

# Quelle interdisciplinarité pour les “ humanités numériques ” ?

Aurélien Bénel

► **To cite this version:**

Aurélien Bénel. Quelle interdisciplinarité pour les “ humanités numériques ”?. Les Cahiers du numérique, Lavoisier, 2014, 10 (4), pp.103-132. 10.3166/lcn.10.4.103-132 . hal-02362479

**HAL Id: hal-02362479**

**<https://hal-utt.archives-ouvertes.fr/hal-02362479>**

Submitted on 13 Nov 2019

**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.

# Quelle interdisciplinarité pour les « humanités numériques » ?

AURELIEN BENEL

Constatant que le mouvement qui prend forme des « humanités numériques » entretient fort peu d'interactions avec la recherche en informatique, nous commençons notre article par l'étude d'exemples d'interdisciplinarité réussie au début des années 1970 entre archéologie et sciences cognitives, et à la fin des années 1950 entre théologie, philologie et mécanographie. Dans la lignée de ces exemples et en nous basant sur notre expérience, nous proposons de refonder les humanités numériques et l'interdisciplinarité STIC-SHS autour de trois pôles : *immersion*, *instrumentation* et *discussion épistémologique*. En outre, en présentant quelques résultats et directions de notre programme scientifique visant à *repenser l'informatique à l'aune de l'herméneutique*, nous explorons la piste des STIC comme pivot du débat interdisciplinaire entre SHS.

## 1. Introduction

Les « humanités numériques » agacent. Le nom, très « marketing », censé réunir les anciens et les modernes, n'y est sans doute pas pour rien. Apparu d'abord dans quelques universités américaines comme intitulé de master (Hockey, 2004), puis en 2004 comme titre d'un manuel d'enseignement rédigé de manière collective (Schreibman et al., 2004), le nom devient en 2006 celui d'un programme du fond américain de financement pour les humanités. Le succès fulgurant du terme aux Etats-Unis encourage alors des collectifs à demander la création de formations et d'infrastructures (donc de postes et de financements) dans les autres pays (Dacos, 2011). Le sujet des postes et des

financements étant toujours délicat dans le monde universitaire (particulièrement dans le domaine des Sciences de l'homme et de la société), l'agacement des chercheurs en place se transforme rapidement en opposition. De leur côté, les partisans de ces « nouvelles humanités » défient les règles de l'évaluation universitaire : se rencontrant dans des non-colloques (*unconference*), sans évaluation par les pairs ni programme préalable, publiés dans des non-actes (*unproceedings*). Les chercheurs en SHS (traditionnelles) commencent à douter publiquement de la contribution des humanités numériques à leurs disciplines. La référence délicieusement subversive à Lewis Carroll<sup>1</sup> n'y change rien, sans évaluation par les pairs il n'y a pas de recherche scientifique.

Le conflit générationnel au sein des SHS en cache un autre, plus difficile à distinguer, avec les « humanities computing », que l'on pourrait traduire imparfaitement par « informatique pour les humanités<sup>2</sup> ». De manière tout à fait surprenante, le manuel des humanités numériques s'ouvre sur un chapitre dédié non à l'histoire des *humanités numériques* mais à celle de *l'informatique pour les humanités* (Hockey, 2004), sans que l'on sache s'il s'agit de la part de Susan Hockey d'un acte de résistance pour sauver de l'oubli le champ auquel elle a consacré tout sa carrière, s'il s'agit de la trace d'un changement de dernière minute du titre de l'ouvrage collectif pour des raisons éditoriales, ou tout autre raison inaccessible au lecteur. Mais de fait, les humanités numériques ont un rapport singulier à leur propre histoire. Sur Wikipédia, son article en anglais et surtout son historique sont à ce sujet très éclairants. Cette page créée et éditée par les acteurs du mouvement<sup>3</sup>, vient rapidement remplacer celle de l'informatique pour les humanités, et réécrire leur histoire commune. De l'histoire passionnante de l'informatique pour les humanités établie par Susan Hockey, l'article de Wikipédia (2013) ne garde que deux figures emblématiques incarnant en quelque sorte le passé et le présent :

– Roberto Busa, dont le travail remarquable (nous le verrons plus loin) est occulté par une forme d'héroïsation – homme des « origines » (Wikipedia, 2013), du « commencement » (Hockey, 2004), ayant donné son nom à un prix honorifique,

– la *Text Encoding Initiative* qui, comme son nom l'indique, tente de promouvoir auprès des chercheurs en SHS (depuis plus de 25 ans) l'encodage

---

<sup>1</sup> De l'autre côté du miroir, Humpty Dumpty souhaite à Alice un joyeux « non-anniversaire » (*unbirthday*).

<sup>2</sup> Ou même « informatique pour les sciences de l'homme », pour éviter la signification trop étroite du terme « humanités » en français.

<sup>3</sup> Ce qui contrevient aux principes de Wikipédia.

des textes qu'ils étudient avec des balises standards censées refléter leur travail de recherche.

Le passage du « computing » au « digital » traduit aussi une disparition du paradigme du calcul, de l'informatique comme expertise, à une informatique de la vie quotidienne. Des mouvements analogues sont déjà mentionnés par Susan Hockey notamment au milieu des années 80, quand les chercheurs en SHS « s'émancipèrent » des centres de calcul suite à l'avènement des ordinateurs personnels et des hypertextes programmables par des autodidactes (HyperCard). Mais ici la rupture est consommée. Malgré les références incessantes à l'interdisciplinarité, à la rencontre des « deux cultures », l'histoire des humanités numériques, plus encore que celle de l'informatique pour les humanités, est une histoire sans informaticiens.

L'interdisciplinarité concerne le transfert de méthodes entre disciplines (Nicolescu, 1996, p. 65). Ces méthodes, au contact d'objets, de questions et d'épistémologies radicalement nouvelles, s'en retrouvent à jamais transformées. Ainsi, il n'y a pas d'interdisciplinarité sans engagement des chercheurs des deux disciplines (ou plus), pas d'interdisciplinarité réussie sans enrichissement mutuel.

Ni l'usage d'un blog en histoire, ni la réalisation d'une base de données en littérature ne sont des projets interdisciplinaires. L'embauche d'un « informaticien » n'y change rien. L'interdisciplinarité nécessite un statut équivalent des représentants des disciplines, un respect de l'autre comme exerçant une discipline scientifique de plein droit. Quand une discipline est éclipsée par ses productions auprès du grand public, l'autre discipline ne doit pas l'enfermer dans ce qu'elle attend de ses productions, mais lui laisser la latitude d'innover. Ce besoin d'innovation signifie aussi qu'un type de projet, comme l'encodage savant de textes avec des balises, peut avoir été interdisciplinaire en 1987 et ne plus l'être aujourd'hui<sup>4</sup>.

