



HAL
open science

Modéliser ce qui résiste à la modélisation : De la sémantique à la sémiotique

Aurélien Bénel

► **To cite this version:**

Aurélien Bénel. Modéliser ce qui résiste à la modélisation : De la sémantique à la sémiotique. Revue Ouverte d'Intelligence Artificielle, 2020, Intelligence Artificielle et Humanités Numériques, 1 (1), pp.71-88. 10.5802/roia.4 . hal-02313535

HAL Id: hal-02313535

<https://hal.science/hal-02313535>

Submitted on 9 Sep 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License



AURÉLIEN BÉNEL

Modéliser ce qui résiste à la modélisation

Volume 1, n° 1 (2020), p. 71-88.

http://roia.centre-mersenne.org/item?id=ROIA_2020__1_1_71_0

© Association pour la diffusion de la recherche francophone en intelligence artificielle et les auteurs, 2020, certains droits réservés.



Cet article est diffusé sous la licence
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*La Revue Ouverte d'Intelligence Artificielle est membre du
Centre Mersenne pour l'édition scientifique ouverte*
www.centre-mersenne.org

Modéliser ce qui résiste à la modélisation

De la sémantique à la sémiotique

Aurélien Bénel^a

^a Tech-CICO, Université de technologie de Troyes, 12 rue Marie Curie, CS 42060, 10004 Troyes, France.

E-mail : aurelien.benel@utt.fr.

RÉSUMÉ. — Pour traiter de l'intelligence artificielle appliquée aux humanités numériques, cet article prend le pari risqué de concentrer son état de l'art sur les années 1970. Nous découvrons que ces premiers travaux de modélisation des objets archéologiques à l'aide de « langages d'analyse » et de « domaines » n'ont rien à envier aux projets actuels à base de RDF et d'OWL, qu'ils recèlent même souvent une finesse d'analyse digne des projets les plus aboutis. Mais ces travaux sont surtout intéressants par les débats qu'ils firent naître dans la communauté des archéologues, débats d'une profondeur théorique suffisante pour rester, nous semble-t-il, applicables plus de 45 ans après. Parmi les critiques de l'époque, la plus intéressante et la plus constructive est probablement celle de l'archéologue Philippe Bruneau : contrairement aux objets de la Nature, les objets des Sciences humaines, parce qu'ils sont déjà porteurs de sens, doivent être décrits avec des méthodes sémiotiques (plutôt que sémantiques). La prise en compte du contexte ne se traduit alors pas par un vague principe ou par un modèle de plus, mais par un refus systématique des modèles à portée générale. Les modèles doivent tenir compte du fait que seuls un petit nombre des traits pouvant caractériser un objet seront pertinents et ce par différence avec les objets qui sont proches de lui dans son univers technique. Consciemment ou inconsciemment, un certain nombre de travaux d'aujourd'hui en ingénierie des connaissances s'inscrivent au moins partiellement dans ces perspectives. Comme nous l'illustrons avec nos propres logiciels et expérimentations, la prise en compte de cette approche sémiotique trace des perspectives prometteuses pour l'instrumentation de la pratique quotidienne des chercheurs en Sciences humaines ainsi que pour la médiation scientifique. Mais, par un juste retour aux sources de l'intelligence artificielle (à l'époque où sa visée était moins substitutive que compréhensive) l'intérêt de la prise en compte de l'approche sémiotique est peut-être plus grand encore dans la foule de questions de conception qu'elle suscite, questions anecdotiques à première vue, mais liées finalement à ce qu'est le sens et à ses modes de construction.

MOTS-CLÉS. — Modélisation des connaissances, archéologie, artefact, trait pertinent.

1. INTRODUCTION

« À la différence des sciences de la nature, les sciences humaines formalisent un objet déjà formalisé » (J. Gagnepain, cité par Bruneau, 1976). Par exemple, « Les

archéologues ne sont pas les premiers à décrire ou à classer le matériel dont ils traitent » [7]. En effet, les concepteurs et les usagers de l'époque, avaient déjà leur propre théorie de leur univers technique.

Ces déclarations sur la nature des sciences humaines, et sur l'archéologie plus particulièrement, furent publiées dans les années 1970, à une époque où archéologues et historiens d'art voulaient moderniser leurs disciplines en usant de « modèles » issus d'autres disciplines et d'imposantes « banques de données » accueillant le « graphe des faits » formalisé dans un « langage documentaire universel ». Aussi stimulantes pour la pensée que furent ces initiatives de modélisation, elles furent identifiées par les critiques de l'époque comme menant à des impasses épistémologiques.

Pour commencer, nous reviendrons avec attention sur ces projets de modélisation des années 1970 en archéologie afin d'établir des parallèles avec les projets actuels visant à appliquer les techniques du Web sémantique aux objets culturels. Ensuite, en étudiant les critiques épistémologiques émises à l'époque, nous verrons que l'échec de ces modèles vient d'un manque de prise en compte de la valeur sémiotique de ces objets. À partir de ces arguments, nous proposerons des pistes pour les modèles et les outils informatiques d'inspiration non plus sémantique mais sémiotique. Enfin, nous verrons ces modèles et outils en action et comment des historiens d'art les utilisent pour « faire sens ».

2. ÉTAT DE L'ART

2.1. RETOUR VERS LE FUTUR

Dans le domaine de la modélisation des connaissances en archéologie, 1972 fut une année décisive en voyant la publication de deux travaux collectifs majeurs : *Models in Archaeology* [11] et *Les banques de données archéologiques* [5]. Dans les deux ouvrages, nombre d'auteurs proposent d'enregistrer sur ordinateur la description des vestiges et de leurs relations dans le temps et l'espace sous forme de modèles statistiques ou logiques. La théorie des ensembles, notamment, est adoptée [23] pour formaliser les *taxonomies* d'objets archéologiques (ex. : « *Tout canthare est un vase à vin* ») et les *méréonomies* spatiales (ex. : « *Paestum est une partie de la Grande Grèce* »).

Pour les chercheurs d'aujourd'hui du domaine du Web sémantique, les travaux les plus intéressants de l'époque sont probablement ceux qui utilisaient « SATIN 1 » [10], un système tout d'abord conçu pour l'Inventaire général des monuments et richesses artistiques de la France. Il se décomposait en un « langage d'analyse » visant à représenter les descriptions des objets culturels et un « langage de requête » permettant de retrouver et d'agrèger ces descriptions.

