

Chapitre 1. Préliminaires méthodologiques

A beginning is a very delicate time.
Frank Herbert, *Dune*, 1965.

La thèse que nous défendons ici est celle d'un informaticien qui a tenté de se laisser « déplacer » par d'autres disciplines au sein du réseau interdisciplinaire constitué et animé par Andrea Iacovella depuis cinq ans.

Pour autant, est-il correct de dire qu'il s'agit d'une « thèse interdisciplinaire » ? Assurément non, une telle expression serait un oxymoron. En effet, rien n'est plus disciplinaire qu'une thèse. Notre thèse ne peut donc qu'être profondément marquée par les méthodes et le discours de la discipline informatique. Autrement dit, si l'on comparait le réseau interdisciplinaire à un polyèdre, notre travail constituerait juste une facette, parmi celles des archéologues, historiens, architectes, historiographes...

Ce travail prendra donc tout son sens une fois accompagné de ceux des autres acteurs du réseau. En attendant, il est toutefois possible, dans ce chapitre préliminaire, de retracer la dynamique du réseau interdisciplinaire, et de replacer ainsi cette thèse dans sa situation d'écriture.

Dans une première partie, nous verrons comment le réseau doit son émergence à une problématique issue des sciences historiques. Ensuite, dans une deuxième partie, nous présenterons l'approche adoptée par le réseau, une approche à la fois interdisciplinaire et opératoire. Par la suite, dans une troisième partie, nous dresserons une liste des difficultés rencontrées puis évitées par le réseau. Enfin, dans une dernière partie, nous tenterons de définir l'objet d'étude commun aux différentes disciplines représentées.

1. Une problématique issue des sciences historiques

Le point de départ de ce réseau réside dans un questionnement théorique sur la création du sens dans les Sciences Historiques. En particulier, Andrea Iacovella a eu l'occasion à plusieurs reprises d'identifier comme une impasse méthodologique le cou-

rant de l'archéologie défendant la définition normative d'une typologie des vestiges (ou l'utilisation de thesaurus pour l'interrogation de « banques de données »).

À la différence du spécialiste des Sciences de l'Information et de la Communication, occupé à affecter des significations non ambivalentes et non ambiguës en regard des corpus documentaires, l'historien manipule un objet qui déborde de sens de toutes parts. Dans un index, un catalogue, un titre d'ouvrage ou dans un corps de texte, le mot « Auschwitz » ne se laisse pas résumer à de l'information. Écrit en toutes lettres ou représenté par son équivalent binaire, Auschwitz reste Auschwitz. Dans le seul fait de prononcer un mot comme celui-là, il y a le risque d'une banalisation qui effraie l'esprit ; le travail de l'historien, n'est pas tant de produire des catégories bien huilées, mais d'élaborer un langage qui permettra à des mots comme « Auschwitz » de se maintenir en toute dignité dans une phrase et de contribuer par là à donner tout son sens à une communauté du dialogue. [IacovellaEtAl03]

Comme le montre la gravité de l'exemple, la question du nécessaire renouvellement des méthodes dans les sciences historiques est relayée par les grands débats de société actuels. Comment se fait-il par exemple que les méthodes de l'Histoire n'aient pu endiguer des dérives telles que celles du négationnisme portant sur la Shoah ? Même s'il est reconnu aujourd'hui un écart entre Méthode et Vérité, peut-on pour autant dire n'importe quoi du moment que l'on adopte le discours de l'érudit ? Puisque l'activité scientifique (quelle que soit la discipline) est incapable de juger du caractère éthique d'une théorie, n'est-il pas nécessaire de faire appel au politique (au sens large) ? Sans recourir forcément à des « comités d'éthique », la solution se trouve peut-être dans un nouveau type de « vulgarisation » qui ôterait à ce mot son caractère péjoratif, autrement dit, dans une meilleure interaction entre le débat scientifique et le débat civique.