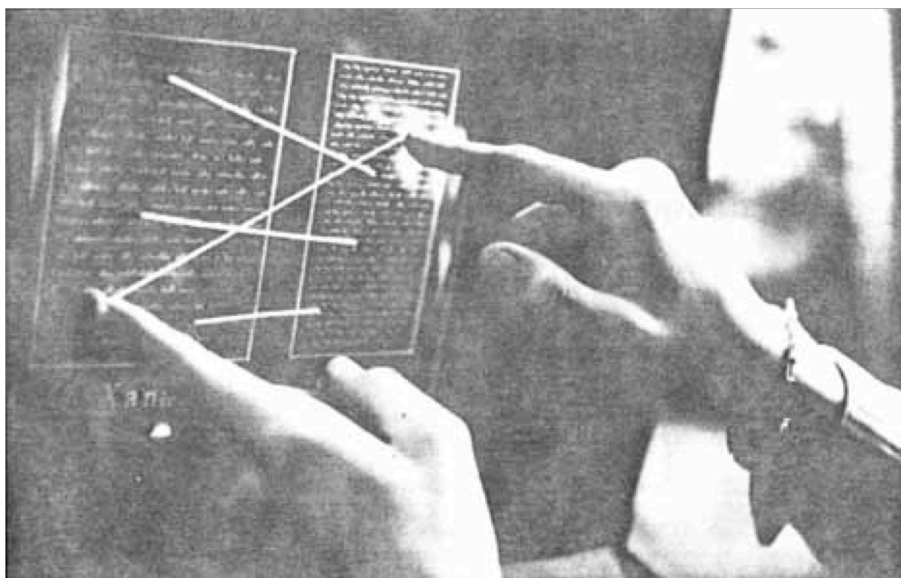
Ce qui précède pourrait sembler bien utopique, pourtant, comme nous le verrons dans une première partie, l'histoire croisée de l'informatique et des SHS recèle de nombreux exemples d'enrichissements mutuels. Informés par cette interdisciplinarité en acte, nous reviendrons alors à Roberto Busa pour discerner les caractéristiques de l'interdisciplinarité qui se déploient dans ses travaux. Enfin, dans le but de construire ce genre d'interdisciplinarité, nous présenterons le type d'interactions entre STIC et SHS que nous avons explorées.

---

<sup>4</sup> À moins de s'atteler par exemple au problème des structures concurrentes, identifié depuis fort longtemps (Hockey, 1996) mais jamais traité encore de manière pleinement satisfaisante (Abascal *et al.*, 2004).

## 2. Il était une fois l'interdisciplinarité

Un premier exemple nous est fourni par Theodor Nelson qui en 1960, alors diplômé de philosophie et étudiant en sociologie, imagine de transformer les « calculateurs » en « machines littéraires ». Ses maquettes (*cf.* Fig. 1) témoignent à la fois d'une compréhension fine de la structure en graphe des données dans un ordinateur, et d'une capacité à détourner cette structure abstraite et les interfaces concrètes de l'époque afin d'offrir un nouveau type de support aux opérations mentales des SHS. Transférées sauvagement du monde de l'érudition à celui d'une liste de courses devenue célèbre<sup>5</sup>, ses notions d'« hypertexte » et de « transpointing » (même mal comprises) vont durablement transformer les interfaces de nos ordinateurs ainsi que notre manière de penser la connaissance et son inscription.



*Fig. 1. Lecture savante sur une « machine littéraire » – Maquettes de Xanadu, 1972*

La France des années 1970, et plus particulièrement le Laboratoire d'informatique pour les sciences de l'homme (LISH), nous fournissent deux autres exemples, pilotés cette fois-ci par l'informatique au sens large.

---

<sup>5</sup> Démonstration du NLS (on-Line System) par Douglas Engelbart : <http://sloan.stanford.edu/MouseSite/1968Demo.html>

Le sujet d'étude du projet AVEROES (Borillo, 1984, 45-65) est le commerce maritime antique en Méditerranée Occidentale tel que l'on peut l'analyser à partir des amphores retrouvées de nos jours. Comme le développé de son acronyme l'indique (« Analyse et validation du raisonnement »), le projet répond avant tout à une problématique des sciences cognitives. L'archéologie est ici un « terrain » d'expérimentation. Il s'agit en effet pour Mario Borillo et son équipe de prendre à contre-pied la tendance des l'intelligence artificielle à ne considérer que des mondes arbitraires (jeux) ou extrêmement simplifiés, déjà très proches de l'univers formel des mathématiciens. L'approche ici revient au contraire à analyser des discours scientifiques réels en formalisant les observations, les connaissances et les modes de raisonnement, jusqu'à ce que l'on arrive à reproduire de manière formelle des raisonnements attestés (cf. Fig. 2).

En somme, ces travaux partagent avec l'intelligence artificielle des débuts l'objectif de compréhension de la nature du raisonnement indépendamment d'une visée immédiatement applicative. Pour autant, dans un souci de mise à l'épreuve, elle anticipe l'ingénierie des connaissances par ses méthodes d'analyse et de modélisation fines d'un terrain.

Le résultat le plus intéressant de cette étude est sans doute d'avoir mis en lumière la sophistication des raisonnements en archéologie. Ce résultat est intéressant pour les deux communautés. Pour l'informaticien, cela contribue d'abord à réfuter les supposées universalité et neutralité des raisonnements formels. Pour l'archéologue, la prise de conscience de la place singulière dans son raisonnement du « groupe » et du contexte d'un artefact, entraîne des aménagements méthodologiques pour assurer le caractère itératif de leur constitution. De manière finalement assez poppérienne, il s'agit de gérer le caractère dialectique entre un savoir qui permet l'observation et lui donne un sens, et les observations qui permettent de dépasser éventuellement le savoir originel. Pris davantage en considération par les archéologues, ces résultats auraient sans doute pu éviter l'impasse épistémologique du « logicisme » en archéologie (Gardin, 1997) qui, bien que se réclamant de Popper, défendait de manière tout à fait contradictoire la vision du Positivisme Logique.

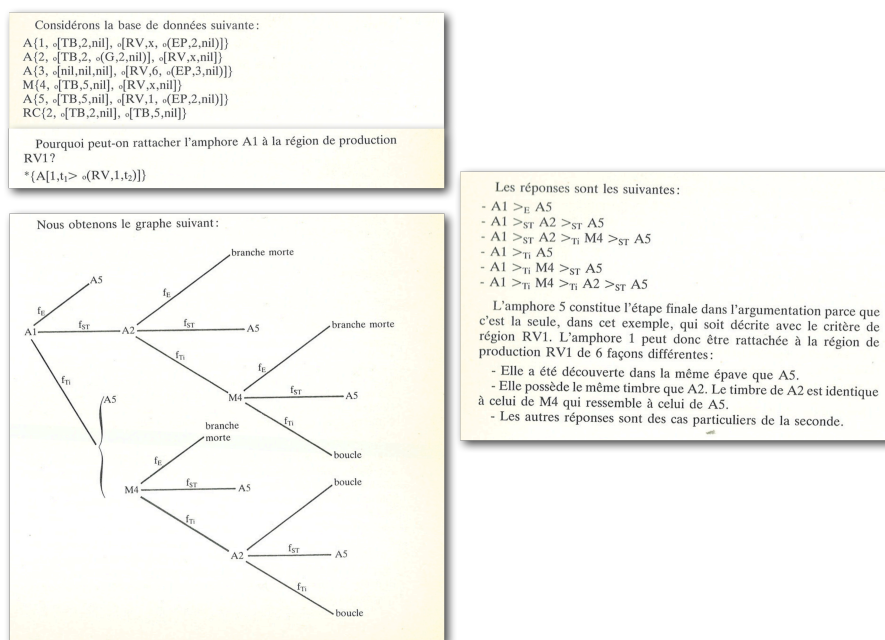


Fig. 2. Simulation du raisonnement permettant de rattacher une amphore à une région de production – Modèle et trace d'exécution d'AVEROES, 1971-1973 (Borillo, 1984)

À la même époque, la vocation du SYCIL, « Système automatisé de consultation d'un corpus d'inscriptions latines » est beaucoup plus instrumentale (Borillo, 1984, 78-95). Le but initial du projet est en effet simplement l'édition et la consultation d'un corpus des inscriptions latines par des moyens informatiques. Cependant, grâce à une collaboration entre informaticiens, épigraphistes et archéologues, il apparaît rapidement qu'il est nécessaire de distinguer l'état initial du texte et ses diverses restitutions (mutilations, abréviations), tout autant que les différents « points de vue » (*vis*) émanant des différentes disciplines pouvant s'intéresser à cet artefact à la fois littéraire (style, genre, contenu, typographie) et archéologique (support, décor, réutilisations, contexte).