Comme RDF de nos jours, le langage d'analyse de SATIN 1 (cf. Fig. 2.1) était suffisamment expressif pour s'attaquer à des descriptions complexes. Par exemple, la figure 2.2 montre la description formalisée [15] d'un petit objet (25 × 15 mm) amygdaloïde⁽¹⁾ en cornaline, trouvé à Vaphio, daté de l'Helladique récent II et représentant une

⁽¹⁾ Amigdaloloïde : en forme d'amande.

silhouette humaine menant un char (avec des rênes, des roues à rayons et un timon) et deux chevaux (à la tête linéaire, debout, de profil, l'un souligné d'un contour, l'autre en arrière-plan, aux pieds griffus). Après cet exemple, comment ne pas ressentir une impression de déjà-vu face aux recommandations actuelles [24, p. 19] pour la description d'objets visuels ?

```

< VERBE > ::= < LETTRE > | < VERBE > < LETTRE >
< CHAINE-ELEMENTAIRE > ::= < DESCRIPTEUR > |
    < DESCRIPTEUR > < ENTIER > |
    < DESCRIPTEUR > < BLOC > |
    < DESCRIPTEUR > < ENTIER > < BLOC >
< CHAINE > ::= < CHAINE-ELEMENTAIRE > |
    < CHAINE > < CHAINE-ELEMENTAIRE >
< PHRASE > ::= < VERBE > < CHAINE >
< BLOC > ::= < PHRASE > | < BLOC > < PHRASE >
< UNITE > ::= < PHRASE > | < UNITE > < PHRASE >
< RACINE > ::= < UNITE >
    
```

FIGURE 2.1. Grammaire formelle du langage d'analyse de SATIN 1 [10]

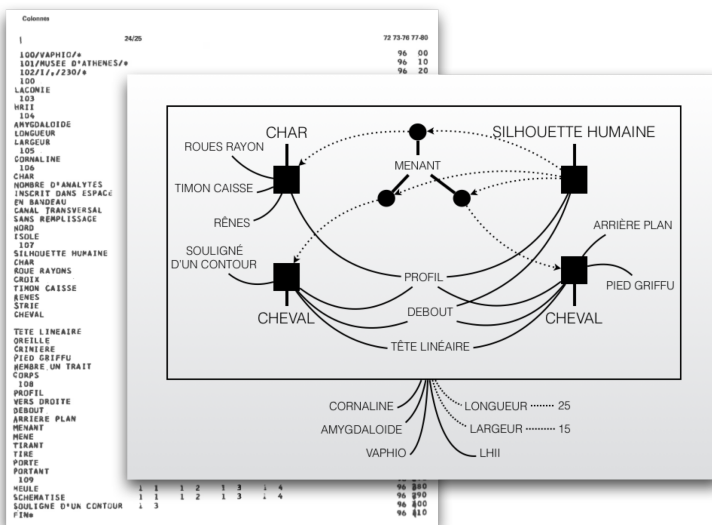


FIGURE 2.2. Modélisation avec SATIN 1 d'un sceau créto-mycénien et de son iconographie : codage [15] et représentation sous forme de graphe

Par ailleurs, de manière similaire à ce qui est fait aujourd'hui dans les ontologies du Web, chaque *descripteur* (« LONGUEUR », « CHEVAL », « MENANT ») devait être défini dans un *domaine* (matériau, lieu de découverte, description, etc.), et dans certains de ces domaines, le lexique pouvait être structuré de manière hiérarchique (« LACONIE > VAPHIO », « PIERRE > CORNALINE », « HELLADIQUE RÉCENT > HR II > HR II B »).

Enfin, de même qu'aujourd'hui les requêtes en SPARQL mettent à profit des inférences issues des ontologies, le langage de requête de SATIN 1 mettait à profit la structure du lexique définie dans les domaines. Pour être plus précis, la nature de ces inférences devait être choisie au niveau de la requête par des « opérateurs lexicaux » [6, 15] permettant par exemple de trouver l'objet précédemment cité en cherchant des objets fabriqués en cornaline, datés de l'helladique récent II B (avec généralisation possible) trouvés en Laconie (avec spécialisation possible).

D'une manière très actuelle, le concepteur de SATIN 1 faisait remarquer que comme les descripteurs de différents domaines pouvaient être combinés dans la même description, l'ajout de nouveaux descripteurs (concernant par exemple le décor) était réalisable de plusieurs manières : soit en l'ajoutant à tous les domaines concernés (ex : sculpture, mobilier, architecture, etc.) soit en créant un nouveau domaine utilisable pour tout type d'objets [10]. Au-delà du bénéfice formel de l'expressivité combinatoire, la réutilisation à l'échelle mondiale de domaines et de descriptions était présentée par un des promoteurs de ces projets comme une « nécessité » et même un « devoir », il s'agissait selon lui de passer d'une « possession égoïste » à un « bien commun » [14].

Une belle illustration de ce courant nous est fournie en 1975 par Anne-Marie Guimier-Sorbets dans sa thèse de doctorat sur l'analyse et la formalisation des ornements géométriques des mosaïques gréco-romaines en vue du traitement automatique. D'une manière très rigoureuse et très formelle, elle définit chaque attribut utilisé et décrit le processus que l'on doit suivre pour affecter la bonne valeur au bon fragment de chaque vestige. Philippe Bruneau, qui était dans le jury de cette thèse, écrivit par la suite un article [7] qui déclencha une polémique sans précédent dans l'univers distingué de l'École française (d'archéologie) d'Athènes. Dans le numéro suivant de la revue, le directeur de publication sentit le besoin d'écrire un prologue [1] pour donner « la définition de ce qu'est et de ce que n'est pas le *Bulletin de correspondance hellénique* [la revue] », précisant même que « [Le Bulletin] ne paraît pas être un endroit approprié pour des exposés de doctrine ou des traités de méthode ». À la parution de son propre manuel, l'auteur de la thèse sentit encore la nécessité de répondre aux arguments de Philippe Bruneau dans son introduction, quatorze ans plus tard [17].

2.2. DES CRITIQUES VENANT DU PASSÉ

Les critiques de Philippe Bruneau sur la manière dont les objets culturels sont modélisés se concentrent autour de la notion de « descripteur » (c'est-à-dire un élément d'une ontologie : une classe, un individu, une propriété, etc.). Philippe Bruneau commence par le simple constat que les descripteurs sont souvent choisis pour leur caractère universel, indépendant de l'époque et de la géographie (ex : « couleur de fond »); des critères qu'il trouve pour le moins surprenants dans le cadre d'une approche historique. Il note ensuite qu'à propos des postes (cf. Fig. 2.3), ces décors (aussi appelés « méandres » ou « grecques ») dans lesquels une frise noire sur fond blanc peut aussi être vue comme la frise complémentaire blanche sur fond noir, Anne-Marie Guimier-Sorbets précise : « nous convenons d'analyser la ligne de postes situées le plus à l'extérieur de la mosaïque (...). L'autre partie, complémentaire, est analysée

comme fond ». Philippe Bruneau trouve regrettable de décider *par convention* de décrire une mosaïque du bord vers le centre, alors que les mosaïques étaient créées à l'époque à partir d'un panneau central puis étendues vers l'extérieur. Enfin, il conclut en faisant remarquer que le terme même de « descripteur » est problématique. Par sa terminaison, ce terme laisse penser qu'il s'agirait d'un agent qui décrit, autrement dit, le terme pourrait faire croire à tort que c'est le dispositif qui décrit l'objet et non l'archéologue. Oublier que l'archéologue est celui qui décrit pourrait faire oublier que les premiers à avoir décrit et classifié l'objet étaient les usagers de l'époque.

