n'enregistre pas tous les articles, négligeant en effet ceux qui sont dits « vides ». On en verra le premier exemple à la figure 1 (ABAISSE une équation), article jugé par Schwab dénué de contenu, alors qu'il comporte d'après nous deux métadonnées : le désignant (*terme d'Algebre*) et le renvoi à un autre article (*Voyez ABAISSEMENT*). Il est vrai que l'*Inventory* de Schwab retient parfois ce genre d'articles avec la légende (V), mais sans être exhaustif. Il paraît donc nécessaire de récupérer ces articles « vides ».

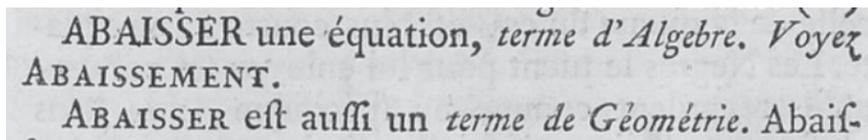


Figure 1

Enc., I, 8. Exemples d'article « vide » et d'article avec vedette en petites capitales. Toutes les images de l'*Encyclopédie* sont tirées de l'exemplaire de la Bibliothèque de l'Université Keio (Tokyo, Japon). Il s'agit de la première édition parisienne, sur laquelle nous menons notre travail.

Nous avons déjà distingué deux sortes d'entrées : les articles ordinaires et les interventions. S'y ajoute une troisième, à savoir les sous-articles marqués en petites capitales⁴. Quand il y a une suite d'articles sous la même vedette, seul le premier est marqué en grandes capitales, tandis que les autres le sont en petites. On voit à la figure 1 deux entrées dont la vedette commune est ABAISSE ; la graphie montre très bien que le premier en grandes capitales est l'article principal, et le second en

3. Certaines éditions omettent cependant l'astérisque.

4. Marie Leca-Tsiomis souligne l'importance de la petite capitale à usage polysémique dans l'*Encyclopédie*. Voir M. Leca-Tsiomis, « Numérisation et exactitude du texte encyclopédique : quelques propositions pour l'avenir », *RDE*, n° 31-32, 2002, p. 295.

petites le sous-article⁵. À l'instar des interventions, les sous-articles en petites capitales sont dans notre base étiquetés « PC ». Grâce au repérage des sous-articles et des interventions, on peut calculer le nombre des articles principaux. Parmi les 5445 entrées du tome premier, on compte 156 interventions et 1380 sous-articles (outre 15 éléments éditoriaux comme la dédicace, l'Avertissement, etc.) ; le tome premier comporte ainsi 3 894 articles principaux⁶.

B. Localisation de l'entrée (page, colonne, ligne)

L'*Encyclopédie* est un dictionnaire. Les articles y sont rangés en principe par ordre alphabétique. Il n'est pourtant pas si aisé d'atteindre un article recherché sur le support imprimé. La tête de chaque colonne ne présente que trois lettres au lieu de la première vedette en entier. Les vedettes elles-mêmes ne sont pas mises en gras. Il paraît donc utile d'informer le lecteur de la localisation de l'article en question. L'idéal serait d'indiquer sa localisation à la ligne près. Mais au cours de nos travaux en groupe, l'enregistrement des lignes n'a pas été sans difficultés, les éléments non littéraires (tableaux, figures, etc.) empêchant de compter les lignes de manière cohérente, ou le comptage des lignes entraînant aisément des erreurs d'inattention ; ce qui mettrait en cause la fiabilité de notre base de données. Pour le moment, tout en cherchant une meilleure solution, nous nous contentons de noter la page et la colonne.

C. Numérotation

Pour l'utilité des recherches, toutes les entrées doivent être numérotées. Puisque nous avons suivi en grande partie le système de Schwab sur la division des entrées, il est naturel de reprendre sa numérotation. Cependant, les entrées que nous ajoutons nécessitent des numéros supplémentaires. A l'exemple des numéros d'adresse, nous leur attribuons un numéro « bis ». Ainsi, à l'entrée ABAISSER une équation, située entre l'intervention commençant par « *En un mot » (numéro 61) et le sous-article ABAISSER (numéro 62), on accorde le numéro « 61,2 ». Quand une deuxième entrée est ajoutée, on lui donne un numéro « ter » (*,3), et ainsi de suite.

