

Rééditer un texte
philosophico-linguistique : la
digitalisation de *An Essay towards a
Real Character, and a Philosophical
Language* de John Wilkins (1668)

Alexis Huchelmann

INTRODUCTION

Comment rendre la littérature d'hier plus accessible ?

Il ne s'agit pas de multiplier des duplicatas de vieux textes et de les distribuer à tous vents. Lorsque l'on parle d'accessibilité, la qualité importe autant que la quantité. Ce qui était lisible hier ne l'est plus forcément aujourd'hui : sans parler des dégradations techniques qu'ont pu subir les supports, il faut prendre en compte les différences langagières entre les époques, et les différences entre les conventions typographiques (ainsi les abréviations). Il importe de capturer avant tout le sens de l'original avant son apparence. À cette fin, on transcrit le texte en caractères informatiques standards ; et pour encoder encore plus d'information, on l'organise dans un schéma standardisé.

Ranger les choses dans un schéma était aussi le but de John Wilkins dans *An Essay Towards a Real Character, and a Philosophical Language* en 1668. Dans cet essai, l'évêque de Chester proposa à ses pairs de la Royal Society une nouvelle langue basée sur une hiérarchie rationnelle du monde, en accord avec les développements de la science de l'époque. Le livre comprend plus de quarante tableaux hiérarchiques de mots classés par sens, un thésaurus regroupant synonymes et paronymes de la langue anglaise de l'époque et une proposition d'écriture idéographique. Un simple duplicata ne rendrait pas justice à la masse d'information présentée.

TABLE DES FIGURES

IV.1	Exemple de caractères chinois	17
IV.2	Le Notre-Père	18

NATURE ET HISTOIRE DE LA DIGITALISATION

I Définition

1.i Digitalisation vs numérisation

La digitalisation est à distinguer de la numérisation. Dans les deux cas, il y a création d'un fichier électronique, mais l'information n'est pas la même : la numérisation est une photographie du texte, la digitalisation convertit l'information de l'original dans un format conservant l'information sémantique : quels sont les caractères composant le texte, quelle est sa hiérarchie interne.

1.ii Différentes encodages

En passant de l'imprimé au numérique, il s'agit de convertir les caractères du texte en entités lisibles par la machine à des fins de manipulation. Aux débuts de l'ère informatique,

2 Emploi

2.i Pour la sauvegarde du patrimoine littéraire

Projet Gutenberg

Le 4 juillet 1971, Michael Hart pose la première pierre du projet Gutenberg en mettant en ligne la Déclaration de l'indépendance des États-Unis depuis les ordinateurs de l'université de l'Illinois. À l'époque, les réseaux informatiques sont limités aux grandes universités américaines, et la saisie et la diffusion des textes sont relativement lentes; mais Hart est vite rejoint par des volontaires, et le rythme de publication augmente d'année en année. De dix textes en 1989, on passe à mille en 1997, puis à dix mille dès 2003 (LEBERT 2010).

Comme tous les textes sont mis à disposition librement sur des serveurs situés aux États-Unis, ils doivent respecter les lois américaines sur le copyright. Du fait de la diffusion mondiale du projet, il peut entrer en conflit avec les lois locales sur le droit d'auteur : ainsi, en 2015, S. Fischer Verlag, un éditeur allemand, demanda le retrait de dix-huit titres de la bibliothèque en ligne que les lois locales protégeaient encore. Un compromis fut trouvé en 2018 d'après lequel la fondation Gutenberg supprima l'accès à toutes les pages de son domaine aux adresses IP allemandes (*Court Order to Block Access from Germany* 2018).

Google Books

En 2002, Google commence à scanner

Internet Archive

L'organisation à but non-lucrative Internet Archive, surtout connue pour son travail de sauvegarde des sites web (Wayback Machine) présente également un volet de stockage de numérisations, dont des livres. En octobre 2019, près de 2500000 livres étaient présents sur le site, une moitié librement consultable par tous (livres hors-copyright), l'autre pouvant être lue en ligne par les inscrits au service (une personne à la fois peut consulter un livre ¹).

Internet Archive propose un service payant de numérisation des collections des institutions. Avec 33 centre de scanners répartis dans le monde entier, l'organisation peut numériser mille pages par jour, qui sont transformées en PDF couleur de qualité haute permettant la recherche de texte (grâce à la reconnaissance optique des caractères) ². Elle permet aussi aux utilisateurs enregistrés de téléverser leurs propres numérisations dans une collection open source « Community Texts » ³.

Wikisource

La fondation Wikimedia, qui chapeaute entre autres l'encyclopédie Wikipédia, lança en 2003 Wikisource pour stocker des textes libres de droits (domaine public ou œuvres libres). Ils sont digitalisés, c'est à dire que le texte brut est mis en forme selon la syntaxe wiki. Comme Wikisource est un projet collaboratif ouvert, il se

1. <https://archive.org/details/books> (consulté le 15 octobre 2019)

2. <https://archive.org/scanning> (consulté le 15 octobre 2019)

3. <https://archive.org/details/opensource> (consulté le 15 octobre 2019)

pose la question de la validité des sources; une extension du logiciel de Wikimedia permet de comparer le texte à une image numérisée servant de source ⁴.

Initiatives nationales

À côté de ces projets d'envergure mondiale (quoique basés aux USA), certains pays fournissent leurs propres efforts pour numériser leurs fonds patrimoniaux. Ainsi, la Bibliothèque nationale de France gère depuis 1997 le portail Gallica, mettant en ligne les images des livres des collections nationales accompagnés d'une notice descriptive ⁵. La plateforme Europeana lancé en 2008 par la Commission européenne sert d'interface de recherche permettant d'accéder aux fonds numériques des institutions culturelles européennes partenaires, dont Gallica ⁶.

2.ii Dans les humanités

La digitalisation est très importante dans les humanités, où elle permet de sauvegarder des textes dont le support a vieilli, et de fournir un matériel de travail interactif aux chercheurs.

4. <https://en.wikipedia.org/wiki/Wikisource> (consulté le 15 octobre 2019)

5. <https://gallica.bnf.fr>

6. <https://www.europeana.eu/portal/fr>

LE SCHÉMA D'ENCODAGE TEI

I Un standard SGML et XML

Le SGML (*Standard Generalized Markup Language*) est un langage de description à balises apparu en 1986 qui permet de définir des éléments et leurs relations hiérarchiques les uns aux autres. Dès 1987, un groupe de chercheurs new-yorkais utilise ce standard comme base sur laquelle créer des éléments de descriptions de toutes sortes de textes : manuscrits, dictionnaires, pièces de théâtres, lettres, etc, de toutes époques et de toutes langues. Ils publient leurs consignes en 1990 en tant que TEI P1 (Text Encoding Initiative Proposal 1), et la P3 publié en 1994 contient plus de 600 éléments définis sur 1292 pages (VANHOUTTE 2004).