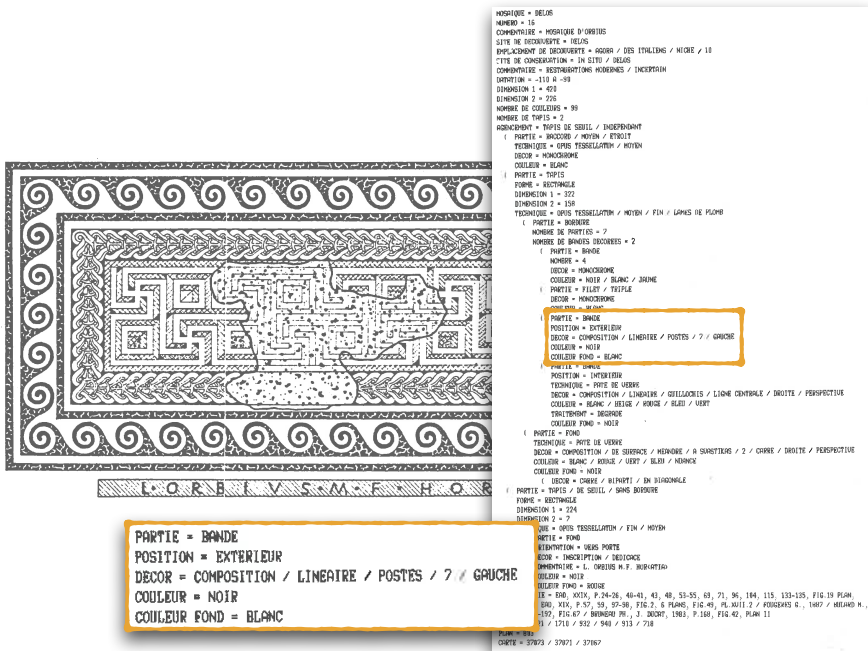


FIGURE 2.3. Modélisation d'une mosaïque à postes noires sur fond blanc [17]

Au-delà de cet exemple, au-delà même des problèmes de modélisation, Philippe Bruneau tente de mettre au jour ce qui fait de l'archéologie une science humaine. L'archéologie s'intéresse aux *artefacts* (*arte factum* c'est à dire « fabriqué par la technique », façonné par les humains et non par la nature). Pour cette raison, l'objet archéologique est sémiotique : tout comme les deux faces du signe selon Saussure, on ne peut séparer sa *configuration* matérielle du *programme* qui lui a été assigné par ses concepteurs (et ses usagers). Par conséquent, contrairement à ce qui est souvent énoncé, commencer par une description factuelle et produire une interprétation dans un second temps n'a aucun sens. Par ailleurs, la signification de l'objet archéologique – parce qu'il est sémiotique – est dépendante des objets qui l'entourent. Philippe Bruneau nous donne un exemple : dans la configuration classique d'un bureau, la fonction (ou *trait pertinent*) d'un crayon est que, contrairement à un stylo, ce que l'on écrit avec est effaçable. En

l'absence de stylo, par contre, son trait pertinent sera de pouvoir écrire avec. Enfin, en l'absence d'un cure-pipe, son trait pertinent sera sa forme allongée. Autrement dit, il est impossible de décrire un crayon, son usage et ses propriétés, sans connaître l'*état de choses* (les objets du même contexte technique). Pour cette raison, la description d'un artefact n'est jamais achevée : elle devra être révisée et révisée encore selon une *démarche en spirale*. Par ailleurs, quoiqu'aussi rationnellement structuré qu'un univers linguistique, un état de choses est toujours idiomatique : il ne peut en aucune manière être universel. Et il ne viendrait à l'idée de personne de décrire une langue étrangère sans faire l'effort d'entrer dans le système de ses locuteurs.

3. APPROCHE ET POSITIONNEMENT

Les considérations sémiotiques exposées plus haut mettent au jour de sérieuses limites aux projets d'hier et d'aujourd'hui de modélisation sémantique (formelle) des objets culturels. Pour autant, elles ouvrent également des perspectives sur la manière dont la modélisation des connaissances portant sur ces objets culturels pourrait se mettre au service de la construction (humaine) du sens.

Pour commencer, au lieu de chercher l'universalité du langage de description, une approche plus sémiotique serait de lier la description de l'objet très fortement à l'état de choses pour lequel il a été créé. Mais, parce qu'un artefact appartient à un nombre indéfini d'états de chose, cette approche devrait permettre le maintien de l'identité de l'artefact dans des états de choses superposés et donc dans les multiples analyses qui s'y rapportent. Par ailleurs, au lieu de se concentrer sur les inférences basées sur des définitions hors-contextes (*type*, *subClassOf*, *partOf*), une telle approche permettrait de « naviguer » à travers les différents états de choses et de voir comment un trait active ou désactive les autres traits, et ce, tant au niveau des artefacts que de leurs fragments. En termes plus informatiques, il s'agirait de proposer une visualisation interactive et multi-niveau des co-occurrences.

Ces pistes quoique très minoritaires dans le paysage mondial de la recherche en informatique peuvent se retrouver sous des noms, des formes et des degrés différents :

- « ontologies locales » [9] par opposition aux habituelles « ontologies réutilisables » [16], ou encore « Web pragmatique » [29] renonçant volontairement à considérer les ontologies comme immuables, hors-contexte et indépendantes des communautés, par opposition au Web sémantique [4];
- « Web sémantique “débrillé” ne nécessitant pas d'ontologies⁽²⁾ » [21];
- « modèle informatique de la sémantique interprétative » [30, 20], « analyse de données interprétative » [19] ou « sémantique légère » [25], modèles pour

⁽²⁾« Although I wouldn't actually express it as a skepticism, I would say that we're enthusiasts for a particular piece of the Semantic Web, which some people are skeptical about, which is a sort of sloppy or scruffy Semantic Web (...) – I guess what we both believe more in is, you know : a little structure goes a long way if you combine it with, for example, a human being that has a lot of intelligence between his or her ears (...) – No, the Semantic Web does not need ontologies. »

l'accès au document et l'exploration de son contenu, explicitement basés sur le concept sémiotique d'isotopie ;

- « collections numériques » [28], comme des alternatives aux « classes » des logiciens et des informaticiens ;
- « Web cognitivement sémantique » [8] (renommé plus tard « Web socio-sémantique ») préférant aux ontologies des cartes de thèmes, modèle initialement destiné à l'index thématique d'un livre, donc à la fois peu formalisé, dépendant d'un contexte précis et pouvant par ailleurs se superposer en autant de « scopes » ou « points de vue ».

C'est à la conception de ce genre d'approches et d'outils que nous avons consacré une partie de nos travaux de ces vingt dernières années, notamment au travers de :

- *LaSuli*, logiciel d'annotation sociale à l'usage des lecteurs-interprètes,
- *Porphyry*, logiciel de confrontation d'analyses de corpus par navigation multidimensionnelle,
- *Argos*, service de gestion de points de vue,
- *Steatite*, service de traitement, émission et archivage d'images.

Parmi les lieux d'expérimentation de ces outils et des méthodes associées, les usages qui nous ont permis le plus de creuser les questions de sémiotique sont probablement ceux du Pr. Jean-Marc Luce (PLH-CRATA de l'Université de Toulouse II), ses collègues et ses étudiants autour de l'étude de vases grecs.

4. MISE À L'ÉPREUVE

4.1. QUAND LES USAGES ILLUSTRONT LA THÉORIE EN DÉPASSANT LES ATTENTES

À l'occasion de la recension d'un nouvel ouvrage sur les fouilles d'une nécropole de l'âge du Fer dans la zone du Céramique (à Athènes), l'équipe toulousaine a modélisé avec nos outils (en totale autonomie vis-à-vis de nous) la description de plusieurs centaines de vases selon cette nouvelle étude – d'une part – et – d'autre part – une étude plus ancienne qui faisait jusqu'alors référence (cf. Fig. 4.1). De manière inattendue, les deux modèles (cf. Fig. 4.2) illustrent parfaitement la différence entre les approches sémiotiques et sémantiques.