5. Il est à noter que l'*Inventory* de Schwab ne contient pas cet article ABAISSER une équation, considéré comme pur renvoi. Mais l'usage des grandes capitales justifie selon nous qu'on l'intègre dans la base de données.

6. Résultat néanmoins susceptible de modification, car certaines entrées restent difficiles à classer soit en articles principaux, soit en sous-articles.

D. *Longueur de l'entrée*

Schwab désigne la longueur des entrées par le nombre de colonnes. Par exemple, à l'article ART, qui commence à la ligne 61, colonne b, page 713 et finit à la ligne 70, colonne b, page 717, Schwab lui donne le chiffre de « 8,1 », à savoir huit colonnes et un dixième. Nous ne sommes pas tout à fait convaincus par cette notation certes pratique, mais imprécise. C'est pourquoi nous avons essayé d'évaluer la longueur par le nombre de lignes. Pour l'article ART, nous avons donné le chiffre de « 602 ». Mais, comme nous l'avons noté, il est difficile de compter les lignes de manière cohérente, et nous tâtonnons encore pour déterminer quelle en serait la meilleure notation.

E. *Partie du discours et genre*

L'*Encyclopédie* n'est pas uniquement un dictionnaire universel des idées et des choses. Elle est aussi un dictionnaire de langue, où des indications précisant la partie du discours et le genre figurent dans plus de quarante pour cent des articles du tome premier. Qu'un article porte cette indication ou non n'aura pas beaucoup d'intérêt en soi-même, mais pour qui mène une étude sur l'ensemble des articles du dictionnaire, les métadonnées sur la partie du discours et le genre pourraient apporter une vue nouvelle.

F. *Désignant*

On sait bien que le mot d'encyclopédie signifie l'enchaînement des connaissances. Pour assurer ce caractère à leur *Dictionnaire raisonné* qui suit l'ordre alphabétique, les éditeurs Diderot et d'Alembert élaborent un double dispositif : développer d'une part l'arbre des connaissances humaines et, de l'autre, désigner à quelle branche de cet arbre chaque article se rattache. C'est grâce aux désignants et aux renvois que le lecteur peut saisir les liens entre les articles.

Il y a plusieurs façons d'indiquer les désignants : la plus typique est de les mettre en italiques et entre parenthèses (Figure 2) ; ensuite, on les rencontre souvent introduits par l'expression « terme de... » (Figure 3) ; il arrive enfin que les désignants soient mêlés dans le texte, alors le plus souvent placés en début d'article avec une initiale en majuscule (Figure 4). On n'a nulle raison de limiter les désignants au cas le plus typique. Les deux expressions « (*Métier à faire des bas.*) » (Figure 2) et « Faiseur de bas au métier » (Figure 4) ne désignent-elles pas le même domaine des activités humaines ? Les deux sous-articles dont la vedette commune est ABATAGE (Figures 3 et 4) ne se distinguent-ils pas l'un de l'autre par ces « désignants » ? Une lecture attentive du texte est donc indispensable pour relever les désignants au sens large du terme.

ABATANT, (*Métier à faire des bas.*) On donne ce nom aux deux parties (85 96) (85 96) fem-

Figure 2

Enc., I, 10. Exemple de désignant typique mis en italique et entre parenthèses : (*Métier à faire des bas.*)

ABATAGE, *terme de Charpentier.* Quand on a une pièce de bois à lever, on pousse le bout d'un

Figure 3

Ibid. Exemple de désignant introduit par l'expression « terme de » : *terme de Charpentier.*

ABATAGE, *fixième manœuvre du Faiseur de bas au métier.* Elle consiste dans un mouvement assez lé-

Figure 4

Ibid. Exemple de désignant mêlé dans le texte : Faiseur de bas au métier.

G. Attribution

On n'entrera pas ici dans le détail des questions délicates concernant l'attribution des articles à leurs auteurs. Quand il y a des astérisques ou des marques d'auteur, nous les reproduisons tels qu'ils apparaissent dans le texte de l'*Encyclopédie*.