Support commun à des hypothèses émanant de différents chercheurs de communautés qui se côtoient peu, le système peut se prêter à leur confrontation (Borillo, 1984, 158-169). L'équipe propose alors de vérifier la cohérence formelle des datations relatives des inscriptions avec les datations « absolues » connues pour un petit nombre d'entre elles (*cf.* Fig. 3).

Série "formelle"	58	51	49	32	53	50	48	59	36	38	34	33	35	31	19	9	23	54	17	13	10	8	7	27	3	18	5
PROCONSULAIRE								□	□																		
NUMIDIE	0													0	0	0	0		0	0	0	0	0	0	0	0	0
MAURETANIE C.	Δ	Δ	Δ		Δ				Δ	Δ	Δ	Δ	Δ					Δ									→
MAURETANIE T.					*																						
Chronologie	210							60																			198- 211

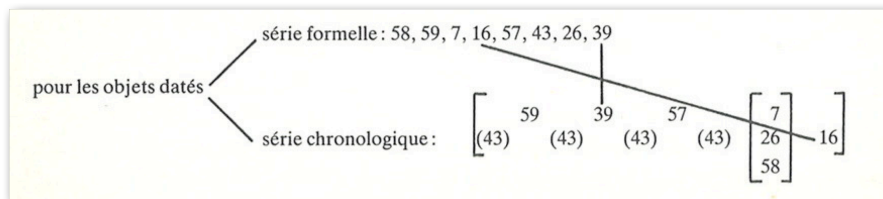


Fig. 3. Confrontation des chronologies relatives absolues d'un corpus d'inscriptions – Données et résultats tirés de SYCIL, 1972-1974 (Borillo, 1984)

L'outil parvient effectivement à détecter plusieurs incohérences, mais plus intéressant encore, ces incohérences semblent être dues à une interprétation chronologique de similarités dans un même espace géographique. En effet, les chronologies relatives de l'archéologie sont construites en fonction du style des artefacts, les ressemblances étant interprétées comme signes de synchronie, et les différences graduelles comme signes d'évolution. Pour le dire autrement, le temps archéologique est un temps des peuples et des cultures qui évidemment n'est pas indépendant des lieux.

Pour l'archéologie, au-delà du bénéfice strictement instrumental, les retombées sont doubles. Sur les contenus scientifiques tout d'abord, des hypothèses ont pu être mises à l'épreuve et certaines être réfutées. Sur la méthode ensuite, l'expérience a permis d'illustrer sur un exemple réel le caractère éminemment périlleux de datations relatives entre des artefacts issus d'aires géographiques différentes. Face à l'incapacité d'établir une chronologie



unique, il est nécessaire d'articuler différentes chronologies locales (Demoule, 1972).

Pour l'informatique, l'intuition selon laquelle dans un système pour la recherche, les différentes interprétations d'un objet devraient être exprimées, les différents points de vue devraient être pris en compte, et selon laquelle l'espace ainsi délimité pourrait devenir un espace de mise à l'épreuve et de confrontation, reste d'une actualité brûlante (Bénel et Lejeune, 2009).

### 3. Retour à Busa

Informé par les illustrations qui précèdent, nous sommes dorénavant mieux préparés à discerner dans l'*Index Thomisticus* de Roberto Busa ce qu'il a d'emblématique pour l'interdisciplinarité entre informatique et SHS.

Tout d'abord, ce projet s'inscrit dans la continuité d'une tradition et d'une expérience personnelle disciplinaires (Busa, 1980 ; Burton, 1981). En 1941, Roberto Busa commence un doctorat canonique de philosophie sur le concept de « présence » chez Thomas d'Aquin. Très rapidement, il se rend compte que chercher les occurrences du nom « *praesentia* » (ou de l'adjectif « *praesens* ») dans les index thématiques ne mène à rien. La fréquentation du corpus lui révèle que la préposition « *in* » est incomparablement plus pertinente. Il entreprend alors la rédaction à la main de 10 000 cartes bristol, une par phrase contenant la préposition (ou un mot apparenté). En 1946, il soutient sa thèse intitulée « La terminologie thomasiennne de l'intériorité : Traité de méthode pour l'interprétation de la métaphysique de la présence »<sup>6</sup>. Ainsi, ancré dans une problématique disciplinaire concrète, Roberto Busa en vient à réévaluer les outils et les méthodes de sa discipline. Bien que dans l'héritage direct des concordances des Massorètes du X<sup>e</sup> s. (Lejeune, 2007), il remet en cause le choix de n'indexer que les termes considérés *a priori* comme signifiants. Cet approfondissement de la méthodologie s'accompagne d'un approfondissement épistémologique. Roberto Busa s'inscrit dans cette tradition de compréhension des mots, non dans leur sens courant, mais dans celui émergent de leurs usages dans le corpus. Cependant, en s'intéressant aux interstices du discours, Roberto Busa vise à l'étude non plus seulement de ce que l'auteur souhaitait consciemment exprimer, mais à ce qui transparait dans son expression (Busa, 1980). Dès la fin de sa thèse, fort de la richesse des résultats de sa première expérience, il imagine de généraliser le principe de ses fiches en réalisant une concordance exhaustive des œuvres de Thomas d'Aquin. Cette

---

<sup>6</sup> « La Terminologia Tomistica dell'Interiorità: Saggi di metodo per una interpretazione della metafisica della presenza ».

exhaustivité, qui se traduit par la prise en compte de mots à forte fréquence comme les conjonctions, prépositions et pronoms, rend la concordance infaisable selon les méthodes traditionnelles. Roberto Busa, qui a entendu parler de la mécanographie<sup>7</sup>, a l'intuition qu'une partie du travail pourrait être automatisée.

