

Slide 1 :

[Remerciements.]

Avec cette intervention je souhaite proposer un bilan réflexif sur le travail de modélisation des données que j'ai effectué dans ma première année de thèse. Il s'agit de la création d'une base de données relationnelle pour le répertoire de la CIP au 18ème siècle, à partir des répertoires existants. Le but est d'étudier l'évolution chronologique de la programmation proposée au public par la compagnie au fil du siècle – on va y revenir.

Actuellement, nous sommes en train de saisir les données avec une application implementée par Camilla Cantiani avec la librairie Django, et utilisant une base de données SQLite. Le travail de saisie est effectué avec deux stagiaires, Chloé Petridis (étudiante en M1 de TAL à Paris) et Chiara Protani (doctorante en théâtre contemporain à l'Université de Bologne, Italie), recrutées pendant 3 mois. Nous nous sommes divisées les données à saisir et vérifier de la manière la plus équitable possible, et nous travaillerons ensemble aux premières visualisations lors de leur dernier mois de stage.

Le bilan que je souhaite vous présenter aujourd'hui sera fait par rapport à des considérations écologiques, et particulièrement par rapport à l'approche minimal computing (que je présenterai par la suite). Il s'agit de considérations plutôt théoriques et qui restent un peu abstraites, mais j'ai essayé de me restreindre la plus strictement possible au travail de modélisation – c'est-à-dire que je n'aborderai presque aucune question concernant le travail en-soi de la chercheuse ou du chercheur (comme les déplacements en avion, la consommation de matériaux, ou l'injonction à la productivité).

Slide 2 :

Je présente tout d'abord rapidement le minimal computing, afin de clarifier la problématique.

Le minimal computing constitue une approche à l'informatique et aux humanités numériques, qui se veut écologique et sociale. De manière générale, cette approche se focalise sur la possibilité d'implémenter les technologies dans des contextes infrastructurels variés, à partir d'une analyse des contraintes et de la situation spécifique dans laquelle les personnes développent les technologies. En d'autres mots, il s'agit d'une approche à l'écologie qui ne se limite pas à la volonté de réduire l'impact du numérique sur les ressources planétaires, et qui propose de réduire au même temps les inégalités sociales, dues à un accès inégal aux ressources, outils et infrastructures du numérique. Donc, selon le minimal computing, le fait qu'une technologie soit appropriable dans des contextes différents n'est pas moins important que réduire la consommation de ressources.

Pour tout cela, dans l'article d'introduction à un numéro de DHQ dédié au minimal computing, Risam et Gil proposent 4 questions heuristiques pour évaluer ce qui est effectivement nécessaire et suffisant dans un projet donné :

1. de quoi est-ce qu'on a besoin ? (what do we need?)
2. qu'est-ce qu'on a ? (what do we have?)
3. qu'est-ce qu'on priorise ? (what do we prioritize?)
4. à quoi sommes-nous prêts à renoncer ? (what are we willing to give up?)

Des chercheurs, se sont par ailleurs penchés spécifiquement sur la modélisation du point de vue du minimal computing. Ainsi, Erik Radio a publié en 2020 un article sur les ontologie et le minimal computing, mentionne différents principes de modélisation minimal computing.

Pour mon travail, répondre à ces questions et prendre en compte ces principes a représenté un outil assez efficace pour évaluer notre modélisation d'un point de vue écologique. Mais est-ce que le minimal computing peut aussi être plus largement un outil heuristique de modélisation dans les HN ? Est-ce que les principes de modélisation du minimal computing (tels que définis dans ER, 2020) permettent-ils d'améliorer l'impact écologique du modèle, tout en respectant les objectifs de notre projet ?

Dans la suite de l'intervention je vais, dans un premier temps, expliciter le contexte et les objectifs du projet. Ensuite, je vais montrer comment le minimal computing a influencé ce travail (en le confirmant et en le changeant), avant d'évaluer les principes de modélisations définis dans l'article d'ER de 2020, par rapport aux objectifs et aux présumées du projet. Je ne vais pas parler des aspects les plus techniques du bilan de modélisation, qui sont bien détaillés dans l'article, parce que je voulais donner plus de place au minimal computing, en espérant en discuter véritablement avec vous, lors des questions ou des pauses café.

Slide 3 :

Je présente d'abord le contexte et les objectifs du projet.

Quand j'ai connu mes directeurs de thèse et commencé mon doctorat, une partie du travail était définie par le travail accompli préalablement : à partir des registres manuscrits du XVIII^e siècle – que la compagnie tenait pour rendre compte au roi de son activité, et qui représentent les sources primaires – des répertoires ont été créés au XX^e et XXI^e siècle.

Ici vous pouvez voir des images des registres à des époques différentes du siècle. On voit que l'écriture change et que les informations transmises ne sont pas les mêmes. Ceci ne sont que 3 exemples, la variabilité est beaucoup plus grande.