Le modèle de la première étude (cf. Fig. 4.2a) permet de décrire les vases selon deux facettes : la forme (cruches, amphoriques, *skyphoi*, etc.) et la datation (hélladique récent III c, submycénien). Chacun de ces descripteurs peut faire l'objet d'une définition dans une encyclopédie ou un dictionnaire spécialisé. Ils font partie de la connaissance du domaine, sont enseignés à l'université. Autrement dit, ils présentent une certaine stabilité et sont pré-existants à l'étude. Plus encore, les descripteurs de plus haut niveau (forme et datation), si on y ajoute la localisation qui est ici la même pour tous les vases, correspondent aux trois dimensions classiques de l'archéologie : l'espace, le temps et les groupes d'artefacts [13]. Autrement dit, le modèle de la première étude est d'un niveau de généralité qui lui permettrait de prendre place dans une « ontologie » réutilisable pour décrire tout artefact archéologique.

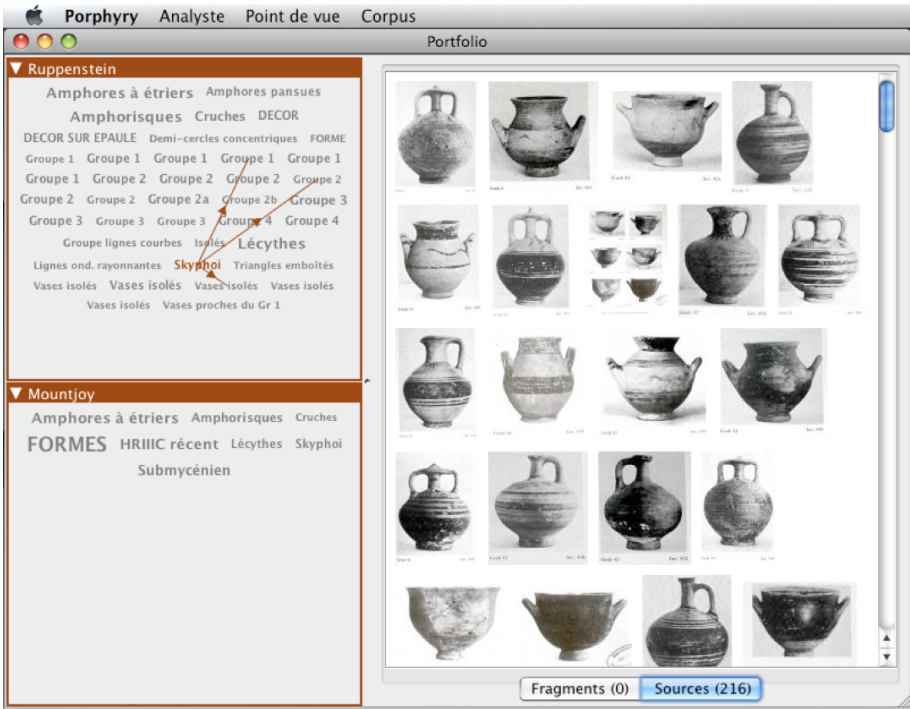


FIGURE 4.1. Description des vases d'une nécropole de l'âge du fer selon deux auteurs (copie d'écran de Porphyry).

Le modèle de la seconde étude (cf. Fig. 4.2b), plus surprenant pour l'ingénierie des connaissances, illustre parfaitement les principes de la sémiotique exposés dans la section 2.2. L'auteur de l'étude a visiblement commencé par regrouper les artefacts qui avaient la même forme (cruches, amphoriques, lécythes). Ensuite, au sein d'un de ces regroupements (par exemple les lécythes), sont identifiés des traits (détails de forme, de décor) permettant de différencier les objets. Les traits, pertinents pour les lécythes, ne sont pas forcément pour les cruches. Plus encore, les traits pertinents pour les lécythes de ce site de fouille, ne seront pas forcément les traits pertinents pour des lécythes d'un autre site. Une fois les objets de ce regroupement décrits selon ces traits, l'auteur combine les traits pour obtenir des groupes aussi cohérents que possible (« groupe 1 », « groupe 2 », etc.). Le caractère itératif de ce processus laisse des traces : certains vases restent « isolés », d'autres, en bordure de groupe (« vases proches du groupe 1 »). Une fois ordonnés selon des heuristiques d'évolution stylistique, ces groupes constituent une forme de datation, mais une datation dépendante du type d'artefact et du site. Bien que locales, ces datations ont toutefois des répercussions globales, en effet, elles suggèrent que, là où les études précédentes et le savoir de référence voyaient une rupture entre la toute fin d'une civilisation et le début d'une nouvelle, la transition se serait déroulée, en fait, par étapes successives.

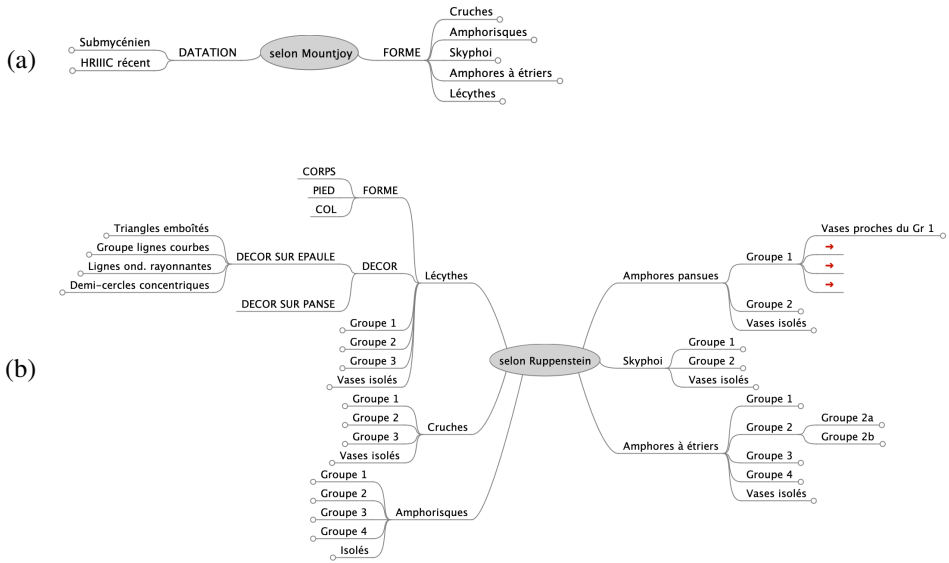


FIGURE 4.2. Vocabulaires de description des vases d’une nécropole de l’âge du fer selon deux auteurs : contraste entre approches (a) sémantique et (b) sémiotique (exports de Porphyry dans Freemind).

De manière un peu provocatrice vis-à-vis de leurs collègues, Patrick Desfarges et Bruno Helly écrivaient : « Contrairement à une opinion bien ancrée dans l’esprit des archéologues, le temps, la chronologie n’est pas une donnée mais un résultat » [12]. Loin de nous l’idée de juger les méthodes et le discours des archéologues. Cependant, en tant qu’informaticiens, nous avons le devoir de nous interroger sur la manière dont sont modélisés ces méthodes et ce discours. Il n’est pas anodin que la quasi-totalité des projets de modélisation en archéologie depuis les années 1970 soient plus *sémantiques* que *sémiotiques*. Cela signifie, pour reprendre la différence introduite par Desfarges et Helly, que ce qui est modélisé et diffusé dans les « banques de données » sont des résultats scientifiques plutôt que des données de recherche ; ou, en des termes de sociologie des sciences [22], une « boîte noire » plutôt que la « boîte blanche » d’une « science en train de se faire ».