Créé en 1998, le XML (*eXtented Markup Language*) est un dérivé de SGML à la syntaxe plus versatile. Moins d'éléments sont définis à la base, et il exige un encodage Unicode là où le SGML laissait les utilisateurs choisir leur jeu de caractères. TEI P4, publié en 2002, est compatible avec cette nouvelle spécification autant qu'avec le SGML; mais en 2007, la version P5 est exclusivement taillée pour le XML. Elle a des exigences plus strictes quand au type de contenus que peuvent présenter les attributs et les balises, introduit de nouveaux éléments permettant de décrire en détail les entités nommées par des noms propres (personnes, lieux, peuples,

etc.), et s'ouvre davantage à la customisation en améliorant la modularité des éléments (WITTERN, CIULA et TUOHY 2009). En date du 16 juillet 2019, les indications TEI P5 sont décrites sur 1934 pages.

2 TEI Lite

La spécification TEI complète n'est jamais employée dans son entièreté dans un projet. Des sous-spécifications moins exhaustives, avec quelquefois des ajouts modulables, sont destinées à des projets à l'envergure bien définie, comme la description des documents épigraphiques (EpiDoc), la mise en forme d'articles de journal (jTEI Article), ou la description de corpus linguistiques (Corpus)⁷. L'une d'entre elle est décrite comme pouvant « couvrir 90% des besoins de 90% de la communauté » : il s'agit de TEI Lite⁸. Comme le laisse entendre son slogan, TEI Lite est souvent la seule version du schéma que rencontre les utilisateurs, sa description ne fait que 294 pages.

3 Potentiel descriptif

Un document TEI, compris entre les balises `<TEI>...</TEI>` a deux éléments principaux : l'en-tête `<teiHeader>` et le texte `<text>`.

7. <https://tei-c.org/guidelines/customization/>

8. <https://tei-c.org/guidelines/customization/Lite/>

LE LIVRE DE JOHN WILKINS

I L'auteur

John Wilkins (1614-1672) était un ecclésiastique anglican et scientifique anglais aux intérêts éclectiques, un des fondateurs de la Royal Society.

1.i Biographie

Né en 1614 dans une famille d'artisans dans le Northamptonshire, Wilkins fut éduqué à Oxford d'où il sorti avec une maîtrise en 1634. En 1637, il entra dans les ordres dans l'Église d'Angleterre, d'abord comme curé dans son village natal de Fawsley, puis réussit à obtenir la place de chapelain auprès de William Fiennes, vicomte de Say and Seale. Lors de l'éclatement de la première révolution anglaise en 1642, Wilkins se réfugia à Londres où il devint chapelain d'abord chez Lord Berkeley, puis chez le prince-électeur palatin Charles Louis, neveu du roi Charles I^{er}. C'est là qu'il participa à l'activité intellectuelle d'un cercle de savants oxfordiens, en exil dans la capitale après la prise de la ville par les forces parlementaires en 1646; on y trouvait entre autres les médecins Charles Scarborough et William Harvey, l'astronome Samuel Foster, et le mathématicien John Wallis. Ce club informel fut la base de ce qui allait devenir, lors du retour à Oxford de ces savants en 1648, la

Royal Society, constituée à Wadham College dont Wilkins était devenu le directeur. En 1659, Wilkins devient directeur de Trinity College à Cambridge, d'après les vœux de Richard Cromwell, le Lord Protecteur, dont il avait épousé la tante Robina en 1656. Mais il ne gardera ce poste que dix mois, puisqu'en 1660 Charles II reprend le pouvoir en Angleterre et les ex-partisans de Cromwell perdent les faveurs des autorités. Wilkins revint vers Londres, et occupa une variété de postes ecclésiastiques : prêcheur à Gray's Inn, vicaire de St Lawrence Jerry, doyen de Ripon, tout en occupant un poste de secrétaire à la Royal Society. En 1668 enfin, il retrouva la pleine faveur du roi et se vit offrir l'évêché de Chester; quatre ans plus tard, en 1672, il mourut, et fut enterré à St Lawrence Jerry.

1.ii *Autres écrits*

Le premier livre de John Wilkins fut un traité d'astronomie titré *The Discovery of a World in the Moone* en 1638.

En 1648, il publia un petit traité de cryptographie intitulé *Mercurie; or the Secret and Swift Messenger*, ainsi que *Mathematical Magick* qui introduit le lecteur aux sciences mécaniques.

2 L'essai

2.i *Contenu*

L'imprimé contient deux livres, le premier étant l'essai à proprement parler, et le second étant le dictionnaire faisant corres-

pondre les mots de la langue anglaise aux tables philosophiques développées précédemment.

Le premier livre est divisé en quatre parties : la première, le prologue, présente une réflexion linguistique et philosophique sur l'état des langues naturelles contemporaines, leurs origines et leurs évolutions (appelées « changements et corruptions »), l'usage des différents systèmes d'écriture et leurs défauts. Wilkins propose devant ces imperfections de bâtir une nouvelle langue sur des bases rationnelles et philosophiques, en commençant par définir quels concepts et objets sont à nommer.

Les chapitres de la deuxième partie présentent les quarante *genres* ou têtes sémantiques, parmi lesquelles six forment des catégories à part (« notions transcendantales générales », « relations transcendantales mixtes », « relations transcendantales d'action », « la communication », « le Créateur », et « le monde ») et le reste est divisé en cinq familles (« substance », « quantité », « qualité », « action », et « relation »). Ces genres comprennent par exemple « les pierres », « les oiseaux », « les mesures », ou « les titres ecclésiastiques » et sont divisés à leur tour en six *différences* chacun, elles-mêmes contenant plusieurs espèces.

La troisième partie de l'essai concerne la grammaire, que Wilkins sépare en deux, la grammaire *naturelle* ou universelle concernant toutes les langues, et la grammaire *instituée* ou *particulière* propre à chacune. Les différentes parties du discours de la grammaire classique sont passées en revue, ainsi que les moyens de les dériver ; puis on se penche sur la phonétique ou *orthographe* comme l'entend Wilkins. Un schéma de transcription phonétique est produit, exemplifié par les versions anglaises du Notre-Père et du Credo.

Enfin, la quatrième partie

2.ii Genèse et rédaction

Le XVII^e siècle en Europe voit des savants se pencher sur la question de la langue.

2.iii Réception

3 Littérature scientifique

George Edmonds

Borges

Okrent

4 Choix de l'ouvrage

4.i Intérêt du sujet

L'intérêt porté par le grand public aux langues construites changea de force et de nature au cours des siècles : de la quête du langage originel, en passant par la création de langues plus rationnelles et la recherche d'un instrument de communication international, on assiste aujourd'hui à la vogue des langues fictionnelles dans les films et séries télévisées (HUCHELMANN 2019).