On aurait pu croire que le problème serait réglé par l'avènement de la « Société de l'Information ». Au contraire, sur la Toile, le problème est accentué par :

- l'explosion du nombre de pages personnelles, reflet d'un certain relativisme postmoderne (prônant que « tout est vrai »),
- la « confiscation » des revues scientifiques par des monopoles éditoriaux imposant peu à peu des modèles économiques rendant leur accès inabordable pour l'amateur « éclairé ».

Ainsi, le but scientifique du réseau vient à la rencontre d'un but plus politique, celui de passer de la « Société de l'information » à une nouvelle dynamique à inventer : la « Société de la Connaissance ».

2. Une approche à la fois interdisciplinaire et opératoire

La double nature (scientifique/éthique) de la problématique nécessitait de « tresser » ensemble deux « fils » : celui de la Recherche et celui du Service (c'est-à-dire de l'action). Ce n'est donc nullement un hasard si cette dynamique a pu voir le jour sous la direction d'Andrea Iacovella, à la fois chercheur en historiographie et responsable du Service Informatique de l'Ecole française d'Athènes (EFA).

La première étape de cette dynamique consista à évaluer la faisabilité de la mise en ligne de la *Chronique des fouilles* publiée par l'EFA. Dans cette perspective, une collaboration fut entreprise entre l'EFA et le Département informatique de l'INSA de Lyon sous la forme d'un certain nombre de stages [Komninidis95, Larcher97, Montocchio97, Benel98]. Il en ressortit que, si les technologies étaient relativement mûres pour la mise en ligne de collections, il subsistait en informatique un « verrou » concernant la gestion de leur sémantique. Andrea Iacovella décida, avec le soutien de Roland Etienne alors directeur de l'EFA, d'associer au projet de mise en ligne des collections, un projet de recherche interdisciplinaire ayant pour but de faire tomber ce « verrou ».

a. Mise en ligne des collections scientifiques de l'EFA

D'abord consacré à la seule *Chronique des fouilles*, le projet de mise en ligne fut rapidement généralisé au *Bulletin de Correspondance Hellénique* [BCH], puis à l'ensemble des collections publiées par l'EFA depuis ses 150 ans d'existence, soit près de 250 000 pages en 570 volumes. Les *Collections de l'Ecole française en ligne* [CEFAEL] constituent ainsi une bibliothèque numérique à la fois exhaustive et publique.

En 2001, CEFAEL fit l'objet d'un prototypage complet, portant sur la chaîne de numérisation et la diffusion sur la Toile du BCH. Le déroulement de cette expérimentation fut rendu possible par la signature d'une convention entre l'EFA et la Maison de

l'Orient et de la Méditerranée [MOM]. Les résultats de cette expérimentation reçurent le franc soutien du Ministère de la Recherche. Ce dernier accorda alors sur le « Fonds National pour la Science » le budget nécessaire pour passer en phase de production. Par la suite, furent étudiées les conditions de l'hébergement de CEFAEL au Centre Informatique National de l'Enseignement Supérieur [CINES].

A mi-parcours du projet, furent organisées des « Journées Bibliothèques Numériques² » afin d'apporter la contribution de l'équipe au débat sur les bibliothèques numériques en France, en particulier en termes de mise en place de modèles de production et de diffusion. A cet effet, un guide pratique³ fut distribué à tous les participants.

Après achèvement de la numérisation, saisie des sommaires, finalisation du modèle des données et transfert sur les serveurs du CINES, le portail CEFAEL⁴ fut inauguré le 9 avril 2003.

b. Recherches sur la gestion de la sémantique des collections scientifiques

Le processus que nous venons de présenter s'est déroulé de pair avec un travail de recherche sur le renouvellement des méthodes dans les disciplines historiques. Plus particulièrement, notre équipe s'est attachée à trouver une solution interdisciplinaire au « verrou » de la discipline informatique évoqué plus haut.

Cette orientation interdisciplinaire s'est affirmée en particulier dans l'organisation d'une table ronde⁵ sur le thème « Sémantique et Archéologie » rassem-

² Journées d'études des 29-30 mai 2002, organisées à Lyon par l'Ecole française d'Athènes, la Maison de l'Orient Méditerranéen et l'Université Lyon 2, avec le soutien de la Sous Direction des Bibliothèques (Ministère de l'Education Nationale).