Autrement dit, la structure très ontologique des résultats de la première étude susmentionnée (cf. Fig. 4.2a) ne signifie pas que la pratique de son auteure n’ait pas été sémiotique. Cette dernière a très bien pu souhaiter publier dans son article ses résultats sans revenir sur la méthode suivie. De la même manière, certains travaux interdisciplinaires significatifs associent une approche très sémiotique de l’Histoire des sciences et des techniques et pour autant des modèles très sémantiques [27].

Une hégémonie des approches sémantiques en archéologie aurait cependant un effet déformant sur la perception de l'archéologie par les publics de même que sur les apports possibles de l'informatique à l'archéologie. Comme les étiquettes d'un musée [2], un modèle sémantique plutôt que sémiotique aura tendance à présenter l'archéologie plus comme un savoir à acquérir que comme une démarche de recherche, toujours en mouvement. De manière plus incarnée, il s'agit finalement de la différence entre la figure du savant et celle du chercheur. Une informatique qui fait le choix de la seconde permet, selon nous, de renouveler à la fois la médiation scientifique en direction des publics et l'instrumentation du travail quotidien du chercheur.

4.2. QUAND LES USAGES INVITENT À APPROFONDIR LA THÉORIE

Plusieurs années avant les travaux que nous venons d'exposer, le tout premier corpus autour duquel nous avons collaboré concernait l'iconographie de Dionysos et des banquets sur les vases de la région de Paestum. L'équipe toulousaine a rassemblé plus de 600 photographies de ces vases, conservés dans des musées à travers le monde, dans des collections particulières, ou parfois même disparus (mais attestés par des photographies). Année après année, le repérage de différents traits (Dionysos barbu ou imberbe, regard en sens inverse de celui de la marche, etc.) ont été assignés à des étudiants dans le cadre de leur master 1, master 2 ou doctorat. Chaque analyse venant se superposer aux précédentes en autant de « points de vue ».

La méthode des co-occurrences [3, 31] s'est révélée particulièrement adaptée à la découverte du sens d'un trait (cf. Fig. 4.3). Par exemple, il est apparu que Dionysos était barbu quand il était représenté en présence d'un canthare (vase utilisé dans les rituels). L'interprétation de la doctorante ayant fait cette étude [26] est que la barbe signifiait que Dionysos était alors le dieu recevant des offrandes, par opposition au personnage des récits mythologiques, représenté imberbe.

Mais le compte-rendu de cette expérimentation serait finalement un peu vain s'il se cantonnait à ce qui a fonctionné comme prévu. De très nombreuses difficultés ont été rencontrées par les usagers, difficultés qui, nous semble-t-il, nous permettent de progresser dans notre réflexion autour de la sémiotique.

4.2.1. *Le problème du nom*

Le premier problème est celui du « nom », question à la fois triviale, mais éminemment philosophique et sémiotique. Au moment de l'import, un *identifiant* est assigné à la photographie. C'est cet identifiant qui permettra à différents experts de savoir qu'ils parlent de la même source. Par conséquent, l'identifiant doit correspondre à une unique source par serveur et être pérenne.

Gênés par la valeur un peu trop objective de ces identifiants, nous avons d'abord choisi de leur donner des valeurs aléatoires. Par la suite, nous avons dû faire face aux imports successifs des mêmes sources (dont les différentes analyses ne pouvaient plus être corrélées) et surtout à l'incapacité d'extraire des rapports qui soient compréhensibles à l'extérieur du système. L'identifiant aléatoire est alors remplacé par le nom du fichier

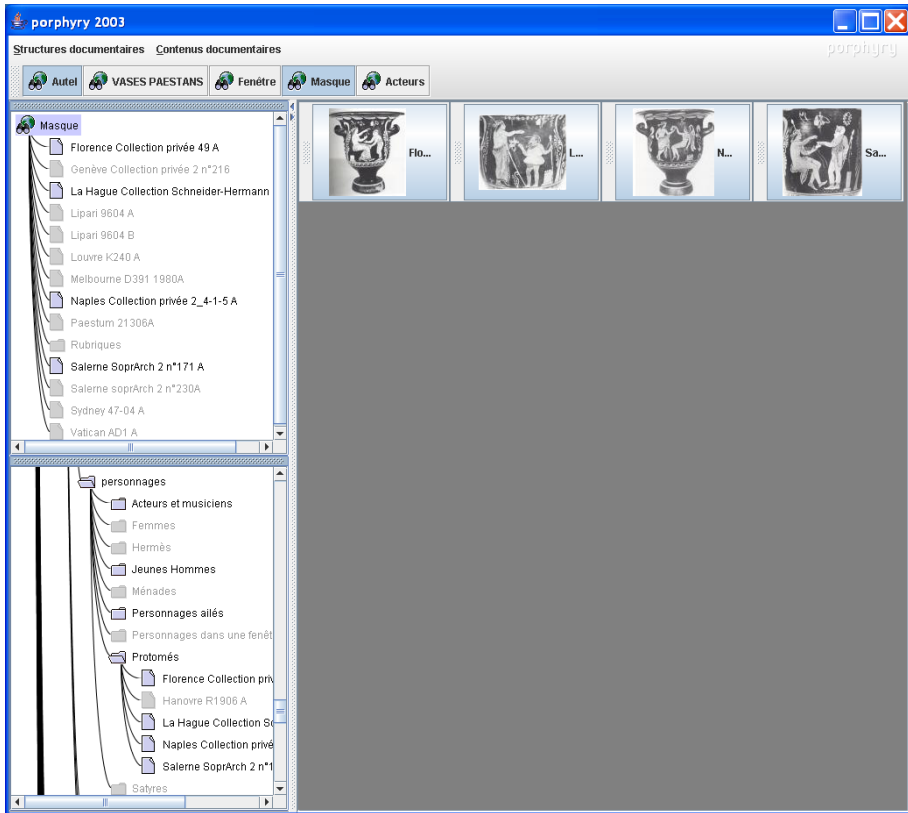


FIGURE 4.3. Statut de co-occurrence entre les masques et les différents types de personnage autour de l'autel (copie d'écran de Porphyry)

importé, souvent significatif, faisant consensus et pérenne (numéro d'archive, référence bibliographique succincte...). Dans notre cas, comme dans d'autres, la nomenclature utilisée reprend une logique de rangement hiérarchique : « Louvre_K217_B » pour la photographie B du vase K217 du Louvre. Cette nomenclature ayant une signification, il devient alors intéressant de la retrouver en tant que structure d'analyse au même titre que les points de vue des experts. La source « Louvre_K217_B » peut alors être décrite par le *réseau de descripteurs* « Louvre > K217 > B » (cf. Fig. 4.4). La création de ce réseau, tâche répétitive et sans grand intérêt pour l'expert, était censée être rendue « transparente » grâce à un algorithme simple sur les chaînes de caractères des noms de fichiers à importer, mais à en croire les doléances remontées du terrain, elle n'en était que plus obscure...