Slide 4 :

Mais moi, je travaille à partir des répertoires (pas des registres), et en particulier à partir d'un répertoire tapuscrit de Clarence Brenner de 1961, et d'un répertoire en format excel, dont Emanuele De Luca a publié une partie en 2010 en format PDF. La base de données serait en fait une fusion de ces deux répertoires : Brenner fournit toutes les dates des représentations mais aucune information sur la nature des pièces, alors que De Luca a fait un travail fin de description des pièces à partir surtout des revues du XVIII^e siècle (genre, nb d'actes etc), mais n'était intéressé qu'à la date de la première représentation de chaque pièce.

Slide 5 :

Cette organisation du travail, m'a posé très vite face à un constat : créer une base de données est un travail interprétatif. Mes documents de départ pour la numérisation ne sont pas des sources primaires, mais contiennent des informations qui ont déjà été interprétées et structurées avec des objectifs spécifiques. À mon tour, je vais reprendre ces informations et les réorganiser encore une fois pour d'autres objectifs spécifiques. C'est-à-dire que le modèle des données pour décrire un objet ne va pas de soi, mais dépend de ce qu'on souhaite obtenir, ainsi que de notre vision de cet objet.

Nous voulons étudier l'évolution du répertoire au fil du temps, c'est-à-dire observer quels types de pièces ont été mises à l'affiches et quels types de pièces étaient produites pendant une certaine période par la compagnie. En d'autres mots, nous voulons analyser quelles pièces étaient proposées au public le long du siècle, et cela a des conséquences sur la structure des données.

Slide 6 :

Pour donner un exemple concret de comment nos objectifs ont guidé la modélisation (différent de celui présent dans l'article publié dans les actes), observons l'objet pièce, présent dans notre modèle. Le concept de pièce pourrait être représenté numériquement de plusieurs façons : une pièce pourrait correspondre au texte édité de la pièce, ou bien au texte que les acteurs jouaient avant qu'il soit édité. Mais puisque les acteurs ne jouent jamais exactement le texte rédigé, et puisqu'un texte peut avoir plusieurs versions jouées mais inédites – sans compter que chaque représentation est différente – une pièce pourrait correspondre aussi à une performance spécifique, ou à un ensemble de performances.

Dans notre cas, une pièce est une entité très large, et plutôt abstraite, qui regroupe les potentielles nombreuses versions d'un texte, dont ses éventuelles éditions. Il s'agit en fait de l'unité spectaculaire reconnaissable par les spectateurs. En d'autres mots, nous ne voulons pas tenir trace

de toutes les variations d'une pièce, à partir du moment où le public la reconnaît comme étant la même pièce.

Slide 7 :

Nous venons donc de voir le contexte et objectifs du projet, je vais maintenant présenter les aspects qui sont à mon avis les plus intéressants d'un point de vue minimal computing.

L'opération la plus concrète induite par les questions proposées par Risam et Gil est sans doute la simplification du modèle. Notre modèle initial prévoyait le traitement d'informations concernant les personnes ayant travaillé dans le théâtre à cette époque. Or ces informations ne sont pas contenues dans les répertoires. De plus, elles ne sont pas strictement nécessaires dans l'étude chronologique de la programmation, même si elles complètent le cadre de ce que attirait le public dans le théâtre. Nous pouvions donc y renoncer et nous les avons pas implémentées.

En revanche, le fait que le modèle ait été créé même pour ces entités facilite une éventuelle prise en main successive par d'autres personnes qui voudraient les implémenter. Cela nous mène à une autre caractéristique intéressante de la modélisation selon le minimal computing : un modèle est une manière non ambiguë et indépendante du contexte pour décrire des phénomènes.

Slide : (citazione sui contesti infrastrutturali del MC + citazione di Mouza)

Or, ce caractère non ambigu et indépendant du contexte du modèle, permet potentiellement de communiquer de manière très précise ce que chaque élément représente, afin qu'il soit possible de s'en approprier. Et le fait qu'une technologie soit appropriables dans des contextes qui ont accès à des infrastructures différentes est un des grands axes de l'approche minimal computing.

Sur un premier niveau d'analyse, communiquer à travers un modèle, en intégrant les aspects interprétatifs du travail de modélisation, serait donc tout-à-fait cohérent avec le minimal computing.

Mais une grande question reste néanmoins ouverte pour moi : pour que cette communication précise soit possible, il est nécessaire que les personnes comprennent le même langage – cela ne veut pas nécessairement dire qu'elles doivent parler la même langue naturelle, cela veut surtout dire partager les codes utilisés dans le modèle, mais aussi dans les pratiques de recherche. Cette question est abordée dans l'article de Cédric du Mouza, Stéphane Lamassé et Philippe Rygiel « De l'histoire numérique à l'histoire données ? ».