³ *Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques* (expérimentés à partir de la numérisation des publications de l'Ecole française d'Athènes), Manuel publié dans le cadre des journées d'études « Bibliothèques Numériques », Lyon, 29-30 mai 2002.

⁴ <http://cefael.efa.gr>

⁵ Sémantique et Archéologie : Aspects expérimentaux (Renouvellements méthodologiques dans les bibliothèques numériques et les publications scientifiques), Table ronde interdisciplinaire, Athènes, 18-19 novembre 2000.

CHAPITRE 1. PRÉLIMINAIRES MÉTHODOLOGIQUES

blant des intervenants issus de l'Archéologie, de l'Historiographie, de la Sémiotique, des Sciences Cognitives, et de l'Informatique. La dynamique entamée par cette table ronde s'est poursuivie par la création d'un réseau interdisciplinaire (EFA, LISI⁶, MOM, GRACO⁷) se retrouvant périodiquement dans des séminaires de travail.

Les aspects méthodologiques du problème ne firent pas oublier la nécessité d'offrir des solutions opératoires. Aussi, une convention entre l'EFA et le LISI fut signée en 2000 pour la création d'un prototype informatique (appelé « Porphyre »). De même, le volet « recherche en informatique » fut encouragé par le lancement de notre thèse sous la codirection d'Andrea Iacovella pour l'EFA et de Jean-Marie Pinon (assisté de Sylvie Calabretto) pour le LISI.

Les travaux du réseau interdisciplinaire furent grandement soutenus par le CNRS et le Ministère de la Recherche. De 1999 à 2002, notre thèse fit l'objet d'une allocation de recherche. De plus, les départements SHS⁸ et STIC⁹ du CNRS, dans le cadre du Programme « Société de l'Information », nous accordèrent en 2001 un financement pour un projet sur la modélisation du temps dans les publications archéologiques mises en ligne¹⁰ et en 2003 sur la sémiotique spatiale du document d'architecture en archéologie¹¹. Il y a peu, le réseau a déposé au département STIC du CNRS une demande pour constituer une équipe projet¹².

⁶ LISI : Laboratoire d'Ingénierie des Systèmes d'Information, Lyon. Aujourd'hui intégré au LIRIS (Laboratoire d'InfoRmatique en Images et Système d'Information).

⁷ GRACO : Groupe de Recherche sur l'Antiquité Classique et Orientale, Toulouse.

⁸ SHS : Sciences Humaines et Sociales.

⁹ STIC : Sciences et Techniques de l'Information et de la Communication.

¹⁰ « Assistance dans la gestion de ressources intertextuelles multiformes. Production et intégration interactives de parcours interprétatifs ».

¹¹ « Modélisation opératoire d'un système d'interrogation du document d'architecture en archéologie à base d'un langage visuel : application aux corpus numérisés et à la modélisation des connaissances ».

¹² « Du partage de corpus de documents structurés à la confrontation de points de vue ».

3. Cartographie des écueils

Si l'on peut se réjouir aujourd'hui du trajet parcouru par le réseau interdisciplinaire face à cette question de la gestion de la sémantique des collections scientifiques, il est cependant nécessaire de noter que tout cela ne s'est pas fait sans mal. L'interdisciplinarité a ceci de particulièrement délicat qu'y faire appel, c'est reconnaître qu'aucune discipline ne possède la réponse à la question posée [IacovellaEtAl03]. Elle se distingue en ceci de la pluridisciplinarité. Par exemple, lorsque l'archéologue demande au physicien de lui dater un vestige au radiocarbone, il s'agit de pluridisciplinarité. Au contraire, si les incohérences de cette mesure amènent le physicien à mettre en doute des hypothèses tacites de sa discipline, il s'agit d'un premier pas vers l'interdisciplinarité.