En outre, nous respectons les résultats de la recherche minutieuse menée par Schwab. Celui-ci intervient dans son *Inventory* chaque fois qu'il rencontre des irrégularités sur l'attribution. L'astérisque, par exemple, est un des signes les plus fiables renvoyant à Diderot. Mais il existe des articles non-étoilés rédigés par lui, et il en existe aussi d'étoilés qui ne sont pas de sa main. Il en est de même pour les autres collaborateurs. Dans tous ces cas, Schwab donne des remarques précises qu'il faut retenir dans notre base de métadonnées.

H. Renvois aux articles du texte

Les renvois à d'autres articles sont une des dispositions qui assurent à ce *Dictionnaire raisonné* son caractère d'enchaînement des connaissances. En principe, les renvois sont mis en petites capitales et précédés d'une indication comme « Voyez », « Voy. » ou tout simplement « V. ». Il est alors très facile de les reconnaître.

Mais, dans certains cas, ils sont donnés d'une manière différente, dont on trouvera quelques exemples dans l'article AGRICULTURE,

non signé mais attribué à Diderot dans le Discours préliminaire (p. xli). Tantôt, les renvois sont désignés en italiques au lieu des petites capitales : « Pour cultiver les terres avec avantage, il importe d'en connoître la nature : [...] On trouvera à l'article *Terre & Terroir* en général ce qui y a rapport, & *aux plantes différentes* le terroir & la culture qu'elles demandent⁷ ». Tantôt, il faut déchiffrer les renvois mêlés dans le texte : « Mais nous renvoyons à l'article GRAIN & à d'autres articles, ce qui a rapport à la récolte, à la vente, au commerce, au transport, à la police des grains⁸ ». On peut considérer que l'auteur renvoie ici le lecteur à l'ensemble des articles RÉCOLTE, VENTE, etc. Voici un autre exemple plus compliqué. Après cette formule : « Voulez-vous connoître le travail de votre année ? le voici. », l'auteur expose sur deux colonnes entières les opérations qui se font de janvier en décembre, et il termine son explication par ces mots : « Voyez le détail de chacune de ces opérations à leurs articles⁹ ». En pareil cas, tous les termes qui sont mentionnés en plus de 150 lignes font l'objet de renvois. Il est bien sûr difficile de préciser de quels articles il s'agit, mais l'important reste de montrer qu'il existe des indications qu'on peut considérer comme des renvois. La saisie de ce type de méta-données réclame un œil expérimenté qu'aucun ordinateur ne semble pouvoir remplacer.

I. Références aux planches

Il est assez facile de reconnaître la présence des références aux planches, puisque l'indication en est explicite, comme dans l'article *A, *caractere alphabétique* : « V. sur ces lettres nos Planches¹⁰ ». D'ailleurs, l'*Inventory* de Schwab ajoute la légende « PL » pour chaque entrée où l'on est invité à se référer aux planches.

D'autre part, il est difficile d'identifier ou parfois même de trouver les planches mentionnées. Par exemple, dans l'article AGRICULTURE, l'auteur Diderot parle de quatre figures des Planches sur l'agriculture : « On brûle les terres qui ne se labourent que tous les dix ans ; & voici comment on s'y prend : on coupe toute la surface en pieces les plus régulières qu'on peut, comme on les voit en *a a a* (fig. 1. Pl. d'agriculture) de huit à dix pouces en quarré [...] », « on les dresse ensuite les unes contre les autres, comme on voit en *b b b* (fig. 2.) », « on commence par élever une petite tour cylindrique, *a f b* (fig. 3.) » et enfin « l'on acheve le

7. *Enc.*, I, 185a.

8. *Ibid.*

9. *Enc.*, I, 187b.

10. *Enc.*, I, 6b.

fourneau avec les mêmes gasons en dôme, comme on voit (*fig. 4.*) en *e d*¹¹ ». Mais à quelles planches Diderot nous renvoie-t-il exactement ? Il mentionne seulement « *Pl. d'agriculture* », comme s'il n'y en avait qu'une qui traite de cet art. En fait, au début du premier tome du *Recueil de Planches*, 83 planches sont rassemblées sous la catégorie d'« AGRICULTURE ET ECONOMIE RUSTIQUE » ; et c'est dans la Planche V du Labourage qu'on trouve les quatre figures qui correspondent au passage de l'article du volume de discours.

