HAL
open science

Tiziana Mancinelli & Elena Pierazzo, Che cos'è un'edizione scientifica digitale, Rome, Carocci, 2020

Chloé Tardivel

To cite this version:

Chloé Tardivel. Tiziana Mancinelli & Elena Pierazzo, Che cos'è un'edizione scientifica digitale, Rome, Carocci, 2020. Encyclo. Revue de l'école doctorale Sciences des Sociétés ED 624, 2020, pp.165-170. hal-02900766

HAL Id: hal-02900766

https://u-paris.hal.science/hal-02900766

Submitted on 16 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chloé Tardivel*

Tiziana Mancinelli & Elena Pierazzo, *Che cos'è un'edizione scientifica digitale*, Rome, Carocci, 2020.

Qu'est qu'une édition numérique savante ? L'expression, peu usuelle en italien comme en français, vient de l'anglais *Digital scholarly edition* et définit les pratiques éditoriales scientifiques réalisées entièrement au moyen du numérique[1]. Les autrices, Tiziana Mancinelli et Elena Pierazzo, en sont des spécialistes : la première est chercheuse postdoctorale au *Cologne Center for eHumanities* et coordinatrice de divers projets éditoriaux numériques[2]; la seconde est professeure de littérature italienne et d'humanités numériques à l'Université de Grenoble et reconnue internationalement pour ses multiples contributions scientifiques sur le sujet[3]. Dans cet ouvrage, les deux expertes font part d'une réflexion passionnante sur les enjeux passés, présents et à venir de la philologie numérique (et non de la philologie numérisée, rappellent-elles, qui n'est que le prolongement numérique d'une édition papier existante)[4]. Leur propos, clair et haletant, s'adresse à tout type de lectorat. Le néophyte trouvera, au cours des sept chapitres (125p.), une approche stimulante sur ce qui pourrait être envisagé comme une discipline à part entière et tirera

* Doctorante en histoire au laboratoire ICT (*Identités, Cultures, Territoires*), Université de Paris.

[1] On trouve aussi les traductions suivantes : « édition scientifique numérique », « édition électronique savante ». J'ai retenu l'expression « édition numérique savante » car elle est utilisée telle quelle en français par l'une des autrices (Elena Pierazzo).

[2] Citons son projet en cours, *Magica Levantina*, qui vise à éditer les inscriptions magiques des villes romaines du Levant des IVe et VIe siècles (https://papyri.uni-koeln.de/magica-levantina/index.html).

[3] À noter sa publication précédente en anglais, Elena Pierazzo, *Digital Scholarly Editing. Theories, Models and Methods*, Routledge, London-New York, 2015.

[4] À ces deux types de philologie, elles en évoquent une troisième qui n'est pas traitée par la suite. Il s'agit de la philologie computationnelle, qui recourt à des méthodes informatiques avancées, telles que l'intelligence artificielle (par ex., les logiciels de reconnaissance et de transcription automatique des manuscrits) ou bien des algorithmes de fouilles de données (*data mining*).

profit du glossaire qui définit les nombreux acronymes et expressions anglaises utilisées (XML-TEI, GIS, HTML, HTR, *distant reading*, *data mining*, *etc*.), rappelant qu'en ce domaine, l'anglais est la *lingua franca*. La personne expérimentée lira avec intérêt un retour critique sur des pratiques en cours et se reportera à la bibliographie et sitographie pour plus de détails. Cependant, quel que soit son profil, il vaut mieux avoir à sa disposition une connexion internet et un ordinateur à portée de main si l'on veut consulter les nombreuses éditions numériques citées et commentées. La lecture de l'ouvrage, invitant à naviguer entre plusieurs médias (papier et web) rend compte, d'une certaine manière, de la démarche scientifique propre aux humanités numériques qui est celle d'un dialogue continuel entre le langage humain et le langage informatique.

Tiziana Mancinelli et Elena Pierazzo entendent montrer que le numérique ne relève pas du simple changement de support par rapport à l'imprimé. Il agit, au contraire, comme un véritable transformateur modifiant les caractéristiques essentielles d'une édition savante. Le numérique bouleverse le paradigme de la philologie traditionnelle, affirment-elles. Dans cette perspective, la thèse défendue s'inscrit dans le sillage de la pensée du théoricien de la communication, Marshall McLuhan (†1980), pour qui chaque moyen de communication (forme) engendre un nouveau contenu (fond)[5]. Celle-ci revendique également l'assertion du pionnier des humanités numériques, Willard McCarty, selon laquelle la vraie révolution née de l'ordinateur est d'ordre qualitatif (nouvelle manière de conceptualiser le monde) et non quantitatif (réduire les temps de travail)[6]. Le premier chapitre constate, ainsi, toutes les étapes du travail éditorial classique altérées par le numérique. Grâce aux catalogues en ligne des bibliothèques ou d'archives, il est désormais possible d'identifier n'importe quelle source en vue de l'éditer. L'appareil photo ou un simple téléphone portable permet, ensuite, la photographie des manuscrits et, donc, la production de fac-similés[7]. La mise en ligne des images numériques ouvre la voie aux transcriptions collectives, à l'instar des projets de *crowdsourcing* qui fleurissent sur le net. Les programmes OCR (*Optical Character Recognition*) permettent, quant

[5] Marshall McLuhan, *Understanding Media, The Extensions of Man*, McGraw-Hill Book Company, New York, 1964. Il est célèbre pour sa formule lapidaire « le médium est le message ».