Si ce projet est remarquable c'est aussi dans le dépassement des technologies de l'époque grâce à la collaboration avec des chercheurs en traitement de l'information (Busa, 1980). En 1949, profitant d'un déplacement aux États Unis, Roberto Busa visite vingt-cinq universités à la recherche de « gadgets » permettant d'automatiser la constitution de la concordance. De recommandations en recommandations (Bibliothèque du Congrès, MIT), il finit par rencontrer le président d'IBM. Ce dernier lui déclare d'abord que ce qu'il souhaite faire est impossible avec ses machines, mais, mis en défaut par un slogan au mur qui pérorait « Le difficile nous le faisons tout de suite, l'impossible prend un peu plus de temps », le président d'IBM finit par accorder la coopération totale de sa firme pour le temps que durera le projet, le tout à titre gracieux. Cette coopération est à la fois technique et scientifique. Les bureaux d'IBM à Milan fourniront l'accès aux machines ainsi que l'accompagnement technique quotidien. Tandis que l'expertise algorithmique sera fournie par Paul Tasman, un ingénieur de la R&D. Des expertises croisées de Tasman et de Busa, va naître un processus particulièrement astucieux (*cf.* Fig. 4). À partir du texte préparé par un spécialiste, une carte est produite pour chaque phrase en la tapant (ainsi que sa référence) sur le clavier d'une perforatrice (avec double saisie pour vérification). Pour chacune de ses cartes est engendrée de manière automatique une carte par mot de la phrase (sont perforés le mot et sa position). Pour une exploitation ultérieure plus facile, le contexte du mot (de plusieurs phrases) est imprimé (à partir des cartes des phrases) au dos de la carte. Ces cartes de mot sont ensuite « dédoublées » automatiquement afin de produire une carte par « forme » (qui précise également la fréquence de la forme dans l'œuvre). Un spécialiste, auquel ces formes sont fournies par ordre alphabétique (ex : « *erat* », « *erit* », « *es* », « *est* ») les regroupe par lemme (dans notre exemple : « *sum* ») et crée une carte pour chacun de ces groupes. Les cartes des phrases sont interclassées automatiquement avec les cartes de formes ainsi groupées par lemme. L'identifiant automatique des cartes de lemmes (perforé préalablement de manière automatique) est reporté automatiquement sur les cartes des formes et

---

<sup>7</sup> Mécanographie : Automatisation du traitement de l'information par des moyens électromécaniques.

des mots (Tasman, 1957 ; Busa, 1980). Le raffinement extrême de l'algorithme<sup>8</sup> nous permet de mesurer l'apport de Paul Tasman pour réaliser ce qui, même aux yeux du directeur d'IBM, était « impossible ».

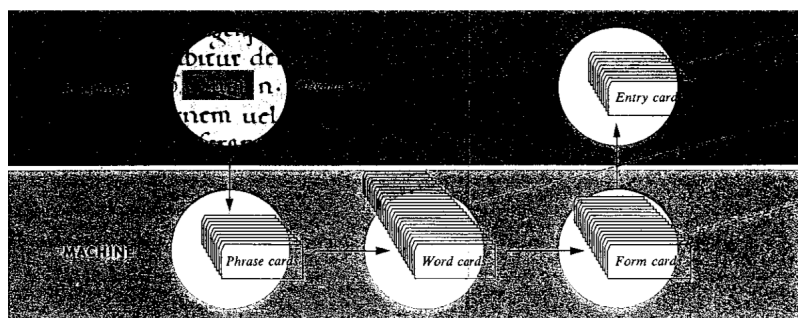


Fig. 4. Coopération de l'homme et de la machine pour fabriquer une concordance – Concordance de Thomas d'Aquin par Roberto Busa (Tasman, 1957)

Plus que tous les projets de concordances automatisées de la même époque (Burton, 1981), la collaboration de Busa et Tasman va avoir un impact décisif sur l'informatique. Publiée en 1957 dans la revue de la R&D d'IBM (Tasman, 1957), ces travaux vont inspirer Hans Peter Luhn, un autre ingénieur de la R&D d'IBM, pour d'autres usages. Luhn va d'abord détourner les familles par lemme pour en faire des familles par notion (Luhn, 1957), et les utiliser comme le cœur d'un dispositif de veille à destination des entreprises (Luhn, 1958). Pour faciliter l'indexation manuelle d'articles de chimie, Luhn va ensuite réaliser une concordance du corpus d'articles, en fixant non plus le nombre de phrases du contexte mais le nombre de caractères. Ces « *keywords in context* » (KWIC), tout en simplifiant les traitements, fournissent au lecteur de manière très visuelle la forme recherchée, les mots qui le précèdent et ceux qui le suivent (Luhn, 1960). En retour, ces deux innovations (familles notionnelles et KWIC) vont être très largement utilisées dans les outils destinés aux SHS.

Les premiers volumes de l'*Index Thomisticus* seront publiés en 1974 (Burton, 1981), soit près de 30 ans après les intuitions premières, et c'est là sans doute la dernière leçon que nous fournit cette expérience sur l'interdisciplinarité entre informatique et SHS. Pour reprendre quasiment les mots de Roberto Busa, l'ordinateur ne devrait jamais être envisagé comme une manière d'aller plus vite dans un travail de recherche ou de réduire l'effort des chercheurs. Cela

<sup>8</sup> Étonnamment proche d'ailleurs de ce que l'on ferait aujourd'hui avec des technologies récentes comme *MapReduce*.

ne serait ni réaliste ni souhaitable. Si un travail peut être fait à la main en un an, écrit-il, plutôt que d'y passer un mois avec un ordinateur pour avoir le même résultat, mieux vaut passer deux ans sur un projet plus vaste mais surtout plus profond. Pendant ces deux ans, le travail ne sera pas moins intense, bien au contraire, mais pourra se concentrer sur les parties les plus intellectuelles et les plus créatives (Busa, 1981).

#### 4. Une autre manière de penser les humanités numériques

##### *Repenser l'interdisciplinarité entre STIC et SHS*

Derrière la figure tutélaire des humanités numériques, se révèle donc une collaboration de 30 ans entre deux chercheurs, une interdisciplinarité exigeante, respectueuse des finalités et des critères disciplinaires de chacun. Si pour nous l'objectif est désormais clair, les méthodes pour l'atteindre restent cependant à (ré)inventer. Elles prennent forme peu à peu au gré de nos réussites et de nos échecs.

Un premier facteur déterminant semble être l'*immersion*. Dans notre cas, le contact quotidien avec des chercheurs en archéologie à Athènes ou en sciences sociales à Troyes nous a fourni de nombreux exemples mobilisables dans nos discussions et expérimentations. Mais l'immersion va au-delà. Roberto Busa raconte que Paul Tasman et son épouse avaient saisi les enjeux de son projet. La référence à Dorothy Tasman laisse deviner une certaine familiarité entre les deux chercheurs, des moments partagés en dehors du cadre strict du travail. Les réunions de travail sont en effet peu propices à l'explicitation des enjeux réels. La méthode scientifique a appris aux chercheurs à garder pour eux leur « métaphysique » (au sens de Duhem, ce qui sous-tend leurs théories mais ne peut être soumis à l'épreuve). Selon notre propre expérience, les valeurs de l'autre discipline (par exemple : « prendre au sérieux les acteurs » en sociologie) ne se révèlent que hors du cadre, autour d'un café ou d'un repas. Il s'agit en quelque sorte de donner à l'interdisciplinarité une dimension « ethnologique ». Mais cette immersion nécessite du temps, le temps de voir l'autre dans une situation de controverse, le temps de voir évoluer la perception qu'il a de sa propre discipline.

Le second facteur déterminant est lié à la définition de l'interdisciplinarité comme transfert de méthodes. Il s'agit de la valeur proprement *instrumentale* de ce transfert. On peut affirmer par exemple que le modèle des aires d'influence des villes, dans les mains d'archéologues à la recherche de villes disparues en Thessalie (Auda et al., 1990), est bien un « instrument » au sens de Pierre Rabardel (1995). En effet, il s'agit d'une entité mixte :

- artefact symbolique produit par ses concepteurs (des géographes voulant expliquer la répartition et la taille des villes dans le territoire),
- schème d'utilisation construit dans une finalité propre à l'utilisateur (prévoir la présence des vestiges d'une ville).