J. Références bibliographiques explicites

Les études sur les sources de l'*Encyclopédie* sont d'une importance primordiale, car il existe un nombre considérable d'articles qui ne sont pas conçus spécialement pour ce dictionnaire, mais tirés de textes antérieurs. Qu'est-ce qui amène l'auteur à choisir tel ou tel texte au détriment d'autres écrits relevant du même domaine ? Quels passages y supprime-t-il ? Quels sont les ajouts de l'encyclopédiste ? C'est seulement à travers ces questionnements qu'on peut évaluer l'« originalité » de chaque article. D'où la nécessité de fournir aux chercheurs une base de données sur les références bibliographiques.

Celles-ci ne sont pas nécessairement explicites ; beaucoup d'articles sont pillés, bien sûr sans aucune indication supplémentaire. Mais pour l'établissement des métadonnées, nous nous contenterons des références explicites. Retrouver les sources implicites ou cachées constitue une étape ultérieure qui pourra être atteinte à partir de notre base de métadonnées.

Une des grandes difficultés de notre travail – la plus grande peut-être mis à part les problèmes d'ordre technologique – consiste à relever d'une manière non arbitraire ces références bibliographiques explicites. Il faut interpréter le texte, mais cela implique que les interprétations peuvent diverger d'un lecteur à l'autre. Nous ne sommes cependant pas si pessimistes à ce sujet : Yoichi Sumi et moi-même, ayant collationné séparément les cinquante premières pages du tome premier en marquant les références bibliographiques (ainsi que toutes les autres métadonnées), sommes quasiment parvenus au même résultat. Il paraît donc possible d'établir un critère objectif.

Une référence bibliographique peut se composer de deux éléments : le titre et l'auteur. Mais les encyclopédistes omettent fréquemment les titres, indiquant seulement un nom d'auteur. Les titres, même difficiles à identifier, ne passent pas inaperçus, tandis que les noms d'auteurs sans

11. *Enc.*, I, 188ab.

titres risquent davantage d'être négligés. Du moins faut-il noter que le risque d'interprétation divergente est plus grand pour ces noms d'auteurs sans titres.

K. Citations

Là encore nous devons nous borner aux références explicites en signalant seulement les passages mis entre guillemets ou en italique.

Quand les citations sont longues, il est embarrassant ou impossible de les enregistrer intégralement ; on doit alors se contenter d'indiquer quelques mots du début et de la fin, quoique ce mode d'enregistrement n'apporte pas de renseignements précis à ceux qui veulent savoir si telle ou telle phrase est citée dans l'*Encyclopédie*. Notre base de métadonnées sert donc seulement à signaler l'existence des citations.

L. Tableaux

Les tableaux ne sont pas réservés aux volumes de planches. On en trouve çà et là dans ceux de discours. Par exemple, l'article ABREGÉ contient un tableau en haut de la page 37 du tome premier (Figure 5), en l'occurrence une partition musicale¹². Comme il n'est pas possible de reproduire les tableaux, nous décrivons la nature de chacun.

A B R A B R 37

Sommier A C

Sommier C B

La disposition des rouleaux pour faire cette re-
partition est représentée dans la Figure.
ABREGER un Fief, terme de Jurisprudence féo-
dale, synonyme à *démembrer*; mais qui se dit fing-

Dans l'antiquité on appelloit les abréviations *notes*.
On les nomme encore de même dans les anciennes
inscriptions latines. (G)
ABRÉVIATIONS. Ce sont des lettres initiales

Figure 5

Enc., I, 37. Exemple de tableau (partition musicale).

12. Schwab ne manque pas de remarquer : « Includes engravings of musical staves ». Voir Schwab, o.c., t. 2, SVEC, 83, 1971, p. 23.

M. Caractères d'imprimerie spéciaux

La version numérisée de Redon ne contenait pas les caractères non latins, voire les caractères grecs, ce qui en gêne la lecture. Nous indiquons donc la présence des caractères d'imprimerie spéciaux qui provoqueraient des difficultés lors de la numérisation du texte de l'*Encyclopédie*. Outre les caractères en usage dans telle ou telle langue (les caractères hébraïques, par exemple), quelques signes sont parfois insérés dans le texte. La figure 6 en montre un exemple, à savoir deux rectangles entre trois majuscules de l'alphabet latins dans l'article ABREGÉ.