[6] Willard McCarty, *Humanities Computing*, Palgrave Macmillan, Basingstoke-New York, 2005.

[7] Les autrices insistent sur l'importance de l'entrée en vigueur de la loi italienne sur « le marché et la concurrence » (29 août 2017) qui a autorisé la libre reproduction des biens archivistiques et bibliothécaires à usage personnel.

à eux, d'entrevoir l'automatisation de cette opération. Ils sont même à l'origine du néologisme « océrisation », qui définit la transformation automatique d'un fichier contenant l'image d'un texte en MRF (*Machine Readable Form*, par exemple le format .txt)[8]. La collation des témoins est, elle-aussi, soumise à des bouleversements d'envergure. Les nouvelles technologies, inspirées par le modèle développé en biologie pour retracer les mutations de l'ADN (méthode cladistique), permettent, de plus en plus et avec des résultats probants, d'automatiser cette étape et de produire le *stemma codicum*. Enfin, la mise en ligne d'un grand nombre de données (*big data*) selon un langage d'encodage standardisé, tel que le XML-TEI (*Extensible Markup Language-Text Encoding Initiative*), offre la possibilité de les interroger à l'infini avec des systèmes computationnels, une démarche intellectuelle impensable et impossible avec une édition papier. L'édition numérique ouvre donc la voie à de nouvelles pistes de recherche en sciences humaines et sociales. Toutefois, elle fait émerger, simultanément, des problèmes méthodologiques et éthiques.

Conscientes qu'il existe peu de littérature scientifique sur le sujet, particulièrement en langue italienne, les autrices conçoivent leur ouvrage comme un guide, une boussole, pour aider à la navigation dans les « contrées » philologiques du XXIe siècle. Ni manuel, ni essai théorique, le livre ne dispense aucune formation technique mais est une tentative de « cartographie d'une *terra incognita* »[9]. Afin de donner des points de repères, Elena Pierazzo dresse, tout d'abord, une typologie des principales éditions numériques actuelles (chap. 2). Il existe, ainsi, divers modèles d'éditions (« documentaire » lorsque le projet consiste à éditer un seul manuscrit comme des lettres ou un journal intime ; « critique » lorsque l'édition est réalisée à partir de témoins ; « génétique » lorsqu'elle se base sur des fragments ; les trois à la fois, soit « composite »), réalisées soit de manière collaborative et/ou ouverte, présentant des données de type hypertextuel ou hypermédia. Chacune des éditions citées a en commun des évolutions techniques afin de faire face au problème de l'obsolescence et de la pérennité des données. La philologue se concentre, ensuite, sur l'univers éditorial italien. Si l'édition des œuvres classiques de la littérature italienne (Dante, Pétrarque, Boccace, etc.) bénéficient d'une grande visibilité,

[8] Les machines ne sont pas encore capables de lire et décoder correctement les abréviations médiévales et modernes, précisent-elles. Elles sont, toutefois, efficaces dans l'examen des textes contemporains.

[9] La citation exacte est « in un certo senso scrivere un libro sulle edizioni scientifiche digitali è come cartografare per la prima volta una *terra incognita* » (p. 9).

grâce aux équipes et fonds engagés, les nombreux mémoires de master ou thèses de doctorat sur le sujet demeurent peu accessibles, faute d'une plateforme scientifique, à l'échelle nationale ou universitaire, permettant la publication de leurs travaux. Les problèmes institutionnels et infrastructurels apparaissent alors comme les premiers freins au développement des éditions numériques. Le chapitre suivant (chap. 3) propose, quant à lui, des repères historiques sur la pratique philologique numérique. Quatre paradigmes, correspondant à des problématiques, des supports et des méthodologies différentes, ont été identifiés. Le premier, baptisé « phase d'expérimentation », s'ouvre dans les années 1950 avec les travaux du père jésuite Roberto Busa (†2011) sur la lemmatisation des œuvres de saint Thomas d'Aquin. Cette période, marquée par le début de l'utilisation de l'informatique en philologie, se conclue en 1978, à l'occasion du colloque de Paris sur *La pratique des ordinateurs dans la critique des textes* qui rend compte de la difficulté à traiter des problèmes fondamentaux de la critique textuelle par les ordinateurs de l'époque. La phase suivante, qualifiée de « consolidation », s'étend de la fin des années 1970 à 1996, date de la publication de la première édition numérique sur CD-ROM réalisée par collation automatique (*The Wife of Bath's Prologue* de Geoffrey Chaucer). La troisième, de 1996 à 2005, se caractérise par le « début du WWW », car elle coïncide avec l'expansion du Web, introduit en 1992. C'est le temps des premières éditions en ligne, celui de l'utilisation du XML (né en 1998) et du premier cadrage théorique sur les « humanités computationnelles » (cf note 6 de ce compte rendu). Depuis 2005, on observe une phase d'« interconnexion et interopérabilité »[10]. Grâce à l'évolution des langages de codifications et de programmation, des standards d'encodage et des infrastructures numériques, les données peuvent être échangées entre logiciels et mises en relation, démultipliant les manières de faire de la recherche non seulement en philologie mais plus généralement en sciences humaines et sociales.