À côté de cela, une communauté d'inventeurs de langues s'est retrouvée sur Internet et produit des échanges en nombre toujours croissant autour de leurs propres créations et celles des temps passés. Leurs efforts sont en majorité orientés vers leur propre plaisir, et ils suivent avec intérêt les travaux des autres. Bien qu'ils aient conscience des réalisations de leurs prédécesseurs avant l'âge digital, seules les plus connues d'entre elles leur sont accessibles dans

une version électronique : c'est le cas par exemple de l'espéranto⁹ et du volapük¹⁰, deux langues à vocation international dont la communauté de locuteur désire la dissémination, ou des langues elfiques de J.R.R. Tolkien. Les projets n'ayant pas eu autant de succès sont parfois, s'ils ont été publiés en tant que livres, numérisés dans diverses collections.

4.ii Précédentes numérisations et digitalisations

Numérisation

Il existe plusieurs numérisations depuis Google Books; seules trois peuvent être lues intégralement, provenant d'originaux de la bibliothèque nationale de la République Tchèque¹¹, de la bibliothèque municipale de Lyon¹² et de la bibliothèque de l'État de Bavière¹³. Elles permettent la recherche de texte, mais la reconnaissance automatique des caractères est très mal adaptée à la police et aux ligatures employées ici, et ne permet de rien extraire qui soit utilisable. Lues en ligne, ces images sont en couleur; une fois téléchargées en tant que PDF, elles sont en noir et blanc, ce qui baisse la qualité de la reproduction.

9. esperanto.net (consulté le 19 octobre 2019)

10. volapuk.com (consulté le 19 octobre 2019)

11. <https://books.google.fr/books?id=Bu7pwpr5qBcC> (consulté le 16 octobre 2019)

12. <https://books.google.fr/books?id=BCctZjBtiEYC> (consulté le 16 octobre 2019)

13. <https://books.google.fr/books?id=Q85TAAAcAAJ> (consulté le 16 octobre 2019)

Digitalisation de la Text Creation Partnership

La Text Creation Partnership est une initiative conjointe de l'Université du Michigan, de la bibliothèque Bodléienne d'Oxford, le portail en ligne ProQuest et le Council on Library and Information Resources pour digitaliser leurs ouvrages déjà numérisés selon un même standard et mettre ces ressources en ligne. Les trois collections disponibles sont Early English Books Online (1473-1700), Eighteenth-Century Collections Online et Early American Imprints. Plutôt que d'employer la reconnaissance optique des caractères, l'initiative employa une centaine de personnes pour encoder les textes manuellement dans un schéma SGML dérivé de TEI P3, qui fut plus tard converti en XML dérivé de TEI P4¹⁴.

An Essay towards a Real Character fut saisi en 2004 par Mona Logarbo. Il est divisé en chapitres et l'organisation interne du texte est assez fidèlement respectée, par exemple les listes hiérarchiques d'éléments apparaissent sous formes de listes HTML imbriquées, les tableaux apparaissent en tant que tel, et les annotations marginales sont intégrées dans le texte comme appels de note. Les changements de pages et de lignes dans l'édition originale sont toujours indiqués. Cependant, les images ne sont pas incluses, pas plus que les caractères non-latins.

5 Fac-similé ou réédition ?

Une digitalisation peut consister en une copie du texte de base, où le nombre et l'organisation des pages restent les mêmes, et le

14. <https://textcreationpartnership.org/about-the-tcp/>
(consulté le 17 octobre 2019)

texte transcrit le plus fidèlement possible.

5.i Apporter quelque chose de neuf

À la digitalisation du Text Encoding Project, il y a plusieurs choses que l'on pourrait ajouter :

1. Une meilleure gestion des caractères non-latins, dont l'écriture philosophique.
2. Des liens externes vers les ouvrages cités dans le texte.
3. Des conversions dynamiques vers le mot en langue philosophique depuis les éléments du thésaurus.

5.ii Les limites de la réédition

DÉFIS TECHNIQUES POSÉS PAR LE TEXTE

1 Caractéristiques techniques du livre originel

Le livre fait plus de 600 pages. Il est imprimé sur du papier et relié de cuir.

2 Caractères spéciaux

Wilkins présente beaucoup de langues dans le cadre de sa démonstration. Certaines de ces langues, comme l'allemand et le danois, ont d'autres caractéristiques typographiques que le texte alentours, par exemple elles sont composées en lettres gothiques. Dans ses descriptions phonétiques, Wilkins utilise aussi des caractères spéciaux, comme la lettre ⟨α⟩ (alpha latin, différent de ⟨α⟩ alpha grec), le digramme ⟨8⟩ (pour ⟨ou⟩) et un ⟨y⟩ agrémenté d'une ligne souscrite. Cette dernière lettre, qui contrairement aux autres n'est pas dans les spécifications Unicode, est remplacée dans la digitalisation EEB par ⟨ỵ⟩.

2.i *Caractères non-latins*

Un des tableaux donne des exemples de caractères chinois avec prononciation (WILKINS 1668a, p. 451). L'encodage d'Early English Books Online, pour chacun d'entre eux, signale in non-

FIGURE IV.1 – Exemple de caractères chinois



Latin alphabet. On perd ainsi de l'information, alors que de nos jours Unicode permet – entre autres – l'encodage de tous les systèmes d'écriture dérivés des logogrammes chinois.

Le problème est qu'à l'époque aucun imprimeur en Europe ne disposait de polices de caractères chinois, et ce qu'on retrouve dans ETRC sont des gravures faites à main levée. Pour retrouver le caractère représenté, les recherches graphiques doivent être accompagnées de recherches phonétiques, en espérant que la romanisation fournie par Wilkins permette de retrouver quelle langue chinoise est ainsi décrite.

2.ii *L'écriture philosophique*

En plus d'employer des systèmes d'écriture non-européens, Wilkins en invente un dans le cadre de sa recherche d'une langue par-

FIGURE IV.2 – Le Notre-Père

The Lords Prayer.