Si nous avons choisi de discuter de ce problème, c'est qu'il révèle selon nous une question fondamentale. Nous nous plaisons à considérer le texte (ou ici l'image) comme le lieu de confrontation des différentes analyses des experts. Cet objet, parce qu'il est toujours à interpréter, nous éviterait les pièges de tous les intégrismes (le positivisme

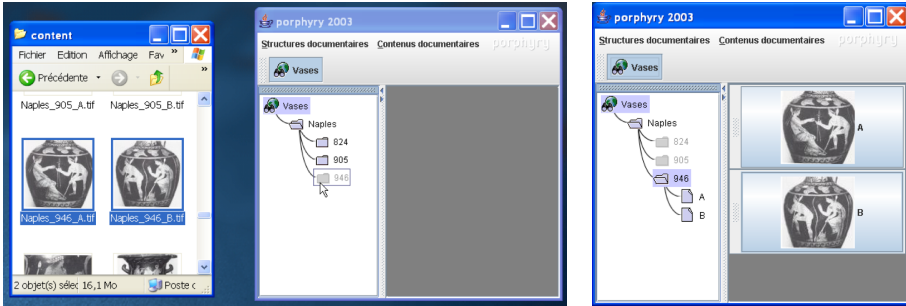


FIGURE 4.4. Séquence d'importation de sources (par un glisser-déposer) lors de la constitution du corpus (copie d'écran de Porphyry)

y compris), et, parce qu'il est accessible à tous et qu'il devient « document » (au sens de « preuve »), nous protégerait du relativisme. Le nom donné au document, toujours relatif à d'autres noms et donc à la structure du corpus, devient alors extrêmement problématique. En effet, si le document joue bien un rôle de « butée ontologique » vis-à-vis des points de vue d'experts, comment ne pas retrouver dans la structure du corpus la naïveté des « ontologies informatiques », supposées référentielles et éternelles [2] ?

4.2.2. L'absence d'usage des fragments

Une fois le corpus établi, même partiellement, le travail sur les sources peut commencer. Parmi les manipulations permises par le système sur les sources, une seule laisse des traces : celle consistant à sélectionner des fragments (cf. Fig. 4.5). Or, l'observation du corpus produit par les usagers nous a révélé une absence totale de fragments. Comment expliquer que les usagers d'un logiciel d'annotation n'utilisent pas une fonctionnalité *a priori* aussi centrale que la gestion des fragments ? Formation à l'outil trop courte ? Ergonomie imparfaite de la création et de l'utilisation des fragments ? Des éléments de réponse nous sont donnés lorsque l'on cherche un exemple dans le corpus de ce que l'emploi de fragments aurait pu apporter.

Tout d'abord, il apparaît que les éléments graphiques intéressant nos experts, par exemple un oiseau et une chèvre, sont tout à fait visibles dans la scène d'ensemble et, dans la majorité des cas, présents une seule fois. Quel intérêt alors de les sélectionner ? Nous nous trouvons sans doute ici devant une singularité du matériel étudié lui-même : les scènes représentées sur les vases semblent utiliser les figures avec parcimonie comme pour accentuer leur valeur symbolique. Ensuite, on serait en droit de s'interroger sur l'intérêt d'afficher, par exemple, la position d'un personnage à la jambe levée sans afficher *en même temps* celle du personnage assis qui lui fait face. Autrement dit, la sélection de fragments ne semble avoir d'intérêt que dans une logique différentielle.

Enfin, on peut douter du bien-fondé d'une interface qui encourage la création de catégories d'analyse, nommées dès le départ, puis la création de fragments dans ces

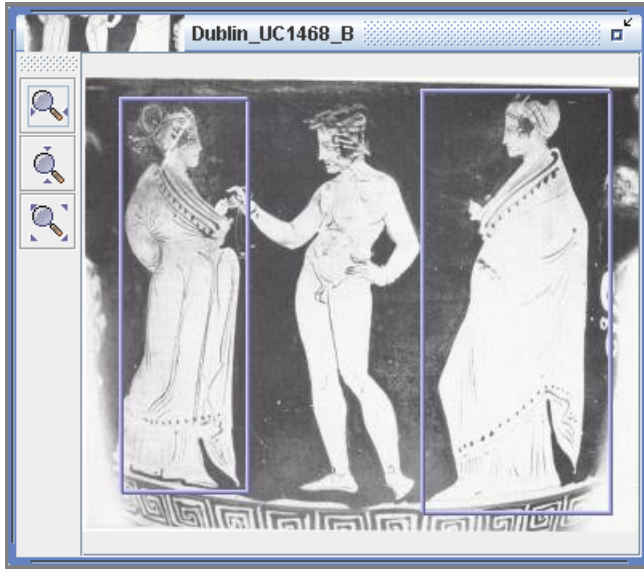


FIGURE 4.5. Cas fictif de sélection de fragments caractérisés par les mêmes traits (copie d'écran de Porphyry)

catégories. Sans doute faudrait-il plutôt permettre dans un premier temps de sélectionner les fragments, de les rassembler ensuite en groupes anonymes et aux limites un peu floues, et de préciser ces limites graduellement (c'est ce comportement que nous avons implémenté par la suite dans LaSuli pour l'annotation des textes).

4.2.3. Où est la signification du descripteur ?

Les ensembles de fragments (ou de sources) allaient ensuite être décrits par l'expert, à l'aide de ce que nous appelons alors des « descripteurs ». Cette appellation avait pour but d'éviter les « concepts » chers à l'intelligence artificielle et d'indiquer que ces simples termes avaient pour seule vocation d'être utilisés dans des descriptions. L'emprunt de ce nom aux sciences documentaires est cependant malheureux puisqu'il évoque des « listes d'autorité », consensuelles et pérennes, là où au contraire notre descripteur est propre au point de vue de l'expert (sorte d'idiolecte ou de sociolecte), et peut y apparaître ou disparaître à tout moment.

Pour tester si le descripteur n'était qu'un terme, si sa signification était uniquement donnée par son nom, nous avons décidé de brider volontairement le logiciel durant quelques mois en n'offrant pas de fonctionnalité de renommage. Les usagers n'ont pas manqué de nous faire savoir que c'était très ennuyeux puisque cela obligeait à en créer un autre et à déplacer les fragments et les sources associées. Ce retour est intéressant car il semble vouloir dire que plus que par son nom le « descripteur » est davantage caractérisé par les fragments et les sources auquel il est associé. Tout porte à croire que nous sommes en présence d'une collection [28], qui, à l'image de la collection

d'œuvres d'art, serait construite, non à partir d'une définition subsumant chacun de ses membres, mais en traçant un parcours à travers leurs différences. La constitution de cette collection pourrait alors se faire dans la continuité des regroupements dont nous parlions dans la section précédente. La collection permettrait de donner une frontière à ce groupe, puis éventuellement un nom, le tout de manière transitoire au gré des découvertes de singularités parmi les sources et les fragments.

4.2.4. *L'envie irrépressible de compter*

L'un des retours des usagers portait sur l'absence, dans les premières versions du logiciel, de comptage automatique. Ce reproche était surprenant pour un outil qui se veut avant tout qualitatif. Ne serait-ce pas par souci de l'exhaustivité ? Le même outil, dans les mains d'historiens de la Chine des années 30, était utilisé au contraire pour relever une ou deux manifestations typiques d'un phénomène (habillement, technique, etc.). Il n'est pas impossible que cette différence de statut donné aux sources dépende de la discipline. Il est intéressant de noter que l'interface homme-machine de l'époque n'affichait pas l'ensemble des objets documentaires mais juste quelques exemples prototypiques. Les usagers avaient cependant réussi à contourner ce comportement du logiciel en dupliquant à la racine du point de vue chaque objet documentaire. Par ailleurs, que s'agissait-il de compter : les fragments, les sources photographiques ou bien les dossiers d'archive et donc, indirectement, les vases ? N'y a-t-il pas alors un surprenant « glissement » référentiel ? Ceci vient rappeler le rôle de *témoin* qu'assure le document pour la communauté scientifique en archéologie vis-à-vis d'objets et de structures rendus à jamais inaccessibles par l'analyse destructive qu'est la fouille [18].