Par exemple, lorsqu'on a implémenté la base de donnée avec l'ingénieure informatique Camilla Cantiani, elle a très rapidement corrigé la normalisation, et crée la base de l'interface, ainsi que la base de données, à partir uniquement de l'image du modèle et de sa description en langage naturel. Pour elle ce langage était compréhensible et opérationnel. Ainsi, un modèle semble un outil particulièrement efficace pour communiquer entre scientifiques (sciences formelles) des informations analytiques et interprétatives concernant des objets littéraires, mais est-ce aussi efficace pour la communication interdisciplinaire ? Suivant l'article de Mouza, Lamassé et Rygiel, le modèle apparaît toujours comme un bon outil, peut-être le meilleurs qu'on a à disposition, mais qui n'est pas suffisant à lui seul.

Slide 8 : (citazione Sayers + principi di ER)

Nous venons de voir comment le minimal computing a influencé le modèle et les questions que je me pose sur l'utilité de la modélisation. Nous allons passer maintenant dans la dernière partie : évaluer les principes de modélisations définis dans l'article d'ER de 2020, par rapport aux objectifs et aux présupposées du projet.

ER défend l'idée que la manière dont on modélise le monde a des conséquences directes sur l'infrastructure nécessaire à implémenter ces visions du monde. Il utilise alors le MC pour réduire la complexité de ces modèles. Concrètement, il mentionne 4 principes proposés par Jenterey Sayers dans un article de 2016, et en ajoute deux nouveaux.

Les principes de Sayers sont définis dans la slide :

1. minimal design
2. minimal maintenance
3. minimal obsolescence
4. maximum justice

ER ajoute deux autres principes pour guider une modélisation minimal computing :

1. maximum materiality, qui sert à ancrer le modèle dans les aspects matériels et physiques des objets modélisés.
2. maximum transparency, pour expliciter l'écart entre les éléments basés dans la réalité concrète et les éléments interprétatifs.

L'interprétation apparaît ainsi comme un élément étranger à la modélisation. Elle doit donc être le plus possible éliminée (maximum materiality). Lorsque cela n'est pas possible, il est nécessaire d'expliquer la divergence entre l'interprétation et la réalité (maximum transparency).

Slide 9 : (citations Mouza)

Or cela n'est pas totalement compatible avec le constat que la création d'une BDD est un travail interprétatif et que toute modélisation représente une certaine interprétation des objets qu'on étudie. Si cela est vrai dans tous les contextes, même les plus scientifiques formels, le constat prend un sens particulier dans les études littéraires, où l'acte interprétatif est le fondement de toute recherche. Si nous avons suivi ces principes de manière stricte, nous aurions du, par exemple, représenter les éditions des pièces dans une entité distincte, car dans la *réalité concrète* l'édition d'une pièce est un objet profondément différent du texte jouée, aussi bien que de la performance.

Est-ce que dans les humanités il existe un modèle qui soit plus *proche de la réalité* que toutes les autres modèles ?

Si dans les humanités l'acte interprétatif est le fondement de toute recherche, nos modèles pourraient le valoriser. Je pense que ce propos est proche des considérations faites pour l'incertitude en histoire dans l'article de Mouza, Lamassé et Rygiel.

Je tiens à ajouter un tout dernier aspect, absent de l'article de Risam et Gil, et qui compléterait à mon avis le cadre des problématiques écologiques liées à la mise en place des projets numériques : ce sont les conditions de travail des personnes qui accomplissent de fait les tâches. Les humanités numériques sont largement confrontées au problème de la pérennisation des projets, et nombreuses personnes font le constat que ce problème technique est empiré par le grand turnover dans les labos de recherche, qui concerne surtout des postes les plus précaires. Or, très souvent, les personnes aux postes les plus précaires sont celles qui assurent de fait l'implémentation des projets (ingénieures de recherches, post-doc, stagiaires, doctorant.e.s). Intégrer les conditions de travail dans nos évaluations ne serait donc pas seulement juste d'un point de vue écologique-social, mais aussi bénéfiques pour la qualité de nos projets.

Slide 10 :

Quelques mots pour conclure : lors du travail de modélisation que j'ai accompli pour le moment, le minimal computing a fourni des principes utiles et opérationnels pour évaluer l'impact écologique et l'appropriabilité du projet. En revanche, on pourrait intégrer la prise en compte des conditions de travail des personnes qui créent le projet – non seulement de celles qui l'utilisent ou qui veulent s'en approprier. L'approche s'est révélée aussi plutôt faible en ce qui concerne la modélisation des aspects interprétatifs. Suivant les principes mentionnés dans ER, 2020, on pourrait même penser que l'interprétation va à l'encontre de l'écologie, en complexifiant le modèle (ce qui n'est pas forcément vrai, voir éditions des pièces). Il serait intéressant, à mon avis, de définir des principes où les deux approches ne seraient pas opposées, voire incompatibles.