En l'absence d'un *manuel de l'interdisciplinarité*, nous avons été contraints de « naviguer à vue » entre les écueils. Certains de ces écueils sont peut-être encore présents dans cette thèse, cependant nous avons tenté à chaque « collision » de les « cartographier » et de les éviter par la suite. En espérant que ceci puisse bénéficier à d'autres projets interdisciplinaires, nous allons maintenant tenter de dresser une liste de ces écueils.

a. Création d'une nouvelle discipline

Le premier danger qui pèse sur l'interdisciplinarité consiste en l'affranchissement à l'égard des disciplines d'origine, en une fondation comme discipline scientifique autonome [IacovellaEtAl03]. En effet, quand une discipline se forme par fusion de deux autres (comme par exemple la biophysique à partir de la biologie et de la physique), on serait en droit de se demander si un débat entre deux disciplines ne serait pas plus stimulant pour la Science que la création d'un consensus local.

Conscient de ce danger, nous avons résolument choisi durant notre thèse de ne pas cantonner nos publications aux seuls colloques interdisciplinaires [BenelEtAl99, BenelEtCalabretto00, Benel00], mais de soumettre également des articles dans des colloques clairement identifiés comme appartenant à la discipline informatique [BenelE-

tAl00a, BenelEtAl00b, BenelEtAl01a, BenelEtAl02]. Ainsi nos pairs, en acceptant nos articles, ont pu nous signifier notre appartenance à la discipline informatique.

b. Instrumentalisation et réification des disciplines

Dans une première version de ce chapitre, nous avons analysé les buts poursuivis par l'EFA, ses flux d'information avec l'extérieur, son organisation interne, etc. En somme, nous avons appliqué naïvement les méthodes d'analyse de systèmes d'information. On nous a reproché, à raison, de réifier par là la discipline archéologique. En effet, dans le cas qui nous occupe, une véritable posture interdisciplinaire nécessite d'instrumenter l'archéologie sans pour autant l'instrumentaliser. La relation entre STIC et SHS que nous tentons de promouvoir dans cette thèse est la suivante :

- les SHS proposant des « grilles de lecture » aux théories des STIC,
- les STIC proposant des instruments aux pratiques en SHS.

Dans tous les cas, il est nécessaire, d'une part, que soit réciproque l'application de méthodes d'une discipline à l'autre, et d'autre part, que cette application soit matière à un questionnement théorique des disciplines.

c. Impérialisme d'une discipline

Au cours de la constitution du réseau, la collaboration avec certaines disciplines fut tellement difficile que le responsable du réseau dut prendre la difficile décision d'y mettre fin. Avec le recul, l'impossibilité de travailler avec ces disciplines s'expliquerait peut être par leur aspiration à devenir des « méta-disciplines », des censeurs des autres disciplines. En effet, comme nous l'évoquions plus haut, la condition *sine qua non* pour qu'un chercheur s'implique dans une démarche interdisciplinaire, c'est de reconnaître que quelque chose « échappe » à sa discipline.

d. Dissymétrie des « produits »

Aujourd'hui lorsque l'on parle du réseau interdisciplinaire, la plupart de nos interlocuteurs nous demandent si *Porphyre* fonctionne et si l'on peut leur en faire une démonstration. Il y a là un grave danger de rester rivé aux seuls aspects opératoires [la-

covellaEtAI03]. Reste dans l'ombre non seulement tout l'aspect théorique des recherches informatiques, mais aussi, ce qui est plus grave encore, la part des Sciences Humaines dans le projet.

Si la part de chacun dans le projet apparaît aujourd'hui de manière aussi faussée, c'est probablement que nous n'avons pas été suffisamment attentifs à respecter une certaine symétrie des produits¹³ non seulement entre théorie et expérimentation¹⁴, mais encore entre SHS et STIC. Ainsi, il pourrait être intéressant qu'à l'issue d'un cycle de séminaires chaque intervenant écrive quelques pages présentant son point de vue, disciplinaire, sur le débat qui a eu lieu. Ces synthèses, accessibles par tous, permettraient ainsi de donner une meilleure transparence de la dynamique interdisciplinaire qui se joue dans le réseau.