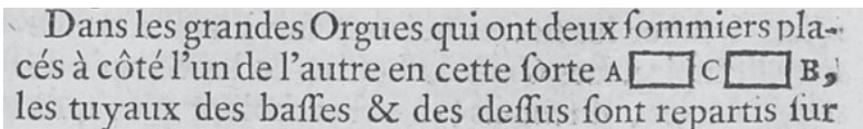


Figure 6

Enc., I, 36. Exemple de caractère d'imprimerie spécial.

N. Remarques Schwab

Les fruits des travaux de Schwab et de Rex sont précieux. Dans la mesure du possible, il vaudrait mieux enregistrer toutes les remarques en anglais de Schwab. On pourra recueillir, au fur et à mesure que le projet avance, toutes sortes d'annotations ou d'interventions de la part des chercheurs du monde entier.

Normalisation des données

Une fois toutes les métadonnées enregistrées, resterait quand même pour ceux qui consultent la base de données la difficulté posée par l'éventuelle incohérence des éléments recueillis. Il faut tout de suite souligner que cette incohérence est propre au texte original. C'est à nous de « normaliser » ces éléments selon des critères assez précis pour que la base de données fonctionne efficacement comme moteur de recherche de l'*Encyclopédie*.

Nous espérons que notre projet aboutira un jour à une édition critique informatisée de l'*Encyclopédie*. Comme les chercheurs ayant accès à la véritable première édition parisienne ne sont pas nombreux, il est important d'en conserver l'orthographe originale dans notre base de métadonnées. D'autre part, trop de fidélité suscitera divers inconvénients qui en rendraient la consultation incommode.

Dans l'article ART, par exemple, l'auteur Diderot parle de « Montagne » à trois reprises. Il s'agit bien sûr de Montaigne, auteur des *Essais* (c'est précisément l'expression utilisée par Diderot dans cet article). Depuis *les Pensées philosophiques* jusqu'à l'*Essai sur Sénèque*, Diderot ne cesse de transcrire le nom du moraliste sans i. La bonne conscience professionnelle exige que nous gardions cette habitude de Diderot. Mais si l'on veut relever tous les passages de l'*Encyclopédie* où est mentionné l'auteur des *Essais*, quelles conséquences aurait le maintien de l'orthographe originale ? Est-ce qu'on entre « Montaigne » pour faire cette recherche ? On ignorera que l'article ART parle de lui. Lancera-t-on alors une deuxième recherche avec le mot « Montagne » ? Mais on ne pourra jamais prévoir toutes les variantes... D'où la nécessité absolue de la normalisation, en sorte que les utilisateurs puissent trouver le bon résultat en une seule recherche à partir de la forme la plus courante. Autrement dit, à nous de trouver une formule qui permette de montrer toutes les occurrences de « Montaigne » et de « Montagne » à partir de l'orthographe normalisée.

Les références bibliographiques ne sont pas, loin de là, les seules métadonnées qu'il faut normaliser, puisqu'en fait la plupart d'entre elles doivent l'être. Nous avons déjà montré que « *Pl. d'agriculture* » correspond à la Planche V du Labourage : il s'agit là d'une normalisation. Les renvois aux autres articles doivent quelquefois être explicités (*GRAIN & d'autres articles*). Quant aux marques d'auteurs, on ne peut pas les supposer connues de tous les utilisateurs de la base de données : il faut convertir la marque (*O*) en d'Alembert, (*G*) en Mallet, (*Z*) en Bellin, etc. Il en va de même pour les désignants, la partie du discours et le genre : ainsi, « *Métier à faire des bas* » et « *Faiseur de bas au métier* » doivent être unis ; « s. m. », « s. masc. », « sub. m. » et « subst. m. » doivent être regroupés.

Au terme de cette normalisation, notre base de métadonnées sera munie de contenus suffisants pour fonctionner comme moteur de recherche de l'*Encyclopédie*. D'ici sa mise en œuvre, subsistent cependant des problèmes d'ordre technologique que nous espérons résoudre en collaborant avec le laboratoire de Shinji Ozawa, spécialiste en sciences informatiques. Au fil de ce travail collectif, nous sommes prêts à recueillir tous les avis des chercheurs pour améliorer et enrichir notre projet.

Takeshi Koseki
Université Mitotsubashi