١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠ ١١ ١٢ ١٣ ١٤ ١٥ ١٦ ١٧ ١٨ ١٩ ٢٠ ٢١ ٢٢ ٢٣ ٢٤ ٢٥ ٢٦ ٢٧ ٢٨ ٢٩ ٣٠ ٣١ ٣٢ ٣٣ ٣٤ ٣٥ ٣٦ ٣٧ ٣٨ ٣٩ ٤٠ ٤١ ٤٢ ٤٣ ٤٤ ٤٥ ٤٦ ٤٧ ٤٨ ٤٩ ٥٠ ٥١ ٥٢ ٥٣ ٥٤ ٥٥ ٥٦ ٥٧ ٥٨ ٥٩ ٦٠ ٦١ ٦٢ ٦٣ ٦٤ ٦٥ ٦٦ ٦٧ ٦٨ ٦٩ ٧٠ ٧١ ٧٢ ٧٣ ٧٤ ٧٥ ٧٦ ٧٧ ٧٨ ٧٩ ٨٠ ٨١ ٨٢ ٨٣ ٨٤ ٨٥ ٨٦ ٨٧ ٨٨ ٨٩ ٩٠ ٩١ ٩٢ ٩٣ ٩٤ ٩٥ ٩٦ ٩٧ ٩٨ ٩٩ ١٠٠ ١٠١ ١٠٢ ١٠٣ ١٠٤ ١٠٥ ١٠٦ ١٠٧ ١٠٨ ١٠٩ ١١٠ ١١١ ١١٢ ١١٣ ١١٤ ١١٥ ١١٦ ١١٧ ١١٨ ١١٩ ١٢٠ ١٢١ ١٢٢ ١٢٣ ١٢٤ ١٢٥ ١٢٦ ١٢٧ ١٢٨ ١٢٩ ١٣٠ ١٣١ ١٣٢ ١٣٣ ١٣٤ ١٣٥ ١٣٦ ١٣٧ ١٣٨ ١٣٩ ١٤٠ ١٤١ ١٤٢ ١٤٣ ١٤٤ ١٤٥ ١٤٦ ١٤٧ ١٤٨ ١٤٩ ١٥٠ ١٥١ ١٥٢ ١٥٣ ١٥٤ ١٥٥ ١٥٦ ١٥٧ ١٥٨ ١٥٩ ١٦٠ ١٦١ ١٦٢ ١٦٣ ١٦٤ ١٦٥ ١٦٦ ١٦٧ ١٦٨ ١٦٩ ١٧٠ ١٧١ ١٧٢ ١٧٣ ١٧٤ ١٧٥ ١٧٦ ١٧٧ ١٧٨ ١٧٩ ١٨٠ ١٨١ ١٨٢ ١٨٣ ١٨٤ ١٨٥ ١٨٦ ١٨٧ ١٨٨ ١٨٩ ١٩٠ ١٩١ ١٩٢ ١٩٣ ١٩٤ ١٩٥ ١٩٦ ١٩٧ ١٩٨ ١٩٩ ٢٠٠ ٢٠١ ٢٠٢ ٢٠٣ ٢٠٤ ٢٠٥ ٢٠٦ ٢٠٧ ٢٠٨ ٢٠٩ ٢١٠ ٢١١ ٢١٢ ٢١٣ ٢١٤ ٢١٥ ٢١٦ ٢١٧ ٢١٨ ٢١٩ ٢٢٠ ٢٢١ ٢٢٢ ٢٢٣ ٢٢٤ ٢٢٥ ٢٢٦ ٢٢٧ ٢٢٨ ٢٢٩ ٢٣٠ ٢٣١ ٢٣٢ ٢٣٣ ٢٣٤ ٢٣٥ ٢٣٦ ٢٣٧ ٢٣٨ ٢٣٩ ٢٤٠ ٢٤١ ٢٤٢ ٢٤٣ ٢٤٤ ٢٤٥ ٢٤٦ ٢٤٧ ٢٤٨ ٢٤٩ ٢٥٠ ٢٥١ ٢٥٢ ٢٥٣ ٢٥٤ ٢٥٥ ٢٥٦ ٢٥٧ ٢٥٨ ٢٥٩ ٢٦٠ ٢٦١ ٢٦٢ ٢٦٣ ٢٦٤ ٢٦٥ ٢٦٦ ٢٦٧ ٢٦٨ ٢٦٩ ٢٧٠ ٢٧١ ٢٧٢ ٢٧٣ ٢٧٤ ٢٧٥ ٢٧٦ ٢٧٧ ٢٧٨ ٢٧٩ ٢٨٠ ٢٨١ ٢٨٢ ٢٨٣ ٢٨٤ ٢٨٥ ٢٨٦ ٢٨٧ ٢٨٨ ٢٨٩ ٢٩٠ ٢٩١ ٢٩٢ ٢٩٣ ٢٩٤ ٢٩٥ ٢٩٦ ٢٩٧ ٢٩٨ ٢٩٩ ٣٠٠ ٣٠١ ٣٠٢ ٣٠٣ ٣٠٤ ٣٠٥ ٣٠٦ ٣٠٧ ٣٠٨ ٣٠٩ ٣١٠ ٣١١ ٣١٢ ٣١٣ ٣١٤ ٣١٥ ٣١٦ ٣١٧ ٣١٨ ٣١٩ ٣٢٠ ٣٢١ ٣٢٢ ٣٢٣ ٣٢٤ ٣٢٥ ٣٢٦ ٣٢٧ ٣٢٨ ٣٢٩ ٣٣٠ ٣٣١ ٣٣٢ ٣٣٣ ٣٣٤ ٣٣٥ ٣٣٦ ٣٣٧ ٣٣٨ ٣٣٩ ٣٤٠ ٣٤١ ٣٤٢ ٣٤٣ ٣٤٤ ٣٤٥ ٣٤٦ ٣٤٧ ٣٤٨ ٣٤٩ ٣٥٠ ٣٥١ ٣٥٢ ٣٥٣ ٣٥٤ ٣٥٥ ٣٥٦ ٣٥٧ ٣٥٨ ٣٥٩ ٣٦٠ ٣٦١ ٣٦٢ ٣٦٣ ٣٦٤ ٣٦٥ ٣٦٦ ٣٦٧ ٣٦٨ ٣٦٩ ٣٧٠ ٣٧١ ٣٧٢ ٣٧٣ ٣٧٤ ٣٧٥ ٣٧٦ ٣٧٧ ٣٧٨ ٣٧٩ ٣٨٠ ٣٨١ ٣٨٢ ٣٨٣ ٣٨٤ ٣٨٥ ٣٨٦ ٣٨٧ ٣٨٨ ٣٨٩ ٣٩٠ ٣٩١ ٣٩٢ ٣٩٣ ٣٩٤ ٣٩٥ ٣٩٦ ٣٩٧ ٣٩٨ ٣٩٩ ٤٠٠ ٤٠١ ٤٠٢ ٤٠٣ ٤٠٤ ٤٠٥ ٤٠٦ ٤٠٧ ٤٠٨ ٤٠٩ ٤١٠ ٤١١ ٤١٢ ٤١٣ ٤١٤ ٤١٥ ٤١٦ ٤١٧ ٤١٨ ٤١٩ ٤٢٠ ٤٢١ ٤٢٢ ٤٢٣ ٤٢٤ ٤٢٥ ٤٢٦ ٤٢٧ ٤٢٨ ٤٢٩ ٤٣٠ ٤٣١ ٤٣٢ ٤٣٣ ٤٣٤ ٤٣٥ ٤٣٦ ٤٣٧ ٤٣٨ ٤٣٩ ٤٤٠ ٤٤١ ٤٤٢ ٤٤٣ ٤٤٤ ٤٤٥ ٤٤٦ ٤٤٧ ٤٤٨ ٤٤٩ ٤٥٠ ٤٥١ ٤٥٢ ٤٥٣ ٤٥٤ ٤٥٥ ٤٥٦ ٤٥٧ ٤٥٨ ٤٥٩ ٤٦٠ ٤٦١ ٤٦٢ ٤٦٣ ٤٦٤ ٤٦٥ ٤٦٦ ٤٦٧ ٤٦٨ ٤٦٩ ٤٧٠ ٤٧١ ٤٧٢ ٤٧٣ ٤٧٤ ٤٧٥ ٤٧٦ ٤٧٧ ٤٧٨ ٤٧٩ ٤٨٠ ٤٨١ ٤٨٢ ٤٨٣ ٤٨٤ ٤٨٥ ٤٨٦ ٤٨٧ ٤٨٨ ٤٨٩ ٤٩٠ ٤٩١ ٤٩٢ ٤٩٣ ٤٩٤ ٤٩٥ ٤٩٦ ٤٩٧ ٤٩٨ ٤٩٩ ٥٠٠ ٥٠١ ٥٠٢ ٥٠٣ ٥٠٤ ٥٠٥ ٥٠٦ ٥٠٧ ٥٠٨ ٥٠٩ ٥١٠ ٥١١ ٥١٢ ٥١٣ ٥١٤ ٥١٥ ٥١٦ ٥١٧ ٥١٨ ٥١٩ ٥٢٠ ٥٢١ ٥٢٢ ٥٢٣ ٥٢٤ ٥٢٥ ٥٢٦ ٥٢٧ ٥٢٨ ٥٢٩ ٥٣٠ ٥٣١ ٥٣٢ ٥٣٣ ٥٣٤ ٥٣٥ ٥٣٦ ٥٣٧ ٥٣٨ ٥٣٩ ٥٤٠ ٥٤١ ٥٤٢ ٥٤٣ ٥٤٤ ٥٤٥ ٥٤٦ ٥٤٧ ٥٤٨ ٥٤٩ ٥٥٠ ٥٥١ ٥٥٢ ٥٥٣ ٥٥٤ ٥٥٥ ٥٥٦ ٥٥٧ ٥٥٨ ٥٥٩ ٥٦٠ ٥٦١ ٥٦٢ ٥٦٣ ٥٦٤ ٥٦٥ ٥٦٦ ٥٦٧ ٥٦٨ ٥٦٩ ٥٧٠ ٥٧١ ٥٧٢ ٥٧٣ ٥٧٤ ٥٧٥ ٥٧٦ ٥٧٧ ٥٧٨ ٥٧٩ ٥٨٠ ٥٨١ ٥٨٢ ٥٨٣ ٥٨٤ ٥٨٥ ٥٨٦ ٥٨٧ ٥٨٨ ٥٨٩ ٥٩٠ ٥٩١ ٥٩٢ ٥٩٣ ٥٩٤ ٥٩٥ ٥٩٦ ٥٩٧ ٥٩٨ ٥٩٩ ٦٠٠ ٦٠١ ٦٠٢ ٦٠٣ ٦٠٤ ٦٠٥ ٦٠٦ ٦٠٧ ٦٠٨ ٦٠٩ ٦١٠ ٦١١ ٦١٢ ٦١٣ ٦١٤ ٦١٥ ٦١٦ ٦١٧ ٦١٨ ٦١٩ ٦٢٠ ٦٢١ ٦٢٢ ٦٢٣ ٦٢٤ ٦٢٥ ٦٢٦ ٦٢٧ ٦٢٨ ٦٢٩ ٦٣٠ ٦٣١ ٦٣٢ ٦٣٣ ٦٣٤ ٦٣٥ ٦٣٦ ٦٣٧ ٦٣٨ ٦٣٩ ٦٤٠ ٦٤١ ٦٤٢ ٦٤٣ ٦٤٤ ٦٤٥ ٦٤٦ ٦٤٧ ٦٤٨ ٦٤٩ ٦٥٠ ٦٥١ ٦٥٢ ٦٥٣ ٦٥٤ ٦٥٥ ٦٥٦ ٦٥٧ ٦٥٨ ٦٥٩ ٦٦٠ ٦٦١ ٦٦٢ ٦٦٣ ٦٦٤ ٦٦٥ ٦٦٦ ٦٦٧ ٦٦٨ ٦٦٩ ٦٧٠ ٦٧١ ٦٧٢ ٦٧٣ ٦٧٤ ٦٧٥ ٦٧٦ ٦٧٧ ٦٧٨ ٦٧٩ ٦٨٠ ٦٨١ ٦٨٢ ٦٨٣ ٦٨٤ ٦٨٥ ٦٨٦ ٦٨٧ ٦٨٨ ٦٨٩ ٦٩٠ ٦٩١ ٦٩٢ ٦٩٣ ٦٩٤ ٦٩٥ ٦٩٦ ٦٩٧ ٦٩٨ ٦٩٩ ٧٠٠ ٧٠١ ٧٠٢ ٧٠٣ ٧٠٤ ٧٠٥ ٧٠٦ ٧٠٧ ٧٠٨ ٧٠٩ ٧١٠ ٧١١ ٧١٢ ٧١٣ ٧١٤ ٧١٥ ٧١٦ ٧١٧ ٧١٨ ٧١٩ ٧٢٠ ٧٢١ ٧٢٢ ٧٢٣ ٧٢٤ ٧٢٥ ٧٢٦ ٧٢٧ ٧٢٨ ٧٢٩ ٧٣٠ ٧٣١ ٧٣٢ ٧٣٣ ٧٣٤ ٧٣٥ ٧٣٦ ٧٣٧ ٧٣٨ ٧٣٩ ٧٤٠ ٧٤١ ٧٤٢ ٧٤٣ ٧٤٤ ٧٤٥ ٧٤٦ ٧٤٧ ٧٤٨ ٧٤٩ ٧٥٠ ٧٥١ ٧٥٢ ٧٥٣ ٧٥٤ ٧٥٥ ٧٥٦ ٧٥٧ ٧٥٨ ٧٥٩ ٧٦٠ ٧٦١ ٧٦٢ ٧٦٣ ٧٦٤ ٧٦٥ ٧٦٦ ٧٦٧ ٧٦٨ ٧٦٩ ٧٧٠ ٧٧١ ٧٧٢ ٧٧٣ ٧٧٤ ٧٧٥ ٧٧٦ ٧٧٧ ٧٧٨ ٧٧٩ ٧٨٠ ٧٨١ ٧٨٢ ٧٨٣ ٧٨٤ ٧٨٥ ٧٨٦ ٧٨٧ ٧٨٨ ٧٨٩ ٧٩٠ ٧٩١ ٧٩٢ ٧٩٣ ٧٩٤ ٧٩٥ ٧٩٦ ٧٩٧ ٧٩٨ ٧٩٩ ٨٠٠ ٨٠١ ٨٠٢ ٨٠٣ ٨٠٤ ٨٠٥ ٨٠٦ ٨٠٧ ٨٠٨ ٨٠٩ ٨١٠ ٨١١ ٨١٢ ٨١٣ ٨١٤ ٨١٥ ٨١٦ ٨١٧ ٨١٨ ٨١٩ ٨٢٠ ٨٢١ ٨٢٢ ٨٢٣ ٨٢٤ ٨٢٥ ٨٢٦ ٨٢٧ ٨٢٨ ٨٢٩ ٨٣٠ ٨٣١ ٨٣٢ ٨٣٣ ٨٣٤ ٨٣٥ ٨٣٦ ٨٣٧ ٨٣٨ ٨٣٩ ٨٤٠ ٨٤١ ٨٤٢ ٨٤٣ ٨٤٤ ٨٤٥ ٨٤٦ ٨٤٧ ٨٤٨ ٨٤٩ ٨٥٠ ٨٥١ ٨٥٢ ٨٥٣ ٨٥٤ ٨٥٥ ٨٥٦ ٨٥٧ ٨٥٨ ٨٥٩ ٨٦٠ ٨٦١ ٨٦٢ ٨٦٣ ٨٦٤ ٨٦٥ ٨٦٦ ٨٦٧ ٨٦٨ ٨٦٩ ٨٧٠ ٨٧١ ٨٧٢ ٨٧٣ ٨٧٤ ٨٧٥ ٨٧٦ ٨٧٧ ٨٧٨ ٨٧٩ ٨٨٠ ٨٨١ ٨٨٢ ٨٨٣ ٨٨٤ ٨٨٥ ٨٨٦ ٨٨٧ ٨٨٨ ٨٨٩ ٨٩٠ ٨٩١ ٨٩٢ ٨٩٣ ٨٩٤ ٨٩٥ ٨٩٦ ٨٩٧ ٨٩٨ ٨٩٩ ٩٠٠ ٩٠١ ٩٠٢ ٩٠٣ ٩٠٤ ٩٠٥ ٩٠٦ ٩٠٧ ٩٠٨ ٩٠٩ ٩١٠ ٩١١ ٩١٢ ٩١٣ ٩١٤ ٩١٥ ٩١٦ ٩١٧ ٩١٨ ٩١٩ ٩٢٠ ٩٢١ ٩٢٢ ٩٢٣ ٩٢٤ ٩٢٥ ٩٢٦ ٩٢٧ ٩٢٨ ٩٢٩ ٩٣٠ ٩٣١ ٩٣٢ ٩٣٣ ٩٣٤ ٩٣٥ ٩٣٦ ٩٣٧ ٩٣٨ ٩٣٩ ٩٤٠ ٩٤١ ٩٤٢ ٩٤٣ ٩٤٤ ٩٤٥ ٩٤٦ ٩٤٧ ٩٤٨ ٩٤٩ ٩٥٠ ٩٥١ ٩٥٢ ٩٥٣ ٩٥٤ ٩٥٥ ٩٥٦ ٩٥٧ ٩٥٨ ٩٥٩ ٩٦٠ ٩٦١ ٩٦٢ ٩٦٣ ٩٦٤ ٩٦٥ ٩٦٦ ٩٦٧ ٩٦٨ ٩٦٩ ٩٧٠ ٩٧١ ٩٧٢ ٩٧٣ ٩٧٤ ٩٧٥ ٩٧٦ ٩٧٧ ٩٧٨ ٩٧٩ ٩٨٠ ٩٨١ ٩٨٢ ٩٨٣ ٩٨٤ ٩٨٥ ٩٨٦ ٩٨٧ ٩٨٨ ٩٨٩ ٩٩٠ ٩٩١ ٩٩٢ ٩٩٣ ٩٩٤ ٩٩٥ ٩٩٦ ٩٩٧ ٩٩٨ ٩٩٩ ١٠٠٠