4. Définition de l'objet interdisciplinaire

Ayant toutes ces précautions à l'esprit, on pourrait nous reprocher d'avoir pris comme objet d'étude interdisciplinaire le verrou existant au sein des STIC concernant la gestion de la sémantique des collections scientifiques. Cependant, il est nécessaire de préciser que le seul fait de prendre un objet d'une discipline et de le livrer à l'étude des autres disciplines change cet objet.

Lorsque nous disons, par exemple, que les *logiques temporelles* sont un thème de recherches interdisciplinaires associé au projet, ça ne signifie pas que par *logiques temporelles* on entend leur sens théorique érigé en dogme de référence [...] et qui donnerait un statut majeur au discours de l'informaticien sur les autres disciplines. L'historien, l'archéologue, l'historiographe sont associés au projet comme « experts » du temps historique et à ce titre ils contribuent à un approfondissement de cette catégorie conceptuelle à l'intérieur des logiques temporelles. [IacovellaEtAI03]

¹³ Rapports, articles scientifiques, prototypes, manuels, etc

¹⁴ A titre d'illustration, *Porphyre*, avec la vingtaine de stagiaires qui ont participé à son développement, représente 18 000 lignes de code source, alors que cette thèse en représente environ 4 000.

Parce que la question de la consultation des documents concerne différentes communautés au sein de la discipline informatique, il devient nécessaire, avant d'aller plus loin, de voir comment ce « verrou » se décline d'une communauté à l'autre.

a. Recherche d'informations

La première communauté, celle de la « recherche d'information », a été popularisée par les moteurs de recherche sur la Toile (Altavista, Google...). Ses techniques consistent en général à indexer les documents à l'aide des mots du texte qui les discriminent le plus du corpus. Notons que d'un point de vue philosophique il n'est nullement question ici de « contenu »¹⁵, seules étant considérées les propriétés statistiques de la « forme » (distribution des mots, des liens hypertextes, etc.).

Cette approche s'explique par l'histoire du concept d'information dans la discipline informatique. En 1948, lorsque Claude E. Shannon, ingénieur chez Bell, établit sa « théorie de l'information » [Shannon48], il quantifia l'information en fonction de la nouveauté d'un message par rapport aux messages reçus, c'est à dire comme ce qui est statistiquement improbable. Son but était d'obtenir un modèle permettant d'augmenter le rendement de la transmission à travers un réseau de communication. Son travail, d'ailleurs, a eu et continue d'avoir des applications spectaculaires en télécommunication et en compression de données.

Cependant, selon Bernard Miège (en Sciences de la Communication), l'application de cette théorie à d'autres domaines, même couronnée d'un certain succès, serait l'affaire d'un terrible quiproquo [Miège95]. Tout d'abord, on reproche souvent à la théorie de ne pas tenir compte de la composante sémantique du message. Plus précisément, le concept d'information peut s'opposer parfois à celui de signification, puisque, comme l'a expliqué Umberto Eco [Eco62], plus un message sera signifiant (non ambigu, redondant), moins il sera porteur d'information. D'autre part, le modèle est

¹⁵ Même si paradoxalement on parle de « recherche par le contenu » en référence au « texte intégral ».

celui d'un canal de communication : ni l'émetteur ni le récepteur ne sont modélisés¹⁶. Les seules ambiguïtés considérées sont celles dues au « bruit » présent sur le canal. Enfin, le modèle requiert un ensemble fini de messages possibles, donné *a priori*, ce qui oblige en pratique à appliquer le modèle à des signaux simples voire simplistes (suites de caractères de longueur limitée).

b. Modélisation des connaissances

La seconde communauté est celle de la « modélisation des connaissances » dont les technologies les plus connues sont les systèmes de bases de données et de connaissances. Dans le domaine documentaire, sa méthode consiste à réaliser des modèles formels des besoins de l'utilisateur et du contenu des documents. Un document est apparié à un besoin si l'on peut déduire à partir du modèle du premier le modèle du second. C'est le cas par exemple de l'approche booléenne utilisée dès les années 60 [Salton68] dans les catalogues informatisés des bibliothèques. Un document indexé par « philosophe ET mortel » sera apparié à la requête « philosophe OU grec ». De même, dans des systèmes plus récents, un document indexé par un graphe modélisant « Le philosophe Socrate est mortel » sera apparié au graphe requête modélisant « Un homme est mortel ».