Dans le cas où l'artefact est explicitement conçu par une discipline pour une autre, cette dualité se complexifie mais demeure. Ainsi, nous avons été agréablement surpris par les propriétés prêtées à l'un de nos logiciels lors de son utilisation dans un master en sciences de l'antiquité (Bénel et al., 2010a) :

*L'outil est intéressant à plusieurs titres : (...) introduire à l'étude de la céramique, apprendre à tout décortiquer puis recomposer. Quand [les étudiants] font leur mémoire dessus, ils sont cadrés. Ça les oblige à faire un travail rigoureux.*

Notre logiciel d'analyse qualitative de documents n'avait pas été spécifiquement conçu pour les historiens d'art. Plus encore, l'idée de « tout décortiquer puis recomposer » ne faisait pas partie des contraintes imposées par le logiciel, mais juste du champ des actions possibles. Ce que décrivait l'historien d'art dans cet extrait n'était pas l'*artefact*, conçu par nos soins, mais l'*instrument*, muni des schèmes d'usage qu'il avait lui-même construits et transmis à ses étudiants. Le moment de la formation aux instruments est donc un moment particulièrement favorable à l'interdisciplinarité. Mais cela ne peut avoir lieu qu'à condition de ne pas cantonner la formation à la recherche vaine d'une virtuosité technique. Pour éviter cela, selon notre expérience, la présentation des contraintes et du champ des possibles introduits par l'artefact doit être accompagnée de celle de leurs raisons d'être, de leurs enjeux épistémologiques, et même du programme de recherche poursuivi par le concepteur. L'atelier proprement dit peut ensuite démarrer sur un cas choisi conjointement par les chercheurs des deux disciplines. Au fur et à mesure de l'exercice, le chercheur de la discipline « source » laisse la place à celui de la discipline « cible », tout comme l'artefact nu laisse la place à l'instrument. Tout générique que soit le logiciel, ses formations (et ses notices) doivent s'adapter à chaque discipline. C'est à cette condition que le logiciel devient le « cadre » loué plus haut par l'enseignant.

On s'interrogera peut-être sur la faisabilité de convier un chercheur de la discipline « source », voire même le concepteur de l'artefact, aux séances de formation. Cependant ces formations peuvent être envisagées pour la discipline source comme des lieux d'expérimentation. De manière plus générale, l'instrumentation interdisciplinaire est une occasion de mise à l'épreuve des hypothèses incorporées à l'artefact. Pour prendre un exemple célèbre, la procédure de datation au carbone 14, et plus particulièrement l'abaque faisant correspondre une datation à une mesure, incorporait l'hypothèse selon laquelle les rayonnements cosmiques à l'origine du radiocarbone dans l'air avaient été

constants à travers les âges. Les incohérences avec les autres sources de datation (textes, cernes des arbres, etc.), décelées par les archéologues, ont permis aux physiciens de réviser leurs hypothèses et en retour de produire de nouvelles abaques.

Pour être tout à fait franc, l'instrumentation interdisciplinaire, malgré (ou peut-être à cause de) son statut d'expérimentation pose cependant un problème de ressources. Seul un petit nombre de « terrains » pourront ainsi être suivis par un chercheur de la discipline « source », d'autant plus que la genèse instrumentale dont nous parlions plus haut se poursuit très au-delà de la formation liminaire. La fidélité nécessaire à ces « happy few » risque de poser des problèmes de compréhension aux « éconduits » de la discipline « cible » tout comme aux évaluateurs de la discipline « source ».

Par ailleurs, le statut expérimental de l'instrumentation pose d'autres difficultés. Le problème le plus trivial est que l'on ne peut attendre d'un prototype innovant le même degré de finition que les produits « sur étagère » auxquels sont habitués les usagers. Dans le domaine des « nouvelles technologies », le logiciel le plus connu des non-informaticiens (un traitement de texte en situation de quasi-monopole) a tout de même trente ans et ses fonctionnalités n'ont que très peu évolué depuis. Un problème plus subtil est dû au fait que, contrairement à un « produit », un prototype expérimental n'est pas conçu pour « satisfaire le client », mais pour tester des hypothèses, certaines pouvant parfois être contre-intuitives. Par exemple, dans nos logiciels, pour explorer la nature des « descripteurs » de documents (Bénel, 2006), nous avons à une époque délibérément empêché de les renommer (obligeant l'utilisateur à en créer d'autres). L'utilisateur n'étant évidemment pas un simple « cobaye », nous avons dû, suite à la réfutation de notre hypothèse, corriger le logiciel afin de continuer notre collaboration. Autre exemple complémentaire : nous avons longtemps résisté à la demande des usagers de nos logiciels de nous « aligner sur la concurrence » et d'ajouter des fonctionnalités *quantitatives* à nos logiciels d'analyse *qualitative* (Lejeune et Bénel, 2012). Il s'agissait pour nous de rester fidèles à notre programme de recherche et à notre quête d'une certaine « pureté épistémologique ». Nous n'avons inséré des comptages que lorsque nous avons découvert des principes d'interaction autour d'eux qui soient réellement qualitatifs. Un artefact est toujours plus ou moins porteur du programme de son concepteur, mais, dans le cas d'un prototype expérimental, cette dimension est exacerbée. Il est donc déterminant que ce programme soit explicité par le concepteur et accepté par l'utilisateur.

Nourri par l'immersion et l'instrumentation, le troisième facteur d'une interdisciplinarité réussie semble être la *discussion épistémologique*. Dans notre cas,

elle prend la forme de séminaires interdisciplinaires<sup>9,10</sup> articulant de manière assez hétérodoxe les choix de conception des logiciels, les cas disciplinaires et les enjeux d'épistémologie générale. La différence entre approches qualitative et inductive, la place de la mise à l'épreuve ou de la singularité en SHS sont autant de sujets épistémologiques que suscitent (ou ressuscitent) l'instrumentation logicielle des SHS. Par les ouvrages de référence qui y sont mobilisés, ces séminaires permettent d'inscrire les pratiques singulières des participants dans l'épaisseur historique de leur discipline. Paradoxalement, dépasser la simple « autorité de l'expert » est aussi une forme de respect : ne pas considérer que nous sommes les premiers à penser sa pratique, mais au contraire le reconnaître comme issu d'une tradition riche de ses controverses et de ses mouvements réflexifs.

Lorsqu'une synthèse semble possible, l'exercice de la co-écriture d'articles permet de pousser plus avant l'explicitation et la compréhension mutuelles. Pour autant, la discipline étant le lieu de l'évaluation par les pairs, il est important que ces articles interdisciplinaires ne se fassent pas aux dépens des publications de chacun dans sa propre discipline.

### ***Repenser la place des STIC dans l'interdisciplinarité entre SHS***

Un autre angle sous lequel aborder la question de l'interdisciplinarité dans les humanités numériques serait de voir les STIC comme « pivot de l'interdisciplinarité », notamment comme « mouvement d'explicitation et de confrontation des modèles » (Garbay, 2010). Les résultats allant dans ce sens sont encore timides. Néanmoins, Catherine Garbay nous ayant fait l'honneur de citer l'un de nos travaux (Accary et al, 2004) comme exemple d'« artefacts infléchissant la recherche vers de nouvelles formes de recherche interdisciplinaire » (Garbay, 2010, p. 167), nous nous enhardirons à présenter non plus nos méthodes de travail mais leur objet.