4.2.5. *L'essentiel n'est peut-être pas modélisé*

Parmi les retours des usagers, on peut noter qu'ils aient particulièrement apprécié avec nos logiciels – à la différence des tableaux qu'ils utilisaient précédemment – d'être maintenus en contact constant avec les photographies. La question que cela pose est si finalement il n'y aurait pas d'autres traits pertinents pour les usagers dans les photographies que ceux qu'ils ont décrits dans le système.

Cette manière de faire, peut-être décevante d'un point de vue informatique, ne serait pas absurde d'un point de vue sémiotique : les traits modélisés permettraient de sélectionner ce qui est commun (le fond isotopique) et c'est alors que l'interprète pourrait repérer parmi les éléments graphiques ce qui diffère sans forcément avoir besoin de l'inscrire dans l'outil.

Une autre récrimination concerne l'absence dans le logiciel de fonctionnalité d'impression. On pourrait s'étonner de cette volonté à tout prix d'imprimer les résultats quand on utilise un environnement visant au contraire une continuité numérique entre la lecture et l'écriture ? Le problème est peut-être plus profond : la diffusion d'une interprétation nécessite probablement un changement de forme, un argumentaire, la production d'un texte plutôt que d'un simple modèle.

Ces deux derniers points inciteraient les informaticiens à garder une certaine humilité : l'outil aurait beau permettre le travail humain d'interprétation, de construction du sens, il n'enregistrerait pas forcément les traces des traits les plus pertinents, ni ne serait forcément le moyen le plus adapté à sa diffusion. Autrement dit, l'essentiel de l'interprétation serait peut-être en dehors de ce qui est modélisé.

5. CONCLUSION

En ne s'intéressant que trop peu à son histoire, l'informatique prend le risque de toujours emprunter les mêmes voies et de se retrouver finalement face aux mêmes obstacles. C'est pourquoi, pour traiter de l'intelligence artificielle appliquée aux humanités numériques, sujet pourtant d'une grande actualité, nous avons fait le pari risqué dans cet article de concentrer notre état de l'art sur les années 1970. Les résultats de cette investigation nous semblent paradoxalement assez « rafraichissants ». Nous découvrons que ces premiers travaux de modélisation des objets archéologiques à l'aide de « langages d'analyse » et de « domaines » n'ont rien à envier aux projets actuels à base de RDF et d'OWL, qu'ils recèlent même souvent une finesse d'analyse digne des projets les plus aboutis. Mais ces travaux sont surtout intéressants par les débats qu'ils firent naître dans la communauté des archéologues, débats d'une profondeur théorique suffisante pour rester, nous semble-t-il, applicables plus de 45 ans après.

Parmi les critiques de l'époque, la plus intéressante et la plus constructive est probablement celle de l'archéologue Philippe Bruneau : contrairement aux objets de la Nature, les objets des Sciences humaines, parce qu'ils sont déjà porteurs de sens, doivent être décrits avec des méthodes sémiotiques (plutôt que sémantiques). La prise en compte du contexte ne se traduit alors pas par un vague principe ou par un modèle de plus, mais par un refus systématique des modèles à portée générale, par une prise en compte du fait que chaque objet pourrait être décrit par un nombre indéfini de traits sémantiques, mais que seuls un petit nombre d'entre eux seront pertinents et ce par différence avec les objets qui sont proches de lui dans son univers technique. Consciemment ou inconsciemment, un certain nombre de travaux d'aujourd'hui en ingénierie des connaissances s'inscrivent au moins partiellement dans ces perspectives. Cependant, elles restent trop minoritaires pour être réellement visibles dans ce champ scientifique et encore moins dans celui des humanités numériques.

Comme nous l'avons illustré avec nos propres expérimentations, la prise en compte de cette approche sémiotique dans les logiciels ainsi que les usages qui en émergent tracent des perspectives assez enthousiasmantes pour l'instrumentation de la pratique quotidienne des chercheurs en Sciences humaines ainsi que pour la médiation scientifique. Les perspectives qui s'ouvrent ne doivent pas pour autant cacher les questions qui se posent plus nombreuses encore sur ce que sont le sens et ses modes de construction en Science humaines. Mais peut-être est-ce là justement un retour aux sources de l'intelligence artificielle, à l'époque où sa visée n'était pas substitutive mais compréhensive, où il ne s'agissait pas tant de simuler ou de remplacer l'intelligence humaine que de tenter humblement de mieux la comprendre.

BIBLIOGRAPHIE

- [1] P. AMANDRY, « Avant-propos », *Bulletin de correspondance hellénique* **101** (1977), n° 1 (français), p. 1-3.
- [2] A. BÉNEL, « Ontologies du Web : Histoire refoulée et perspectives paradoxales », *Intellectica. Revue de l'Association pour la Recherche Cognitive* **61** (2014), n° 2014/1 (français), p. 123-141.
- [3] A. BÉNEL, S. CALABRETTO, A. IACOVELLA & J. PINON, « Vers un outil documentaire unifié pour les chercheurs en archéologie », in *Actes du 18^e Congrès INFORSID, Lyon, France, 16-19 mai, 2000*, 2000, p. 133-145.
- [4] T. BERNERS-LEE, J. HENDLER & O. LASSILA, « The Semantic Web », *Scientific American* **284** (2001), n° 5, p. 34-43.
- [5] M. BORILLO & J.-C. GARDIN, *Les banques de données archéologiques. Marseille, 12-14 juin 1972*, Éditions du CNRS, 1974.
- [6] L. BOURRELY & E. CHOURAQUI, *Le système documentaire SATIN 1 : description générale et manuel d'utilisation*, Éditions du CNRS, 1974.
- [7] P. BRUNEAU, « Quatre propos sur l'archéologie nouvelle », *Bulletin de correspondance hellénique* **100** (1976), n° 1, p. 103-135.
- [8] J. CAUSSANEL, J.-P. CAHIER, M. ZACKLAD & J. CHARLET, « Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique ? », in *Actes des 13^e journées francophones d'ingénierie des connaissances (IC)*, 2002.
- [9] J. CHARLET, B. BACHIMONT, J. BOUAUD & P. ZWEIGENBAUM, « Ontologie et réutilisabilité : expérience et discussion », in *Acquisition et ingénierie des connaissances* (N. Ausseac-Gilles, P. Laublet & C. Reynaud, eds.), Cepadués Éditions, 1996, p. 69-87.
- [10] E. CHOURAQUI, « Le système d'exploitation automatique de l'inventaire général des monuments et richesses artistiques de la France (Formalisation du langage d'analyse) », in *Les banques de données archéologiques. Marseille, 12-14 juin 1972* (M. Borillo & J.-C. Gardin, eds.), Éditions du CNRS, 1974.
- [11] D. L. CLARKE, *Models in Archaeology*, Methuen, London, 1972.
- [12] P. DESFARGES & B. HELLY, « L'archéologie : système d'information scientifique », in *Actas coloquio internacional de arqueologia a informatica, Saint-Germain-en-Laye, 1991*, 1991.
- [13] R. ÉTIENNE, Y. AUDA & A. IACOVELLA, « Spécificité des problèmes d'analyse des données en archéologie : Application à l'analyse des nécropoles », in *Actas coloquio internacional de arqueologia a informatica, Saint-Germain-en-Laye, 1991*, 1991.
- [14] R. GINOUVÈS, « Projets de banques de données archéologiques à l'Université de Paris X », in *Les banques de données archéologiques. Marseille, 12-14 juin 1972* (M. Borillo & J.-C. Gardin, eds.), Éditions du CNRS, 1974.
- [15] R. GINOUVÈS & A.-M. GUIMIER-SORBETS, *La constitution des données en archéologie classique*, Éditions du CNRS, 1978.
- [16] T. R. GRUBER, « The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases », in *Principles of Knowledge Representation and Reasoning : Proceedings of the Second International Conference* (J. A. Allen, R. Fikes & S. E., eds.), Morgan Kaufmann, Cambridge, 1991, p. 601-602.
- [17] A.-M. GUIMIER-SORBETS, *Les bases de données en archéologie : Conception et mise en œuvre*, Éditions du CNRS, 1990.
- [18] A. IACOVELLA, A. BÉNEL, S. CALABRETTO & B. HELLY, « Assistance à l'interprétation dans les bibliothèques numériques pour les sciences historiques », in *La société de l'information et ses enjeux, Actes du colloque de bilan du programme interdisciplinaire « Société de l'information »* (J.-L. Lebrave, éd.), CNRS, Cambridge, 2005, p. 167-179.
- [19] I. KANELLOS & C. MAUCERI, « Une conscience interprétative face à un univers de textes : arguments en faveur d'une analyse de données interprétative », *Syntaxe et sémantique* **9** (2008), p. 37-52.
- [20] I. KANELLOS, T. THLIVITIS & A. LÉGER, « Indexation et recherche d'information imagée par le contenu : catégories de textes et catégories d'images dans le projet SemIndex », in *Actes du troisième Colloque International sur le Document Électronique CIDE'2000, Lyon, 4-6 juillet 2000*, Europa Productions, Paris, 2000, p. 171-183.
- [21] D. KARGER & A. BERNSTEIN, « Interview », 2010, In : Kate Ray, *Web 3.0*, documentary.