faite, qu'il nomme *real character*. Il s'agit d'un système de nature idéographique : ses « philosophical letters » sont composées de trois parties indiquant le *genre*, la *différence* et l'*espèce* en accord avec le système de classification développé dans les chapitres précédents. Des diacritiques viennent préciser la classe grammaticale, les inflexions verbales, le nombre. De plus, des caractères spéciaux sont prévus pour les mots grammaticaux tels que la copule, les pronoms, les prépositions, les interjections (WILKINS 1668a, part. IV, chap. I).

Ce système n'a pas encore été proposé à l'encodage chez Unicode. De ce fait, il n'est pas possible de l'employer comme n'importe quel autre écriture dans un document électronique. Une solution serait de trouver une police faisant correspondre les lettres d'un système existant aux caractères philosophiques. Si pour l'impression d'ETRC en 1668 une police mécanique semble bien avoir été créée par le typographe Joseph Moxon (BRYDEN 2004), il n'en existe pas d'équivalent numérique à ce jour. Il faudra donc partir de zéro.

La création typographique est en principe accessible à tous : à

côté des logiciels d'édition commerciale comme FontLab¹⁵ utilisés par les professionnels, il existe des libres comme FontForge¹⁶ avec les mêmes fonctionnalités. Cependant, le design d'une police demande une certaine expérience, non seulement pour obtenir un aspect visuellement harmonieux, mais aussi, dans le cas de l'écriture philosophique de Wilkins, pour construire les ligatures qui rassemblent les éléments basiques en caractères uniques.

3 Images

L'ouvrage de base est illustré (présence de gravures) et décoré (en-têtes de chapitres, lettrines). On désire que ces images, ou des approximations remplissant la même fonction, soient présentes dans l'édition finale.

La page 378, par exemple, présente un tableau consistant en trente-cinq vignettes représentant un visage d'homme de face et en coupe latérale articulant les consonnes et les voyelles décrites dans la partie III. Les distinctions sont très fines, malheureusement les PDF de la numérisation par Google Books utilisent un contraste noir-blanc trop violent qui efface les détails; et même l'image en ligne n'a pas de grain assez fin pour être satisfaisant. L'exemple ci-devant est tiré de la numérisation depuis l'exemplaire de la bibliothèque de l'État de Bavière.