À l'occasion d'un projet de navigation « sémantique » dans un corpus de documents archéologiques, il nous était apparu que la seule manière d'obtenir un index adapté aux chercheurs consistait à le leur faire construire au fil de leurs lectures (Bénel et al., 1999). En somme, il s'agissait, bien avant le « Web 2.0 » et

---

<sup>9</sup> Séminaire ARTCADHi (*Atelier de Recherches Transdisciplinaires sur la Construction du sens en Archéologie et autres Disciplines Historiques*) rassemblant à Lyon (2003-2005) des chercheurs en informatique, histoire, historiographie, archéologie, histoire de l'art.

<sup>10</sup> Séminaire HYPERTOPIC (*Infrastructure pour la co-construction du sens*), rassemblant à Troyes (depuis 2007) ou Liège des chercheurs en informatique, sociologie, philosophie, psychologie et sciences de l'information, venant de Troyes, Dijon, Paris et Liège.

ses « folksonomies » (O'Reilly, 2005), de transposer le fonctionnement des communautés de l'*open source* aux communautés scientifiques, et de l'écriture de logiciels à l'interprétation de documents. Loin d'être le modèle postmoderne de dissolution de l'auteur que certains lui prêtent, le modèle de l'*open source* a donné naissance au contraire à des dispositifs techniques et légaux pour que chaque contribution soit datée, authentifiée, publiée et préservée<sup>11</sup>. Clairement inspiré du modèle universitaire, le modèle de l'*open source* joue sur la dualité de la *publication*, assurant à l'auteur la preuve d'origine de ses idées, et aux lecteurs leur diffusion. Dans cet esprit, l'indexation que nous proposons n'était pas une « donnée » anonyme, mais un élément de discours porté par un chercheur. Il s'agissait d'articuler autour d'un même objet plusieurs points de vue, éventuellement contradictoires, correspondant à différentes expertises ou écoles de pensée (Bénel et al., 1999). Pour reprendre les termes de Catherine Garbay, le chercheur qui utilise nos artefacts s'engage donc bien dans un mouvement d'*explicitation* et de *confrontation* des modèles.

Pour éviter tout malentendu, cette explicitation ne se traduit pas par le passage des SHS sous les fourches caudines de la logique et de l'informatique. Les éléments de modélisation sont ici suffisamment simples et familiers (analogues à la relation entre un terme générique et un terme spécifique dans un thésaurus) pour que le chercheur en SHS puisse les utiliser sans craindre de faire des « fautes » logiques. L'enjeu n'est plus, comme à la grande époque de l'intelligence artificielle, de simuler le raisonnement humain, mais de permettre à l'écriture de se déployer dans de nouvelles dimensions comme autrefois avec l'invention des alinéas, des listes ou des tableaux. Comme il y a une « raison graphique », il s'agit de développer chez le lecteur une « raison computationnelle » (Bachimont, 1999).

De l'instrumentation de la confrontation des points de vue à celle du dialogue interdisciplinaire, il semble n'y avoir qu'un pas :

*Dans le cas d'une pratique interdisciplinaire, où le même objet est étudié par des experts de diverses disciplines, les agoras virtuelles produites dans les divers univers, qualifiées par les méthodes propres à chaque discipline, peuvent être mis à contribution dans leurs recoupements en générant un espace de dialogue intercommunautaire (Iacovella, 2002).*

Cependant, comme le nuance ici Andrea Iacovella, ce pas est moins évident qu'il n'y paraît. Il nécessite des confrontations préalables entre points de vue

---

<sup>11</sup> À comparer avec les développements informatiques réalisés pour le compte d'une entreprise, rare exception du code de la propriété intellectuelle où l'auteur doit renoncer à ses droits patrimoniaux.



des mêmes disciplines (il n'y a pas de dialogue interdisciplinaire sans dialogues disciplinaires), ainsi qu'un objet commun entre disciplines. Or, comme l'ont montré des collègues (Atifi et al., 2006) dans le cadre d'un projet du laboratoire, même autour d'un terrain et d'un sujet communs, des chercheurs en sociologie, en linguistique et en psychologie ne travaillent pas sur les mêmes objets : retranscriptions d'entretiens pour les uns, d'interactions langagières naturelles pour les autres, captation de situations pour d'autres. Même dans le cas où l'organisation du projet les amènerait à analyser les documents produits par les autres disciplines, leur définition de la représentativité, ainsi que les niveaux d'analyse sont extrêmement hétérogènes, et par conséquent le niveau nécessaire de détail des transcriptions l'est également (Atifi et al., 2006).

Le problème d'absence d'objets partagés n'est pas sans rappeler les difficultés rencontrées par d'autres projets de « données scientifiques ouvertes ». En génétique, par exemple, lors d'une réunion visant à constituer une base de données mondiale des publications relatives à chaque gène, la directrice d'un laboratoire spécialisé dans le système immunitaire avait fait la surprise amère de l'inadéquation entre les objets partagés dans la base de données (gènes) et l'objet des travaux de son équipe (combinaisons complexes de fragments de gènes). Sans accroche aux objets partagés, les points de vue de son équipe étaient de fait exclus de l'espace numérique de débat.

### ***Repenser les STIC à l'aune des SHS... pour repenser les SHS***

À la différence des sciences de la nature, un peu dépourvues face à ce genre de difficultés relatives à la manière de gérer un débat. Les sciences de l'homme et de la société possèdent une tradition particulièrement riche sur le sujet. Lorsqu'à la fin du XIX<sup>e</sup> siècle, Dilthey pose les bases épistémologiques et méthodologiques de ce qu'il appelle les « sciences de l'esprit », il définit une « herméneutique philosophique » par généralisation des méthodes multiséculaires d'interprétation de la bible. Comme nous l'avons détaillé dans d'autres travaux (Bénel et Lejeune, 2009), l'herméneutique se distingue des mirages du positivisme par la prise en compte de *documents* plutôt que de « données », d'*interprétations* plutôt que d'inférences et de *débats intersubjectifs* plutôt que de « preuves ».

Ainsi, dans la tradition herméneutique, l'existence d'un corpus partagé est un préalable à la bonne tenue des débats. Le problème se pose d'ailleurs dès l'époque d'Origène, au III<sup>e</sup> siècle, à Alexandrie : juifs et chrétiens s'opposent sur l'interprétation de la bible, mais plus encore sur ce qu'*est* la bible, rendant tout débat stérile. Le premier désaccord concerne la liste des livres la constituant. Le corpus servant de référence aux interprétations des chrétiens inclue en effet

d'autres livres que ceux transmis par les juifs (outre, bien sûr, les écrits néotestamentaires). Pour commencer, Origène, dans ses lettres et ses commentaires, met de l'ordre dans ce canon chrétien en cours d'élaboration, en incluant certains textes (comme le fragment de « Suzanne au bain ») et en rejetant d'autres (comme le Livre d'Hénoch), utilisant comme critère explicite la lecture ou non de ces textes par toutes les communautés chrétiennes (Daniélou, 1948, p. 141). Le corpus interne ayant été délimité, Origène établit ensuite le corpus externe, utilisable dans le débat :

*Nous nous sommes efforcés de ne pas ignorer [leurs] éditions, afin que, discutant avec eux, nous n'apportions pas des textes qui ne se trouvent pas dans leurs exemplaires et qu'inversement nous utilisions les textes qui se rencontrent chez eux, même s'ils ne sont pas dans nos livres.* Origène (Daniélou, 1948, p. 141).