Au cours des chapitres suivants, Tiziana Mancinelli et Elena Pierazzo posent le cadre méthodologique et éthique nécessaire à la réalisation d'une édition numérique savante. Le qualificatif « savante » prend alors tout son sens car il s'agit de sensibiliser les personnes engagées dans un processus éditorial aux bonnes pratiques et réflexes. Il faut savoir, expliquent-elles, que toute édition numérique repose sur le modèle conceptuel et technologique dit *input-output* (« entrée-sortie »).

[10] Cette phase correspond, ainsi, aux principes FAIR (*Findable, Accessible, Interoperable, Reusable*, c'est-à-dire « Facile à trouver, Accessible, Interopérable et Réutilisable ») qui régissent actuellement la recherche en humanités numériques.

Celui-ci prévoit un point de départ, à savoir l'encodage du manuscrit en langage informatique (chap. 4), et un point final, correspondant à la publication du texte à travers une interface utilisateur (chap. 5). Avant de se lancer dans l'entreprise éditoriale, il est bon de définir précisément l'objectif poursuivi (quel type d'édition souhaité ? édition documentaire ? critique ?), d'évaluer le temps et les ressources (financières, matérielles) à disposition. Il ne faut pas négliger aussi les ressources humaines car une édition numérique appelle presque toujours un travail d'équipe, dans le sens où elle requiert plusieurs compétences et connaissances (philologiques, numériques, informatiques, etc.), rarement réunis en une seule et même personne. La phase de publication et de distribution du texte suscite, ainsi, généralement une collaboration entre éditeurs et développeurs, contrairement au paradigme de l'édition papier dans lequel le travail scientifique se termine à la remise du texte à la maison d'édition, responsable exclusive de sa distribution commerciale. Il est aussi indispensable de passer par la modélisation du texte, c'est-à-dire choisir en amont les caractéristiques qui seront effectivement publiées et de quelle manière. En effet, l'édition numérique n'est pas une simple reproduction électronique du texte, c'est-à-dire des lettres, espaces, signes de ponctuation et dispositions spatiales (niveau sémiotique). Elle rend compte également du sens du texte (niveau sémantique). Dans la phrase « dans le *Devisement du Monde* Marco Polo raconte son voyage en Chine » (p. 61), « Devisement du Monde » est un titre, « Marco Polo » un nom de personne (et ici, de surcroît, l'auteur), « Chine » est un pays. L'encodage en XML-TEI permet de préciser ces nuances à travers un langage de balisage aussi appelé « tag » (du type <auteur>Marco Polo</auteur>), mais encore faut-il veiller à respecter scrupuleusement sa syntaxe. Les expertes prennent soin d'identifier les erreurs à ne pas commettre. À chaque étape du travail, Tiziana Mancinelli et Elena Pierazzo amènent les éditeurs et éditrices à se poser les bonnes questions et rappellent que tout choix comporte des implications scientifiques et éthiques. Notons, toutefois, que la personne familiarisée avec les éditions numériques trouvera certainement rébarbatif le chapitre 6, qui revient sur des éléments présents en filigrane dans les chapitres précédents. Pour conclure (chap. 7), les autrices abordent la question de la qualité scientifique des projets éditoriaux numériques. Comment la juger ? Si elles citent les diverses grilles d'évaluation validées par la communauté scientifique, elles expliquent qu'une bonne édition numérique doit contenir sur son site les éléments suivants : des informations générales (noms des éditeurs, institutions, *etc*.) ; une définition des objectifs et du type d'édition retenue ; une présentation du rôle de chaque personne

dans le projet ; une documentation classique sur la collation, la sélection et la transcription des manuscrits et une autre technique sur le format des données numériques, les logiciels utilisés et le type d'interface ; et, enfin, une licence précisant les droits d'utilisation.

L'ouvrage de Tiziana Mancinelli et Elena Pierazzo est intéressant à plus d'un titre. Il vient combler, tout d'abord, un vide scientifique dans le domaine. Il fait, ensuite, le pari de l'italien dans un milieu anglophone, rendant hommage à la longue tradition philologique italienne. Il laisse, enfin, entrevoir au chercheur et chercheuse le potentiel heuristique des éditions numériques et celui, *de facto*, des humanités numériques tout court. Parce que ces dernières ne sont pas une autre ou une sous discipline des sciences humaines et sociales, mais bien leur présent et leur avenir, nous recommandons la lecture de cet ouvrage à toute personne intéressée par ces domaines de recherche.