15. <http://www.fontlab.com>

16. <https://fontforge.github.io/en-US/>



Dans ces conditions, comment obtenir des images de meilleure qualité? On pourrait réaliser une nouvelle reproduction à partir de l'original avec des appareils de meilleur qualité. Mais peut-être que recréer ces images, les moderniser tout en conservant les principes de construction, rendrait la version électronique plus harmonieuse.

4 Liens

Des renvois sont faits dans le texte vers l'intérieur et l'extérieur de celui-ci. Il est possible d'en faire des liens dynamiques.

4.i Liens internes

Tant que chaque division du livre est correctement signalée à l'aide de balises imbriquées, il est possible de leur attribuer un identifiant unique avec l'attribut `@xml:id` auquel on peut renvoyer par un attribut `@target` se trouvant dans une balise `<ref>`, laquelle encadre la référence à la division dans le texte.

4.ii Liens externes

Wilkins cite beaucoup d'ouvrages et chapitres d'ouvrages qu'il a pu consulter pour la rédaction de son essai, sur des sujets linguistiques et philosophiques. Lorsque des versions électroniques existent en ligne, par exemple sur une des plate-formes mentionnée en section ??2, on aimerait pouvoir y renvoyer le lecteur.

MÉTHODOLOGIE DE TRAVAIL

I Choix des sources

Pour réaliser le livre électronique, on prendra pour base la digitalisation préexistante du Texte Encoding Project, en se référant aux numérisations Google Books pour vérification, et enfin à un exemplaire imprimé original.

1.i Sources physiques

Il en existe trois exemplaires physiques du livre à la Bibliothèque Nationale Universitaire de Strasbourg : le texte original imprimé en 1668, qui est uniquement consultable sur place; et deux fac-similés produits par le service d'impression à la demande à partir de numérisations.

1.ii Sources numériques : les images

Il existe plusieurs numérisations depuis Google Books; seules trois peuvent être lues intégralement, provenant d'originaux de la bibliothèque nationale de la République Tchèque¹⁷, de la biblio-

17. <https://books.google.fr/books?id=Bu7pwpr5qBcC> (consulté le 16 octobre 2019)

thèque municipale de Lyon¹⁸ et de la bibliothèque de l'État de Bavière¹⁹. Elles permettent la recherche de texte, mais la reconnaissance automatique des caractères est très mal adaptée à la police et aux ligatures employées ici, et ne permet de rien extraire qui soit utilisable. Lues en ligne, ces images sont en couleur; une fois téléchargées en tant que PDF, elles sont en noir et blanc, ce qui baisse la qualité de la reproduction.

1.iii Sources numériques : l'Université du Michigan

La digitalisation d'ETRC par le Text Encoding Project de l'université du Michigan (cf 4.ii) a été saisie originellement dans un schéma SGML ressemblant à la spécification TEI 3, converti par la suite en un schéma XML proche de la spécification TEI 4. À partir de là, on recherchera les différences entre ce schéma et le schéma cible (TEI 5) et on établira un tableau de correspondances qui servira à la conversion automatique de l'un à l'autre grâce à une feuille de style XSLT (voir section 3).

18. <https://books.google.fr/books?id=BCCtZjBtiEYC> (consulté le 16 octobre 2019)

19. <https://books.google.fr/books?id=Q85TAAAcAAJ> (consulté le 16 octobre 2019)

2 Travail éditorial : recherches à faire et décisions à prendre

2.i *Omissions dans la digitalisation du Text Encoding Project*

En plus de la non prise en compte des caractères non-latins,

2.ii *Errata*

Wilkins donne une page d'errata en début du livre, signalant les erreurs découvertes après l'impression. La question est de savoir si l'on conserve cette page telle que dans la source, ou si l'on intègre les corrections dans le corps du texte. L'élément `<choice>` de la spécification TEI permet de signaler une modification éditorial dans le texte, en encadrant un élément `<sic>` correspondant au texte (fautif) original, et un élément `<corr>` correspondant à la correction. Ce dernier peut également préciser l'origine de la correction avec un attribut `@resp`, renvoyant à un élément `respStmt` définissant qui a pris cette décision de correction. On peut définir plusieurs sources de correction ainsi.

2.iii *Particularités orthographiques et typographiques*

Wilkins n'emploie pas les conventions typographiques actuelles. Quelles sont celles qu'une réédition doit conserver, si le but est de respecter la source tout en étant lisible par les lecteurs d'aujourd'hui?

Capitalisation

Certains mots, en majorité des substantifs et adjectifs commencent par une majuscule à l'intérieur des phrases, comme on peut le voir par exemple en p. 353 :

Noun *Adjectives* need not have any note to express *Number, Gender, Case*, because in all these they agree with their Substantives; unless such Adjectives as are used Substantively, by reason of their composition with the Transcendental marks of *Person, Thing, Time, Place, &c.*

Comme cela n'entraîne pas de problèmes de lecture, on les conservera telles quelles.

s long ⟨f⟩

ETRC emploie deux caractères là où l'anglais moderne n'utilise que le ⟨s⟩ minuscule : le second est le s long ⟨f⟩, apparaissant dans tous les contextes excepté à la fin des mots. Cette lettre a cessé d'être d'un usage courant vers le début du XIX^e siècle, pour des raisons d'économie auprès des imprimeurs, et parce qu'elle pouvait se confondre à la lecture avec un ⟨f⟩ (WEST 2006). Pour cette dernière raison, il semble préférable d'utiliser le ⟨s⟩ rond par défaut; mais comme certaines polices informatiques, comme EB Garamond, proposent des variantes de caractères dont le s long, que l'on peut activer à loisir.

Orthographe alternatives

Réclame

Le premier mot de chaque page est répété au bas de la page précédente; il s'agissait d'une pratique indiquant au relieur quelles pages devait se suivre (ROBERTS et ETHERINGTON 1982). Cette particularité purement technique n'a pas besoin d'être répétée dans une édition électronique.

3 Conversion dans les formats HTML et EPUB

À partir du fichier encodé en XML, il est possible de générer automatiquement un fichier HTML, et à partir de celui-ci un fichier EPUB. Pour cela, on emploiera une feuille de style XSLT. Celle-ci permet de transformer un fichier XML en manipulant les balises et les attributs. Sa cible peut être un nouveau fichier XML, un fichier HTML, ou même un fichier \LaTeX qui servira de base à un PDF.

CONCLUSION

La réalisation d'un ouvrage digital est compliquée.