- [22] B. LATOUR, *La Science en action*, La Découverte, 1989.
- [23] J. LITVAK KING & R. GARCÍA MOLL, « Set Theory Models : An approach to taxonomic and locational relationships », in *Models in Archaeology* (D. Clarke, éd.), Methuen, London, 1972, p. 735-755.
- [24] D. OLDMAN, « The CIDOC Conceptual Reference Model (CIDOC-CRM) : Primer », Technical paper, version 1.1., 2014.
- [25] V. PERLERIN, « Sémantique légère pour le document : Assistance personnalisée pour l'accès au document et l'exploration de son contenu », *Texto!* (2004).
- [26] V. POUYADOU, « Dionysos barbu : Le sens du poil », *Dionysos barbu : Le sens du poil* 57 (2001), p. 169-183.
- [27] B. ROHOU, S. LAUBÉ & S. GARLATTI, « L'ontologie PHO en Histoire des sciences et techniques », in *Actes des 28^e journées francophones d'ingénierie des connaissances (IC)*, 2017.
- [28] F. ROUSSEAU, « La collection, un lieu privilégié pour penser ensemble singularité et synthèse », *EspacesTemps.net* (2006).
- [29] M. SCHOOP, A. DE MOOR & J. L. G. DIETZ, « The pragmatic web : a manifesto », *Communications of the ACM* 49 (2006), n° 5, p. 75-76.
- [30] L. TANGUY, « Traitement automatique de la langue naturelle et Interprétation : Contribution à l'élaboration d'un modèle informatique de la Sémantique Interprétative », *Texto!* (1998).
- [31] C. ZHOU & A. BÉNEL, « From the crowd to communities : New interfaces for social tagging », in *Proceedings of the 8th international conference on the design of cooperative systems (COOP)*, 2008.

ABSTRACT. — To deal with Artificial intelligence applied to Digital humanities, this article boldly focuses its state of the art on the 1970s. We discover that modelling archaeological artifacts at that time with “analysis languages” and “domains” competes with current projects based on RDF and OWL, even with subtleties that only the best ones achieve today. But projects from the 1970s are especially interesting because of the debates they inspired in the archaeologists community, debates with such a theoretical depth that they still could have an impact, more than 45 years later. Among the critics of that time, the most insightful and constructive is P. Bruneau (an archaeologist). According to him, the objects of Human Sciences, because they are already meaningful (unlike the objects of Nature), must be described with semiotic (rather than semantic) methods. Coping with context cannot be reduced to an elusive statement, neither to an additional model: it requires the systematic refusal of broad scope models. Models should take into account that among all the possible distinctive features of an object, only a few are pertinent: they are highlighted by the differences with the neighbouring objects of the “technical universe”. Several current works in Knowledge engineering, consciously or not, fit into those perspectives at least partially. As illustrated with our own software and experiments, taking into account this semiotic approach offers promising prospects for the instrumentation of the daily practice of researchers in Humanities as well as for scientific mediation. But a greater prospect could be in the bunch of design open issues that this approach raises: mundane issues at first sight, but related indeed with a better comprehension of sense and sense making. This could lead us back to the roots of Artificial intelligence and Cognitive sciences, when their main aim was less to replace human intelligence than to better understand it.

KEYWORDS. — Knowledge modeling, archaeology, artefact, pertinent feature.

RESUMEN. — Para abordar la inteligencia artificial aplicada a las humanidades digitales, este artículo toma la arriesgada apuesta de centrar su estado del arte en los años 70. Descubrimos que los primeros trabajos de modelización de objetos arqueológicos utilizando “lenguajes de análisis” y “dominios” no tienen nada que envidiar a los proyectos actuales basados en el Marco de Descripción de Recursos (RDF) y en el Lenguaje de Ontologías Web (OWL), los cuales a menudo incluso contienen una finura de análisis digna de los proyectos más exitosos. Sin embargo, estos trabajos son especialmente interesantes por los debates que han suscitado en la comunidad arqueológica, debates de suficiente profundidad teórica para seguir siendo aplicables, a nuestro parecer, más de 45 años después. Entre las críticas de la época, la más interesante y constructiva es probablemente la del arqueólogo Philippe Bruneau: a diferencia de los objetos de la Naturaleza, los objetos de las Ciencias Humanas, por ser ya portadores de significado, deben ser descritos mediante métodos semióticos (más que semánticos). El hecho de tener en cuenta el contexto no se traduce a un principio vago o a un modelo más, sino a un rechazo sistemático de los modelos de alcance general. Los modelos deben tener en cuenta el hecho de que sólo un pequeño número de los rasgos que pueden caracterizar a un objeto serán relevantes, y es por medio de la diferencia con los objetos que están cerca de él en su universo técnico. Consciente o inconscientemente, un cierto número de los trabajos actuales de ingeniería del conocimiento están al menos parcialmente en línea con estas perspectivas. Como lo ilustramos con nuestros propios software y experimentos, tomar en cuenta este enfoque semiótico traza perspectivas prometedoras para la instrumentación de la práctica cotidiana de los investigadores en las Humanidades, así como para la mediación científica. Pero por justo retorno a las fuentes de la inteligencia artificial (en una época en que su objetivo era menos sustitutivo que comprensivo), el interés por tener en cuenta el enfoque semiótico es tal vez aún mayor en la multitud de asuntos de concepción que ésta suscita, preguntas que son anecdóticas a primera vista, pero que en última instancia están vinculadas a lo que es el significado, y a sus modos de construcción.

PALABRAS CLAVES. — Ingeniería del conocimiento, arqueología, artefacto, rasgos relevantes.

Manuscrit reçu le 27 septembre 2018, révisé le 20 janvier 2019, accepté le 15 mars 2019.