Le second désaccord concerne la version des livres référencée dans le débat. La version qui fait autorité pour les chrétiens est alors la septante (LXX), traduction grecque réalisée par des savants juifs pour le compte de Ptolémée II, cinq siècles plus tôt. Mais depuis le II<sup>e</sup> siècle, une nouvelle édition hébraïque ayant été établie, de nouvelles traductions grecques, plus fidèles, ont la préférence des juifs d'Alexandrie. Origène entreprend alors la compilation de toutes les traductions grecques, certaines d'usage courant, d'autres moins connues, une autre encore, qu'il avait découverte, dit-il, quelques années auparavant à Jéricho dans un tonneau (*sic*).

*[Origène] rassembla toutes ces versions dans un même livre qu'il divisa en côla et il les mit en face l'une de l'autre, avec aussi le texte hébreu ; ainsi il nous a laissé l'exemplaire appelé Hexaples.* Eusèbe (Daniélou, 1948, p. 140).

Au-delà de sa vocation de support au débat, cette édition de la bible en 6 colonnes se révèle être un outil précieux pour l'étude des textes. Origène s'en sert d'abord pour déterminer dans la septante les passages manquants ou douteux et les marquer avec les signes critiques que les grammairiens de l'époque utilisent pour les textes d'Homère. Par ailleurs, Origène mobilise les Hexaples dans ses commentaires pour expliciter les différents sens qui ressortent des traductions (Daniélou, 1948, p. 141).

Au fil des siècles, les études bibliques vont se doter de nouveaux outils : les concordances ainsi que la numérotation des chapitres et des versets. Là encore, le débat entre juifs et chrétiens sera le moteur de leur adoption (Lejeune, 2007). Au XV<sup>e</sup> siècle, Mordecai Nathan va par exemple réaliser une concordance, pourtant inutile pour les exégètes juifs qui connaissaient la bible par cœur, dans le seul but de faciliter le débat théologique avec les chrétiens. À cette fin, il adoptera même les numéros de chapitre utilisés par les chrétiens

(Weinberg, 2004). Toutes ces innovations ont contribué à fournir au débat des objets partagés plus fins.

L'histoire de notre plateforme logicielle est comparable (bien qu'heureusement plus rapide). La modélisation de la structure des objets partagés entre points de vue (et de ses conséquences sur les algorithmes de cooccurrence) nous aura finalement donné beaucoup plus de difficultés que celle des points de vue eux-mêmes. Alors que notre plateforme, depuis le début (2000), permet aux points de vue de ne pas porter strictement sur le même corpus, il faudra attendre 2004 (cf. Fig. 5), pour qu'elle gère explicitement plusieurs corpus, pour que les documents du « portfolio » soient l'union des documents décrits par les points de vue actifs (Bénel, 2006).

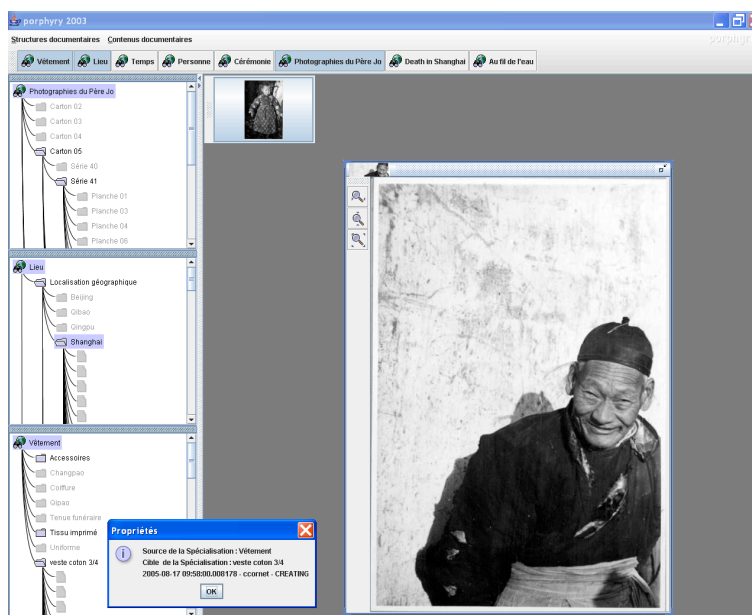


Fig. 5. Confrontation d'analyses portant sur deux corpus. Copie d'écran de Porphyry (en 2005).

Dans notre plateforme, le mode de navigation est basé sur des calculs complexes de cooccurrence. Par conséquent, il nous faudra encore plusieurs années (2007) pour trouver les algorithmes et les principes d'interaction capables de gérer des analyses de niveau hétérogène – comme l'analyse (cf. Fig. 6) d'un vase grec selon sa forme et selon l'iconographie de ses faces (Zhou et Bénel, 2008) – ou s'appuyant sur des fragments dont les contours peuvent se

superposer – comme dans le cas (cf. Fig. 7) de l'apparition dans le même contexte de documents produits par des ONG de différentes familles notionnelles de la société civile (Bénel et Lejeune, 2009 ; Bénel et al., 2010).



Fig. 6. Confrontation d'analyses portant sur des niveaux documentaires différents. Copie d'écran de Porphyry (en 2005).



Fig. 7. Confrontation d'analyses portant sur des fragments de textes – Copie d'écran de Porphyry+Cassandra (en 2008)

La prochaine étape consisterait à partager entre points de vue des unités identifiables dans les différentes *versions*<sup>12</sup> d'un document, comme l'enregistrement vidéo d'une conversation et sa transcription, un manuscrit et sa transcription diplomatique, une œuvre littéraire et ses différentes traductions. Dans ce but nous participons à la conception d'une plateforme pour les traducteurs de textes culturels, sorte d'hexaples et de concordance multilingue participatives (Bénel et Lacour, 2011).

## 5. Conclusion

Comme nous l'avons vu dans cet article, les *humanités numériques* se définissent comme la rencontre des « deux cultures ». Pourtant l'histoire qu'elles donnent d'elles-mêmes semble être une histoire sans informaticiens. À l'inverse, celle de *l'informatique pour les sciences de l'homme* regorge de collaborations interdisciplinaires fructueuses entre informatique et SHS, de récits de détournements réciproques, de prises de conscience épistémologiques mutuelles, de mises à l'épreuve et de confrontations d'hypothèses. Derrière l'aura un peu aveuglante de précurseur des humanités numériques, Roberto Busa, lui-même, s'est révélé, dans sa collaboration avec Paul Tasman (ingénieur/chercheur en informatique) comme l'artisan d'une interdisciplinarité exigeante, fidèle et respectueuse de chacune des disciplines. Pour refonder les humanités numériques sur une interdisciplinarité digne de son précurseur, nous avons proposé trois pôles complémentaires :

– *l'immersion* du chercheur en STIC auprès des chercheurs en SHS, afin notamment de percevoir leurs « valeurs » et leurs enjeux réels,

– *l'instrumentation*, dans ses rapports complexes de détournement par les SHS de l'artefact conçu par les STIC, et de ce que cela révèle des hypothèses des STIC et des méthodes des SHS,

– *la discussion épistémologique*, permettant de percevoir au-delà des pratiques des chercheurs en présence, la dimension disciplinaire et interdisciplinaire des questions soulevées par les deux premiers pôles.

Pour finir, nous avons exploré la piste des STIC comme pivot du débat interdisciplinaire entre SHS. Pour cela, nous avons montré les difficultés rencontrées et les pistes ouvertes par une informatique repensée à l'aune de l'herméneutique.