BIBLIOGRAPHIE

- BLEIER, Roman et al. (2018). *Digital Scholarly Editions as Interfaces*. OCLC : 1079409488. URL : http://www.bod.de/index.php?id=296&objk_id=2544467 (visité le 28/01/2019).
- BORGES, Jorge Luis (1952). « El idioma analítico de John Wilkins ». In : *Otras inquisiciones*, p. 158–159.
- BRYDEN, D. J. (2004). *Moxon, Joseph (1627–1691), printer and globe maker*. In : *Oxford Dictionary of National Biography*. Oxford University Press. URL : <https://www-oxforddnb-com.scd-rproxy.u-strasbg.fr/view/10.1093/ref:odnb/9780198614128.001.0001/odnb-9780198614128-e-19466> (visité le 14/10/2019).
- BURNARD, Lou et C M SPERBERG-McQUEEN (2012). *TEI Lite : Encoding for Interchange : an introduction to the TEI — Final revised edition for TEI P5*. URL : https://tei-c.org/release/doc/tei-p5-exemplars/pdf/tei_lite.doc.pdf (visité le 20/09/2019).
- CAMPS, Jean-Baptiste (2019). *TEI, La TeX et les éditions critiques sur papier : I. Les différents packages*. Sacré Gr@@l. URL : <https://graal.hypotheses.org/484> (visité le 23/01/2019).
- Court Order to Block Access from Germany* (2018). Project Gutenberg Library Archive Foundation. URL : <https://cand.pglaf.org/germany/index.html> (visité le 07/10/2019).

- DELMOTTE, Stéphanie (2009). « Publications scientifiques en sciences humaines ». In : *Les Cahiers du numérique* Vol. 5,2, p. 53–84. URL : <https://www.cairn.info/revue-les-cahiers-du-numerique-2009-2-page-53.htm> (visité le 23/01/2019).
- HENDERSON, P. A. Wright (2008). *The Life and Times of John Wilkins Warden of Wadham College, Oxford; Master of Trinity College, Cambridge; and Bishop of Chester*. URL : <http://www.gutenberg.org/ebooks/26674> (visité le 23/09/2019).
- HUCHELMANN, Alexis (2019). « Les langues construites : délimitation, historique et typologie suivies d'une illustration du processus de création d'une langue naturaliste nommée « tüchte » ». In : *Fiat Lingua* (fl-00005B-00). URL : <http://fiatlingua.org/2019/04/> (visité le 25/09/2019).
- LEBERT, Marie (2010). *Le Projet Gutenberg (1971-2009)*. Project Gutenberg. URL : <http://www.gutenberg.org/cache/epub/31634/pg31634-images.html>.
- OKRENT, Arika (2009). *In the land of invented languages Esperanto rock stars, Klingon poets, Loglan lovers, and the mad dreamers who tried to build a perfect language*. New York : Spiegel & Grau. (Visité le 08/04/2016).
- PATONNIER, Jérémie (2011). *Les ligatures dans les navigateurs*. Typographisme. URL : <http://www.typographisme.net/post/Les-ligatures-dans-les-navigateurs#note7b> (visité le 29/01/2019).
- PIERAZZO, Elena (2014). *Digital Scholarly Editing: Theories, Models and Methods*. URL : <http://hal.univ-grenoble-alpes.fr/hal-01182162/document> (visité le 28/01/2019).

- ROBERTS, Matt et Don ETHERINGTON (1982). *Catchword*. In : *Bookbinding and the Conservation of Books. A Dictionary of Descriptive Terminology*. Washington : Library of Congress. URL : <http://cool.conservation-us.org/don/don.html> (visité le 17/10/2019).
- SHILLINGSBURG, Peter L. (2006). *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge : Cambridge University Press. URL : <http://ebooks.cambridge.org/ref/id/CB09780511617942> (visité le 28/01/2019).
- SMITH, Courtney Weiss (2017). « Rhyme and Reason in John Wilkins's Philosophical Language Scheme ». In : *Modern Philology* 115.2, p. 183–212. URL : <https://www.journals.uchicago.edu/doi/full/10.1086/692809> (visité le 16/04/2019).
- VANHOUTTE, Edward (2004). « An Introduction to the TEI and the TEI Consortium ». In : *Literary and Linguistic Computing* 19.1, p. 9–16. URL : <https://academic.oup.com/dsh/article/19/1/9/928389> (visité le 23/01/2019).
- WEST, Andrew (2006). *The Rules for Long S*. BabelStone Blog. URL : <http://www.babelstone.co.uk/Blog/2006/06/rules-for-long-s.html> (visité le 16/10/2019).
- WILKINS, John (1668a). *An Essay towards a Real Character and a Philosophical Language*. URL : <https://books.google.fr/books?id=Q85TAAAAcAAJ> (visité le 16/04/2019).
- (1668b). *An essay towards a real character, and a philosophical language by John Wilkins ...* 2015^e éd. URL : <http://name.umd.umich.edu/A66045.0001.001>.
- WITTERN, Christian, Arianna CIULA et Conal TUOHY (2009). « The making of TEI P5 ». In : *Literary and Linguistic Com-*

puting 24.3, p. 281–296. URL : <https://academic.oup.com/dsh/article/24/3/281/968658> (visité le 17/10/2019).

TABLE DES MATIÈRES

Introduction	I
Table des figures	2
I Nature et histoire de la digitalisation	3
1 Définition	3
1.i Digitalisation vs numérisation	3
1.ii Différentes encodages	3
2 Emploi	4
2.i Pour la sauvegarde du patrimoine littéraire	4
2.ii Dans les humanités	6
II Le schéma d'encodage TEI	7
1 Un standard SGML et XML	7
2 TEI Lite	8
3 Potentiel descriptif	8
III Le livre de John Wilkins	9
1 L'auteur	9
1.i Biographie	9
1.ii Autres écrits	10
2 L'essai	10

2.i	Contenu	10
2.ii	Genèse et rédaction	12
2.iii	Réception	12
3	Littérature scientifique	12
4	Choix de l'ouvrage	12
4.i	Intérêt du sujet	12
4.ii	Précédentes numérisations et digitalisations	13
5	Fac-similé ou réédition?	14
5.i	Apporter quelque chose de neuf	15
5.ii	Les limites de la réédition	15
IV	Défis techniques posés par le texte	16
1	Caractéristiques techniques du livre originel	16
2	Caractères spéciaux	16
2.i	Caractères non-latins	16
2.ii	L'écriture philosophique	17
3	Images	19
4	Liens	21
4.i	Liens internes	21
4.ii	Liens externes	21
V	Méthodologie de travail	22
1	Choix des sources	22
1.i	Sources physiques	22
1.ii	Sources numériques : les images	22
1.iii	Sources numériques : l'Université du Michigan	23

2	Travail éditorial : recherches à faire et décisions à prendre	24
2.i	Omissions dans la digitalisation du Text Encoding Project	24
2.ii	Errata	24
2.iii	Particularités orthographiques et typographiques	24
3	Conversion dans les formats HTML et EPUB . .	26
	Bibliographie	28