Cette approche, participant au projet de « l'intelligence artificielle », est calquée sur la démonstration automatique de théorèmes. Elle consiste à démontrer des « buts » en appliquant successivement des règles de transformation à des « faits » de base.

Aujourd'hui, les plus vives critiques de l'approche viennent de la communauté elle-même. Ainsi, John F. Sowa, inventeur des célèbres « graphes conceptuels », dénonce-t-il l'idéologie dominante de sa communauté comme étant devenu une « nuisance obstructive » [Sowa01a]. Pour conforter son propos, il nous présente le parcours de Terry Winograd. Celui-ci écrivit en 1972 un livre intitulé « Comprendre le langage naturel » où il exposait le système SHRDLU capable de dialoguer avec un être humain au sujet de cubes empilés les uns sur les autres. En 1986, après avoir discuté avec des phé-

¹⁶ Il est représentatif de voir qu'à un colloque de recherche d'information comme SIGIR, la « recherche d'information interactive » est souvent tenue à l'écart des conférences plénières et reléguée aux « ateliers ».

noménologues, il expliquait les faiblesses de sa première approche : l'oubli du contexte, du cadre de référence, du besoin de négociation et des inévitables exceptions. Des aspects que le courant philosophique à l'origine du domaine avait volontairement mis de côté. Convaincu que l'intelligence artificielle menait à une impasse, Terry Winograd travaille aujourd'hui dans le domaine des interactions homme-machine.

c. Hypertexte

La troisième communauté à considérer est celle de l'hypertexte (ou hypermédia). Chacun de nous a tendance à l'associer au « World Wide Web », c'est-à-dire à une version simplifiée inventée en 1989-90 par le CERN pour ses besoins propres [Berners-Lee89].

La notion, quant à elle, remonte aux années 1945 [Bush45]¹⁷. Il s'agissait à l'époque de construire une machine (mécanique) permettant de « feuilleter » des microfilms. Le principe était en effet le suivant : associer à un *contenu* documentaire interprétable uniquement par l'homme, une *structure* (ou parcours) gérable par une machine.

Si l'approche a un succès retentissant, certains chercheurs remarquent néanmoins que de nombreux lecteurs d'hypertextes se trouvent désorientés, « perdus dans l'hyperespace ». En effet, contrairement à un texte, il est souvent difficile de remettre en situation un hypertexte, qu'il s'agisse de sa situation d'écriture (Qui a écrit quoi et quand ?) ou de lecture (Où suis-je ? D'où viens-je ? Où puis-je aller ?). En somme, bien souvent, l'hypertexte serait moins qu'un texte : il serait un « hypotexte » [Bachimont99b].

d. Quelle grille de lecture ?

Dans les trois communautés considérées, nous avons vu qu'il existait des critiques fondamentales portant non sur les techniques elles-mêmes, mais plus sur le cadre

¹⁷ Le terme, lui, n'apparaît que vingt ans plus tard (THEODOR H. NELSON, « The Hypertext », in *Proceedings of the World Documentation Federation*, 1965).

philosophique dans lequel elles sont mises en œuvre. Il serait donc vain de s'en tenir à des typologies purement informatiques.

On mesure alors en quoi les débats menés au sein du réseau interdisciplinaire nous ont été indispensables. Néanmoins, ces apports n'apparaîtront pas en tant que tel puisqu'il nous était impossible d'appuyer notre argumentaire sur des échanges verbaux. Nous avons donc tenté de trouver des articles ou des monographies y faisant écho. Les prises de position des participants restant très avant-gardistes dans la discipline archéologique, nous avons été contraints à généraliser notre « état de l'art » aux Sciences Humaines voire même pour certains aspects aux Sciences en général. En termes purement méthodologiques, ceci revient juste à prendre plus de risques, étant donné que plus un discours scientifique prétend à l'universalité et plus il sera facile à réfuter...