---

<sup>12</sup> Le terme « versions » est utilisé ici au sens de « version courte / version longue » ou « version originale / version française », c'est-à-dire de variantes d'une œuvre selon l'audience ciblée, à ne pas confondre avec « révisions » qui désignent les différents états d'une œuvre dans le temps.

## Bibliographie

- Abascal R., Beigbeder M., Bénel A., Calabretto S., Chabbat B., Champin P.-A., Chatti N., Jouve D., Prié Y., Rumpler B., Thivant E. (2004). Documents à structures multiples. *Conférence internationale des sciences de l'électronique, des technologies de l'information et des télécommunications (SETIT)*, 2004.
- Accary T., Bénel A., Calabretto S., Iacovella A. (2004). Confrontation de points de vue sur des corpus documentaires : Le cas de la modélisation du temps archéologique. *Congrès francophone AFRIF-AFLA de reconnaissance des formes et intelligence artificielle (RFLA)*, 2004. p. 197-205.
- Atifi H., Lejeune C., Ninova G., Zacklad M. (2006). Méthodologie transdisciplinaire de gestion de corpus pour les disciplines de l'interaction : Recherche de principes directeurs. *Corpus en lettres et sciences sociales : des documents numériques à l'interprétation, Supplément à Textol*, vol. 11, n° 2, p. 167-172.
- Auda Y., Darmezine L., Decourt J.-C., Helly B., Lucas G. (1990). Espace géographique et géographie historique en Thessalie. *Archéologie et espaces, Rencontres internationales d'archéologie et d'histoire*, 1989. Éditions APDCA.
- Bachimont B. (1999). L'intelligence artificielle comme écriture dynamique : De la raison graphique à la raison computationnelle. *Au nom du sens*. Grasset, Paris, p. 290-319.
- Bénel A., Calabretto S., Pinon J.-M. (1999). Indexation "sémantique" de documents archéologiques. *L'indexation à l'heure d'Internet, Colloque du chapitre français de l'ISKO*, 1999. Éditions ISKO-France, 2001, p. 145-152.
- Bénel A. (2006). Porphyry au pays des Paestans : usages d'un outil d'analyse qualitative de documents par des étudiantes de maîtrise en iconographie grecque. *Corpus en lettres et sciences sociales : des documents numériques à l'interprétation, Supplément à Textol*, vol. 11, n° 2, p. 173-179.
- Bénel A., Lejeune C. (2009). Partager des corpus et leurs analyses à l'heure du Web 2.0. *Degrés, revue de synthèse à orientation sémiologique*, vol. 136-137, article 'm'.
- Bénel A., Zhou C., Cahier J.-P. (2010). Beyond Web 2.0... and beyond the Semantic Web. *From CSCW to Web 2.0: European Developments in Collaborative Design*, Springer, p. 155-171.
- Bénel A., Lejeune C., Zhou C. (2010). Éloge de l'hétérogénéité des structures d'analyse de textes. *Document numérique, RSTI*, vol. 13, n° 2, p. 41-56.
- Bénel A., Lacour P. (2011). Towards a participative platform for cultural texts translators. *Virtual Community Building and the Information Society: Current and Future Directions*, IGI Global, p. 153-162.
- Borillo M. (1984) *Informatique pour les Sciences de l'Homme : Limites de la formalisation du raisonnement*, Pierre Mardaga Éditeur, Bruxelles.

- Burton D. M. (1981). Automated concordances and word indexes: The fifties. *Computers and the humanities*, vol. 15, p. 1-14.
- Busa R. (1980). The Annals of Humanities Computing: the Index thomisticus. *Computers and the Humanities*, vol. 14, p. 83-90.
- Dacos M. (Ed.) (2011). Manifeste des Digital Humanities. *Non-conférence sur les digital humanities* (ThatCamp), 2010.
- Daniélou J. (1948). *Origène*. Éditions du Cerf, Paris, 2012 (Première édition : La Table Ronde, 1948).
- Demoule J.-P. (1972). Projet de bibliographie automatique en préhistoire et protohistoire européenne. *Colloque "Banques de données archéologiques"*, Marseille.
- Garbay C. (2010). Les sciences du traitement de l'information comme pivot de l'interdisciplinarité : une vision systémique. *Repenser l'interdisciplinarité*, Éditions Slatkine, p. 145-170.
- Gardin J.-C. (1997). « Quand on voit c'qu'on voit, quand on sait c'qu'on sait ». *L'Homme*, vol. 37, n° 143, p. 83-90.
- Hockey S. (1996). Knowledge Representation. *Research Agenda for Cultural Heritage on Information Networks*. Getty art history information program, p. 31-34.
- Hockey S. (2004). The history of humanities computing. *A companion to digital humanities*, Blackwell, Oxford.
- Iacovella A. (2002). Modèle opératoire de navigation pour les experts : Appropriation sémantique et délimitation de l'espace documentaire. *La navigation. Les cahiers du Numérique*, vol. 3, n° 3.
- Iacovella A., Bénel A., Calabretto S., Helly B. (2005). Assistance à l'interprétation dans les bibliothèques numériques pour les sciences historiques. *Colloque de bilan du programme interdisciplinaire "Société de l'information"*, Lyon, p. 167-179.
- Iacovella A. (2007). Éléments pour une historiographie digitale. *Humanités numériques, Tome 1 : Nouvelles technologies cognitives et épistémologie*, Lavoisier, p. 33-51.
- Lejeune C. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. *Humanités numériques, Tome 1 : Nouvelles technologies cognitives et épistémologie*, Lavoisier, p. 197-214.
- Lejeune C., Bénel A. (2012). Lexicométrie pour l'analyse qualitative : Pourquoi et comment résoudre le paradoxe. *Journées internationales d'analyse statistique de données textuelles (JADT)*, 2012.
- Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, vol. 1, n° 4, p. 309-317 .
- Luhn, H. P. (1958). A business intelligence system. *IBM Journal of Research and Development*, vol. 2, n° 4, p. 314-319.

- Luhn, H. P. (1960). Keyword in context index for technical literature. *American Documentation*, vol. 11, n° 4, p. 288-295.
- Nelson, T. H. (1981). *Literary Machines*. Mindful Press.
- Nelson, T. H. (1999). Xanalogical Structure Needed Now More Than Ever. *ACM Computing Surveys*, vol. 31, n° 4.
- Nicolescu B. (1996). Une nouvelle vision du monde : La transdisciplinarité. *La Transdisciplinarité : Manifeste*, Éditions du Rocher, Monaco, p. 60-71.
- O'Reilly T. (2005). *What is web 2.0: Design patterns and business models for the next generation of software*, Blog post, Sept. 30, 2005.
- Rabardel P. (1995). Qu'est-ce qu'un instrument ? Appropriation, conceptualisation, mises en situation. *Les dossiers de l'ingénierie éducative*, mars 1995. CNDP.
- Schreibman S., Siemens R., Unsworth J. (2004). *A Companion to Digital Humanities*. Blackwell, Oxford.
- Tasman P. (1957). Literary data processing. *IBM Journal of Research and Development*, vol. 1, n° 3, p. 249-256.
- Weinberg, B. H. (2004). Predecessors of scientific indexing structures in the domain of religion, *Conference on the history and heritage of scientific and technical information systems*, 2004. p. 126-134.
- Wikipedia contributors (2013). Digital humanities. *Wikipedia, The Free Encyclopedia*. August 9, 2013.
- Zhou C., Bénel A. (2008). From the crowd to communities: New interfaces for social tagging. *International conference on the design of cooperative systems (COOP)*, 2008.